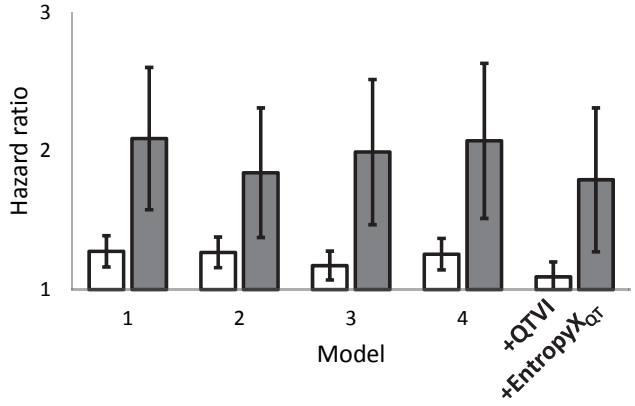
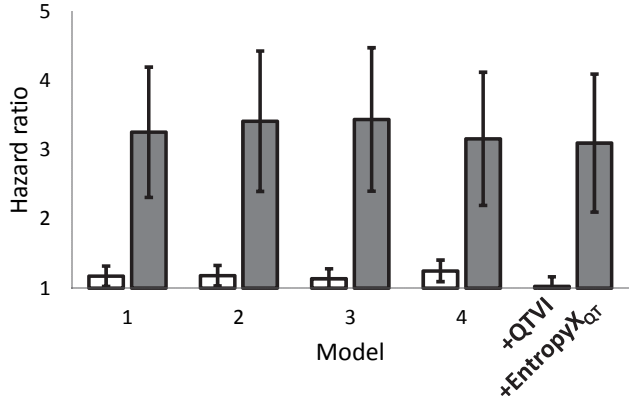
SUPPLEMENTARY FIGURE 1: Effect of EntropyX$_{QT}$ and QT variability index (QTVI) on incrementally adjusted proportional hazards ratio in models 1-4.

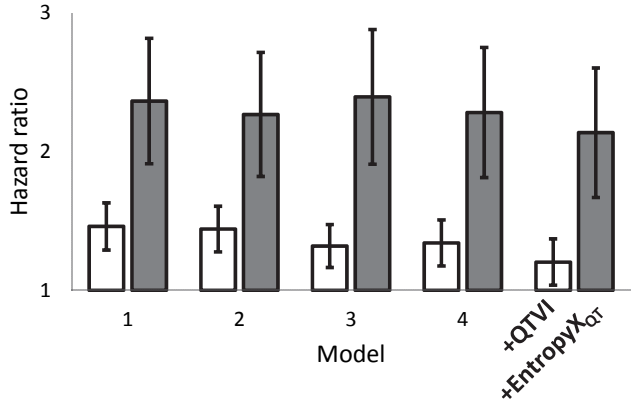
A. Composite events



B. Appropriate ICD shocks



C. All-cause mortality



The hazard ratios corresponding to hierarchical models for composite events, appropriate ICD shock and all-cause mortality are shown for QT variability index (QTVI, unshaded bars) and EntropyX$_{QT}$ (shaded). This analysis was based on 45 ± 24 months of follow-up of 816 patients; there were 134 events for appropriate ICD shocks, 168 events for all-cause mortality and 300 composite events.

Model 1 consisted of QTVI (unshaded) or EntropyX$_{QT}$ (shaded).

Model 2 consisted of QTVI or EntropyX$_{QT}$, as well as demographics (age at implant, gender, race), medical history (paroxysmal atrial fibrillation, diabetes mellitus, ischemic cardiomyopathy, NYHA class) and prescribed medication (beta blocker).

Model 3 consisted of all covariates in model 2 as well as mean arterial pressure, left ventricular ejection fraction, serum BUN levels, and biomarkers (hsCRP and NT-proBNP).

Model 4 consisted of all covariates in model 3 as well as 5-min ECG time and frequency domain measures of heart rate variability (heart rate SDNN and LF:HF ratio).

