Which Heart Rate Variability (HRV) Indices Should I Use for Psychophysiological Research? A Data-driven Approach to Identifying Clusters of HRV Indices

Tam Pham[1, 2], Carly J. Johnco[1,2], Zen J. Lau[3], Dominique Makowski[4,5, *], & Miriam K. Forbes[1,2, *]

[1] School of Psychological Sciences, Macquarie University
[2] Macquarie University Lifespan Health & Wellbeing Research Centre, Macquarie University
[3] School of Social Sciences, Nanyang Technological University
[4] School of Psychology, University of Sussex
[5] Sussex Centre for Consciousness Science, University of Sussex

[*] Dominique Makowski and Miriam K. Forbes are co-last authors and contributed equally to the senior leadership and oversight of this project.

Corresponding author: Tam Pham, School of Psychological Sciences, Macquarie University, Australian Hearing Hub, Sydney 2109, Australia. Email: phamttam17@gmail.com.
Co-authors: Carly J. Johnco (carly.johnco@mq.edu.au); Zen J. Lau (laujenzuen@gmail.com); Dominique Makowski (d.makowski@sussex.ac.uk); Miriam K. Forbes (miri.forbes@mq.edu.au).

Short title: Heart Rate Variability Structure.

Financial Support: This research is supported by an Australian Government Research Training Program (RTP) Scholarship – awarded to Tam Pham**.**

Conflict of Interest: None**.**

Authorship: Tam Pham: Conceptualization (Equal), Methodology (Equal), Software (Equal), Formal Analysis (Lead), Writing – Original Draft (Lead), Writing – Review and Editing (Lead), Project Administration. **Carly J. Johnco**: Data Curation (Equal), Writing – Review and Editing (Supporting), Supervision (Supporting). **Zen J. Lau**: Software (Equal), Writing – Original Draft (Supporting). **Dominique Makowski**: Conceptualization (Equal), Software (Lead), Data Curation (Equal), Writing – Review & Editing (Supporting), Supervision (Supporting). **Miriam K. Forbes**: Methodology (Equal), Writing – Review and Editing (Equal), Supervision (Lead).

Ethical Standards Disclosure: Not Applicable.

Abstract

Heart Rate Variability (HRV) can be quantified using a myriad of mathematical indices, but the lack of systematic and empirical comparison between these indices complicates the evaluation and interpretation of HRV data. This study assessed the reliability, consistency, and generalizability of structural relationships among 89 HRV indices using a consensus-clustering approach. We analyzed 635 short-term resting-state electrocardiogram (ECG) recordings from two samples of college students with differing psychological profiles. Results from a sample with elevated internalizing symptoms (N=233)—collected across two sessions, one week apart—were compared to evaluate the test-retest reliability of the HRV clusters. To further assess the stability and generalizability of these HRV clusters beyond individuals with elevated internalizing symptoms, these results were compared with a second sample not selected based on psychological symptoms (N=203). We identified 21 clusters of 70 HRV indices with cross-method, test-retest, and cross-sample robustness. Based on the robust empirical convergence and the relative popularity of some HRV indices in the extant literature, we recommend 13 HRV indices for short-term recordings of resting-state HRV (under 10 minutes): RMSSD, SDNN, RSA (Porges-Bohrer or Peak-to-Trough method), RSA (Gates method), SD1/SD2 or CSI, SampEn, HF or LnHF, DFA α1, DFA α2, one of the MDFA α1 features, one of the MDFA α2 features, one of the heart rate asymmetry indices, and one of the heart rate fragmentation indices. This approach mitigates the biases that can arise from redundant or highly correlated indices, facilitates clearer interpretation, and enhances the validity of conclusions drawn from HRV analyses.

Keywords: heart rate variability, electrocardiogram, resting, methodology, clustering, psychopathology,

Which HRV Indices Should I Use? A Data-driven Approach to Identifying Clusters of HRV Indices

Heart Rate Variability (HRV) is a measure that quantifies the variation in time intervals between consecutive heartbeats. In psychophysiological research, HRV serves as a multifaceted index for exploring the interactions between the autonomic nervous system (ANS) and other physiological systems—such as the central nervous system, the endocrine system, and the metabolic processes—that collectively contribute to the regulation of heart rate (Bigger, 1997; Pham et al., 2021; Quigley et al., 2024; Shaffer & Ginsberg, 2017; Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). The Neurovisceral Integration Model (Thayer & Lane, 2000) and Polyvagal Theory (Porges, 2009, 2011) propose that HRV abnormalities (e.g., significantly lower HRV at rest) reflect a diminished capacity for self-regulation, or the body's ability to regulate and adapt to internal and external environmental changes, which is associated with higher risks for mental and physical health issues (Austin et al., 2007; Porges, 2011). Correspondingly, reduced HRV at rest has been linked to a wide range of mental and physical health outcomes in the extant literature, including diabetes (Schroeder et al., 2005), cardiovascular diseases (Thayer et al., 2010), poor cognitive functioning (Forte et al., 2019; Holzman & Bridgett, 2017), emotional dysregulation (Fantini-Hauwel et al., 2020; Pinna & Edwards, 2020; Sloan et al., 2017), and a variety of mental disorders (e.g., posttraumatic stress disorder: Campbell et al., 2019; anxiety disorders: Chalmers et al., 2014; Cheng et al., 2022; schizophrenia: Clamor et al., 2016; depression: Kemp et al., 2010; Koch et al., 2019).

In the early days, HRV was quantified over a long period of recording (e.g., 24 hours) primarily as a risk assessment tool for heart disease (Bigger, 1997; Cripps et al., 1991; Kleiger et al., 1987). With advancements in digital signal processing techniques, new methodologies have been introduced to capture subtle beat-to-beat variations in shorter recordings (e.g., ≤ 5 minutes; Pham et al., 2021; Shaffer & Ginsberg, 2017). The increased accessibility of HRV measurement methods has likely fueled the exponential increase in the popularity of HRV research in the past few decades, alongside the rapid development of new indices aimed at capturing the dynamic nature of HRV more effectively. For instance, the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (Task Force, 1996) initially identified 15 indices and their derivatives (i.e., normalized or logarithmic forms). Within 25 years, more than 40 indices were commonly used in research (Pham et al., 2021; Shaffer & Ginsberg, 2017) and a wide range of software packages—such as ARTiiFACT (Kaufmann et al., 2011), Kubios HRV (Tarvainen et al., 2014), HRVanalysis (Pichot et al., 2016), NeuroKit2 (Makowski et al., 2021), PyBios (Silva et al., 2020), and PhysioZoo (Behar et al., 2018)—now enable researchers to compute dozens to hundreds of HRV indices across time-, frequency-, and nonlinear domains, greatly enhancing the accessibility and scope of HRV analysis.

While the application of HRV indices in research has become increasingly popular, there is a limited empirical and conceptual understanding of the relationships among the indices. This knowledge gap, coupled with the proliferation of new indices, creates challenges for research progress, including issues with replicability, collinearity, and interpretability. The remainder

of this introduction provides an overview of existing HRV indices, discusses problems arising from their proliferation, and outlines the present study, which aims to address these challenges by empirically examining the associations among these indices and identifying clusters of similar measures to streamline research efforts.

Overview of HRV Indices

Generally, the multitude of HRV indices is categorized based on the mathematical approach underpinning them, with categories conventionally named *time-domain*, *frequency-domain*, and *nonlinear-domain* (Quigley et al., 2024; Shaffer & Ginsberg, 2017; Task Force, 1996). In parallel, other metrics were developed to specifically capture the natural variation in heart rate during the breathing cycle, referred to as the *respiratory sinus arrhythmia* (*RSA*) indices (Task Force, 1996).

Time-domain indices (**Table 1**) are among the most straightforward methods for quantifying the variability of intervals between successive normal heartbeats (i.e., excluding abnormal beats such as ectopic beats), or otherwise often referred to as normal-to-normal intervals. Some commonly derived indices include the standard deviation of all normal-to-normal intervals (*SDNN*), the root mean square of the sum of successive differences of normal-to-normal intervals (*RMSSD*), and the percentage of adjacent normal-to-normal intervals separated by more than 50ms (*pNN50*). These measures are computationally simple and are widely used in research and clinical settings. Several, such as *RMSSD* and *pNN50*, have been shown to be reliable markers of parasympathetic (vagal) activity, particularly in short-term recordings where they correlate strongly with high-frequency (*HF*) spectral power (Laborde et al., 2017; Shaffer & Ginsberg, 2017). However, while these specific indices predominantly capture rapid, high-frequency variability, time-domain methods as a broader group do not distinguish between contributions from different frequency bands. This limits their ability to disentangle the influence of multiple autonomic and non-autonomic systems, particularly slower mechanisms such as circadian rhythms or delayed baroreflex responses (Berntson et al., 1997; Billman, 2011; Task Force, 1996).

Table 1

A summary of HRV time-domain indices.

Sub Domain	Index	Description
	MeanNN	Mean of normal-to-normal intervals
	MedianNN	Median of normal-to-normal intervals
	MinNN	Minimum normal-to-normal intervals
	MaxNN	Maximum normal-to-normal intervals
	MadNN	Median absolute deviation of the normal-to-normal intervals
	CVNN	Coefficient of variation of normal-to-normal intervals, calculated by dividing SDNN by MeanNN
	MCVNN	Coefficient of variation of normal-to-normal intervals, calculated by dividing MadNN by MedianNN

Deviation-based approach	IQRNN	Interquartile range of normal-to-normal intervals
	Prc20NN	20th percentile of the normal-to-normal intervals
	Prc80NN	80th percentile of the normal-to-normal intervals
	SDNN	Standard deviation of normal-to-normal intervals
	SDNN Index (SDNNI)	Mean of the standard deviations of normal-to-normal intervals
	SDRMSSD	SDNN divided by RMSSD
Difference-based approach	SDSD	Standard deviation of successive normal-to-normal intervals differences
	RMSSD	Root mean square of successive normal-to-normal intervals differences
	CVSD	Coefficient of variation of successive normal-to-normal intervals differences, calculated by dividing RMSSD by MeanNN
	pNN20	Proportion of successive normal-to-normal intervals differences larger than 20ms
	pNN50	Proportion of successive normal-to-normal intervals differences larger than 50ms
Geometric approach	HTI	HRV Triangular Index, derived from the diving the integral of the density of the normal-to-normal intervals histogram by its height
	TINN	Triangular interpolation of normal-to-normal intervals, derived from the baseline width of the normal-to-normal intervals histogram

Frequency-domain indices (**Table 2**) offer partial improvements over time-domain measures by providing greater temporal resolution by separating the total variability into components according to the speed or frequency of heart rate fluctuations. These indices are often used to infer activity in specific branches of the ANS. However, caution is warranted, as frequency-domain measures do not fully disentangle the contributions of sympathetic and parasympathetic influences, and some components may also reflect non-autonomic physiological processes. For instance, the high frequency (*HF*) index is an estimation of power within the high frequency band, between 0.15 to 0.40 Hz, and is widely considered as a marker of parasympathetic activity, particularly reflecting heart rate variabilities related to respiratory sinus arrhythmia (Task Force, 1996; Berntson et al., 1997; Shaffer & Ginsberg, 2017). Low frequency (*LF*) typically defined between 0.04 and 0.15 Hz, was historically associated with sympathetic activities. However, this interpretation has been challenged by evidence suggesting that LF reflects mixed autonomic influences. Similarly, the LF/HF ratio, once widely considered as an index of sympathovagal balance, is now regarded with skepticism, as its physiological interpretation lacks consistent empirical support (Billman, 2013; Laborde et al., 2017). Very-low-frequency (VLF; <0.04 Hz) and ultra-low-frequency (ULF; <0.003 Hz) components are thought to be influenced by slower-acting regulatory

systems, such as thermoregulation, hormonal fluctuations, and circadian rhythms (Shaffer & Ginsberg, 2017). In addition to absolute power measures, frequency-domain analysis may include relative metrics, including the normalized forms of the absolute powers, such as normalized *LF* (*LFn*) and HF (*HFn*), or their natural logarithmic variants (e.g., *LnHF*).

Table 2

A summary of HRV frequency-domain indices.

Sub Domain	Index	Description
Absolute Power	ULF	Ultra-low frequency, power spectrum in the frequency range of < 0.003 Hz
	VLF	Very low frequency, power spectrum in the frequency range of 0.0033-0.04 Hz
	LF	Low frequency, power spectrum in the frequency range of 0.04-0.15 Hz
	HF	High frequency, power spectrum in the frequency range of 0.15-0.4 Hz
	TP	Total Power, sum of power spectrum in all frequency bands
	LnHF	Natural logarithm of HF
Normalized / Relative Power	HFn	Normalized HF, calculated by dividing HF by TP
	LFn	Normalized LF, calculated by dividing LF by TP
	LF/HF	Ratio of LF to HF power

While time- and frequency-domain indices remain the most commonly used approaches for quantifying HRV, they are inherently linear measures and assume stationarity of the underlying signal. As a result, they are limited in their ability to fully capture the complex, dynamic, and often non-linear nature of autonomic regulation. To this end, *non-linear* indices (**Table 3**), grounded in the concepts in the non-linear dynamics and chaos theory literature (Goldberger, 1996; Lau et al., 2022a; Makowski et al., 2022), have gained increasing attention for their potential to more accurately characterize the complexity and adaptability of physiological system. Early indices of this type included measures obtained from a Poincaré plot where all normal-to-normal intervals are plotted against the preceding intervals. An ellipse is subsequently fitted to circumvent all plotted points and is used to derive Poincaré plot non-linear indices such as *SD1* (ellipse's width) and *SD2* (ellipse's length) which represent the standard deviation of the short-term and long-term normal-to-normal interval variability respectively (Karmakar et al., 2009).

More recently, advanced methods have been introduced to quantify the predictability and regularity of the heart rate signal—two important properties of the complex heart rate variability system. Predictability describes the amount of correlation present in the evolution of the signals over time whereas regularity describes the general amount of repetitions of patterns in the signal (Lau et al., 2022b). To assess predictability, several indices have been developed based on fractal analysis, which captures self-similar patterns at multiple scales (Captur et al., 2017; Katz, 1988; Mandelbrot et al., 1997). Some *fractal-based indices* include the Detrended Fluctuation Analysis (*DFA*), multi-fractal DFA (*MFDFA*), and the correlation dimension (*CD*). On the other hand, regularity is commonly assessed using entropy-based measures, which quantity the amount of uncertainty or randomness in the signal (Cover,

1999; Voss et al., 2008). Some *entropy-based indices* include approximate entropy (ApEn), sample entropy (*SampEn*), and multiscale entropy (*MSE).*

In addition, newer non-linear methods include the heart rate fragmentation indices (Costa et al., 2017a), which quantify abrupt changes or disorganization in the heartbeat pattern, and the heart rate symmetry indices (Yan et al., 2017a), which assess the balance of accelerations and decelerations in heart rate dynamics. Building on the rapid development of non-linear HRV methods, recent approaches have extended non-linear HRV analysis into dynamic, multimodal domains—such as the time-varying indices of cardiac-autonomic coupling proposed by Candia-Rivera (2023)—highlighting emerging interest in embedding HRV within broader network physiology mechanisms.

