Supplementary Material

Section numbers correspond to those in the main text.

2.4 RR interval series outlier detection, PTBDB record exclusion, and SF_{Ex} exclusion

The primary focus of this paper is how SF itself (rather than SF-induced outliers in an RR interval series) influences HRV measures. For this reason, only those PTBDB records and SF_{Ex} values that yielded outlier-free RR interval series (i.e., no missing or spurious R-peaks) for all subjects were of interest. Two types of outliers were considered.

First, an RR interval series may contain "implausibly" long or short intervals (resulting in inflated estimates of HRV), even at the original 1000-Hz sampling rate. Such outliers were detected using the widely used algorithm developed by Berntson *et al* (1990) and implemented by Kaufmann *et al* (2011). Briefly, this algorithm uses thresholds derived from various percentile-based statistics calculated from each subject's RR interval distribution, and was shown to have excellent detection accuracy (and a low incidence of falsely detected outliers) in a sample of 60 healthy subjects (each with an RR interval series of 256 beats, and with ECGs sampled at 1000 Hz). Any PTBDB record that had an outlier at the original 1000-Hz sampling rate was eliminated, leaving a set of 47 unique records (listed below in **Table S1**).

Patient	Record	Patient	Record	Patient	Record	Patient	Record	Patient	Record
104	s0306lre	169	s0329lre	242	s0471_re	243	s0472_re	263	s0499_re
105	s0303lre	170	s0274lre	229	s0453_re	244	s0473_re	264	s0500_re
116	s0302lre	172	s0304lre	233	s0483_re	245	s0474_re	267	s0504_re
117	s0291lre	174	s0325lre	234	s0460_re	246	s0478_re	276	s0526_re
122	s0312lre	180	s0374lre	235	s0461_re	247	s0479_re	277	s0527_re
131	s0273lre	182	s0308lre	236	s0462_re	248	s0481_re	279	s0534_re
155	s0301lre	184	s0363lre	237	s0465_re	251	s0486_re	284	s0552_re
156	s0299lre	185	s0336lre	238	s0466_re	252	s0487_re		
165	s0322lre	198	s0402lre	240	s0468_re	255	s0491_re		
166	s0275lre	214	s0436_re	241	s0469_re	260	s0496_re		

Table S1. PTDBD patient numbers and unique records used in the present study.

Second, for each of the 47 subjects records that were outlier-free at 1000 Hz, two statistics were calculated (cf. Kohler et al., 2002) for each subject at each SF_{Ex}: the *true positive rate* [TPR = *TP* / (*TP* + *FN*)] and the *positive predictive value* [PPV = *TP* / (*TP* + *FP*)]). With respect to R-peak detection accuracy, *TP* is the number of true positives (i.e., R-peaks in R_{Ex} with a corresponding R-peak in R_{Or}), *FP* is the number false positives (i.e., R-peaks in R_{Ex} without a corresponding R-peak in R_{Or}), and *FN* is the number of false negatives (i.e., R-peaks in R_{Or} without a corresponding R-peak in R_{Ex}). Here, "corresponding" is arbitrarily defined as falling within window of ± 50 ms. (As explained above, the method of R-peak

refinement used here will yield identical false positive and false negative counts in both the nonrefined and refined R-peak series.)

Figure S1 summarizes the result of TPR and PPV analyses, plotting the number of subjects (out of 47) whose data achieved perfect (= 1.0) sensitivity (dashed black line) or positive predictive value (solid gray line) for R-peak detection at each SF_{Ex}. The results are straightforward: for this particular sample of healthy subjects, R-peak detection was 100% accurate in all subjects down to SF_{Ex} = 71.43 Hz. (To be doubly safe, the Berntson *et al* 1990 outlier detection algorithm was run on each R-peak series at the lowest SF_{Ex}; no outliers were detected.) Thus, further exploration of the relationship between HRV measures and SF_{Ex} values > 71 Hz can be performed (in the present data set) without RR interval outliers influencing the results.

