Sample Entropy in Electrocardiogram During Atrial Fibrillation

Takuya Horie,* Naoto Burioka,† Takashi Amisaki‡ and Eiji Shimizu§

*Division of Clinical Laboratory, Tottori University Hospital, Yonago 683-8504, Japan, †Department of Pathological Science and Technology, School of Health Science, Tottori University Faculty of Medicine, Yonago 683-8503, Japan, ‡Department of Biological Regulation, School of Health Science, Tottori University Faculty of Medicine, Yonago 683-8503, Japan, §Division of Medical Oncology and Molecular Respirology, Department of Multidisciplinary Internal Medicine, School of Medicine, Tottori University Faculty of Medicine, Yonago 683-8504, Japan

ABSTRACT

Background Atrial fibrillation (AF) is an arrhythmia commonly encountered in clinical practice. There is a high risk of thromboembolism in patients with AF. Nonlinear analyses such as electroencephalogram (EEG), electrocardiogram (ECG), and respiratory movement have been used to quantify biological signals, and sample entropy (SampEn) has been employed as a statistical measure to evaluate complex systems. In this study, we examined the values of SampEn in ECG signals for patients with and without AF to measure the regularity and complexity.

Methods ECG signals of lead II were recorded from 34 subjects without arrhythmia and 15 patients with chronic AF in a supine position. The ECG signals were converted into time-series data and SampEn was calculated.

Results The SampEn values for the group without arrhythmia were 0.252 ± 0.114 [time lag (τ) = 1] and 0.533 ± 0.163 (τ = 5), and those for the chronic AF group were 0.392 ± 0.158 (τ = 1) and 0.759 ± 0.246 (τ = 5). The values of SampEn were significantly higher in the group with chronic AF than in the group without arrhythmia (P < 0.01 for τ = 1, P < 0.004 for τ = 5). The constructed three-dimensional vectors were plotted in time-delayed three-dimensional space. We used time lags of τ = 5 and τ = 1. The shape of the loops of the three-dimensional space was better for τ = 5.

Conclusion The values of SampEn from ECG for chronic AF patients were higher than for subjects without arrhythmia, suggesting greater complexity for the time-series from chronic AF patients. SampEn is considered a new index for evaluating complex systems in ECG.

Key words atrial fibrillation; electrocardiogram; sample entropy; time-series data

Atrial fibrillation (AF) is caused by a combination of many mechanisms such as multiple-circuit reentry and ectopic tachycardia. AF is diagnosed by confirming characteristic electrocardiogram (ECG) findings such as the appearance of f waves and the absolute irregularity of the R-R interval. AF is categorized as paroxysmal, persistent, or permanent depending on the cause and duration of the attack and is a major risk factor for ischemic stroke. Treatments for AF include drug therapy, electrical defibrillation, surgery, and catheter ablation. However, treatment depends on the patient’s background and the type of AF. Therefore, appropriate and comprehensive treatment is necessary.

Analytical methods to examine the nonlinear characteristics of biomedical signals were recently developed. The correlation dimension, maximal Lyapunov exponent, approximate entropy (ApEn), sample entropy (SampEn), etc. are used as statistics to evaluate complex systems. SampEn can be evaluated from time-series data with irregular fluctuations and has been used to examine the complexity of time-series data for biological signals such as electroencephalogram (EEG) and respiratory movement. For example, SampEn for EEG was proposed to monitor the depth of anesthesia for patients during surgery, and SampEn has been found to be more feasible for real-time detection.

In this study, we analyzed SampEn to investigate the irregularity of ECG signals for AF. Additionally, we developed a new method for plotting ECG signals with time-delayed coordinates.

MATERIALS AND METHODS

Thirty-four subjects (21 men) without arrhythmia aged between 60 and 86 years (72.2 ± 8.0, mean ± SD), and 15 patients (11 men) with AF aged between 69 and 91 years (79.7 ± 7.0) participated in this study (Table 1). Nine patients with chronic AF were taking warfarin, and 1 patient was taking dabigatran. Eight patients received
antiarrhythmic medication, specifically 6 patients were taking methyl digoxin, 4 patients were taking bisoprolol, 1 patient was taking disopyramide, 1 patient was taking verapamil (with duplication). β-blockers were not used by any of the subjects without arrhythmia. This study was approved by the Tottori University ethical board (#1899), and written informed consent was obtained from each subject and patient.

