**Editor:**

As you'll see both expert reviewers are concerned with the validity of the data used to develop your empirical taxonomy of hrvmeasures. Both reviewers were also supportive of your aim of reviewing the variety of measures and classifying the relevant interpretations from different classes of measures. Reviewer 1 is very clear that resting and active state data should not be considered equivalent. Reviewer 2 agrees and adds concerns with the low sampling frequency of a number of the datasets. Other methodological concerns relevant to the studies are also mentioned. I won't further paraphrase the reviews as they are clearly written. My reading confirmed the points of the reviewers. The Journal would be interested in examining the variety of hrv measures and your approach seems interesting. The need to have appropriate underlying datasets renders this ms unacceptable. The time/effort required to assemble an appropriate database suggests that any revision decision would be inappropriate. To reiterate, this ms must be rejected but we encourage further work that takes into account the reviews. 

**Reviewer: 1**

Comments to the Author  
Using data from 6 databases on heart rate variability (HRV), the authors attempted to classify indices of HRV using a consensus-based meta-clustering method. The analysis is sophisticated and the results are presented in a detailed and effective manner. However, there is a fundamental problem with this study, and the resulting findings cannot be used for future researches.  
They used a mixture of HRV data obtained in controlled laboratory conditions and in free-running walking conditions. The similarity of HRV indices under controlled conditions may reflect similarities in physiological characteristics, whereas the similarity under free-running conditions may contain similarities that are due to responses to physical and mental activities. Two indices that are correlated at rest may be inversely correlated in data that includes a responses to some stimulus. Suppose that index A, which reflects the level of sympathetic nerve activity at rest, has the property of increasing with standing independently of the sympathetic nervous system. If the sympathetic nerve activity level is related to the severity of heart failure, A may be high in mild cases because of the long standing time during daily activities, and A may be low in severe cases with long lying time. As a result, A under free activity is inversely proportional to the sympathetic nerve activity level. The characteristics of HRVindices including their mutual similarity that are obtained under controlled conditions are not applicable directly to the interpretations of HRV under free-running conditions. Therefore, the results of this study with mixed data cannot be applied to HRV under either condition. It is necessary to separate them and analyze only the database for the same measurement conditions.  
  
Specific comment  
There is a study that analyzed the interrelationship of HRV indicators in free-running long-term monitoring using big data, and I think it is related to this study.   
<https://onlinelibrary.wiley.com/doi/full/10.1111/anec.12790>

**Reviewer: 2**

Comments to the Author  
In the current study, Pham et al. evaluated the interrelationship of existing HRV indices using a meta-clustering approach. The authors found that the selected HRV indices all fall into one of three main clusters based upon their features, namely distribution-related, harmony-related, and frequency/complexity related. In general, I found the paper well written and the topic interesting and timely. However, I do have several concerns and suggestions and have listed them below.   
  
Comments:          
1.      The introduction and discussion would benefit from further elaboration of how the various HRV measures are used in psychophysiological research and what they are believed to measure. Additionally, a more nuanced description of HRV would be helpful. The current operationalization of HRV as “reflecting the heart’s ability to effectively regulate and adapt to internal and external environmental changes” is unnecessarily vague and fails to reflect the contribution of the CNS and ANS in this regulation and adaptation.  
  
2.      There is a considerable amount of heterogeneity in experimental conditions and demographics across the various datasets used in this study. What selection criteria were used when selecting databases (i.e. why these specific 6)?   
  
3.      More specificity is needed on how data were collected, pre-processed, and reduced. For example, in the methods the authors describe the raw ECG recordings from each database but later suggest there was not reliable access to the raw ECG recordings. For example:   
  
“online database was not in the form of ECG recordings but sample locations of annotated heartbeats (R-peaks).”  
  
Similarly  
  
“The NeuroKit2 software (Makowski et al., 2021) was used to preprocess the raw ECG  
146 signals (when available).”   
  
This lack of clarity makes it very difficult to evaluate the findings of this study. Furthermore, each data set was comprised of different conditions and recording lengths which have known impacts on many of the HRV measures evaluated in this manuscript, how did the authors take this into account? Did the authors compile the data in standardized epoch lengths or variable epoch lengths?  How did the authors deal with discontinuous segments? Did the authors control for respiration rate in any standardized way?   
  
4.      The authors describe using ECG data with various sampling rates ranging from 128 Hz – 4000 Hz. The authors should address concerns that a considerable amount of their data was sampled at or below the minimum recommended rate for many HRV measures.