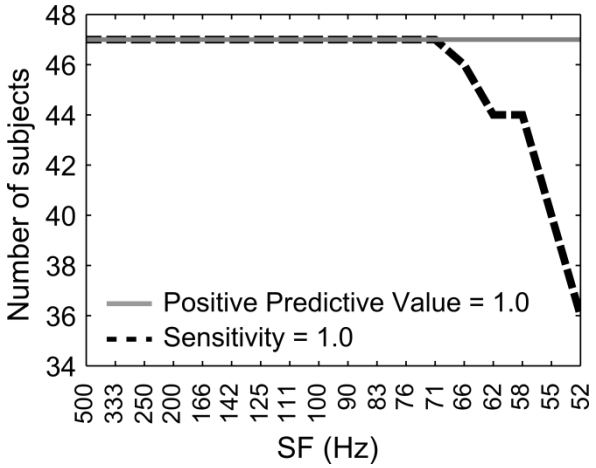

Figure S1. A tally of subjects with perfect sensitivity (dashed black line) or positive predictive value (solid gray line) as a function of SF_{Ex}. All 47 subjects had R-peak detection rates that perfectly matched that of the original data (at SF_{Or} = 1000 Hz) down to SF_{Ex} = 71.42 Hz.

2.5.1 Time-domain outcome measures

The six time-domain measures here exhibit varying degrees of relation to each other, as is made clear by an examination of their formulas (adapted from Tarvainen and Niskanen 2012, Chapter 3). Given a series of N R-to-R intervals (RR), let $\Delta RR_j = RR_{j+1} - RR_j$, and

$$\mathrm{AVNN} = \frac{1}{N}\sum_{j=1}^{N} \mathrm{RR}_j \qquad \text{(Eq. S1)}$$

$$\mathrm{SDNN} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N}\left(\mathrm{RR}_j - \mathrm{AVNN}\right)^2} \qquad \text{(Eq. S2)}$$

$$\mathrm{SDSD} = \sqrt{E\left\{\Delta \mathrm{RR}_j^2\right\} - E\left\{\Delta \mathrm{RR}_j\right\}^2} \qquad \text{(Eq. S3)}$$

$$\mathrm{RMSSD} = \sqrt{\frac{1}{N-1}\sum_{j=1}^{N-1}\left(\mathrm{RR}_{j+1} - \mathrm{RR}_j\right)^2} \quad \text{(Eq. S4)}$$

$$\mathrm{SD1} = \sqrt{\frac{1}{2}\mathrm{SDSD}^2} \quad \text{(Eq. S5)}$$

$$\mathrm{SD2} = \sqrt{2\mathrm{SDNN}^2 - \frac{1}{2}\mathrm{SDSD}^2} \quad \text{(Eq. S6)}$$

In particular, under the "ideal" condition of RR interval stationarity being met, SDSD, RMSSSD, and SD1 are mathematically equivalent (Brennan *et al* 2001). In the past, some authors have preferred to report SDSD and RMSSD (e.g., Acharya *et al* 2004, Dekker *et al* 2000, McNames and Aboy 2006, Kim *et al* 2007) whereas others have preferred to report RMSSD and SD1 (e.g., Gamelin *et al* 2006, Guzik *et al* 2007, Mourot *et al* 2004). We included all three measures in the current study so as to make our methods and results to be applicable to both cases.

2.5.2 Frequency-domain outcome measures

A note about the "simplified" definition of LF_{nu} and HF_{nu}. For five-minute estimates of spectral power, the Task Force 1996 paper (Table 2) defines the denominator for normalized units as "Total Power – VLF Power", where "Total Power" is "the variance of NN intervals over the temporal segment" (i.e., SDNN) or "approximately ≤ 0.4 Hz"; and "VLF Power" is spectral power "≤ 0.04 Hz" (by convention, 0.003 to 0.04 Hz; cf. Tarvainen & Niskanen 2012, p. 14). However, the PTBDB records have a duration of slightly less than *two* minutes (~ 116 s on average) rather than five minutes, the lowest resolvable frequency (per Clifford 2006) is roughly 1 / 116 ≈ 0.009 Hz, thus introducing systematic error into the lower portion of the VLF band. For this reason, the "simplified" denominator used by Smith *et al* 2013 (i.e., "LF_{au} + HF_{au}") was deemed better suited for the present study.