In the last step, EntropyX$_{QT}$ or QTVI were added to model 4 for QTVI (unshaded) or EntropyX$_{QT}$ (shaded), respectively.

SUPPLEMENTARY METHODS

I. Conceptual description of EntropyX$_{QT}$

The concept of **entropy** and **entropy rate** are related to thermodynamic entropy[1] and can be used to nonlinearly quantify the dynamics of physiological signals under the framework of continuous random variables and stochastic processes[2, 3]. We optimized the **sample entropy** measure[4, 5] and developed **EntropyX$_{QT}$**, which is the negative natural logarithm of the conditional probability that a sequence of QT intervals of length ***m***, which repeats itself over the QT interval time series by matching within an arbitrary tolerance ***r***, will also repeat if the length of the sequence was increased by incorporating the next QT interval, ***m+1***. **EntropyX$_{QT}$** is calculated as the **-ln(B/A) + ln(2*r*)**, where **A** denotes the total number of matches of length ***m***, **B** is the subset of **A** that matches for both ***m*** and ***m+1***, and **B/A** represents the conditional probability that subsequent points of a set of closely matching ***m*** intervals also remain close and match.

For example, in a time series consisting of 300 QT intervals (QT_1,...,QT_{300}), consider that the first sequence of *four* consecutive QT intervals, i.e., QT_1, QT_2, QT_3 and QT_4 (***m***=4) recurs ten times (**A_1**=10) elsewhere in the time series within a certain optimal tolerance (***r***=2 msec); the second sequence, QT_2,...,QT_5, does not repeat (**A_2**=0); and the third sequence, QT_3,...,QT_6 repeats twenty times (**A_3**=20).

Consider also that the first sequence of *five* consecutive QT intervals, i.e., QT_1,..., QT_5 (***m+1***=5) in the same time series recurs twice (**B_1**=2) within the same ***r***; the second sequence, QT_2,..., QT_6, does not repeat (**B_2**=0; as expected from A_2=0); and the third set, QT_3,..., QT_7, repeats three times (**B_3**=3).

Up to this point, the calculated conditional probability [-ln(**B_1**+ **B_2**+ **B_3**)/(**A_1**+**A_2**+**A_3**) = -ln(5/30)] is 1.79 and EntropyX$_{QT}$ [1.79 + ln(4)] is 3.18. For the final calculation of EntropyX$_{QT}$, this process would continue up to the last set of QT intervals, i.e., QT_{96},...,QT_{100}.

For regularly repeating QT intervals, the probability of recurrence will approach 1 and EntropyX$_{QT}$ will approach ln(2r). For nonrepeating QT intervals, the probability of recurrence will approach zero, and EntropyX$_{QT}$ will become infinitely large. The accuracy of the conditional probability B/A is dependent on the magnitude of both **A** and **B**. Generally, smaller values of ***r*** lead to higher and less confident entropy estimates because of falling numbers of matches of length ***m*** and, to an even greater extent, matches of length ***m+1***. Whereas this becomes an increasing concern for very short records, this is less of a concern for the record lengths used in this study[2]. In addition, inaccurate probability estimates are typically avoided by EntropyX$_{QT}$ because ***r*** is optimally varied for each time series until a specified number of matches is attained for **B**, analogous to varying the bin widths of histograms to optimally depict the distribution of a particular data set. Because EntropyX$_{QT}$ normalizes for ***r***, estimates made with different values of ***r*** measure the same inherent dynamics and can be directly compared between time series (see next section).

II. Calculation of EntropyX_{QT}

Mathematically expressed[2-5], the entropy (H) of a continuous random variable X with possible values $\{x_1, x_2, \dots, x_m\}$ and probability density function f is defined as

$$H(X) = \sum -\ln[f(X)] = \int_{-\infty}^{+\infty} -\ln[f(x)] \cdot f(x) \cdot dx \quad \textbf{(1)}$$

If X has variance of σ^2, then $Y = x \div \sigma$ has variance of 1 and density of $\sigma f(\sigma y)$, and the entropy of Y will be related to the entropy of X by

$$H(Y) = \int_{-\infty}^{+\infty} -\ln[\sigma f(\sigma y)] \cdot \sigma f(\sigma y) \cdot dy = H(X) - \ln[\sigma] \quad \textbf{(2)}$$

and as such, demonstrates that lower entropy indicates lower variance or higher uncertainty.