Table 3

A summary of HRV non-linear indices.

Sub Domain	Index	Description
Poincaré Plot (Brennan et al., 2002; Denton et al., 1990)	SD1	Poincaré plot standard deviation perpendicular the line of identity, describing short-term variability of HR
	SD2	Poincaré plot standard deviation along the line of identity, describing short- and long-term variability of HR
	SD1/SD2	Ratio of SD1 to SD2
	S	Area of ellipse described by SD1 and SD2
	CSI	Cardiac Sympathetic Index (Toichi et al., 1997)
	CVI	Cardiac Vagal Index (Toichi et al., 1997)
	CSI_Modified	Modified CSI (Jeppesen et al., 2015)
Heart Rate Asymmetry (Piskorski & Guzik, 2011; Yan et al., 2017b)	GI	Guzik's Index, distance of points above line of identity (LI) to LI divided by that of all points in Poincaré plot to LI
	SI	Slope Index, phase angle of points above LI divided by that of all points in Poincaré plot
	AI	Area Index, cumulative area of the sectors of points above LI divided by that of all points in the Poincaré plot
	PI	Porta's Index, number of points below LI divided by the total number of points in Poincaré plot
	SD1a	Contributions of acceleration of HR (shortening of NNIs) to SD1
	SD1d	Contributions of deceleration of HR (prolongations of NNIs) to SD1
	SD2a	Contributions of acceleration to SD2
	SD2d	Contributions of deceleration to SD2
	SDNNa	Contributions of accelerations to SDNN
	SDNNd	Contributions of decelerations to SDNN
	C1a	SD1a divided by SD1 (normalized)
	C1d	SD1d divided by SD1 (normalized)

	C2a	SD2a divided by SD2 (normalized)
	C2d	SD2d divided by SD2 (normalized)
	Ca	Contribution of acceleration to total variability
	Cd	Contribution of deceleration to total variability
Heart Rate Fragmentation (Costa et al., 2017a)	PIP	Percentage of inflection points of NNIs
	IALS	Inverse of average lengths of acceleration segment divided by deceleration segment
	PSS	Percentage of short segments
	PAS	Percentage of NNIs in alternation segments
Entropy (Lau et al., 2022b)	ApEn	Approximate Entropy
	SampEn	Sample Entropy
	FuzzyEn	Fuzzy Entropy
	ShanEn	Shannon Entropy
	MSEn	Multiscale Entropy (Wu et al., 2013)
	CMSEn	Composite Multiscale Entropy (Wu et al., 2013)
	RCMSEn	Refined CMSEn (Wu et al., 2013)
Fractal	CD	Correlation Dimension or D2 (Bolea et al., 2014)
	HFD	Higuchi's Fractal Dimension (Higuchi, 1988)
	KFD	Katz's Fractal Dimension (Katz, 1988)
	LZC	Lempelziv Complexity (Lempel & Ziv, 1976)
	DFA_α1	Monofractal Detrended Fluctuation Analysis (DFA), corresponding to short-term correlations (Faini et al., 2021)
	DFA_α2	DFA, corresponding to long-term correlations
	MDFA_α1	Multifractal Detrended Fluctuation Analysis (MDFA), corresponding to short-term correlations. MDFA indices describe features of the multifractal singularity spectrum, which is usually in the shape of an inverted parabola (Makowski et al., 2022), to quantify the correlation property of a signal. These features include *Fluctuation*–amount of fluctuation in signals, *Mean*–average fluctuation, *Maximum* – maximum fluctuation, *Width*–degree of multifractality, *Peak*–self-affinity level, *Delta*–range of fluctuations, *Asymmetry*–symmetry of fluctuation, *Increment*–robust index of distribution.
	MDFA_α2	MDFA, corresponding to long-term correlations

Lastly, several methods have been developed to calculate RSA indices (**Table 4**). The Peak-to-Trough (P2T) method (Grossman, 1992) calculates RSA by measuring the difference between the highest (peak) and lowest (trough) heart rate within each breathing cycle. This method is straightforward and directly captures the extent of heart rate variability associated with respiration. The Porges-Bohrer (PB) method (Porges, 1985) is another commonly used approach that involves the use of spectral analysis to assess RSA. This method focuses on

isolating the frequency of spontaneous respiration to quantify the magnitude of RSA more precisely. The Gates method (Gates et al., 2015) is the most recent RSA index to be developed. It offers a more sophisticated approach by applying a peak detection algorithm to identify the most significant heart rate changes associated with inhalation and exhalation. While the P2T and PB methods assume that the RSA index is non-varying across the recording period, the Gates method provides a more dynamic (i.e., time-varying) estimate of RSA. Lastly, given the frequency of spontaneous breathing is close to the frequency-domain index of *HF* (Table 2), the *HF* component has also been frequently used in past studies as an RSA index (Grossman et al., 1990).

Table 4

A summary of HRV indices of respiratory sinus arrhythmia (RSA).

Sub Domain	Index	Description
Peak-to-Trough (P2T) method (Grossman, 1992)	RSA Mean (P2T)	Mean of RSA across breath cycles
	RSA Mean Log (P2T)	Logarithm of RSA Mean (P2T)
	RSA SD (P2T)	Standard Deviation of RSA (P2T)
Porges-Bohrer (PB) method (Porges, 1985)	RSA (PB)	PB estimate of RSA
Gates method (Gates et al., 2015)	RSA Mean (Gates)	Mean of RSA across breath cycles, calculated with Gates method
	RSA Mean Log (Gates)	Logarithm of RSA Mean (Gates)
	RSA SD (Gates)	Standard Deviation of RSA (Gates)

Although the same HRV index may be computed across different durations using identical formulas, its physiological meaning can differ substantially depending on the timescale of measurement. Short-term (typically 5-minute) and ultra-short-term (< 5-minute) recordings are widely used in psychophysiological research due to their feasibility and ability to capture meaningful autonomic dynamics. These short-term recordings are particularly sensitive to parasympathetically mediated influences—such as RSA and baroreflex activity—which operate on rapid timescales, ranging from sub-seconds to a few seconds (Eckberg & Sleight, 1992; Feher, 2012; Nunan et al., 2010). Past studies have found indices such as RMSSD, pNN50, and HF power to be especially robust in these contexts, so these indices are often recommended for short-term assessments (Shaffer et al., 2014; Laborde et al., 2017).

While short-term recordings can also reflect broader autonomic interactions, they are less sensitive to slower-acting sympathetic contributions, which typically emerge over longer latencies (e.g., 2–5 seconds or more) and are more reliably quantified in longer recordings (Berntson et al., 1997; Billman, 2011; Laborde et al., 2017; Task Force, 1996). In particular, LF power, VLF components, and derived ratios such as LF/HF require recordings of at least 5 minutes—and often much longer—to be estimated reliably and interpreted meaningfully (Berntson et al., 1997; Task Force, 1996). These slower components are thought to reflect a broader set of physiological processes, including thermoregulation, circadian variation, and

hormonal regulation, which become more prominent in long-term (e.g., 24-hour) recordings (Shaffer & Ginsberg, 2017). Since the physiological interpretation of HRV indices is inherently dependent on the duration of ECG recordings, selecting indices appropriate to the recording duration is critical not only for ensuring measurement validity but also for avoiding misinterpretation of the underlying autonomic dynamics.

Problems Caused by the Proliferation of HRV Indices

Despite their widespread use, the relationships among the existing HRV indices, the specific mechanisms underlying HRV, and pathways linking HRV to the associated mental and physical health conditions are not yet fully understood. This limited understanding, along with the continued addition of new indices, poses challenges to research progress and may exacerbate issues with replicability. For instance, although these indices are often used interchangeably to describe HRV, evidence suggests that they may not represent the same processes (Lewis et al., 2012). As new methods to quantify HRV continue to be added to the existing pool, it is not uncommon for different studies to rely on different sets of indices to examine the same phenomenon (e.g., psychopathology, Chalmers et al., 2014; cognitive functioning, Forte et al., 2019; self-control, Zahn et al., 2016). This variability between studies increases noise in the literature complicates the interpretation and synthesis of findings and limits the generalizability of results—particularly when indices yield conflicting results for the same outcome.

The overlap in statistical properties across HRV indices poses additional pragmatic issues. Early structural studies have demonstrated strong correlations between certain indices. For example, *RMSSD* and *pNN50* correlate above .9, and also show strong correlations with *HF* power (Bigger et al., 1989; Kleiger et al., 2005; Otzenberger et al., 1998). Ciccone et al. (2017) highlighted that *RMSSD* and *SD1* are mathematically equivalent, and studies that report both of these HRV indices in short-term recordings often independently arrive at identical statistical results (e.g., Leite et al., 2015; Peng et al., 2015; Rossi et al., 2015). Other studies have also noted similarities between *SD1/SD2* and *LF/HF*, noting these pairs of indices reflect shared physiology in the balance between short- and long-term HRV (Guzik et al., 2007; Hoshi et al., 2013). When these similarities are not taken into account in analyses, they can lead to statistical issues such as inflated confidence in the results when multiple indices replicate a given trend; collinearity and suppression of effects if multiple correlated indices are used as simultaneous predictors; potential over-correction for multiple comparisons; and needlessly complex and difficult to interpret patterns of results (Dormann et al., 2013; Mela & Kopalle, 2002; Næs & Mevik, 2001). Further, the overwhelming number of options poses a practical challenge to researchers in identifying which indices to use in their HRV research. In studies without a pre-registered data analysis plan, researchers may be inclined to selectively report indices that support their hypotheses (Andrade, 2021). Overall, the proliferation of HRV indices without a good understanding of their interrelationships creates challenges for standardization, complicates comparisons across studies, and may obscure meaningful research on the link between HRV and health outcomes.

The Present Study

The present study aims to quantify the relationships among HRV indices by employing a consensus-based clustering approach to identify patterns of association across a comprehensive set of indices derived from short-term (3–8 minute) resting-state recordings. Our primary goal is to empirically identify reliable clusters of statistically similar HRV indices and describe the statistical and physiological characteristics underlying these groupings to support more systematic selection and interpretation of HRV metrics in research.

To ensure the robustness and generalizability of the clustering solution, we systematically examined its stability across multiple dimensions: clustering method, time, recording procedure, and sample context. First, we applied a consensus-based approach that integrated results from two clustering algorithms within a single sample. We then assessed test-retest reliability by determining whether the consensus cluster structure was preserved across two sessions conducted one week apart in the same individuals. To evaluate generalizability across methodological and population differences, we replicated the consensus-based clustering procedure in a second sample that differed from the first in both recording parameters (e.g., device, duration, sampling rate) and psychological profile. Specifically, the first sample comprised individuals with elevated internalizing symptoms, whereas the second sample consisted of participants recruited from the community (i.e., community sample). The choice of these datasets was intended to capture the range of psychological presentations commonly encountered in psychophysiological research and to support broader generalization of results. We then compared the clustering solutions between the two samples to assess convergence in the resulting structures. The final result is based on the clusters of HRV indices that showed cross-method, test-retest, and cross-sample robustness.

Methods

Participants and Procedure

Data for this study were drawn from two samples who completed resting-state physiological assessments. Participants in **Sample 1** were selected based on elevated public speaking anxiety, as part of a broader clinical study (Johnco et al., 2025). Eligibility required elevated levels of public speaking anxiety, scoring ≥5 on two 9-point Likert scale screening items for *fear* (“How anxious would you feel giving a formal speech before a live audience?”) and *avoidance* (“How likely is it that you would avoid taking a (elective) class that required an oral presentation?”) *of public speaking*. Although the selection focused on social anxiety, assessment data indicated that participants also exhibited elevated symptoms across broader internalizing domains, including generalized anxiety and depression. Participants in **Sample 2** were not selected based on any psychological criteria. No formal screening or inclusion criteria related to internalizing symptoms was used in this group; as such, Sample 2 is treated as a community sample with a lower likelihood of elevated internalizing symptoms than Sample 1.

Sample 1

Participants in Sample 1 completed the resting-state assessment in two sessions, at baseline (T1) and follow-up (T2) a week later. Sample 1 (T1) included 233 adults (74.2% female) aged 18-59 (M_{age} = 21.04, SD = 6.73). As some participants did not return for the follow-up session, Sample 1 (T2) included 199 adults (74.9% female) aged 18-59 (M_{age} = 21.40, SD = 7.32). All Sample 1 participants were recruited from a first-year undergraduate psychology course in Sydney, Australia, or via flyers on the university campus and social media advertisements. Participants recruited from undergraduate university courses received course credit, and participants recruited from other sources were eligible to receive AUD$10 for their participation. All participants gave their written consent, and the research protocol was approved by the Macquarie University Human Research Ethics Committee.

At both baseline (T1) and follow-up (T2) sessions, Sample 1 participants completed a 3-minute resting state (eyes opened) procedure, during which they were asked to sit comfortably and breathe spontaneously. Resting-state electrocardiogram (ECG) was sampled at 265Hz and recorded using the Equivital two-lead sensor belt and the LabChart Software (ADI, Colorado Springs, CO, USA). Participants in Sample 1 also completed a battery of questionnaires at baseline, which assessed for public speaking anxiety, social anxiety, generalized anxiety, and depression.

Personal Report of Public Speaking Anxiety (McCroskey, 1970). This 34-item self-report measures public speaking anxiety using a 5-point Likert scale (1 = *Strongly Disagree* to 5 = *Strongly Agree*), with higher scores indicating greater public speaking anxiety (> 98 indicating moderate and >131 indicating severe levels). Internal consistency was excellent in this sample (α = .91). On average, Sample 1 participants scored above the severe level threshold for public speaking anxiety. **Table 5** summarizes the sample characteristics, including their scores on the battery of clinical measures.