2.6.2 Quantifying FPR

In order for the proposed Monte Carlo based analysis of observed false positive rates (FPRs) for two-sample *t*- and *U*-tests to be statistically valid, it is expected that repeated subsamples (here, two samples of 20) from a single pool of data (here, 47 total subjects) should yield a FPR close to the theoretical $\alpha = .05$ (for *t*-test) or $\alpha = .0491$ (i.e., using the *U*-test's exact critical value when both $n = 20$). Put in terms of the present experimental question, if both samples of HRV values were obtained at the *same* sampling frequency

(SF), the FPR should fall near α. (If the two samples were at *different* SFs, higher FPRs would not be surprising.) Since the FPR itself is calculated as the number of false positives out of the total number of iterations, more iterations should yield a more accurate FPR. But how many iterations is enough?

To answer this question, the following simulation was performed. 1000 independent draws of 47 random values from the normal distribution (mean = 0, standard deviation = 1) were taken. Each set of 47 values was submitted to 100,000 random permutations (iterations). For each iteration, two samples of 20 values were extracted, a *t*- and *U*-test were performed, and the binary significance of the test (i.e., $p < .05$ or $p \geq .05$) was recorded.

Figure S2 shows a percentile summary of the distribution of observed FPRs (*y*-axis), as calculated using an increasing number of subsampling iterations (*x*-axis: from 1,000 to 100,000) for *t*-tests (left panel) and *U*-tests (right panel). As expected, as the observed FPR was calculated using more iterations, its value gravitated closer to the theoretical *t*-test or *U*-test α . With 100,000 iterations, 90% of all *t*-test FPR values fell within the interval [.0479, .0513], and 90% of all *U*-test FPRs fell within the interval [.0478, .0505]. Errors in FPR estimation of roughly ± .002 from the expected α values for both *t*- and *U*-tests was deemed acceptable; thus, 100,000 random subsamples were used to compare differences in HRV values at the original SF (1000 Hz) with each downsampled SF.

