

# Use of Approximation Entropy for Stratification of Risk in Patients With Chagas Disease

M Vizcardo<sup>†</sup>, A Ravelo<sup>‡</sup>

<sup>†</sup>Departamento de Física, Universidad Nacional San Agustín, Arequipa, Perú

<sup>‡</sup>Instituto for Technological Development and Innovation in Communications, Universidad de Las Palmas de Gran Canaria, Spain

## Abstract

*According to the World Health Organization, the number of people infected with Trypanosoma Cruzi is estimated between 6 and 7 million, the causative agent of Chagas disease, and in 550000 people exposed to the risk of affectation. The approximate entropy was used to quantify the regularity of the tachograms of patients with Chagas disease. The study population consisted of three groups of volunteers: 92 controls (C), 102 patients with positive serology without cardiac involvement diagnosed by conventional non-invasive methods (CH1) and 107 patients with positive serology and mild to moderate incipient heart failure (CH2). We analyzed RR segments of 5 minutes, 288 segments, corresponding to 24 hours per patient. We found significant differences between the Control and CH2 groups, which is used to stratify risk in the CH1 group.*

## 1. Introduction

Chagas disease is caused by a flagellated parasite: Trypanosoma cruzi, transmitted by an insect of the genus Triatoma and also by blood transfusions. In Latin America the number of people infected is approximately 6 million, with a population exposed to the risk of infection of 568,000[1]. In 40 % of the population infected with Trypanosoma cruzi (T. cruzi) there is cardiac involvement [2, 3, 11]. These estimates explain why this disease is a serious public health problem in the countries where it is endemic. In the evolution of Chagas disease we can distinguish an initial acute phase of infection and a prolonged intermediate chronic phase, in which the disease is often clinically silent, and the usual diagnostic techniques do not provide a robust criterion to predict whether a seropositive asymptomatic patient will suffer cardiac involvement. It is our interest to develop a non-invasive methodology low cost, that allows to see the dysautonomia or dysfunction in the course of the 24 hours and with this it could be used to detect early any cardiac alterations produced by the T. cruzi.

## 2. Database

We have used the Electrocardiogram (ECG) database of the Institute of Tropical Medicine (IMT) of the Central University of Venezuela. Tachograms were obtained by processing the electrocardiogram (ECG) and obtaining the RR interval. The following test is carried out: clinical evaluation, serological Machado-Gerreiro test, chest x-rays, echocardiogram, electrocardiogram and ambulatory Holter registration (24 hours). The volunteers are classified into three groups: 92 healthy people called group control C; 102 patients infected with only positive serology ( clinical evaluation, chest x-rays, echocardiogram, electrocardiogram and Holter were normal ) called CH1 group; and 107 seropositive patients with incipient cardiac involvement first-degree atrioventricular block (BAV), sinus bradycardia (BS) or right bundle branch block of the bundle of His (BRDHH), that were not being treated with medications, called the CH2 group. All were outpatients, and informed consent was obtained from all of them. The ECG signals were recorded at 500 Hz with 12 bits of resolution, a set of 288 frames of 5 minutes was obtained.

## 3. Method

### 3.1. Preprocessed

For the detection of the QRS complex the program based on the Pan-Tomkips algorithm[9] was used, then the 288 tachograms of the RRs of 5 minutes were generated. A post processing based on an adaptive filtering procedure and details in [10], was also carried out.

### 3.2. Feature extraction

The approximate entropy (ApEn) was developed by Steve Pincus[4, 5] is based on the entropy of Kolmogorov-Sinai KS, proposed by Grassberger and Procaccia[6] and modified by Takens[7]. A modification of ApEn was made by Richman[8] which is the sample entropy (SampEn). both of which measures the irregularity and complexity of

a time series of data.

ApEn and SampEn are based on the comparisons of the component-to-component embedding vectors ( $m = 2$ ) and with a threshold of 20% of the standard deviation of the intervals  $RR$  ( $r = 0.20$ ). The difference between the entropies is that ApEn does not take into account the comparisons of the embedment vectors with itself and the way to calculate the logarithm. To calculate ApEn:

$$ApEn(m, r, N) = -\frac{1}{N - m} \sum_{i=1}^{N-m} \log\left(\frac{A_i}{B_i}\right) \quad (1)$$

where  $B_i$  are the vectors of embedding dimension  $m$  and  $A_i$  are the vectors of dimension of embedding  $m + 1$ ,  $r$  is the threshold that is typically 20% of the value of the standard deviation and  $N$  is the number of elements.

SampEn is calculated as:

$$SampEn(m, r, N) = -\log\left(\frac{\sum_{i=1}^{N-m} A_i}{\sum_{i=1}^{N-m} B_i}\right) = -\log\left(\frac{A}{B}\right) \quad (2)$$

similar a ApEn  $m, r$  and  $N$ .