Another important property of entropy is provided by the inequality

$$H(X) \leq \frac{\ln[2\pi e] + \ln[\sigma^2]}{2} = H(\sigma Z) \quad \textbf{(3)}$$

where Z is a random variable with a standard Gaussian distribution and this Gaussian distribution has maximum entropy among all random variables with the same variance. As such, an estimate of entropy that is substantially lower than this upper bound for a random sample (with sample variance used as an estimate of σ^2) would be consistent with a non-Gaussian underlying distribution, a characteristic often seen with repetitive patterns in multimodal cardiac arrhythmias, e.g., bigeminy, trigeminy.

By letting X denote a random sequence $x_1, x_2, \dots, x_n$, the entropy rate of X is defined as

$$H(X) = \lim_{n \to \infty} \frac{H(x_1, x_2, \dots, x_n)}{n} \quad \textbf{(4)}$$

where the joint entropy of m random variables $x_1, x_2, \dots, x_m$ is defined as

$$H(x_1, x_2, \dots, x_m) = E[-\ln(f(x_1, x_2, \dots, x_m))] \quad \textbf{(5)}$$

where f is the joint probability density function. For stationary processes, an equivalent definition is

$$H(X) = \lim_{m\to\infty} H_m(X) = \lim_{m\to\infty} H_m(x_{m+1}|x_1, x_2, \dots, x_m) \quad \textbf{(6)}$$

Therefore, the entropy rate is the entropy of the conditional distribution of the present observation given the past. For independent and identically distributed random sequences, the entropy rate reduces to the entropy of the common distribution. Estimating the entropy rate for sequences depends on estimates of its densities of order m, also referred to as the embedding dimension. For example, if $x_1, x_2, \dots, x_n$ denotes a stationary random sequence, $x_i(m)$ denotes the template consisting of the $m \times 1$ vector $x_{i-m+1}, x_{i-m}, \dots, x_i$, the whole sequence is denoted by $x_n = x_n(n)$, and the limiting infinite sequence is denoted by $x = x_\infty$, then the sequence $x_m(m), x_{m+1}(m), \dots, x_n(m)$ is not independent and several existing methods for analysis of independent vectors can be applied. More specifically, the m^{th} order probability density function of the sequence f and entropy (equation 5) can be empirically estimated.

In order to be able to assess significant differences in entropy rate values for different time series and conduct meaningful statistical inferences, the standard error of the entropy rate estimates is an important consideration. In general, if $\hat{\sigma}^2$ denotes the sample variance and $\hat{c}_k$ denotes the sample correlation coefficient at lag k calculated using a divisor of $n_k = n - k$, the number of pairs of observations at lag k, and K is the maximum lag at which the random process has significant correlation, then the variance of the entropy estimate can be estimated by

$$\sigma_{\hat{H}}^2 = \frac{\hat{\sigma}^2}{n}\left(1 + 2\sum_{k=1}^{K} n_k \hat{c}_k\right) \quad \textbf{(7)}$$

The calculation of **EntropyX$_{QT}$** is based upon these fundamental calculations (equations 1-7). More specifically, EntropyX$_{QT}$ is calculated by first forming a set of vectors u_j of length m

$$u_j = \left(QT_j, QT_{j+1}, \dots, QT_{j+m-1}\right), \quad j = 1,2, \dots N - m + 1$$

where m represents the embedding dimension and N is the number of measured QT intervals. The distance between these vectors is defined as the maximum absolute difference between the corresponding elements

$$d(u_j, u_k) = max\left[\left|QT_{j+n} - QT_{k+n}\right| | n = 0, \dots, m-1\right]$$