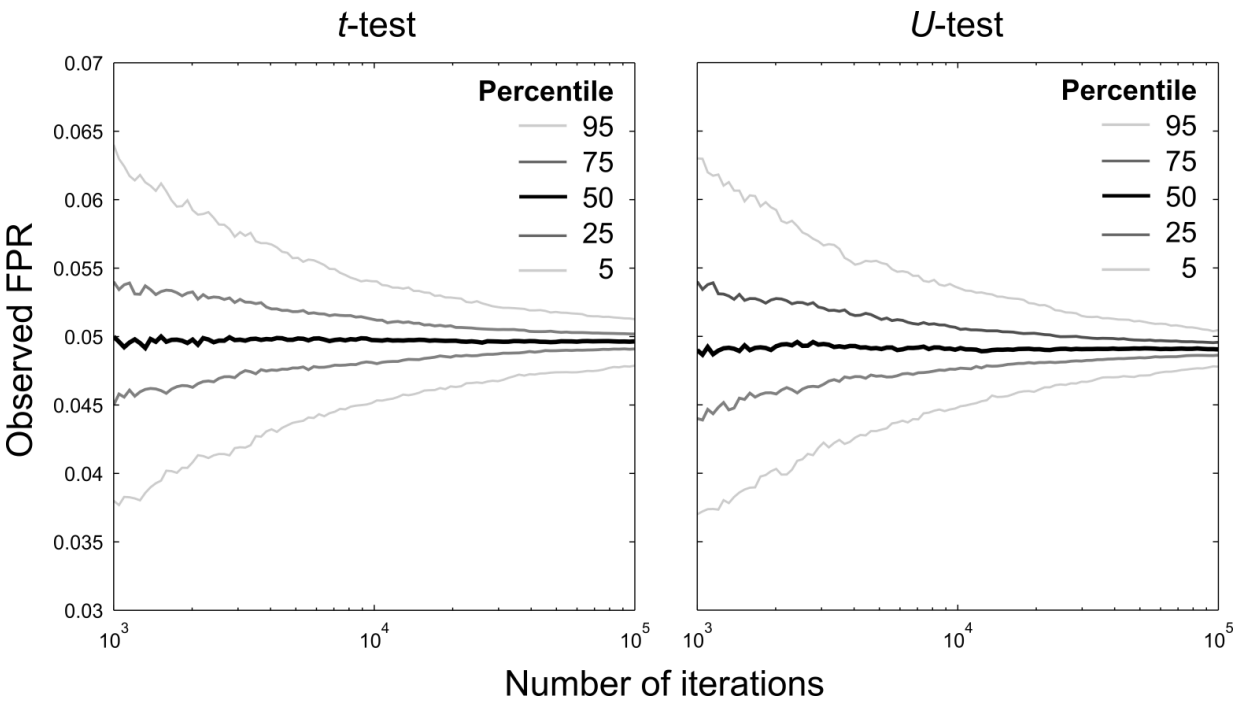

Figure S2. Distributions of observed FPR values (summarized as percentiles along the *y*-axis) calculated from an increasing number of subsampling iterations (*x*-axis), and separately for *t*-tests (left panel) and *U*-tests (right panel).

3.1 SF, R-peak interpolation, and HRV measures

Figure S3 visualizes the relationship between SF and R-peak temporal error (i.e., the error between R-peak timestamps at each SF_{Ex} relative to timestamps at SF_{Or}). Histogram outlines, pooled across all 47 subjects (roughly 6200 values per histogram) are shown for distributions of R-peak temporal errors at five representative SF_{Ex} (500 Hz, 200 Hz, 100 Hz, 66.67 Hz, 52.63 Hz). Separate histograms are shown for nonrefined R-peak errors (gray lines) and refined R-peak errors (black lines).

When R-peak refinement was not utilized, the "spread" of temporal error values along the *x*-axis increases as SF_{Ex} decreases, with a fairly uniform distribution of signed error values. The width of the "flat" portion of each histogram roughly corresponds to the duration (in ms) between successive samples at SF_{Ex}; for example, 19 ms between samples when SF_{Ex} = 52.63 Hz. By contrast, when R-peak refinement *was* utilized, the amount of temporal error spread was greatly reduced.