**Electrocardiogram measurement**

ECG measurement (Auto Cardiner FCP-2201, FUKUDA Denki, Tokyo, Japan) was performed for 3 min in the supine position after the subjects rested for 5 min at Tottori University Hospital and Hitachi Kinen Hospital from May 2012 to December 2015. The ECG signals from lead II were digitized with an A/D converter (PowerLab, ADInstruments, Tokyo, Japan) at a sampling rate of 200 Hz (0.005 s), without filtering. The SampEn for ECG signals was calculated for 34 subjects without arrhythmia and 15 patients with chronic AF, MatLab software (MathWorks, Natick, MA) was used to compute the SampEn values. We selected five different 20 s artifact-free epochs of ECG for each subject and patient. The SampEn of five epochs was averaged and used for analysis.

**Calculation of SampEn**

SampEn is a statistical measure used to evaluate the complexity of physiological time-series signals. SampEn was introduced by Richman and Moorman as a measure of system complexity and has been used to analyze biomedical signals that pick up noise easily. It is derived from the correlation integral $C_m,i(r)$, which represents the proportion of points within a distance $r$ from the $i$th point when the signal is embedded in an $m$-dimensional space. The correlation integral $C_m,i(r)$ is defined as

$$C_m,i(r) = \left( N - (m-1) \right)^{-1} \cdot \sum_{j=1}^{N} \Theta(r - |X_i - X_j|)$$

where $\Theta(t)$ is the Heaviside function: $\Theta(t) = 1$ if $t \geq 0$, and $\Theta(t) = 0$ otherwise.

The vectors $X_i$ and $X_j$ are vectors constructed from the time-series $\{x(1), x(2), ..., x(N)\}$ as

$$X_i = (x(i), x(i + \tau), ..., x(i + (m-1)\tau)),$$

$$X_j = (x(j), x(j + \tau), ..., x(j + (m-1)\tau))$$

where $\tau$ is the time lag. In the time-delayed coordinates system, $\tau$ is a positive integer. The actual time of delay is $\tau \times \Delta t$, where $\Delta t$ is the sampling time. A modified version of SampEn with $\tau$ has been reported.

SampEn is given by:

$$SampEn = -\ln \left[ \frac{1}{N - m \tau} \sum_{i=1}^{N-m\tau} C_m,i(\tau) \right] - \ln \left[ \frac{1}{N-m \tau} \sum_{i=1}^{N-m\tau} C_m,i(1) \right]$$

SampEn depends on two parameters, the “filter factor” $r$ and the embedding dimension $m$. It is important to choose adequate values for the above parameters. We set $r$ to be 0.2 times the standard deviation of the original data series, as previously suggested, and used $m = 2$, in accordance with previous reports. In order to facilitate the interpretation of SampEn values for EEG signals, for $N = 4000$, the SampEn of a sine wave was estimated to be 0.0737 as an example of a regular (linear) signal. The SampEn of a series of uniformly random numbers was estimated to be 2.2802. The original report used $\tau = 1$. On the other hand, the modified SampEn used $\tau \geq 1$. These values represented the average time lag determined for each time-series as the time lag ($\tau$) at

### Table 1. Characteristics of subjects and patients

<table>
<thead>
<tr>
<th></th>
<th>Subjects without AF</th>
<th>Patients with AF</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years old)</td>
<td>72.2 ± 8.0</td>
<td>79.7 ± 7.0</td>
<td>0.004*</td>
</tr>
<tr>
<td>Male (Female)</td>
<td>21 (13)</td>
<td>11 (4)</td>
<td>0.433†</td>
</tr>
<tr>
<td>Mean (R-R)</td>
<td>0.865 ± 0.133</td>
<td>0.744 ± 0.145</td>
<td>0.0181*</td>
</tr>
<tr>
<td>SDNN</td>
<td>0.041 ± 0.035</td>
<td>0.197 ± 0.194</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>BUN</td>
<td>17.2 ± 4.8</td>
<td>20.5 ± 0.2</td>
<td>0.138*</td>
</tr>
<tr>
<td>Creatinine</td>
<td>0.79 ± 0.22</td>
<td>0.88 ± 0.19</td>
<td>0.178*</td>
</tr>
<tr>
<td>Hypertension</td>
<td>18 (52.9%)</td>
<td>12 (80.0%)</td>
<td>0.012‡</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>10 (29.4%)</td>
<td>2 (13.3%)</td>
<td>0.228‡</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>1 (2.9%)</td>
<td>3 (20.0%)</td>
<td>0.044‡</td>
</tr>
<tr>
<td>Cerebral infarction</td>
<td>1 (2.9%)</td>
<td>6 (40.0%)</td>
<td>0.0006†</td>
</tr>
</tbody>
</table>