### 3.3. Testing samples

We will use the Kruskal Wallis test that matches the 288 frames corresponding to the Control-CH1, Control-CH2 and CH1-CH2 groups to test differences between groups. Also a logistic regression is used to evaluate circadian profiles of the average values of ApEn, SampEn. Finally boxplot representations are used to analyze a frame in special.

## 4. Results

### 4.1. Approximate Entropy (ApEn)

ApEn was calculated at 288 tachograms of 5 minutes and 24-hour circadian profiles were obtained (Fig 1a).

- Between 00:00 hours and 06:00 the control group C has slightly higher values than CH1, CH1 ApEn values are slightly greater than CH2.
- Between 06:30 and 11:20 there is no greater differences between the three groups.
- We noticed that between 11:20 and 21:30 there is a sustained difference of group C
- At 14:30 it is the biggest difference between groups C and CH2.

We have used the logistic regression to discriminate groups C and CH2. For this task 50% of frames are selected randomly (144) for training and 50% specificity and 80% of sensitivity between 14:00 h. and 15:40 h.

We applied the Kruskal-Wallis test where we can see in figure 1b that around frame 175 (14:30 hours)  $pvalue < 0.05$ , for the groups: Control-CH1 and Control-CH2. This not the case for CH1-CH2 as shown in the figure 1.

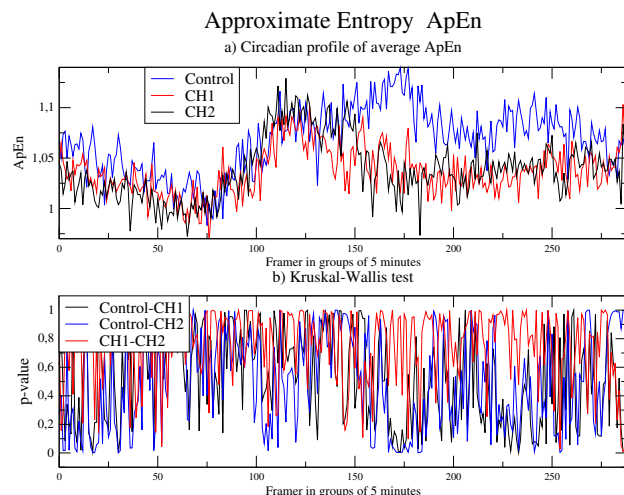


Figure 1. a) ApEn values of the 24 hours (288 frames) b) p-value of the groups: Control-CH1, Control-CH2, CH1-CH2

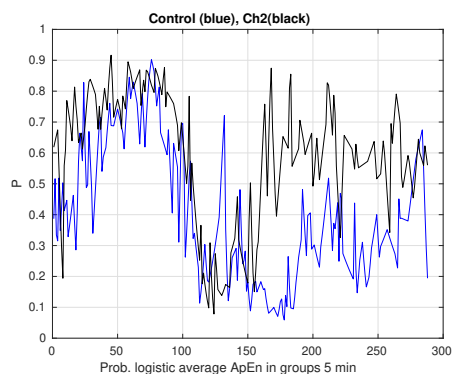


Figure 2.

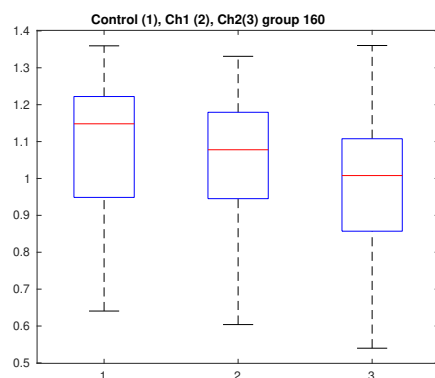


Figure 3. Boxplot frame 160 ApEn

### 4.2. Sample Entropy (SampEn)

We calculated the SampEn to the 288 segments of 5 minutes to the three groups and we obtain the average val-

ues of the  $RR$  intervals. This can be seen in Fig 4a,

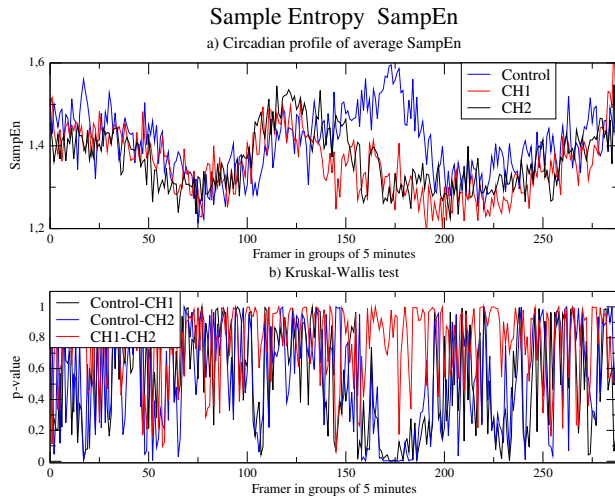


Figure 4. a) SampEn values of the 24 hours (288 frames)  
b) p-value of the groups: Control-CH1, Control-CH2, CH1-CH2