Table 5

Sample Characteristics

	Sample 1 (T1) (N=233)	**Sample 1 (T2) (N=199)**	**Sample 2 (N=203)**		
			Makowski et al. (2023)	Makowski et al. (2024)	Makowski & Neves (2024)
N	233	199	44	101	58
Sex					
Female (%)	173 (74.2)	149 (74.9)	21 (47.7)	52 (51.5)	41 (70.7)
Male (%)	60 (25.6)	50 (25.2)	23 (52.3)	49 (48.5)	17 (29.3)
Age (years)					
Mean (SD)	21.04 (6.73)	21.40 (7.32)	25.18 (4.92)	26.25 (3.71)	22.98 (6.15)
Psychological assessment					

PRPSA	135.11 (14.56)[##]	135.16 (14.20)[##]	-	-	-
SPS	38.05 (15.75)[^]	37.66 (15.92)[^]	-	-	-
SIAS	44.77 (15.51)[^]	44.25 (15.69)[^]	-	-	-
GAD-7	10.03 (5.35)[^]	9.92 (5.43)[^]	-	-	-
PHQ-9	11.02 (5.47)[^]	11.05 (5.42)[^]	-	-	-

[^]above clinical threshold, [##]above severe level of severity. PRPSA = Personal Report of Public Speaking Anxiety, SPS = Social Phobia Scale, SIAS = Social Interaction Anxiety Scale, GAD-7 = 7-item Generalized Anxiety Disorder Scale, PHQ-9 = 9-item Patient Health Questionnaire.

Social Interaction Anxiety Scale (SIAS) and Social Phobia Scale (SPS) (Mattick & Clarke, 1998). The SIAS is a 20-item measure of anxiety during social interactions, and the SPS is a 20-item measure of fear of being scrutinized during routine activities. Both measures are rated on a 5-point Likert scale (0 = *Not at all characteristic or true of me* to 4 = *Extremely characteristic or true of me*), with higher scores indicating greater anxiety severity. Scores above 36 on the SIAS and 26 on the SPS are indicative of social phobia (Peters, 2000). Internal consistency was excellent for the SIAS and SPS in this sample (α = .93 and .92). On average, Sample 1 participants scored above the clinical cut-off for social phobia.

Generalized Anxiety Scale (GAD-7; Spitzer et al., 2006). This 7-item self-report measures the severity of general anxiety symptoms using a 4-point Likert scale (0 = *Not at all* to 3 = *Nearly Every Day*). Scores above 7 are indicative of a probable anxiety disorder diagnosis (Plummer et al., 2016). Internal consistency was good in this sample (α = .89). On average, Sample 1 participants scored above the clinical cut-off for anxiety.

Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001). This 9-item self-report measures the severity of depression symptoms using a 4-point Likert scale (0 = *Not at all* to 3 = *Nearly Every Day*). Scores above 10 are indicative of probable Major Depressive Disorder (Kroenke et al., 2001). Internal consistency was good in this sample (α = .85). On average, Sample 1 participants scored above the clinical cut-off for depression.

Sample 2

Sample 2 included 203 adults (55.2% female) aged 18-50 (M_{age} = 25.10, SD = 5.24). The data for Sample 2 is pooled from three separate studies that utilized a shared resting-state procedure. Specifically, Makowski et al. (2023) included 44 participants (M_{age} = 25.18, SD = 4.92; 47.7% female), Makowski et al. (2024) included 101 participants (M_{age} = 26.25, SD = 3.71; 51.5% female), and Makowski & Neves (2024) included 58 participants (M_{age} = 22.98, SD = 6.15; 70.7% female). All participants gave their written consent. Both Makowski et al. (2023) and Makowski et al. (2024) recruited participants from the general university student population in Singapore, and the research protocol was approved by the Nanyang Technological University Institution Review Board. Makowski & Neves (2024) recruited participants from the general university student population in Sussex, United Kingdom, and the research protocol was approved by the University of Sussex Research Ethics Committee.

Participants in Sample 2 completed an 8-minute resting state (eyes closed) procedure during which they were asked to sit comfortably and breathe spontaneously. In Makowski et al. (2023), resting-state ECG was sampled at 4000Hz and recorded using three Ag/AgCl electrodes following chest-mounted configuration (i.e., one electrode under each clavicle and the third on the lower left rib cage). The signals were recorded via the BioPac MP160 system and the AcqKnowledge 5.0 software (BioPac Systems Inc., USA). In Makowski & Neves (2024) and Makowski et al. (2024), resting-state ECG was sampled at 1000Hz and recorded using the same three Ag/AgCl electrode configurations via the PLUX OpenSignals software and BITalino Toolkit (PLUX Biosignals, Portugal). The ECG data for all participants in Sample 2 was made available on the OpenNeuro open-access platform (Makowski et al., 2023, 2024; Makowski & Neves, 2024).

Data Analysis

Figure 1 illustrates the data analysis pipeline. Prior to the ECG processing step, 54 recordings were removed as the R-peaks could not be reliably located due to excessive levels of noise present. Signal noises were possibly caused by physical movements during data collection or improper placement of acquisition hardware. Specifically, 16 ECG recordings (approximately 6%) were removed from Sample 1 (T1), 36 recordings (approximately 15%) were removed from Sample 1 (T2), and two recordings were removed from Sample 2 (less than 1%).

ECG Processing

All ECG recordings were processed uniformly using the NeuroKit2 toolbox (Makowski et al., 2021). Prior to analysis, recordings from both samples were re-sampled to 1000Hz to ensure consistency in temporal resolution. Signal preprocessing involved the application of a 0.5Hz high-pass Butterworth filter (5^{th} order) to remove baseline wander, followed by a notch filter of 50Hz were applied to remove powerline interference. QRS complexes were detected based on the steepness of the absolute gradient of the ECG signal, followed by the localization of R-peaks as local maxima within the QRS complexes (Brammer, 2020). All detected R-peaks were visually inspected for false detections or missed beats; ectopic beats and artifacts were identified and corrected using an automatic artifact-detection algorithm that employs time-varying thresholds derived from the distribution of successive interval differences, in combination with a novel beat classification scheme designed to distinguish physiologically implausible intervals (Lipponen & Tarvainen, 2019).

The default Neurokit2 setting was subsequently used to compute three datasets of 89 HRV indices (see **Tables 1-4** for the description of all HRV indices). For Fourier-based spectral analysis, interbeat interval series were interpolated using a cubic spline method at the default rate of 100 Hz. The resulting signals were mean-centered (constant detrending), segmented with 50% overlapping windows, and tapered with a Hann window. Power spectral density was estimated using Welch's periodogram method. To ensure sufficient frequency resolution, the window length was dynamically determined such that each segment included at least two full cycles of the lowest frequency of interest within a given spectral band. Note that the four low-frequency components (i.e., *ULF*, *VLF*, *LF,* and *LF/HF* indices) were not computed

because reliable estimation of these indices requires substantially longer recording durations (≥5 minutes for LF; ≥24 hours for ULF) than the 3-minute ECG segments used in Sample 1 (Nussinovitch et al., 2011; Shaffer et al., 2014). Thus, these four indices were not computed in any dataset and were not included in the cluster analysis.

Prior to the consensus-based clustering procedure, we removed the outliers on each index in each dataset; outliers were determined as values with absolute deviation from the median that is three times or greater than the MAD value (median absolute deviation; Leys et al., 2019). On average, 3.3% of data from the Sample 1 (T1), 3.5% of data from the Sample 1 (T2), and 2.6% of data from the Sample 2 were detected as outliers and removed.

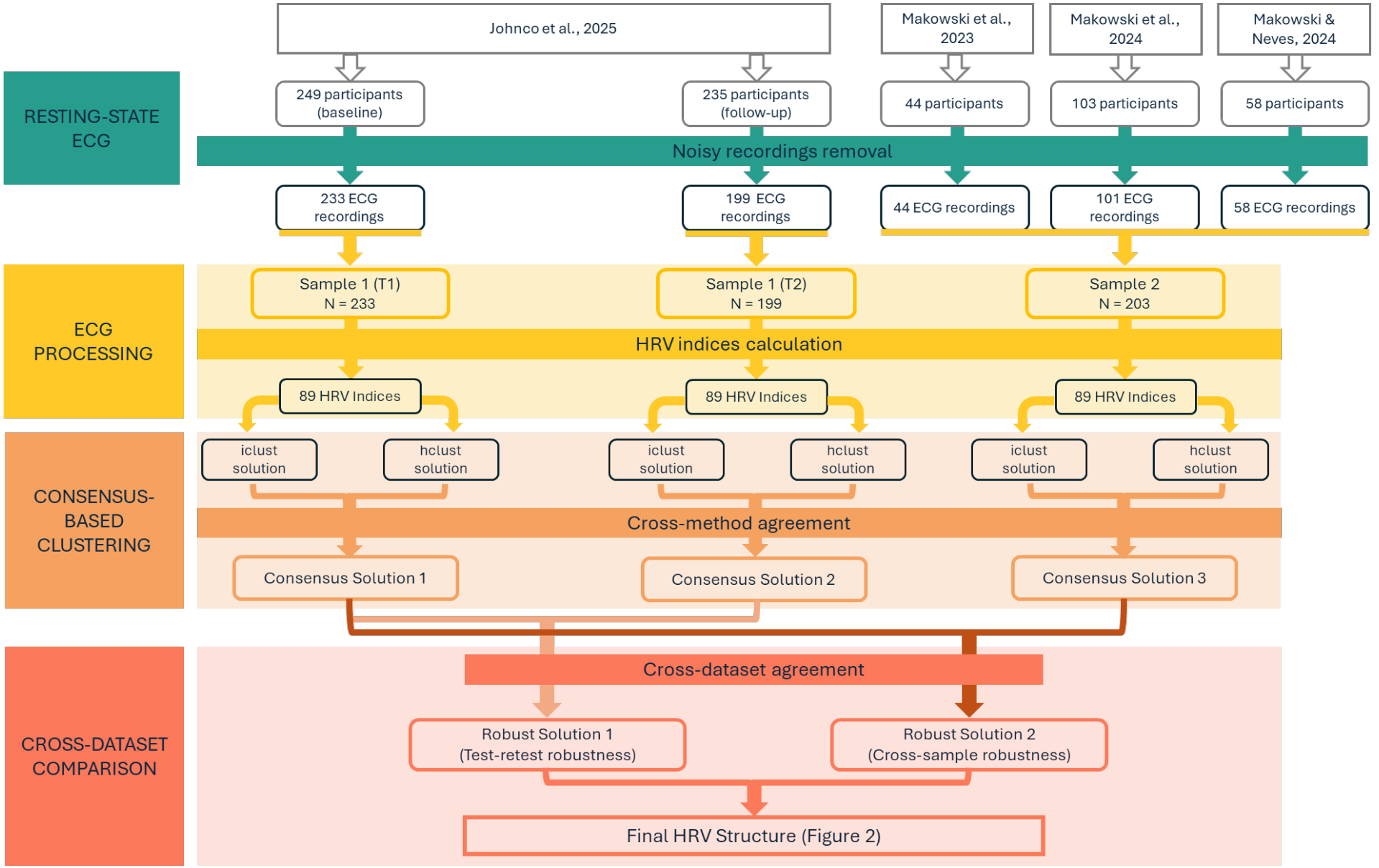

Figure 1. Data analysis pipeline. First, noisy electrocardiogram (ECG) recordings were removed. Second, clean ECG recordings were processed to calculate HRV indices. Third, in consensus-based clustering, consensus solutions were identified to include indices assigned to the same clusters by both clustering methods (iclust = item clustering; hclust = hierarchical clustering). Fourth, in cross-dataset comparisons, robust solutions were identified based on indices assigned to the same clusters in both datasets. Final HRV structure summarizes clusters with cross-method, test-retest and cross-sample robustness.

Consensus-based Clustering

There is no gold standard or clear guidelines to determine the most appropriate method for grouping these physiological indices. As such, choosing one method and presenting its solution as definitive can be misleading. To avoid identifying clusters idiosyncratic to a specific method, we adapted the consensus-based clustering procedure in Forbes et al. (2021), using two different clustering algorithms: item clustering (*iclust*; Revelle, 1978) and hierarchical agglomerative clustering (*hclust*; Ward, 1963), to identify clusters of closely associated HRV indices. Data analysis was conducted in R 4.2.2 (R Core Team, 2021).

To identify *iclust* solutions, the *iclust* function from *psych* package (Revelle, 2024) was used. The *iclust* algorithm works iteratively based the Spearman's correlation matrix of all indices; it begins with the strongest pair of correlated items or indices and gradually adds more indices to form clusters. The criteria for clusters to form include two tests of internal consistency. The first is the increase-in-coefficient-alpha criterion, which requires the Cronbach's alpha (Cronbach, 1951) of the new cluster to be greater than that of its two components (i.e., pair of indices or pair of clusters). The second is the increase-in-coefficient-beta criterion which requires the coefficient beta (Revelle, 1979) of the new cluster to be greater than the average beta of the two components. In this study, the criteria for clusters to form are set to be highly conservative; both criteria must be met, and the coefficient beta is at least .8 (i.e., at least 80% of the variance in the components was associated with a shared general factor).

To identify the *hclust* solutions, the *hclust* R function (R Core Team, 2021) was used. The *hclust* algorithm also works iteratively but based on a dissimilarity matrix to progressively merge the two most similar indices or clusters (agglomerative clustering). Unlike *iclust*, which focuses on maximizing internal consistency and stops clustering when the criteria are not met, *hclust* focuses on the overall structure of the data and how indices are related at different levels of similarity or distance (i.e., hierarchical level). Following the procedure in Forbes et al. (2021), the chosen stopping criterion for the *hclust* solutions is at the last unitary cluster, which means that there is only one cluster in the final *hclust* solutions consisting of one HRV index (single-item cluster).

Cross-method agreement: HRV indices assigned to the same cluster in both *iclust* and *hclust* solutions were identified as receiving cross-method agreement and assigned to the same cluster in the consensus solutions. As illustrated in **Figure 1**, we performed the consensus-based clustering procedure separately on Sample 1 (T1), Sample 1 (T2), and Sample 2 datasets to quantify a consensus solution for each dataset, respectively. Each dataset was clustered independently to ensure that any convergence in cluster structure across samples or time points reflected true structural similarity rather than methodological artifacts. To evaluate the degree of cross-method agreement, we reported the percentage of indices being assigned to the same clusters.

Cross-dataset Comparisons

To investigate the test-retest robustness of the HRV clusters, the cluster solutions from Sample 1 (T1) and Sample 1 (T2) were compared to identify the clusters of indices that were

consistent from baseline to follow-up. Subsequently, to examine the generalizability of the HRV clusters (i.e., cross-sample robustness), the cluster solutions between Sample 1 (T1) and Sample 2 were compared. Since the resting-state data in these two datasets was acquired with different recording procedures and in samples with different psychiatric presentations, the comparison enabled us to assess whether the HRV clusters are robust to variations in methodology and sample characteristics. In this way, the sample and method heterogeneity across datasets provided a more stringent test of the robustness and generalizability of the identified HRV clusters.

Cross-dataset agreement: In each cross-dataset comparison, HRV indices assigned to the same cluster by both clustering methods in both datasets were identified as receiving cross-dataset agreement and assigned to the same cluster in the robust solution. As illustrated in **Figure 1**, only clusters with indices that showed both test-retest and cross-sample robustness were summarized in the final HRV structure.