* Mann–Whitney U test; † χ² test. AF, atrial fibrillation; BUN, blood urea nitrogen; mean R-R, mean R-R interval of electrocardiogram; SDNN, standard deviation of the RR intervals.
which the autocorrelation function best approximates $1/e$ in ECG. The method was implemented in F-Basic (Fu-
jitsu, Tokyo). We examined $\tau$ in every segment and used the average value. We determined $\tau = 5$. To embed the time-series in time-delayed space, we used a time lag $\tau$ of 5 (0.025 s) for ECG.

**Plotting ECG signals in time-delayed space in three-dimensional coordinates**

To plot reconstructed vectors from ECG signals in three-dimensional space with time-delayed coordinates, we reconstructed $N-2\tau$ vectors from the time series $\{x(1), x(2), \ldots, x(N)\}$ as

$$X_1 = (x(1), x(1+\tau), x(1+2\tau)),$$

$$X_2 = (x(2), x(2+\tau), x(2+2\tau))$$

$$\vdots$$

$$X_{N-2\tau} = (x(N-2\tau), x(N-\tau), x(N)).$$

We plotted these reconstructed vectors, $\{X_1, X_2, X_3, \ldots, X_{N-2\tau}\}$ in three-dimensional space with time-delayed co-
ordinates using DeltaGraph (Japan Poladigital, Tokyo).

**Statistical analysis**

The values of SampEn were presented as mean $\pm$ SD. We used the Mann–Whitney $U$ test to compare Samp-
En values and assess the differences between subjects without arrhythmia and patients with AF. Associations among SampEn, age, and the standard deviation of the RR intervals (SDNN) were evaluated using Pearson’s correlation coefficient. Differences and correlations were considered to be statistically significant at $P < 0.05$ (Stat View, SAS Institute, Cary, NC).

**RESULTS**

**Calculated SampEn results**

ECG was measured from the group of subjects without arrhythmia and the group of patients with AF, and each SampEn was calculated. The mean SampEn values in the group without arrhythmia were $0.252 \pm 0.114$ ($\tau = 1$) and $0.533 \pm 0.163$ ($\tau = 5$), and those for the chronic AF group were $0.392 \pm 0.158$ ($\tau = 1$) and $0.759 \pm 0.246$ ($\tau = 5$) (Fig. 1). The mean SampEn with a time lag of $\tau = 1$ was significantly higher in the group with chronic AF than in the group without arrhythmia ($P < 0.01$) (Fig. 1A). The mean SampEn with a time lag of $\tau = 5$ was significantly higher in the group with chronic AF than in the group without arrhythmia ($P < 0.004$) (Fig. 1B). We assessed the intra-variation of the computed SampEn from 5 segments by examining the coefficient of variation (COV). There was no significant difference between the mean COV of SampEn with $\tau = 1$ (11.06 $\pm$ 3.19%) and that with $\tau = 1$ for the chronic AF group (10.33 $\pm$ 4.13%). There was also no significant difference between the mean COV of SampEn with $\tau = 5$ (8.97 $\pm$ 5.17%) and that with $\tau = 5$ for the chronic AF group (7.62 $\pm$ 5.82%).

In this study, there was a significant difference at the age of subjects without AF and patients with AF. We analyzed the values of SampEn in subjects without AF and patients with AF above 70 years old in order to except the effect on SampEn due to age difference (Table 2). There was no significant difference between the two groups in age. The mean SampEn values in the group

<table>
<thead>
<tr>
<th>Subjects without arrhythmia</th>
<th>Patients with AF</th>
</tr>
</thead>
<tbody>
<tr>
<td>SampEn ($\tau = 1$)</td>
<td></td>
</tr>
<tr>
<td>SampEn ($\tau = 5$)</td>
<td></td>
</tr>
</tbody>
</table>

![Fig. 1. SampEn for subjects without arrhythmia and patients with AF. (A) The mean SampEn with a time lag of $\tau = 1$ was significantly higher in the group with chronic AF than in the group without arrhythmia ($P < 0.01$). Mean values are represented with short line. (B) The mean SampEn with a time lag of $\tau = 5$ was significantly higher in the group with chronic AF than in the group without arrhythmia ($P < 0.004$). Mean values are represented with short line. AF, atrial fibrillation; SampEn, sample entropy.](image-url)
without arrhythmia were 0.243 ± 0.122 (τ = 1) and 0.392 ± 0.158 (τ = 5), and those for the chronic AF group were 0.527 ± 0.186 (τ = 1) and 0.759 ± 0.246 (τ = 5). The mean SampEn values with time lags of τ = 1 and τ = 5 were significantly higher in the group with chronic AF than those in the group without arrhythmia (P < 0.02).