For each u_j, the relative number of vectors u_k for which $d(u_j, u_k) \leq r$ is calculated as

$$C_j^m(r) = \frac{\text{number of}\left[u_k \mid d(u_j, u_k) \leq r\right]}{N - m + 1} \forall k \neq j$$

with values of $C_j^m(r)$ ranging between 0 and 1. Average of $C_j^m(r)$ yields

$$C^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} C_j^m(r)$$

and the general form of the equation for calculating the entropy from the QT time series is

$$\mathrm{EntropyX_{QT}}\ (m, N, n, r) = \left[-\ln \frac{C_{QT}^{m+1}(r_{QT})}{C_{QT}^{m}(r_{QT})} + \ln(2 \cdot r_{QT}) \right] \quad \textbf{(8)}$$

where the embedding dimension or template length (m) ≥ 3, the number of sampled intervals per bin of the time series (N) is ≥20, the sufficient number of matches n ≥ (N ÷ 5), r is the calculated tolerance for a given N that satisfies the specified n but without perfect matches, r represents the tolerance designated for subsequent iterative calculations, $C_{QT}^{m}(r_{QT})$ represents the total number of matches within r of length m in the QT time series, $C_{QT}^{m+1}(r_{QT})$ represents the total number of matches within r of length $m+1$ in the QT time series.

The value of $\boldsymbol{r}$ is an important factor for determining the underlying dynamics of a segment of intervals. If $\boldsymbol{r}$ is too small (i.e., smaller than the typical noise amplitude), then a group of $\boldsymbol{m}$ intervals that are similar shall fail to match. However, if $\boldsymbol{r}$ is too large, there will be a loss in discriminating power simply because the group of intervals will look similar to one another given sufficiently lax matching conditions. The ideal condition would be to vary $\boldsymbol{r}$ with the scale of signal noise such that $\boldsymbol{r}$ is as small as possible for searching for order in the dynamics while ensuring the number of matches remains large enough to ensure precise statistics. This is performed using an iterative algorithm and is analogous to varying the bin widths of a histogram to optimally describe its distribution.

Because r is varied to ensure $n = N \div 5$ matches, the calculation of $\mathrm{EntropyX_{QT}}$ depends primarily on two parameters: $\boldsymbol{m}$, the length of vectors considered in the calculation, given ***N*** data points. We performed extensive sensitivity analyses to validate the stability and reproducibility of the EntropyX$_{QT}$ measure in 6,131 asymptomatic adult human subjects in the Sleep Heart Health Study (data not shown). Sensitivity analyses in PROSe-ICD revealed the optimal parameters for calculating EntropyX$_{QT}$ for the prediction of SCD is m = 4 and N = 40 (**Figure 1**). Results of these sensitivity analyses also indicated a large margin of safety for the adjusted hazard ratio. These data also indicated that as m approaches 0 (lower complexity), there is a sharp decrease in the adjusted hazard ratio, consistent with the lack of independent predictive value of conventional measures of variability in this cohort.