Results

Demographic Characteristics

As compared to Sample 2, there was a significantly higher proportion of female participants in Sample 1 (T1) ($\chi^2(1) = 14.5$, $p < .001$; Cramer's $V = 0.18$), and participants in Sample 1 were also significantly younger ($t(429.81) = 6.99$, $p < .001$; Cohen's $d = 0.67$, 95% *CI* [0.47, 0.86]).

Consensus-Based Clustering

In the consensus solution of Sample 1 (T1), the *iclust* method identified 15 clusters with cluster sizes ranging from 3 to 23 indices. The *hclust* method identified 20 clusters with cluster sizes ranging from 1 to 13 indices. In the consensus solution, 87 out of 89 (97.8%) indices were assigned to the same clusters by both methods (i.e., cross-method agreement). Two indices, *LZC* and *TINN*, were assigned to different clusters by the two solutions. **Table 1** in **Supplementary Materials** details the iclust and hclust solutions for Sample 1 (T1).

In the consensus solution of Sample 1 (T2), the *iclust* method identified 20 clusters with cluster sizes ranging from 1 to 26 indices. The *hclust* method identified 20 clusters with cluster sizes ranging from 1 to 16 indices. In the consensus solution, 82 out of 89 (92.1%) indices were assigned to the same clusters by both methods. Seven indices, four indices of *MFDFA* features (*MFDFA α2 Maximum, MFDFA α2 Fluctuation., MFDFA α1 Peak, MFDFA α1 Mean*), *SDANN1*, *PAS,* and *TINN*, were assigned to different clusters by the two solutions. **Table 2** in **Supplementary Materials** details the iclust and hclust solutions for Sample 1 (T2)**.**

In the consensus solution of Sample 2, the *iclust* method identified 16 clusters with cluster sizes ranging from 1 to 25 indices. The *hclust* method identified 19 clusters with cluster sizes ranging from 1 to 13 indices. In the consensus solution, 85 out of 89 (95.5%) indices were assigned to the same clusters by both methods. Four indices, *MSEn*, *CMSEn*, *RCMSEn*, and *TINN*, were assigned to different clusters by the two solutions. **Table 3** in **Supplementary Materials** details the iclust and hclust solutions for Sample 2**.**

Cross-dataset Comparison

In the first robust solution, which summarizes the convergence of Sample 1 (T1) and Sample 1 (T2) consensus solutions, 80 out of 89 HRV indices were assigned to the same clusters in both solutions (i.e., cross-dataset agreement), showing test-retest robustness. Nine indices did not show test-retest robustness were *LZC*, *TINN, MFDFA α2 Maximum, MFDFA α1 Mean*, *MFDFA α2 Fluctuation, MFDFA α1 Peak, SDANN1*, *PAS,* and *CSI (Modified).*

In the second robust solution, which summarizes the convergence of Sample 1 (T1) and Sample 2 consensus solutions, 72 out of 89 HRV indices were assigned to the same clusters in both solutions, showing cross-sample robustness. 17 indices did not show cross-sample robustness were *LZC, TINN*, *MFDFA α2 Maximum*, *MFDFA α1 Mean, SDANN1*, *PAS, CSI (Modified), CVNN, MCVNN, ShanEn, S, CVI, MSEn, CMSEn, RCMSEn, RSA SD (P2T),* and *RSA SD (Gates)*. Ten of the indices that showed test-retest robustness–*CVNN, MCVNN, ShanEn, S, CVI, RSA SD (P2T), RSA SD (Gates), MSEn, CMSEn, RCMSEn*–did not reliably converge in the comparison of consensus solutions between different samples (i.e., failed to show cross-sample robustness). **Table 6** summarizes the cluster memberships of 89 HRV indices across three consensus solutions and two robust solutions.

Table 6.

Summary of cluster memberships for each HRV index in three consensus solutions and two robust solutions.

Indices	Consensus Solution Sample 1 (T1)	Consensus Solution Sample 1 (T2)	Consensus Solution Sample 2	Robust Solution 1 (Sample 1 T1 – T2)	Robust Solution 2 (Sample 1 – Sample 2)
MeanNN	I1_H1	I1_H1	I1_H1	Cluster A	Cluster 1
MedianNN	I1_H1	I1_H1	I1_H1	Cluster A	Cluster 1
Prc20NN	I1_H1	I1_H1	I1_H1	Cluster A	Cluster 1
Prc80NN	I1_H1	I1_H1	I1_H1	Cluster A	Cluster 1
MinNN	I1_H1	I1_H1	I1_H1	Cluster A	Cluster 1
SDNN	I2_H2	I2_H2	I2_H2	Cluster B	Cluster 2
SDNNI1	I2_H2	I2_H2	I2_H2	Cluster B	Cluster 2
SD2	I2_H2	I2_H2	I2_H2	Cluster B	Cluster 2
SD2d	I2_H2	I2_H2	I2_H2	Cluster B	Cluster 2
SD2a	I2_H2	I2_H2	I2_H2	Cluster B	Cluster 2
SDNNd	I2_H2	I2_H2	I2_H2	Cluster B	Cluster 2
SDNNa	I2_H2	I2_H2	I2_H2	Cluster B	Cluster 2
CVNN	I2_H2	I2_H2	**I3_H3**	Cluster B	-
MadNN	I2_H3	I2_H3	I2_H2	Cluster C	Cluster 3
IQRNN	I2_H3	I2_H3	I2_H2	Cluster C	Cluster 3
HTI	I2_H3	I2_H3	I2_H2	Cluster C	Cluster 3
MCVNN	I2_H3	I2_H3	**I3_H3**	Cluster C	Cluster 4

Indices	Consensus Solution Sample 1 (T1)	Consensus Solution Sample 1 (T2)	Consensus Solution Sample 2	Robust Solution 1 (Sample 1 T1 – T2)	Robust Solution 2 (Sample 1 – Sample 2)
ShanEn	I2_H3	I2_H3	**I3_H3**	Cluster C	Cluster 4
RMSSD	I3_H2	I3_H2	I4_H2	Cluster D	Cluster 5
SDSD	I3_H2	I3_H2	I4_H2	Cluster D	Cluster 5
CVSD	I3_H2	I3_H2	I4_H2	Cluster D	Cluster 5
pNN50	I3_H2	I3_H2	I4_H2	Cluster D	Cluster 5
pNN20	I3_H2	I3_H2	I4_H2	Cluster D	Cluster 5
SD1	I3_H2	I3_H2	I4_H2	Cluster D	Cluster 5
SD1d	I3_H2	I3_H2	I4_H2	Cluster D	Cluster 5
SD1a	I3_H2	I3_H2	I4_H2	Cluster D	Cluster 5
S	I3_H2	I3_H2	**I2_H2**	Cluster D	Cluster 6
CVI	I3_H2	I3_H2	**I2_H2**	Cluster D	Cluster 6
MaxNN	I3_H1	I3_H2	I1_H1	Cluster E	Cluster 7
RSA Mean (Gates)	I3_H1	I3_H2	I1_H1	Cluster E	Cluster 7
RSA Mean Log (Gates)	I3_H1	I3_H2	I1_H1	Cluster E	Cluster 7
RSA Mean (P2T)	I4_H2	I4_H2	I5_H2	Cluster F	Cluster 8
RSA Mean Log (P2T)	I4_H2	I4_H2	I5_H2	Cluster F	Cluster 8
RSA (PB)	I4_H2	I4_H2	I5_H2	Cluster F	Cluster 8
RSA SD (P2T)	I4_H2	I2_H2	**I2_H2**	Cluster G	-
RSA SD (Gates)	I4_H2	I2_H2	**I3_H3**	Cluster G	-

Indices	Consensus Solution Sample 1 (T1)	Consensus Solution Sample 1 (T2)	Consensus Solution Sample 2	Robust Solution 1 (Sample 1 T1 – T2)	Robust Solution 2 (Sample 1 – Sample 2)
SDRMSSD	I5_H4	I5_H4	I6_H4	Cluster H	Cluster 9
SD1SD2	I5_H4	I5_H4	I6_ H4	Cluster H	Cluster 9
CSI	I5_H4	I5_H4	I6_ H4	Cluster H	Cluster 9
HF	I6_H5	I6_H5	I7_H5	Cluster I	Cluster 10
VHF	I6_H5	I6_H5	I7_ H5	Cluster I	Cluster 10
TP	I6_H5	I6_H5	I7_ H5	Cluster I	Cluster 10
LnHF	I6_H5	I6_H5	I7_ H5	Cluster I	Cluster 10
HFn	I7_H4	I7_H4	I6_ H4	Cluster J	Cluster 11
DFA α1	I7_H4	I7_H4	I6_ H4	Cluster J	Cluster 11
HFD	I7_H4	I7_H4	I6_ H4	Cluster J	Cluster 11
MFDFA α1 Peak	I7_H4	**I7_H19**	I6_ H4	-	Cluster 11
PIP	I8_H6	I8_H6	I8_H6	Cluster K	Cluster 12
IALS	I8_H6	I8_H6	I8_ H6	Cluster K	Cluster 12
PSS	I8_H6	I8_H6	I8_ H6	Cluster K	Cluster 12
GI	I9_H7	I9_H7	I9_H7	Cluster L	Cluster 13
SI	I9_H7	I9_H7	I9_ H7	Cluster L	Cluster 13
AI	I9_H7	I9_H7	I9_ H7	Cluster L	Cluster 13
PI	I10_H8	I10_H8	I10_H8	Cluster M	Cluster 14
C1d	I10_H8	I10_H8	I10_ H8	Cluster M	Cluster 14

Indices	Consensus Solution Sample 1 (T1)	Consensus Solution Sample 1 (T2)	Consensus Solution Sample 2	Robust Solution 1 (Sample 1 T1 – T2)	Robust Solution 2 (Sample 1 – Sample 2)
C1a	I10_H8	I10_H8	I10_ H8	Cluster M	Cluster 14
C2d	I11_H8	I11_H9	I10_ H8	Cluster N	Cluster 15
C2a	I11_H8	I11_H9	I10_ H8	Cluster N	Cluster 15
Cd	I11_H8	I11_H9	I10_ H8	Cluster N	Cluster 15
Ca	I11_H8	I11_H9	I10_ H8	Cluster N	Cluster 15
MFDFA α1Width	I12_H9	I12_H10	I11_H9	Cluster O	Cluster 16
MFDFA α1Fluctuation	I12_H9	I12_H10	I11_H9	Cluster O	Cluster 16
MFDFA α1 Increment	I12_H9	I12_H10	I11_H9	Cluster O	Cluster 16
MFDFA α1 Maximum	I13_H10	I13_H11	I12_H10	Cluster P	Cluster 17
MFDFA α1 Delta	I13_H10	I13_H11	I12_H10	Cluster P	Cluster 17
MFDFA α1 Asymmetry[a]	I13_H10	I13_H11	I12_H10	Cluster P	Cluster 17
DFA α 2	I14_H11	I14_H12	I13_H11	Cluster Q	Cluster 18
MFDFA α2 Peak	I14_H11	I14_H12	I13_H11	Cluster Q	Cluster 18
MFDFA α2 Mean	I14_H11	I14_H12	I13_H11	Cluster Q	Cluster 18
MFDFA α2 Width	I15_H12	I15_H13	I14_H12	Cluster R	Cluster 19
MFDFA α2 Increment	I15_H12	I15_H13	I14_H12	Cluster R	Cluster 19
MFDFA α2 Fluctuation	I15_H12	**I20_H20**	I14_H12	-	Cluster 19
MFDFA α2 Delta	I16_H13	I16_H14	I15_H13	Cluster S	Cluster 20
MFDFA α2 Asymmetry	I16_H13	I16_H14	I15_H13	Cluster S	Cluster 20

Indices	Consensus Solution Sample 1 (T1)	Consensus Solution Sample 1 (T2)	Consensus Solution Sample 2	Robust Solution 1 (Sample 1 T1 – T2)	Robust Solution 2 (Sample 1 – Sample 2)
ApEn[a]	I17_H14	I17_H15	I6_ H4	Cluster T	Cluster 21
SampEn	I17_H14	I17_H15	I6_ H4	Cluster T	Cluster 21
FuzzyEn[a]	I17_H14	I17_H15	I6_ H4	Cluster T	Cluster 21
CD	I17_H14	I17_H15	I6_ H4	Cluster T	Cluster 21
KFD	I17_H14	I17_H15	I6_ H4	Cluster T	Cluster 21
MSEn	I18_H15	I18_H16	**I16_H14**	Cluster U	-
CMSEn	I18_H15	I18_H16	**I17_H15**	Cluster U	-
RCMSEn	I18_H15	I18_H16	**I18_H16**	Cluster U	-
CSI (Modified)	I19_H3	**I5_H17**	**I2_ H5**	-	-
TINN	I20_H1	**I19_H18**	**I17_H2**	-	-
LZC	I17_H4	**I5_H4**	**I6_ H4**	-	-
SDANN1	I19_H3	**I19_H3**	**I3_H3**	-	-
PAS	I8_H6	**I4_H6**	**I19_ H4**	-	-
MFDFA α1 Mean	I13_H10	**I13_H21**	**I11_H9**	-	-
MFDFA α2 Maximum	I16_H13	**I15_H22**	**I19_ H4**	-	-

Note. In each Consensus Solution column, the consensus cluster names (e.g., I1_H1) were created by concatenating the results of the item clustering method (e.g., I1) and the hierarchical clustering method (e.g., H1). Unique numbers were assigned to unique clusters of the respective dataset (e.g., I1 and I2 are two unique clusters in item clustering results). HRV indices that share the same consensus cluster name show cross-method robustness. Robust Solution columns summarize the cross-dataset comparisons. Robust cluster names were created with unique alphabet (e.g., Cluster A) or unique number (e.g., Cluster 1); HRV indices that did not cluster with another index were left blank. Both alphabets and numbers were assigned in ascending order, by the order of appearance of clusters. HRV indices that share the same robust cluster name in Robust Solution 1 and 2 show test-

retest and cross-sample robustness respectively. Indices that failed to show test-retest or cross-sample robustness are bolded and excluded from the final HRV structure in Figure 2.

[a] While these indices demonstrated consistent cluster assignment across test-retest and cross-sample validation in the primary analyses, sensitivity analyses on subgroups, females and young adults (18-24 years old) only, revealed inconsistent cluster membership for these two indices. See Supplementary Materials for sensitivity analyses.