- Between 00:00 hours and 08:30 the value of SampEn there is no greater difference between the three groups. The control group C has slightly higher values than CH1, CH1 is slightly greater than CH2.
- Between 08:30 and 10:00 the groups CH1 and CH2 have a higher value than group C.
- However, from 12:30 to 16:15 we observe that SampEn values are higher in group C than groups CH1 and CH2
- At 14:10 it is the biggest difference between groups C and CH2.

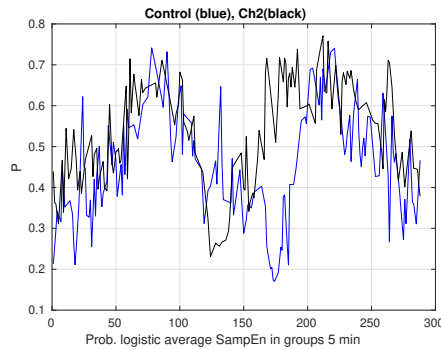


Figure 5.

Now we will use the logistic classifier for groups C and CH2, for this we will use 50% of frames at random (144) for training and 50% for validation, with this it is possible to classify up to 80% of specificity of group C and 70% of sensitivity in the CH2 group between 14:00 h. at 15:40 h. We applied the Kruskal-Wallis test to the values of SampEn to the groups: Control-CH1, Control-CH2, where

we noticed that around frame 175 the p-values lower than 0.05.

Now we will use the logistic regression classifier for groups C and CH2, with this it is possible to reach 80% of specificity and 70% of sensitivity between 14:00 h. at 15:40 h. We applied the Kruskal-Wallis test to the values of SampEn to the groups: Control-CH1, Control-CH2, where we noticed that around frame 175  $p - values < 0.05$ . A similar result is obtained with ApEn.

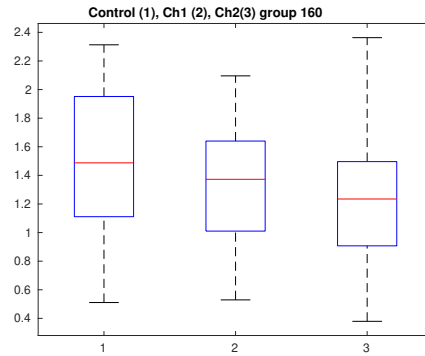


Figure 6. Boxplot frame 160 SampEn

We analyze the segment 160 of the result of the SampEn values that corresponds to the 13:30 hours, using the boxplot. We noticed that it is possible to show the risk to the CH1 group, just like ApEn.

## 5. Discussion and conclusions

The average ApEn values between 11:20 and 21:30 show (Fig. 1a) that group Control presents higher values than the values corresponding to group CH1 and CH2. This could be due to a decrease or enervation in its barometric response, this is also observed in SampEn (Fig 4a) at shorter time intervals from 12:30 to 16:15.

The approximation entropy (ApEn) and the sampling entropy (SampEn) measure the irregularity and the complexity, when we apply these features to the circadian profiles of the Control, CH1 and CH2 groups, we observe that in the afternoon hours the values of the Control group has greater irregularity and complexity than the values of the CH1 and CH2 groups that show decrease or enervation of these groups, this could be used as a measure of the variability of the heart rate.

The logistic classifier and Kruskal-Wallis test allows us to differentiate the both groups (control and CH2) both in ApEn and SampEn in the afternoon hours (18 frames), indicating that it can be used in the stratification of risk of group CH1, therefore it is possible to use to classifier to can be used in the stratify group CH1.

The logistic regression and the Kruskal-Wallis test allow

us to differentiate both groups (control and CH2) both in ApEn and in SampEn in the evening hours 2:35 pm (18 frames or 1:30 h), which can be used to stratify the CH1 group risk.

## Acknowledgements

Virrektorado de Investigación de la Universidad Nacional San Agustín de Arequipa, contrato de subvención 16-2018-UNSA.