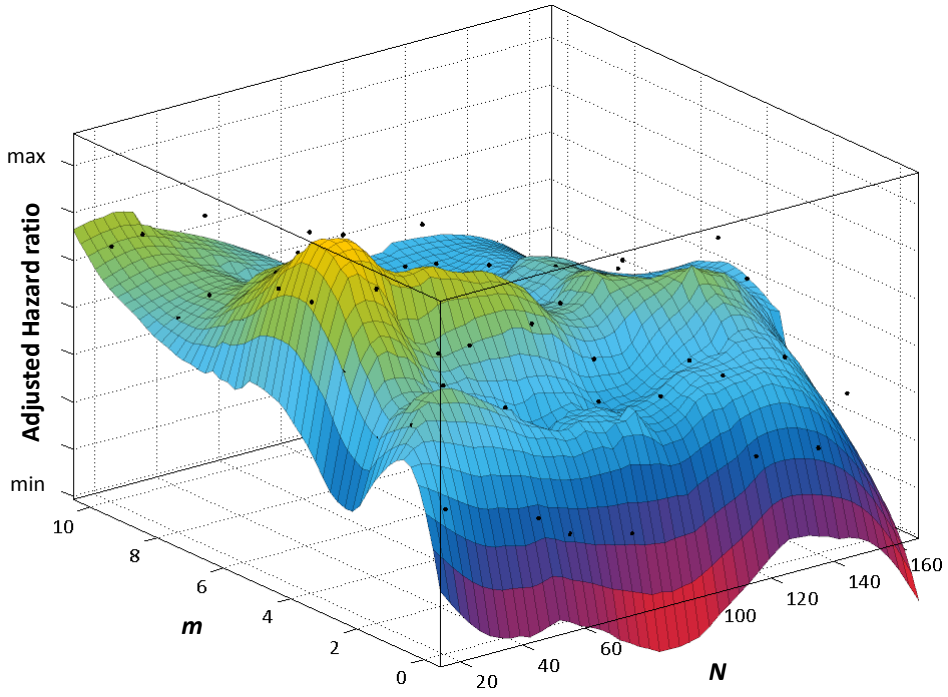


FIGURE 1. Topographic representation of the adjusted hazard ratios for EntropyX$_{QT}$ as a function of m and N.

EntropyX$_{QT}$ provides a global measure of the probabilistic self-similarity of the QT time series data. Smaller values of EntropyX$_{QT}$ indicate a greater likelihood that similar patterns of measurements will be followed by additional similar measurements. If the time series is highly irregular, the occurrence of similar patterns will not be predictive for the following measurements and the EntropyX$_{QT}$ value will be relatively large. The novel features of EntropyX$_{QT}$ include nonlinear quantification of the dynamics of cardiac repolarization while ensuring confident probability estimates and interpreting quadratic entropy rate as a measure of Gaussian white noise[6]. Unlike prior strategies such as SampEn[4, 7] or ApEn[2, 3, 7], EntropyX$_{QT}$ is insensitive to both the degree of tolerance allowed for matching templates and to the presence of outlying points[8]. Unlike ApEn, frequency domain measures or geometric measures such as Poincaré plots[9], EntropyX$_{QT}$ is accurate in short time series. Unlike conventional measures of variability[9, 10], there is no requirement that the data be stationary or preprocessed for the calculation of EntropyX$_{QT}$.

References

1. Toyabe S, Sagawa T, Ueda M, Muneyuki E, Sano M. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. Nat Phys 2010;**6**(12):988-992.

2. Pincus SM. Approximate entropy as a measure of system complexity. Proc Natl Acad Sci U S A 1991;**88**(6):2297-301.

3. Pincus SM, Goldberger AL. Physiological time-series analysis: what does regularity quantify? Am J Physiol 1994;**266**(4 Pt 2):H1643-56.

4. Richman JS, Lake DE, Moorman JR. Sample entropy. Methods Enzymol 2004;**384**:172-84.

5. Demazumder D, Lake DE, Cheng A, Moss TJ, Guallar E, Weiss RG, Jones SR, Tomaselli GF, Moorman JR. Dynamic analysis of cardiac rhythms for discriminating atrial fibrillation from lethal ventricular arrhythmias. Circ Arrhythm Electrophysiol 2013;**6**(3):555-61.

6. Lake DE. Improved entropy rate estimation in physiological data. Conf Proc IEEE Eng Med Biol Soc 2011;**2011**:1463-6.

7. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000;**278**(6):H2039-49.

8. Lake DE. Renyi entropy measures of heart rate Gaussianity. IEEE Trans Biomed Eng 2006;**53**(1):21-7.

9. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation 1996;**93**(5):1043-65.

10. Monfredi O, Lyashkov AE, Johnsen A-B, Inada S, Schneider H, Wang R, Nirmalan M, Wisloff U, Maltsev VA, Lakatta EG, Zhang H, Boyett MR. Biophysical Characterization of the Underappreciated and Important Relationship Between Heart Rate Variability and Heart Rate. Hypertension 2014;**64**(6):1334-1343.