Final HRV structure

The final HRV structure (**Figure 2**) summarizes the robust clusters that showed cross-method agreement in all three datasets as well as both test-retest and cross-sample robustness in the cross-dataset comparisons. The final HRV structure includes 70 HRV indices that consistently formed 21 robust clusters. Additionally, the final HRV structure summarizes the marked similarities in the higher-order structure of the three hierarchical clustering (*hclust*) solutions (**Figures 1-3** in **Supplementary Materials**). Specifically, across the hierarchical structures of the three datasets, Sample 1 (T1), Sample 1 (T2), and Sample 2, we observed two overarching clusters, with one mainly consisting of the time-domain and RSA indices and the other consisting of frequency-domain and non-linear indices. The hierarchical structure in **Figure 2** summarizes this convergence across the hierarchical clustering solutions.

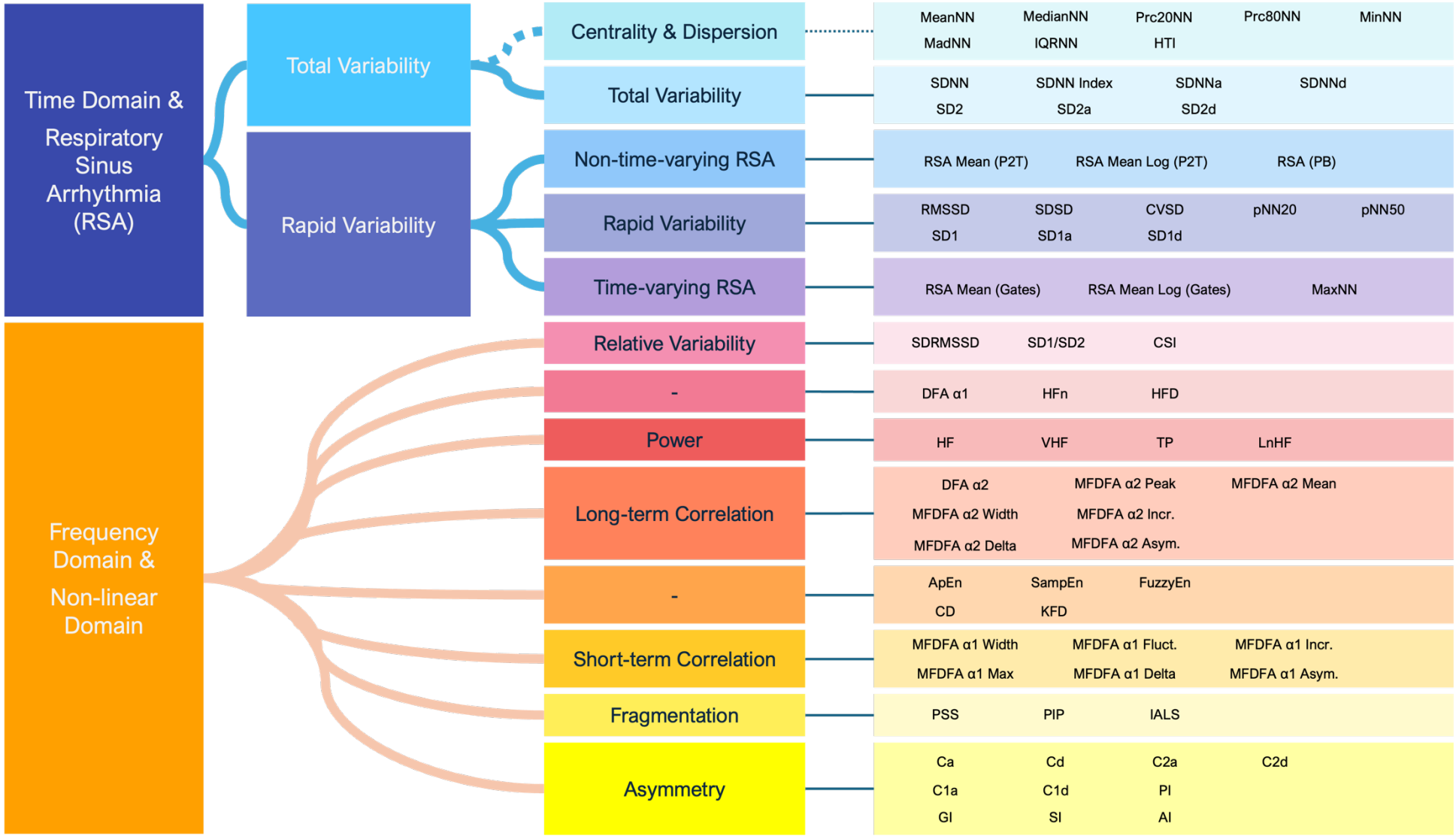

Figure 2. Summary of the final HRV structure. The lowest level (most right) in the hierarchical model represents the 21 robust clusters of 70 indices that received cross-sample agreement across three datasets. The cluster labels (third column from the left) were assigned according to the mathematical or hypothesized physiological underpinning of the indices. Two clusters were not assigned a label (" - ") as the relationships between the members were not clear. The mapping between the clusters across the hierarchical levels represents the convergence of the hclust solutions. Solid lines represent the mapping that was observed across all three hclust solutions. The dotted line represents the mapping that diverged in at least one hclust solution.

Discussion

In this study, we sought to identify a robust structure of HRV indices that consistently capture the interrelationships among 89 HRV indices. To maximize the stability of the final HRV structure, we used two clustering methods to examine the robustness of the clustering solutions across time and samples. This study has four main findings. First, we found that *item clustering* and *hierarchical clustering* yielded highly consistent results, with more than 90% of indices showing cross-method agreement in all three datasets. Second, the structure of HRV indices showed good test-retest robustness (i.e., stability), with 89.9% of indices being assigned to the same clusters in the item clustering and hierarchical clustering solutions at both baseline and a second-time point a week later. Third, the structure of HRV yielded good cross-sample robustness (i.e., generalizability). Despite the substantial differences in methodologies (i.e., different resting durations, eye-closed versus eye-open procedure, acquisition equipment, sampling rate) and sample characteristics (i.e., psychiatric symptom severity, gender distribution, age) between Sample 1 and Sample 2, we found substantial convergence in 87.6% of indices being assigned to the same clusters in the clustering solutions of both samples. Finally, we found clusters of HRV clusters that reinforce the extant literature highlighting specific mathematical and physiological relationships between some indices, as well as observing several clusters of indices for the first time. In the remainder of this discussion, we describe each robust HRV cluster and discuss in-depth their potential shared statistical and physiological characteristics in the context of extant research on their relationships (see **Table 7** for a summary).

Table 7

Summary of clusters and their physiological and mathematical relationship.

Cluster Labels	Indices	Physiological and Mathematical Relationship
Centrality & Dispersion	MeanNN, MedianNN, Prc20NN, Prc80NN, MinNN	Time-domain indices that describe central tendency of the heart rate distribution.
	MadNN, IQRNN, HTI	Time-domain indices that describe the dispersion of the heart rate distribution.
Rapid Variability	RMSSD, SDSD, CVSD, pNN20, pNN50, SD1, SD1a, SD1d	Time-domain indices that reflect rapid, short-term heart rate variability.
Total Variability	SDNN, SDNN Index, SDNNa, SDNNd, SD2, SD2a, SD2d	Time-domain indices that reflect both short-term and long-term heart rate variability. In short recording periods, total variability is likely to predominantly consists of short-term fluctuations.
Non-time-varying RSA	RSA Mean (P2T), RSA Mean Log (P2T), RSA (PB)	Respiratory sinus arrhythmia (RSA) indices that provide non-time-varying estimate of RSA
Time-varying RSA	RSA Mean (Gates), RSA Mean Log (Gates), MaxNN	Respiratory sinus arrhythmia (RSA) indices that provide time-varying estimate of RSA
Relative Variability	SDRMSSD, SD1SD2, CSI	Time-domain and non-linear domain indices that quantify the ratio of short-term and long-term heart rate variability
Power	HF, VHF, TP, LnHF	Frequency-domain indices that provide absolute estimation of power components.
Long-term Correlation	DFA α2, MFDFA α2 Peak, MFDFA α2 Mean	Hurst exponent measures that quantify the self-similarity structure by estimating long-term (>11 heartbeats) temporal correlation of heart rate timeseries.

Cluster Labels	Indices	Physiological and Mathematical Relationship
	MFDFA α2 Width, MFDFA α2 Incr.	
	MFDFA α2 Delta, MFDFA α2 Asym.	
Short-term Correlation	MFDFA α1 Width, MFDFA α1 Fluct., MFDFA α1 Incr.	Hurst exponent measures that quantify the self-similarity structure by estimating short-term (4-11 heartbeats) temporal correlation of heart rate timeseries.
	MFDFA α1 Max, MFDFA α1 Delta, MFDFA α1 Asym.	
Fragmentation	PIP, PSS, PAS	Heart rate fragmentation indices that quantify the abrupt and high-frequency switching between the accelerations and decelerations of heat rate.
Asymmetry	GI, SI, AI	Heart rate asymmetry indices that quantify the asymmetric contribution of heart rate acceleration and deceleration to heart rate variability.
	PI, C1a, C1a	
	Ca, Cd, C2a, C2d	
-	DFA α1, HFn, HFD	Cluster of Hurst exponent (DFA α1), normalized high-frequency component (HFn) and fractal dimension measure (HFD). Observed for the first time in this study.
-	ApEn, SampEn, FuzzyEn, CD, KFD	Cluster of monoscale entropy indices ApEn, SampEn, FuzzyEn) and two fractal dimension measures (CD and KFD). Observed for the first time in this study.

RSA and Time-domain Indices

RSA Indices

In the final HRV structure, five RSA indices (**Table 3**) formed two tight-knit clusters: One cluster comprised the non-time-varying RSA measures *RSA (PB), RSA Mean (P2T)* and *RSA Mean Log (P2T)* (i.e., the non-time-varying RSA cluster in Figure 2) and the other cluster included two time-varying RSA indices from Gates et al. (2015), *RSA Mean (Gates)* and *RSA Mean Log (Gates),* (i.e., the time-varying RSA cluster). The close associations between the RSA indices are consistent with the existing literature (Gates et al., 2015; Lewis et al., 2012). Nevertheless, the resulting clusters also highlight a specificity that potentially differentiates between the non-time-varying quantification (Grossman, 1992; Porges, 1985) and the novel time-varying quantification of RSA (Gates et al., 2015).

Time-domain Indices

The time-domain indices (see **Table 1**) that focus on the variability of NN intervals were mainly split into two clusters; one predominantly consisted of indices that are sensitive to rapid changes or short-term variability of HR (i.e., the rapid variability cluster), and the other included indices that potentially reflect both short-term and long-term variability (i.e., the total variability cluster). The remaining time-domain indices formed three robust clusters that were consistently identified across methods and datasets. Based on the shared characteristics of their cluster members, two of these robust clusters were labeled centrality and dispersion clusters, and the remaining robust cluster of time-domain indices was labeled as relative variability.

Rapid Variability. The rapid variability cluster comprised eight indices, including five time-domain indices (*RMSSD, SDSD, CVSD, pNN20, and pNN50)* and three indices from Poincaré plot (*SD1, SD1a,* and *SD1d).* The close associations between these HRV indices align with existing literature. Mathematically, *SD1*, which measures the short-term HRV, is equivalent to *RMSSD* and *SDSD* (Brennan et al., 2002; Ciccone et al., 2017; Kamen & Tonkin, 1995; Shaffer & Ginsberg, 2017). *CVSD* is a derivative of *RMSSD* and *pNN50* has shown a strong correlation with *RMSSD* in many past studies (Bigger, 1997; Shaffer & Ginsberg, 2017).

From a physiological perspective, studies suggest that the indices in this cluster likely reflect rapid changes in HR (Brennan et al., 2002; Ciccone et al., 2017; Shaffer et al., 2014; Toichi et al., 1997), influenced by physiological processes such as the baroreceptor reflex (Eckberg & Sleight, 1992) or the parasympathetic effects of the autonomic nervous system (ANS). The parasympathetic branch of the ANS impacts HR more rapidly—in less than one second—than sympathetic effects, which can take up to three seconds (Feher, 2012; Nunan et al., 2010). On the higher level in the hierarchical clustering solutions (**Figure 2**), the rapid variability cluster also displayed proximity with both clusters of parasympathetically-mediated RSA indices (non-time-varying RSA and time-varying RSA clusters), further supporting a possible shared physiological relationship with the parasympathetic influence.

Total Variability. The total variability cluster included seven indices: two time-domain indices (*SDNN* and *SDNN Index)*, two heart rate asymmetry (HRA) indices (*SDNNa* and *SDNNd)*, and three indices from Poincaré plot (*SD2, SD2a*, and *SD2d)*. The close

associations between these indices could be attributed to their mathematical relationships, as *SDNN Index, SDNNa* and *SDNNd* are derivates of *SDNN*, while *SD2a* and *SD2d* are derivates of *SD2*. Regarding their physiological association, both *SDNN* and *SD2* have been observed to be influenced by both short-term and long-term variability. They are therefore likely to capture the total variability in the HR signals (Brennan et al., 2002; Task Force, 1996). It is important to note that in short recording periods such as the one used in this study, the total variance in the HR signal would predominantly consist of short-term rapid fluctuations influenced by parasympathetic contributions; the slower variance is captured to a greater extent in longer recordings (e.g., 24 hours).

Centrality and Dispersion. The two robust clusters that subsumed remaining time-domain indices were labeled centrality and dispersion clusters (**Figure 2**) because they mainly consist of time-domain indices that describe the statistical features of a variable distribution (Cardinal, 2015). In particular, *MeanNN*, *MedianNN, Prc20NN,* and *Prc80NN* describe the central tendency of the HR distribution, while *IQRNN* and *MadNN* are mathematical descriptions of dispersion (e.g., the interquartile range or the median absolute deviation of NN intervals). Across the clustering solutions, the centrality and dispersion cluster that includes the indices that describe the spread of the distribution (*IQRNN* and *MadNN*) was consistently observed to join the total variability cluster at the higher-order level in the hierarchies.

Relative Variability. The last robust cluster of time-domain indices included those that describe the ratio of short-term and long-term NN interval variability (i.e., the relative variability cluster). The close association between the indices in this cluster (*SDRMSSD*, *SD1/SD2*, and *CSI*) is mainly attributed to their mathematical relationship. Specifically, *SDRMSSD* is equivalent to *SDNN/RMSSD* ratio, and as previously discussed, and there is a close statistical relationship between *SDNN* and *SD2* and between *RMSSD* and *SD1*. Further, *CSI* and *SD1/SD2* are derived using almost identical mathematical equations (Toichi et al., 1997). Interestingly, the relative variability cluster was observed to consistently join other frequency/non-linear clusters at the higher-order level in the hierarchical structure (**Figure 2**), which suggests that the relative variability indices might capture the unpredictability and irregularities of non-linear changes in the HR signals more so than other time-domain indices.