**New method to plot ECG signals in time-delayed space with three-dimensional coordinates**

Figure 2 is a sample ECG from a 65-year-old man without arrhythmia. The constructed three-dimensional vectors were plotted in time-delayed three-dimensional space (Fig. 3). We used time lags of τ = 5 (Fig. 3A) and τ = 1 (Fig. 3B). We found that in the case of τ = 5, QRS complex plotted large triangular, and P, T waves plotted like a circle or ellipsoidal. On the other hand, in the case of τ = 1, each waveform plotted like a circle or ellipse collapsed. In the case of τ = 5, the loop of the three-dimensional space is larger than that of τ = 1, and it is easy to distinguish each waveform. Therefore, the shape of the loops was better for the former (τ = 5). We also extracted the P wave, QRS complex, and T wave from the electrocardiogram time-series data and displayed it in the three-dimensional space (τ = 5) (Fig. 3C). Figure 4A shows the original ECG data and the new plotting method for the ECG of an 86-year-old man with AF and three-dimensional plotting with a time lag of 0.025 sec (τ = 5). We confirmed the fluctuation of the loops from ECG with AF (Fig. 4).

**Correlation between age and SampEn for subjects without arrhythmia and patients with AF**

Figures 5A and B show the correlation between age and SampEn for subjects without arrhythmia and patients with AF with τ = 1. No significant correlation was observed in simple linear regression analysis. Also, there was no significant correlation between age and SampEn for subjects without arrhythmia with τ = 5 (Fig. 6A). There was a correlation for patients with AF (r = 0.438, P < 0.02, Fig. 6B).

**Correlation between SDNN and SampEn for subjects without arrhythmia and patients with AF**

Figure 7 shows the correlation between SDNN and SampEn for subjects without arrhythmia and patients with AF (τ = 1 and τ = 5). A correlation was observed for both subjects without arrhythmia and patients with AF (τ = 1: r = 0.331, P < 0.0001, Fig. 7A; τ = 5: r = 0.279, P < 0.0001, Fig. 7B).
Fig. 3. New method for plotting ECG in time-delayed three-dimensional space with time lag of 0.025 sec ($\tau = 5$) and 0.005 sec ($\tau = 1$). (A) Three-dimensional plotting with time delay ($\tau = 5$). (B) Three-dimensional plotting with time delay ($\tau = 1$). (C) We extracted the P wave, QRS complex, and T wave from ECG time-series data and displayed them in three-dimensional space ($\tau = 5$). ECG, electrocardiogram.

Fig. 4. New method of plotting ECG in time-delayed three-dimensional space with time lag of 0.025 sec ($\tau = 5$) in a patient with AF. (A) Original ECG data from an 86 year old subject with AF sampled at 200 Hz (0.005 sec). (B) Three-dimensional plot with time lag of 0.025 sec ($\tau = 5$). Fluctuations in the loops from the QRS complex, f wave and T wave were observed. AF, atrial fibrillation; ECG, electrocardiogram.
Fig. 5. Correlation between age and the SampEn of subjects without arrhythmia and patients with AF (τ = 1). (A) Simple linear regression analysis of age and the SampEn of subjects without arrhythmia (τ = 1). There was no significant linear correlation between age and SampEn for subjects without AF in simple linear regression analysis. (B) Simple linear regression analysis of age and SampEn for patients with AF (τ = 1). There was no significant linear correlation between the two parameters for patients with AF in simple linear regression analysis. AF, atrial fibrillation; n.s., not significant; SampEn, sample entropy.