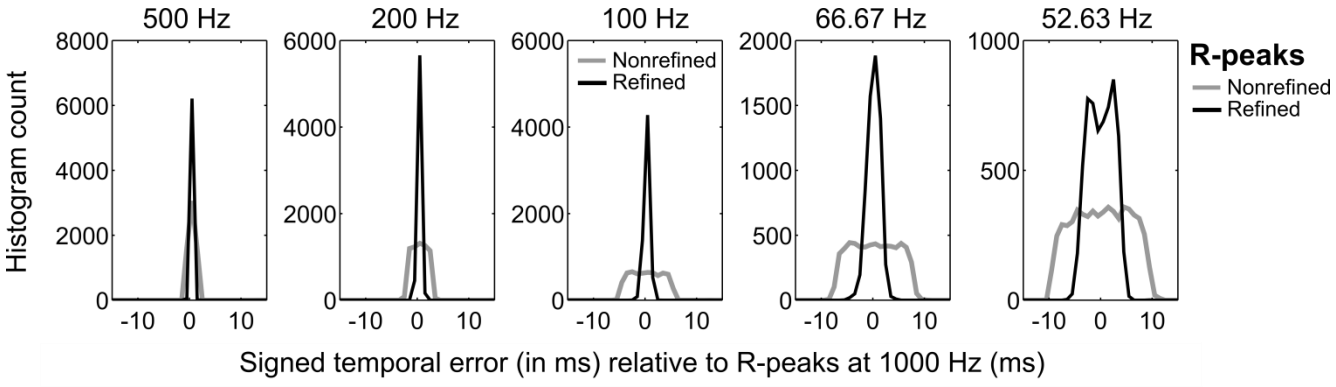

Figure S3. Histogram outlines of signed temporal errors of R-peak timestamps at different SFs (separate subplots) relative to R-peak timestamps at the original 1000 Hz SF. Separate plots are made for nonrefined R-peaks (gray) and refined R-peaks (black).

Supplemental References

Acharya U R, Kannathal N, Sing O W, Ping L Y, Chua T and others 2004 Heart rate analysis in normal subjects of various age groups *Biomed Eng Online* **3** 24

Berntson G G, Quigley K S, Jang J F and Boysen S T 1990 An Approach to Artifact Identification: Application to Heart Period Data *Psychophysiology* **27** 586–98

Brennan M, Palaniswami M and Kamen P 2001 Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Trans. Biomed. Eng.* **48** 1342–7

Clifford G D 2006 ECG statistics, noise, artifacts, and missing data *Advanced Methods and Tools for ECG Data Analysis* ed G D Clifford, F Azuaje and P E McSharry (Boston: Artech House) pp 55–99

Online: http://ecgcode.googlecode.com/git-history/119bec96b306e4eeb0fb3da1f07c77bfba1d54ad/ECG_book_ch3.pdf

Dekker J M, Crow R S, Folsom A R, Hannan P J, Liao D, Swenne C A and Schouten E G 2000 Low heart rate variability in a 2-minute rhythm strip predicts risk of coronary heart disease and mortality from several causes The ARIC Study *Circulation* **102** 1239–44

Gamelin F X, Berthoin S and Bosquet L 2006 Validity of the polar S810 heart rate monitor to measure RR intervals at rest *Med. Sci. Sports Exerc.* **38** 887

Guzik P, Piskorski J, Krauze T, Schneider R, Wesseling K H, Wykretowicz A and Wysocki H 2007 Correlations between the Poincare plot and conventional heart rate variability parameters assessed during paced breathing *J. Physiol. Sci.* **57** 63–71

Kaufmann T, Sütterlin S, Schulz S M and Vögele C 2011 ARTiiFACT: a tool for heart rate artifact processing and heart rate variability analysis *Behav. Res. Methods* **43** 1161–70

Kim K K, Lim Y G, Kim J S and Park K S 2007 Effect of missing RR-interval data on heart rate variability analysis in the time domain *Physiol. Meas.* **28** 1485

Kohler B-U, Hennig C and Orglmeister R 2002 The principles of software QRS detection *IEEE Eng. Med. Biol. Mag.* **21** 42–57

McNames J and Aboy M 2006 Reliability and accuracy of heart rate variability metrics versus ECG segment duration *Med. Biol. Eng. Comput.* **44** 747–56

Mourot L, Bouhaddi M, Perrey S, Rouillon J-D and Regnard J 2004 Quantitative Poincare plot analysis of heart rate variability: effect of endurance training *Eur. J. Appl. Physiol.* **91** 79–87

Smith A-L, Owen H and Reynolds K J 2013 Heart rate variability indices for very short-term (30 beat) analysis. Part 1: survey and toolbox *J. Clin. Monit. Comput.* 1–8

Tarvainen M P and Niskanen J-P 2012a Kubios HRV version 2.1 user's guide *Dep. Phys. Univ. Kuopio Kuopio Finl.* Online: http://kubios.uef.fi/media/Kubios_HRV_2.1_Users_Guide.pdf