Frequency-domain and Non-linear Indices

The higher-order clusters in the hierarchical clustering solutions suggest a close association between frequency-domain measures—encompassing traditional frequency-based indices and their normalized derivates (**Table 2)** and non-linear measures—encompassing heart rate fragmentation, heart rate asymmetry, entropy-based, and fractal-based indices (**Table 4**). The high level of similarity between non-linear and frequency-weighted spectral indices aligns with previous literature that has theoretically demonstrated and empirically verified their proximity (Captur et al., 2017; Francis et al., 2002; Lensen et al., 2020; Young & Benton, 2015). Importantly, the association between frequency-domain and non-linear indices appears to be stronger than the association of either of these domains with time-domain indices. In the hierarchical structure (**Figure 2**), the time-domain clusters only merge with these clusters at the highest level in the hierarchy. The indices in the frequency/non-linear higher-order cluster are mainly grouped with those from the same operational domains (e.g., entropy-based or

fractal-based). Nevertheless, there are also intriguing close associations between indices across the operational domains observed for the first time in this analysis.

Frequency-domain Indices

Out of the five frequency-domain indices included, *HFn* is the only relative quantification of frequency components, calculated by dividing the individual component by the total power. Across all the consensus solutions, *HFn* is not assigned to the same cluster with the other four absolute estimations of power components (*HF*, *LnHF*, *VHF* and *TF*) which consistently formed one robust cluster (i.e., the power cluster). From a physiological perspective, the absolute frequency indices are believed to quantify the amount of variation attributable to different physiological mechanisms. In contrast, relative frequency indices quantify the relative contribution of these mechanisms to the total variability of HR. As the low-frequency components (i.e., *ULF*, *VLF*, *LF,* and *LF/HF* indices) could not be reliably quantified in this study due to the constraint of recording lengths (Nussinovitch et al., 2011; Shaffer et al., 2014), the associations between different power components and the implied relative contributions to HRV could not be examined in this study.

Non-linear Indices

Asymmetry. The heart rate asymmetry (HRA) indices measure the asymmetric contribution of HR acceleration and deceleration to HRV (Guzik et al., 2007; Piskorski & Guzik, 2011; Yan et al., 2017b). These indices split into three clusters, labeled as asymmetry clusters, following their differences in physiological and mathematical meanings. The first asymmetry cluster comprised *C1a* and *C1d*, indexing the asymmetrical contribution of accelerations or decelerations to the short-term HRV. The second cluster comprised *C2a* and *C2d*, indexing the asymmetrical contributions to long-term HRV, and *Ca* and *Cd*, indexing the asymmetrical contributions to total HRV. While the physiological origin of HRA remains unclear, the resulting clusters are consistent with existing data that highlight similar asymmetry patterns in long-term and total HRV. Specifically, while HR decelerations tend to dominate accelerations in short-term HRV, decelerations often demonstrate smaller contributions to long-term and total HRV (Piskorski & Guzik, 2011; Sibrecht et al., 2023). The last HRA cluster comprised *AI*, *GI*, and *SI*, which index the asymmetrical contribution of decelerations to HRV using geometrical information from the Poincaré plot (Yan et al., 2017b). Overall, the HRA indices formed exclusive clusters across methods and datasets, highlighting their unique characterization of HRV and physiological processes that warrant further investigation.

Fragmentation. The heart rate fragmentation (HRF) indices measure the “erratic” behaviours in heart rhythm, manifesting as abrupt and high-frequency switching between the accelerations and decelerations of HR (Costa et al., 2017b). Three out of four of these indices (*PSS*, *PIP*, and *IALS*) formed a robust cluster (i.e., the fragmentation cluster) that was reliably identified across methods and datasets. Similar to HRA indices, even though the diagnostic values of HRF indices have been examined (Costa et al., 2017b, 2021), the specific physiological mechanisms underlying them remain largely unknown.

Complexity-based Indices. The entropy-based and fractal-based indices quantify the overall complexity of the HR signals using different mathematical approaches. The entropy-

based indices split into two clusters depending on whether they quantify the complexity or irregularity of the HR signal using one timescale (monoscale-based) or multiple timescales (multiscale-based). This study found that while the cluster of monoscale entropy indices (*ApEn*, *SampEn*, and *FuzzyEn)* were consistently assigned to the same cluster across methods and samples, multiscale entropy indices (*MSEn*, *CMSEn*, and *RCMSEn)* formed a cluster in Sample 1 but not in Sample 2 dataset.

The fractal-based indices included measures of fractal dimension (*CD*, *HFD*, and *KFD*) and measures of Hurst exponent (*DFA*, *MDFA,* and their derivatives). Both fractal dimension and Hurst exponent quantify the complexity of an HR signal using their correlation properties, defined as the degrees of self-similarity of the heartbeat sequence over time (Rogers & Gronwald, 2022). Generally, an HR signal with higher self-similarity will be more irregular and thus represent more complex variability (Henriques et al., 2020; Lau et al., 2022b). While these indices share a mathematical origin, they yielded an intriguing pattern of associations and groupings that warrant further investigation.

Most of the Hurst exponent measures consistently split into clusters depending on the time scales used. Specifically, the majority of *DFA* and *MDFA* indices that use the short-term time scale (i.e., α1 exponent with the scale of 4-11 heartbeats) formed two clusters (i.e., short-term correlation clusters). These clusters joined each other at the higher-order level in the hierarchical structures, as shown in **Figure 2**. Similarly, the Hurst exponent indices that use the long-term time scale (i.e., α2 exponent with the scale of >11 heartbeats) formed two clusters (i.e., long-term correlation clusters); they also joined at the higher-order level. This cluster structure aligned with the existing literature that associated two scaling exponents with different control mechanisms acting at distinct time scales. The short-term scaling exponent α1 potentially reflects the influence of rapid physiological processes on HR dynamics such as the parasympathetic (vagal) activities or baroreceptor reflex. Conversely, the long-term scaling exponent could reflect the influence of slower regulatory processes such as sympathetic activities or thermoregulatory and endocrine processes (Pham et al., 2021; Shaffer & Ginsberg, 2017). Nevertheless, the mathematical quantification and physiological meanings of the DFA and MDFA indices remain an active field of research. Their validity in characterizing the complexity of HR, especially in short recordings such as the one used in this study has recently been questioned (Carpena et al., 2021). This quickly prompted the development of the modified DFA to improve their validity in short physiological recordings (Gong & Fu, 2023).

Even more intriguing is the close association between the fractal dimension measures with other measures of different mathematical origins. While two of the fractal dimension measures (*CD* and *KFD*) were consistently clustered together with the monoscale entropy indices, *HFD* showed a reliably close association with *HFn* and *DFA α1*. These associations are observed for the first time in this study, and therefore, the clusters were not labeled in the final HRV structure. Further investigation is needed to replicate and explore the underlying reason for their proximity.

Recommendations for selection of HRV indices

In this study, 21 clusters of 70 HRV indices were consistently identified across two clustering methods and three datasets (**Figure 2**). Based on the robust empirical associations found and

the relative popularity of some HRV indices in the extant literature, we provided three recommendations on the selection of HRV indices in future research. **Table 8** summarizes the recommendations.

First, given the potential statistical issues (e.g., inflated Type I error rates) and possible systematic research biases (e.g., inflated confidence in results) caused by the use of redundant or highly correlated indices (Dormann et al., 2013; Mela & Kopalle, 2002; Næs & Mevik, 2001), we recommend future analysis avoid including more than one HRV index from the same cluster. For instance, if *RMSSD* is used, other indices in the same cluster—such as *SDSD*, *CVSD,* or *SD1*—should not be interpreted.

Second, if researchers do decide to include multiple indices from the same cluster, we recommend commenting on their close association and avoiding overemphasis on the apparent robustness of findings when indices from the same cluster show similar patterns of significant results. Notably, this recommendation is also relevant for data extraction in meta-analysis, where researchers are often required to decide which effects to extract when primary studies report multiple HRV indices. If meta-analysis researchers decided to extract effects with more than one HRV index, we recommend choosing indices from different clusters found in this study to improve construct validity.

Finally, based on our results and the relative popularity of some HRV indices in the extant literature, for short recordings of rest (under 10 minutes) we recommend researchers use the following indices: *RMSSD*, *SDNN*, *RSA* (Porges-Bohrer or Peak-to-Trough method), *RSA* (Gates method), *SD1*/*SD2* or *CSI*, *SampEn*, *HF* or *LnHF*, *DFA α1*, *DFA α2*, one of *MDFA α1* features, one of *MDFA α2* features, one of the heart rate asymmetry indices, and one of the heart rate fragmentation indices. For longer recordings, researchers can consider including low-frequency components such as *LF* or *VLF*, and *LF/HF*. Avoiding the inclusion of redundant or substantially similar indices could help to facilitate a more nuanced interpretation of HRV data and enhance the validity of conclusions drawn about the relationships between HRV and various physiological or psychological constructs.

Table 8

Summary of recommendations on the selection of HRV indices for resting state.

#	Recommendations on selection of HRV indices
1	To avoid biases from redundant or correlated indices, we recommend authors of empirical studies and meta-analyses *avoid including more than one HRV index from the same cluster in their analysis*.
2	If researchers decided to include multiple indices from the same cluster, we recommend commenting on their close association and avoiding overemphasis of any apparent 'robustness' of findings related to indices of the same cluster showing similar patterns of significant results.
3	Based on the relative popularity of the indices, we recommend authors to consider these indices for short recording of rest (under 10 minutes): *RMSSD*, *SDNN*, *RSA* (Porges-

Bohrer or Peak-to-Trough method), *RSA* (Gates method), *SD1*/*SD2* or *CSI*, *SampEn*, *HF* or *LnHF*, *DFA α1*, *DFA α2*, one of *MDFA α1* features, one of *MDFA α2* features, one of the heart rate asymmetry indices, and one of the heart rate fragmentation indices.

Even though reduced HRV at rest has been associated with a wide range of mental and physical health outcomes, the dynamic interactions between mechanisms underlying HRV and specific biological pathways linking them to physical and mental health conditions remain poorly understood. The robust clusters identified in this study may capture distinct aspects of heart rate dynamics, each potentially reflecting different underlying mechanisms of autonomic regulation during rest. Given these possibilities, it would be particularly valuable for future research to explore how each cluster of HRV indices relates to specific physical (e.g., cardiovascular function, diabetes) or psychological outcomes, rather than treating all indices as interchangeable markers of the same process. This targeted approach could enhance our understanding of the specific pathways linking HRV to health conditions, leading to more precise interpretations of HRV data in both clinical and research contexts.

Importantly, our recommendations are not based on and are not intended to suggest that certain HRV indices inherently provide greater mechanistic validity. Determining which indices most accurately reflect specific autonomic or physiological processes requires dedicated validation studies (e.g., comparisons with pharmacological blockade, Lewis et al., 2012; see also Quigley et al., 2024). Instead, the present study emphasizes empirical structure and redundancy among HRV indices and advocates for a more systematic and transparent approach to their selection. We encourage researchers to include indices from different clusters to avoid redundancy and to capitalize on the potentially unique information carried by distinct features. This approach can serve as a foundation for future work aimed at evaluating the mechanistic validity and predictive utility of both established and emerging HRV indices.

Lastly, while our analysis and recommendations focus on existing HRV indices, we do not intend to discourage the development of novel metrics. On the contrary, we recognize the importance of continued innovation in HRV methodology, particularly within complexity science, where emerging indices may provide complementary insights into the physiological mechanisms underlying HRV or offer increased robustness to noise and artifacts compared to some traditional linear methods (Quigley et al., 2024; Sassi et al., 2015). However, it is essential that these novel indices are evaluated systematically and in parallel with established measures. Direct comparisons with well-characterized indices—both statistically and physiologically—can help determine their added value, clarify which underlying processes they may uniquely capture (if any), and assess their reliability and interpretability.

Limitations

This study, while providing valuable insights into the clustering of HRV indices, has several limitations that should be considered when interpreting the findings. A limitation of this study is the use of relatively short recording durations (3 minutes in Sample 1 and 8 minutes in Sample 2), which constrain the physiological processes that can be reliably quantified. While these short-term recordings are adequate for assessing high-frequency parasympathetic

activity and short-latency autonomic dynamics, they preclude accurate estimation of lower-frequency components (LF, VLF, and ULF, which require longer recording durations (Shaffer & Ginsberg, 2017; Nussinovitch et al., 2011). Consequently, these lower-frequency indices were excluded from clustering in the current study, and the findings may not generalize to long-term HRV dynamics that may reflect different regulatory systems (e.g., hormonal, circadian, or thermoregulatory influences).

Relatedly, the clustering structure identified in this study is specific to HRV data collected during short-term, resting-state recordings in a controlled laboratory context. The relationships among HRV indices may vary under different conditions, including the length of the ECG recordings (e.g., 10 minutes versus 24 hours), the participant's position during the recording (e.g., sitting, standing, or supine), and the activity being performed (e.g., controlled rest versus free-running). Prolonged recordings may reveal rhythmic physiological changes or stress-related hormonal shifts that alter the strength or direction of associations between HRV indices. Thus, while our results provide a useful framework for understanding HRV index relationships in typical short-term resting protocols, future research should evaluate the extent to which these structures generalize across different recording lengths and resting-state protocols.

Another limitation of the present study concerns the demographic heterogeneity within individual samples, particularly with respect to age and sex. Prior research has consistently shown that HRV differs systematically across these variables (e.g., Stein et al., 1997; Zhang, 2007), suggesting the possibility that some clustering patterns may reflect shared sensitivity to demographic factors rather than shared physiological or statistical properties. To assess the extent of this potential confounding, we conducted exploratory sensitivity analyses in more demographically homogeneous subgroups—specifically, female-only and young-adult participants within each sample. These subgroup solutions closely mirrored the main findings: on average, over 97% of HRV indices were assigned to the same clusters as in the final consensus structure (see Tables 4–8 in Supplementary Materials). Only a small set of indices—*MFDFA α1 Asymmetry, ApEn* and *FuzzyEn*—exhibited inconsistent cluster membership with the final structure in one or more subgroup analyses. This high degree of consistency across subgroups provides reassurance that the observed clustering structure is unlikely to have been substantially influenced by demographic heterogeneity. Nonetheless, we acknowledge that the subgroup analyses were limited by reduced statistical power, and future studies with larger, demographically stratified samples will be important for further validating these findings and assessing the generalizability of clustering-based approaches to HRV interpretation.