Fig. 6. Correlation between age and SampEn for subjects without arrhythmia and patients with AF (τ = 5). (A) There was no significant linear correlation between age and SampEn for subjects without AF with τ = 5 in simple linear regression analysis. (B) There was also a linear correlation between age and SampEn for patients with AF with τ = 5. AF, atrial fibrillation; n.s., not significant; SampEn, sample entropy.
Sample entropy of ECG

Fig. 7. Correlation between SDNN and SampEn for subjects without arrhythmia and patients with atrial fibrillation (τ = 1 and τ = 5). (A) Simple linear regression analysis of SDNN and SampEn for subjects without arrhythmia and patients with AF (τ = 1). Open circles indicate subjects without arrhythmia, and closed circles indicate patients with AF. There was a linear correlation between SDNN and SampEn for both groups in simple linear regression analysis. (B) When SampEn was calculated by τ = 5, there was a linear correlation between SDNN and SampEn for both groups in simple linear regression analysis. AF, atrial fibrillation; SampEn, sample entropy; SDNN, standard deviation of the RR intervals.

DISCUSSION

The ECG data for subjects without arrhythmia and patients with chronic AF were converted into time-series data, and the SampEn for each subject was calculated. Comparison of SampEn in the group without arrhythmia and the group with chronic AF showed that SampEn in the latter group was significantly higher. This suggests that the ECG time-series data for patients with chronic AF are more complicated than for those without arrhythmia.

In recent years, several methods for examining nonlinear characteristics by assuming biological activity as a complex system have been studied. ApEn and SampEn are frequently used in clinical physiology.22–33 ApEn was introduced by Pincus,7 whereas SampEn was proposed by Richman and Moorman8 as a measure of system complexity. ApEn and SampEn are applied in time-series data analysis such as ECG and EEG.3, 24 Richman and Moorman developed SampEn to reduce the bias of ApEn and reported the physiological time-series analyses using ApEn and SampEn.8 ApEn is particularly susceptible to short or noisy time-series and SampEn is a less biased statistical measure.8 Therefore, SampEn was evaluated in this study.

Conventionally, analysis of ECG calculations is performed by extracting a section such as the R-R interval. Research on heart rate variability in obstructive sleep apnea syndrome12, 14 and that on the characteristics of heart rate and blood pressure dynamics12, 14 using SampEn have been reported in a wide range of fields. However, there have been no studies calculating the SampEn of consecutive ECG time-series data without filtering, and no reports comparing subjects without arrhythmia and patients with chronic AF.

In this study, the time-series data for ECGs were displayed in three-dimensional space using the time-delayed coordinates system with a time delay of τ = 5 (Figs. 3A and C, Fig. 4B). By extracting the P wave, QRS wave, and T wave from the time-series data and displaying them in the three-dimensional space (Fig. 3C), it became clear which part of the loop from ECG without arrhythmia display P wave, QRS wave, T wave. From comparing SampEn in the group without arrhythmia and in the group with chronic AF, a significant difference was obtained in the case of τ = 1 and τ = 5 (Fig. 1). On the plot for all segments of ECG in time-delayed coordinates, it was confirmed that τ = 5 is suitable for displaying time-series data in three-dimensional space. In addition, we found that in the case of τ = 5, the loop of the three-dimensional space is clearer than that of τ = 1 (Fig. 3). While SampEn has been conventionally calculated with τ = 1, a modified calculation method using
\( \tau = 5 \) is also useful.

Although the conventional method of extracting a specific section such as the R-R interval has been used to compute SampEn, we calculated the SampEn using all the measured waveforms from ECG in this study. SampEn can also detect irregularities in the R-R interval because there was a significant correlation between SDNN and SampEn for subjects without arrhythmia and patients with AF. There was a significant linear correlation between age and SampEn for patients with AF (\( \tau = 5 \)). Some patients with AF had particularly high values of SampEn. The values of SampEn for patients with AF might be affected by antiarrhythmic drugs. However, in this study, we could not find out why the values of SampEn of subjects without arrhythmia and patients with AF were overlapped.

SampEn is effective for evaluating the irregularity of ECG signals and may become a new index that enables the evaluation of the randomness of chronic AF. For example, there is a possibility that the onset of disease can be detected in advance by complex system analysis. Conventional studies have reported that ApEn and SampEn reductions are characteristic of heart rate variability preceding paroxysmal atrial fibrillation.\(^4\) Also, nonlinear analysis of respiratory movement and EGG during eye-closed awake and different sleep stages has been reported.\(^2\) Complex system analysis may be useful in the clinical physiology. However, the present study is associated with several limitations, with the most obvious being its relatively small sample size. Furthermore, we did not investigate other types of arrhythmia. Further studies involving more number of patients will be needed to clarify the effectiveness of SampEn.

The authors declare no conflicts of interest.

REFERENCES


Sample entropy of ECG