This study also lacks information on relevant health and lifestyle factors such as smoking status, medication use, hypertension, or other cardiovascular or metabolic conditions, which have potential to influence HRV. However, the clustering structure was remarkably consistent across two independent samples differing in psychological characteristics and recording procedures, bolstering our confidence that the observed groupings are likely to reflect relatively stable organizational patterns among HRV indices. However, future studies should aim to replicate and extend these findings in larger, demographically and medically diverse samples, with detailed assessments of health status, medication use, and lifestyle behaviors

(e.g., smoking, physical activity) to examine the robustness and potential modulators of HRV index relationships.

Conclusion

HRV is a complex phenomenon shaped by the dynamic interplay of various physiological systems including the sympathetic and parasympathetic branches of the ANS, and regulatory mechanisms governing respiration, blood pressure, vascular tone, and longer-term processes such as circadian rhythms and body temperature. While HRV indices are all designed to quantify variability in HR, it is crucial not to assume that they share the same physiological significance. Individual measures or clusters of related indices may reflect distinct underlying mechanisms or specific interactions between them. In this study, we applied a consensus-based clustering approach to enhance our understanding of the relationships among 89 HRV indices derived from resting-state assessments. Our findings provide a framework that can aid in the selection of HRV indices and interpretation of HRV data, facilitating a deeper and more comprehensive examination of the underlying dynamic nature of HRV.

References

Andrade, C. (2021). HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. *The Journal of Clinical Psychiatry*, *82*(1), 25941. https://doi.org/10.4088/JCP.20f13804

Austin, M. A., Riniolo, T. C., & Porges, S. W. (2007). Borderline personality disorder and emotion regulation: Insights from the Polyvagal Theory. *Brain and Cognition*, *65*(1), 69–76. https://doi.org/10.1016/j.bandc.2006.05.007

Behar, J. A., Rosenberg, A. A., Weiser-Bitoun, I., Shemla, O., Alexandrovich, A., Konyukhov, E., & Yaniv, Y. (2018). PhysioZoo: A Novel Open Access Platform for Heart Rate Variability Analysis of Mammalian Electrocardiographic Data. *Frontiers in Physiology*, *9*. https://doi.org/10.3389/fphys.2018.01390

Berntson, G. G., Thomas Bigger, J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., & 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

Bigger, J. T. (1997). The Predictive Value of RR Variability and Baroreflex Sensitivity in Coronary Heart Disease. *Cardiac Electrophysiology Review*, *1*(1), 198–204. https://doi.org/10.1023/A:1009902022073

Bigger, J. T., Albrecht, P., Steinman, R. C., Rolnitzky, L. M., Fleiss, J. L., & Cohen, R. J. (1989). Comparison of time- and frequency domain-based measures of cardiac parasympathetic activity in Holter recordings after myocardial infarction. *The American Journal of Cardiology*, *64*(8), 536–538. https://doi.org/10.1016/0002-9149(89)90436-0

Billman, G. E. (2011). Heart Rate Variability – A Historical Perspective. *Frontiers in Physiology*, *2*, 86. https://doi.org/10.3389/fphys.2011.00086

Bolea, J., Laguna, P., Remartínez, J. M., Rovira, E., Navarro, A., & Bailón, R. (2014). Methodological Framework for Estimating the Correlation Dimension in HRV Signals. *Computational and Mathematical Methods in Medicine*, *2014*, 129248. https://doi.org/10.1155/2014/129248

Brammer, J. C. (2020). biopeaks: A graphical user interface for feature extraction from heart- and breathing biosignals. *Journal of Open Source Software*, *5*(54), 2621. https://doi.org/10.21105/joss.02621

Brennan, M., Palaniswami, M., & Kamen, P. (2002). Poincaré plot interpretation using a physiological model of HRV based on a network of oscillators. *American Journal of Physiology-Heart and Circulatory Physiology*, *283*(5), H1873–H1886. https://doi.org/10.1152/ajpheart.00405.2000

Campbell, A. A., Wisco, B. E., Silvia, P. J., & Gay, N. G. (2019). Resting respiratory sinus arrhythmia and posttraumatic stress disorder: A meta-analysis. *Biological Psychology*, *144*, 125–135. https://doi.org/10.1016/j.biopsycho.2019.02.005

Candia-Rivera, D. (2023). Modeling brain-heart interactions from Poincaré plot-derived measures of sympathetic-vagal activity. *MethodsX*, *10*, 102116. https://doi.org/10.1016/j.mex.2023.102116

Captur, G., Karperien, A. L., Hughes, A. D., Francis, D. P., & Moon, J. C. (2017). The fractal heart—Embracing mathematics in the cardiology clinic. *Nature Reviews Cardiology*, *14*(1), 56–64. https://doi.org/10.1038/nrcardio.2016.161

Cardinal, L. J. (2015). Central tendency and variability in biological systems. *Journal of Community Hospital Internal Medicine Perspectives*, *5*(3), 10.3402/jchimp.v5.27930. https://doi.org/10.3402/jchimp.v5.27930

Carpena, P., Gómez-Extremera, M., & Bernaola-Galván, P. A. (2021). On the Validity of Detrended Fluctuation Analysis at Short Scales. *Entropy*, *24*(1), 61. https://doi.org/10.3390/e24010061

Chalmers, J. A., Quintana, D. S., Abbott, M. J.-A., & Kemp, A. H. (2014). Anxiety Disorders are Associated with Reduced Heart Rate Variability: A Meta-Analysis. *Frontiers in Psychiatry*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyt.2014.00080

Cheng, Y.-C., Su, M.-I., Liu, C.-W., Huang, Y.-C., & Huang, W.-L. (2022). Heart rate variability in patients with anxiety disorders: A systematic review and meta-analysis. *Psychiatry and Clinical Neurosciences*, *76*(7), 292–302. https://doi.org/10.1111/pcn.13356

Ciccone, A. B., Siedlik, J. A., Wecht, J. M., Deckert, J. A., Nguyen, N. D., & Weir, J. P. (2017). Reminder: RMSSD and SD1 are identical heart rate variability metrics. *Muscle & Nerve*, *56*(4), 674–678. https://doi.org/10.1002/mus.25573

Clamor, A., Lincoln, T. M., Thayer, J. F., & Koenig, J. (2016). Resting vagal activity in schizophrenia: Meta-analysis of heart rate variability as a potential endophenotype. *British Journal of Psychiatry*, *208*(1), 9–16. https://doi.org/10.1192/bjp.bp.114.160762

Costa, M. D., Davis, R. B., & Goldberger, A. L. (2017a). Heart Rate Fragmentation: A New Approach to the Analysis of Cardiac Interbeat Interval Dynamics. *Frontiers in Physiology*, *8*. https://doi.org/10.3389/fphys.2017.00255

Costa, M. D., Davis, R. B., & Goldberger, A. L. (2017b). Heart Rate Fragmentation: A New Approach to the Analysis of Cardiac Interbeat Interval Dynamics. *Frontiers in Physiology*, *8*, 255. https://doi.org/10.3389/fphys.2017.00255

Costa, M. D., Redline, S., Soliman, E. Z., Goldberger, A. L., & Heckbert, S. R. (2021). Fragmented sinoatrial dynamics in the prediction of atrial fibrillation: The Multi-Ethnic Study of Atherosclerosis. *American Journal of Physiology - Heart and Circulatory Physiology*, *320*(1), H256–H271. https://doi.org/10.1152/ajpheart.00421.2020

Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.

Cripps, T. R., Malik, M., Farrell, T. G., & Camm, A. J. (1991). Prognostic value of reduced heart rate variability after myocardial infarction: Clinical evaluation of a new analysis method. *British Heart Journal*, *65*(1), 14–19.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Denton, T. A., Diamond, G. A., Helfant, R. H., Khan, S., & Karagueuzian, H. (1990). Fascinating rhythm: A primer on chaos theory and its application to cardiology. *American Heart Journal*, *120*(6, Part 1), 1419–1440. https://doi.org/10.1016/0002-8703(90)90258-Y

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Eckberg, D. L., & Sleight, P. (1992). Human Baroreflexes in Health and Disease. In D. L. Eckberg & P. Sleight (Eds.), *Human Baroreflexes in Health and Disease* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198576938.002.0001

Faini, A., Parati, G., & Castiglioni, P. (2021). Multiscale assessment of the degree of multifractality for physiological time series. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2212), 20200254. https://doi.org/10.1098/rsta.2020.0254

Fantini-Hauwel, C., Batselé, E., Gois, C., & Noel, X. (2020). Emotion Regulation Difficulties Are Not Always Associated With Negative Outcomes on Women: The Buffer Effect of HRV. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00697

Feher, J. (2012). 5.13—Regulation of Arterial Pressure. In J. Feher (Ed.), *Quantitative Human Physiology (Second Edition)* (pp. 608–618). Academic Press. https://doi.org/10.1016/B978-0-12-800883-6.00058-6

Forbes, M. K., Sunderland, M., Rapee, R. M., Batterham, P. J., Calear, A. L., Carragher, N., Ruggero, C., Zimmerman, M., Baillie, A. J., Lynch, S. J., Mewton, L., Slade, T., & Krueger, R. F. (2021). A Detailed Hierarchical Model of Psychopathology: From Individual Symptoms up to the General Factor of Psychopathology. *Clinical Psychological Science*, *9*(2), 139–168. https://doi.org/10.1177/2167702620954799

Forte, G., Favieri, F., & Casagrande, M. (2019). Heart Rate Variability and Cognitive Function: A Systematic Review. *Frontiers in Neuroscience*, *13*, 710. https://doi.org/10.3389/fnins.2019.00710

Francis, D. P., Willson, K., Georgiadou, P., Wensel, R., Davies, L. C., Coats, A., & Piepoli, M. (2002). Physiological basis of fractal complexity properties of heart rate variability in man. *The Journal of Physiology*, *542*(Pt 2), 619–629. https://doi.org/10.1113/jphysiol.2001.013389

Gates, K. M., Gatzke-Kopp, L. M., Sandsten, M., & Blandon, A. Y. (2015). Estimating time-varying RSA to examine psychophysiological linkage of marital dyads. *Psychophysiology*, *52*(8), 1059–1065. https://doi.org/10.1111/psyp.12428

Goldberger, A. L. (1996). Non-linear dynamics for clinicians: Chaos theory, fractals, and complexity at the bedside. *Lancet (London, England)*, *347*(9011), 1312–1314. https://doi.org/10.1016/s0140-6736(96)90948-4

Gong, H., & Fu, Z. (2023). Modified DFA for a robust discrimination between short-term and long-range correlations in short time series. *Physica A: Statistical Mechanics and Its Applications*, *625*, 128997. https://doi.org/10.1016/j.physa.2023.128997

Grossman, P. (1992). Breathing Rhythms of the Heart in a World of No Steady State: A Comment on Weber, Molenaar, and van der Molen. *Psychophysiology*, *29*(1), 66–72. https://doi.org/10.1111/j.1469-8986.1992.tb02013.x

Grossman, P., Van Beek, J., & Wientjes, C. (1990). A Comparison of Three Quantification Methods for Estimation of Respiratory Sinus Arrhythmia. *Psychophysiology*, *27*(6), 702–714. https://doi.org/10.1111/j.1469-8986.1990.tb03198.x

Guzik, P., Piskorski, J., Krauze, T., Schneider, R., Wesseling, K. H., Wykretowicz, A., & Wysocki, H. (2007). Correlations between the Poincaré Plot and Conventional Heart Rate Variability Parameters Assessed during Paced Breathing. *The Journal of Physiological Sciences*, *57*(1), 63–71. https://doi.org/10.2170/physiolsci.RP005506

Henriques, T., Ribeiro, M., Teixeira, A., Castro, L., Antunes, L., & Costa-Santos, C. (2020). Nonlinear Methods Most Applied to Heart-Rate Time Series: A Review. *Entropy*, *22*(3), 309. https://doi.org/10.3390/e22030309

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, *31*(2), 277–283. https://doi.org/10.1016/0167-2789(88)90081-4

Holzman, J. B., & Bridgett, D. J. (2017). Heart rate variability indices as bio-markers of top-down self-regulatory mechanisms: A meta-analytic review. *Neuroscience & Biobehavioral Reviews*, *74*, 233–255. https://doi.org/10.1016/j.neubiorev.2016.12.032

Hoshi, R. A., Pastre, C. M., Vanderlei, L. C. M., & Godoy, M. F. (2013). Poincaré plot indexes of heart rate variability: Relationships with other nonlinear variables. *Autonomic Neuroscience*, *177*(2), 271–274. https://doi.org/10.1016/j.autneu.2013.05.004

Jeppesen, J., Beniczky, S., Johansen, P., Sidenius, P., & Fuglsang-Frederiksen, A. (2015). Detection of epileptic seizures with a modified heart rate variability algorithm based on Lorenz plot. *Seizure*, *24*, 1–7. https://doi.org/10.1016/j.seizure.2014.11.004

Johnco, C. J., Norberg, M., Wuthrich, V. M., & Rapee, R. M. (2025). Cognitive restructuring before exposure therapy or behavioral experiments? How the timing of expectancy violation and magnitude of expectancy change influence exposure therapy outcomes. *Journal of Consulting and Clinical Psychology*, *93*(5), 369–381. https://doi.org/10.1037/ccp0000953

Kamen, P. W., & Tonkin, A. M. (1995). Application of the Poincaré plot to heart rate variability: A new measure of functional status in heart failure. *Australian and New Zealand Journal of Medicine*, *25*(1), 18–26. https://doi.org/10.1111/j.1445-5994.1995.tb00573.x

Karmakar, C. K., Khandoker, A. H., Gubbi, J., & Palaniswami, M. (2009). Complex Correlation Measure: A novel descriptor for Poincaré plot. *BioMedical Engineering OnLine*, *8*, 17. https://doi.org/10.1186/1475-925X-8-17

Katz, M. J. (1988). Fractals and the analysis of waveforms. *Computers in Biology and Medicine*, *18*(3), 145–156. https://doi.org/10.1016/0010-4825(88)90041-8

Kaufmann, T., Sütterlin, S., Schulz, S. M., & Vögele, C. (2011). ARTiiFACT: A tool for heart rate artifact processing and heart rate variability analysis. *Behavior Research Methods*, *43*(4), 1161–1170. https://doi.org/10.3758/s13428-011-0107-7

Kemp, A. H., Quintana, D. S., Gray, M. A., Felmingham, K. L., Brown, K., & Gatt, J. M. (2010). Impact of Depression and Antidepressant Treatment on Heart Rate Variability: A Review and Meta-Analysis. *Biological Psychiatry*, *67*(11), 1067–1074. https://doi.org/10.1016/j.biopsych.2009.12.012

Kleiger, R. E., Miller, J. P., Bigger, J. T., & Moss, A. J. (1987). Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. *The American Journal of Cardiology*, *59*(4), 256–262. https://doi.org/10.1016/0002-9149(87)90795-8

Kleiger, R. E., Stein, P. K., & Bigger, J. T. (2005). Heart rate variability: Measurement and clinical utility. *Annals of Noninvasive Electrocardiology: The Official Journal of the International Society for Holter and Noninvasive Electrocardiology, Inc*, *10*(1), 88–101. https://doi.org/10.1111/j.1542-474X.2005.10101.x

Koch, C., Wilhelm, M., Salzmann, S., Rief, W., & Euteneuer, F. (2019). A meta-analysis of heart rate variability in major depression. *Psychological Medicine*, *49*(12), 1948–1957. https://doi.org/10.1017/S0033291719001351

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data

Analysis, and Data Reporting. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00213

Lau, Z. J., Pham, T., Chen, S. H. A., & Makowski, D. (2022a). Brain entropy, fractal dimensions and predictability: A review of complexity measures for EEG in healthy and neuropsychiatric populations. *European Journal of Neuroscience*, *56*(7), 5047–5069. https://doi.org/10.1111/ejn.15800

Lau, Z. J., Pham, T., Chen, S. H. A., & Makowski, D. (2022b). Brain entropy, fractal dimensions and predictability: A review of complexity measures for EEG in healthy and neuropsychiatric populations. *European Journal of Neuroscience*, *56*(7), 5047–5069. https://doi.org/10.1111/ejn.15800

Leite, M. R., Ramos, E. M. C., Kalva-Filho, C. A., Rodrigues, F. M. M., Freire, A. P. C. F., Tacao, G. Y., de Toledo, A. C., Cecílio, M. J., Vanderlei, L. C. M., & Ramos, D. (2015). Correlation between heart rate variability indexes and aerobic physiological variables in patients with COPD. *Respirology (Carlton, Vic.)*, *20*(2), 273–278. https://doi.org/10.1111/resp.12424

Lempel, A., & Ziv, J. (1976). On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory*, *22*(1), 75–81. IEEE Transactions on Information Theory. https://doi.org/10.1109/TIT.1976.1055501

Lensen, I. S., Monfredi, O. J., Andris, R. T., Lake, D. E., & Moorman, J. R. (2020). Heart rate fragmentation gives novel insights into non-autonomic mechanisms governing beat-to-beat control of the heart's rhythm. *JRSM Cardiovascular Disease*, *9*, 2048004020948732. https://doi.org/10.1177/2048004020948732

Lewis, G. F., Furman, S. A., McCool, M. F., & Porges, S. W. (2012). Statistical strategies to quantify respiratory sinus arrhythmia: Are commonly used metrics equivalent? *Biological Psychology*, *89*(2), 349–364. https://doi.org/10.1016/j.biopsycho.2011.11.009

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to Classify, Detect, and Manage Univariate and Multivariate Outliers, With Emphasis on Pre-Registration. *International Review of Social Psychology*, *32*(1). https://doi.org/10.5334/irsp.289

Lipponen, J. A., & Tarvainen, M. P. (2019). A robust algorithm for heart rate variability time series artefact correction using novel beat classification. *Journal of Medical Engineering & Technology*, *43*(3), 173–181. https://doi.org/10.1080/03091902.2019.1640306

Makowski, D., & Neves, A. (2024). *Interoception Primals* [Dataset]. https://doi.org/doi:10.18112/openneuro.ds005434.v1.0.0

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & 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

Makowski, D., Tam, P., & Zen Juen, L. (2023). *Deception Data* [Dataset]. https://doi.org/doi:10.18112/openneuro.ds004511.v1.0.2

Makowski, D., Te, A. S., Jiayi, Stephanie, K., & Zi Liang. (2024). *Illusion Game Study* [Dataset]. https://doi.org/doi:10.18112/openneuro.ds004587.v1.0.0

Makowski, D., Te, A. S., Pham, T., Lau, Z. J., & Chen, S. H. A. (2022). The Structure of Chaos: An Empirical Comparison of Fractal Physiology Complexity Indices Using NeuroKit2. *Entropy*, *24*(8), Article 8. https://doi.org/10.3390/e24081036

Mandelbrot, B. B., Blumen, A., Fleischmann, M., Tildesley, D. J., & Ball, R. C. (1997). Fractal geometry: What is it, and what does it do? *Proceedings of the Royal Society of London*.

A. Mathematical and Physical Sciences, *423*(1864), 3–16. https://doi.org/10.1098/rspa.1989.0038

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety1. *Behaviour Research and Therapy*, *36*(4), 455–470. https://doi.org/10.1016/S0005-7967(97)10031-6

McCroskey, J. C. (1970). Measures of communication-bound anxiety. *Speech Monographs*, *37*(4), 269–277. https://doi.org/10.1080/03637757009375677

Mela, C. F., & Kopalle, P. K. (2002). The impact of collinearity on regression analysis: The asymmetric effect of negative and positive correlations. *Applied Economics*, *34*(6), 667–677. https://doi.org/10.1080/00036840110058482

Næs, T., & Mevik, B.-H. (2001). Understanding the collinearity problem in regression and discriminant analysis. *Journal of Chemometrics*, *15*(4), 413–426. https://doi.org/10.1002/cem.676

Nunan, D., Sandercock, G. R. H., & Brodie, D. A. (2010). A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. *Pacing and Clinical Electrophysiology: PACE*, *33*(11), 1407–1417. https://doi.org/10.1111/j.1540-8159.2010.02841.x

Nussinovitch, U., Elishkevitz, K. P., Katz, K., Nussinovitch, M., Segev, S., Volovitz, B., & Nussinovitch, N. (2011). Reliability of Ultra-Short ECG Indices for Heart Rate Variability. *Annals of Noninvasive Electrocardiology : The Official Journal of the International Society for Holter and Noninvasive Electrocardiology, Inc*, *16*(2), 117–122. https://doi.org/10.1111/j.1542-474X.2011.00417.x

Otzenberger, H., Gronfier, C., Simon, C., Charloux, A., Ehrhart, J., Piquard, F., & Brandenberger, G. (1998). Dynamic heart rate variability: A tool for exploring sympathovagal balance continuously during sleep in men. *The American Journal of Physiology*, *275*(3), H946-950. https://doi.org/10.1152/ajpheart.1998.275.3.H946

Peng, R.-C., Zhou, X.-L., Lin, W.-H., & Zhang, Y.-T. (2015). Extraction of Heart Rate Variability from Smartphone Photoplethysmograms. *Computational and Mathematical Methods in Medicine*, *2015*(1), 516826. https://doi.org/10.1155/2015/516826

Pham, T., Lau, Z. J., Chen, S. H. A., & Makowski, D. (2021). Heart Rate Variability in Psychology: A Review of HRV Indices and an Analysis Tutorial. *Sensors (Basel, Switzerland)*, *21*(12), 3998. https://doi.org/10.3390/s21123998

Pichot, V., Roche, F., Celle, S., Barthélémy, J.-C., & Chouchou, F. (2016). HRVanalysis: A Free Software for Analyzing Cardiac Autonomic Activity. *Frontiers in Physiology*, *7*. https://doi.org/10.3389/fphys.2016.00557

Pinna, T., & Edwards, D. J. (2020). A Systematic Review of Associations Between Interoception, Vagal Tone, and Emotional Regulation: Potential Applications for Mental Health, Wellbeing, Psychological Flexibility, and Chronic Conditions. *Frontiers in Psychology*, *11*, 1792. https://doi.org/10.3389/fpsyg.2020.01792

Piskorski, J., & Guzik, P. (2011). Asymmetric properties of long-term and total heart rate variability. *Medical & Biological Engineering & Computing*, *49*(11), 1289–1297. https://doi.org/10.1007/s11517-011-0834-z

Porges, S. W. (1985). *Method and apparatus for evaluating rhythmic oscillations in aperiodic physiological response systems* (United States Patent No. US4510944A). https://patents.google.com/patent/US4510944A/en

Porges, S. W. (2009). The polyvagal theory: New insights into adaptive reactions of the autonomic nervous system. *Cleveland Clinic Journal of Medicine*, *76*(Suppl 2), S86–S90. https://doi.org/10.3949/ccjm.76.s2.17

Porges, S. W. (2011). *The polyvagal theory: Neurophysiological foundations of emotions, attachment, communication, and self-regulation (Norton series on interpersonal neurobiology)*. WW Norton & Company.

Quigley, K. S., Gianaros, P. J., Norman, G. J., Jennings, J. R., Berntson, G. G., & de Geus, E. J. C. (2024). Publication guidelines for human heart rate and heart rate variability studies in psychophysiology—Part 1: Physiological underpinnings and foundations of measurement. *Psychophysiology*, *61*(9), e14604. https://doi.org/10.1111/psyp.14604

R Core Team. (2021). *R: A language and environment for statistical computing*. [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Revelle, W. (1978). ICLUST: A cluster analytic approach to exploratory and confirmatory scale construction. *Behavior Research Methods & Instrumentation*, *10*(5), 739–742. https://doi.org/10.3758/BF03205389

Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*(1), 57–74. https://doi.org/10.1207/s15327906mbr1401_4

Revelle, W. (2024). *psych: Procedures for Psychological, Psychometric, and Personality Research*. (Version R package version 2.4.6) [Computer software]. https://cran.r-project.org/package=psych

Rogers, B., & Gronwald, T. (2022). Fractal Correlation Properties of Heart Rate Variability as a Biomarker for Intensity Distribution and Training Prescription in Endurance Exercise: An Update. *Frontiers in Physiology*, *13*, 879071. https://doi.org/10.3389/fphys.2022.879071

Rossi, R. C., Vanderlei, L. C. M., Gonçalves, A. C. C. R., Vanderlei, F. M., Bernardo, A. F. B., Yamada, K. M. H., da Silva, N. T., & de Abreu, L. C. (2015). Impact of obesity on autonomic modulation, heart rate and blood pressure in obese young people. *Autonomic Neuroscience: Basic & Clinical*, *193*, 138–141. https://doi.org/10.1016/j.autneu.2015.07.424

Sassi, R., Cerutti, S., Lombardi, F., Malik, M., Huikuri, H. V., Peng, C.-K., Schmidt, G., Yamamoto, Y., Document Reviewers:, Gorenek, B., Lip, G. Y. H., Grassi, G., Kudaiberdieva, G., Fisher, J. P., Zabel, M., & Macfadyen, R. (2015). Advances in heart rate variability signal analysis: Joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society. *EP Europace*, *17*(9), 1341–1353. https://doi.org/10.1093/europace/euv015

Schroeder, E. B., Chambless, L. E., Liao, D., Prineas, R. J., Evans, G. W., Rosamond, W. D., Heiss, G., & Atherosclerosis Risk in Communities (ARIC) study. (2005). Diabetes, glucose, insulin, and heart rate variability: The Atherosclerosis Risk in Communities (ARIC) study. *Diabetes Care*, *28*(3), 668–674. https://doi.org/10.2337/diacare.28.3.668

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*, 258. https://doi.org/10.3389/fpubh.2017.00258

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology*, *5*, 1040. https://doi.org/10.3389/fpsyg.2014.01040

Sibrecht, G., Piskorski, J., Krauze, T., & Guzik, P. (2023). Heart Rate Asymmetry, Its Compensation, and Heart Rate Variability in Healthy Adults during 48-h Holter ECG

Recordings. *Journal of Clinical Medicine*, *12*(3), 1219. https://doi.org/10.3390/jcm12031219

Silva, L. E. V., Fazan, R., & Marin-Neto, J. A. (2020). PyBioS: A freeware computer software for analysis of cardiovascular signals. *Computer Methods and Programs in Biomedicine*, *197*, 105718. https://doi.org/10.1016/j.cmpb.2020.105718

Sloan, E., Hall, K., Moulding, R., Bryce, S., Mildred, H., & Staiger, P. K. (2017). Emotion regulation as a transdiagnostic treatment construct across anxiety, depression, substance, eating and borderline personality disorders: A systematic review. *Clinical Psychology Review*, *57*, 141–163. https://doi.org/10.1016/j.cpr.2017.09.002

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stein, P. K., Kleiger, R. E., & Rottman, J. N. (1997). Differing Effects of Age on Heart Rate Variability in Men and Women. *The American Journal of Cardiology*, *80*(3), 302–305. https://doi.org/10.1016/S0002-9149(97)00350-0

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability: Standards of measurement, physiological interpretation and clinical use. *Circulation*, *17*, 354–381.

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, *61*(3), 201–216. https://doi.org/10.1016/s0165-0327(00)00338-4

Thayer, J. F., Yamamoto, S. S., & Brosschot, J. F. (2010). The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *International Journal of Cardiology*, *141*(2), 122–131. https://doi.org/10.1016/j.ijcard.2009.09.543

Toichi, M., Sugiura, T., Murai, T., & Sengoku, A. (1997). A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of R–R interval. *Journal of the Autonomic Nervous System*, *62*(1), 79–84. https://doi.org/10.1016/S0165-1838(96)00112-9

Voss, A., Schulz, S., Schroeder, R., Baumert, M., & Caminal, P. (2008). Methods derived from nonlinear dynamics for analysing heart rate variability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1887), 277–296. https://doi.org/10.1098/rsta.2008.0232

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

Wu, S.-D., Wu, C.-W., Lee, K.-Y., & Lin, S.-G. (2013). Modified multiscale entropy for short-term time series analysis. *Physica A: Statistical Mechanics and Its Applications*, *392*(23), 5865–5873. https://doi.org/10.1016/j.physa.2013.07.075

Yan, C., Li, P., Ji, L., Yao, L., Karmakar, C., & Liu, C. (2017a). Area asymmetry of heart rate variability signal. *BioMedical Engineering OnLine*, *16*(1), 112. https://doi.org/10.1186/s12938-017-0402-3

Yan, C., Li, P., Ji, L., Yao, L., Karmakar, C., & Liu, C. (2017b). Area asymmetry of heart rate variability signal. *BioMedical Engineering OnLine*, *16*(1), 112. https://doi.org/10.1186/s12938-017-0402-3

Young, H., & Benton, D. (2015). We should be using nonlinear indices when relating heart-rate dynamics to cognition and mood. *Scientific Reports*, *5*(1), 16619. https://doi.org/10.1038/srep16619

Zahn, D., Adams, J., Krohn, J., Wenzel, M., Mann, C. G., Gomille, L. K., Jacobi-Scherbening, V., & Kubiak, T. (2016). Heart rate variability and self-control—A meta-analysis. *Biological Psychology*, *115*, 9–26. https://doi.org/10.1016/j.biopsycho.2015.12.007

Zhang, J. (2007). Effect of Age and Sex on Heart Rate Variability in Healthy Subjects. *Journal of Manipulative and Physiological Therapeutics*, *30*(5), 374–379. https://doi.org/10.1016/j.jmpt.2007.04.001