## Conflicts of Interest

The authors declare no conflict of interest

## References

- [1] World Health Organization, The World Health Report, 2017, 2:75,9–16, <http://www.who.int/wer>
- [2] Moleiro F. and I. Mendoza, *Circulation*, 1978, 113.
- [3] Moleiro F, Mendoza I, Miocardiopatía Chagásica Crónica. Estudio Epidemiológico, *Acta Cient. Venez.* 1980, 31:66
- [4] Pincus S, Gladston I, A Regularity Statistic for Medical Data Analysis, *Journal of Clinical Monitoring* October, 1991. 1980, Vol 7 No 4
- [5] Pincus S Approximate entropy as a measure of system complexity, *Proc. Natl. Acad. Sci. USA*, 1991, Vol. 88, 2297-2301, *Mathematics*
- [6] Grassberger P, Procaccia I. Estimation of the Kolmogorov entropy from a chaotic signal, *Phys Rev A* 1983, 28:2591-2593
- [7] Takens S, Invariants related to dimension and entropy, In: *Atas do 13 Rio de Janeiro, Col. Brasileiro de Matematicas*, 1983
- [8] Richman J., Moorma J., Physiological time-series analysis using approximate entropy and sample entropy, *Am J Physiol Heart Circ Physiol*, 278: H2039H2049, 2000.
- [9] Pan J. and Tompkins W., A Real-Time QRS Detection Algorithm *IEEE Transactions on Biomedical Engineering*, Vol. BME-32, NO. 3, March, 1985.
- [10] Wessel, N., et.al., Nonlinear analysis of complex phenomena in cardiological data, *Herzsch. Elektrophys.*, 11(3), 2000, 159-173
- [11] Hagar J. M. and Rahimtoola S. H., *N. Eng. J. Med.*, 325, 763, 1991.
- [12] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, *Circulation*, 73, 5, 1044, 1996.
- [13] Guzzetti S., et.al., Impaired heart rate variability in patients with chronic Chagas' disease., *Am. Heart J.* 1991, 121:1727
- [14] Loyo J, et.al., DV:Effect of postural changes in the autonomic balances in latent or indeterminate phase of Chagas Disease, *Chest* 1995, 3:108
- [15] Liebovitch L. S., et.al., *Phys. Rev. E*, 59, 3, 1999.
- [16] Vizcardo M., et.al, Analysis of the heart rate variability and stratification of the risk of cardiac patients with Chagas disease, *Computers in Cardiology*, 2007.
- [17] Wales D. J., Calculating the rate of loss of information from chaotic time series by forecasting, *Nature* 350, 485, 1991.
- [18] Ravelo-Garca A. G., et.al. Symbolic dynamics marker of heart rate variability combined with clinical variables enhance obstructive sleep apnea screening, *Chaos* 24, 024404, 2014.
- [19] Campos Junqueira de Souza A. and et.a Development of a risk score to predict sudden death in patients with Chaga's heart disease,l., *International Journal of Cardiology* 187, 2015, 700704
- [20] Martín-González S., et.al., Heart rate variability feature selection in the presence of sleep apnea: An expert system for the characterization and detection of the disorder, , *Computers in Biology and Medicine* 91, 2017, 4758
- [21] Ravelo-García A. G., et.al., Application of the Permutation Entropy over the Heart Rate Variability for the Improvement of Electrocardiogram-based Sleep Breathing Pause Detection, , *Entropy*, 2015, 17, 914-927; doi:10.3390/e17030914
- [22] Merejo Pea C. M. and et.al., Dysautonomia in different death risk groups (Rassi score) in patients with Chagas heart disease, *Pacing Clin Electrophysiol*, 2018, 41:238245.
- [23] Caeiro A. , Iosa T., Chronic Chagas' disease: possible mechanism of sinus bradycardia. *Can J Cardiol*, 1994, 10:765-8
- [24] Goin J. C., Borda E. S., Auger S, Storino R, Sterin-Borda L. Cardiac M(2) muscarinic cholinceptor activation by human chagasic autoantibodies: association with bradycardia, *Heart Sep*, 1999, 82(3):2738.
- [25] Sterin-Borda L, Borda E. Role of neurotransmitter autoantibodies in the pathogenesis of chagasic peripheral dysautonomia. *Ann N Y Acad Sci*, 2000, 917:27380.
- [26] Pedrosa R. C., et.al., Sera from chronic chagasic patients with complex cardiac arrhythmias depress electrogenesis and conduction in isolated rabbit hearts. *Circulation*, 1997, 96:20317
- [27] Octavio J, et.al., Perfiles circadianos de la frecuencia cardíaca y de su variabilidad instantánea en una población de pacientes con infección chagásica crónica. *Rev Esp Cardiol.*, 2004, 57(2):130-7
- [28] Ribeiro F., et.al, Parasympathetic dysautonomia precedes left ventricular systolic dysfunction in Chagas disease. *Am Heart J*, 2001, 141:260-5.
- [29] Ravelo-Garca, et.al, Cepstrum feature selection for the classification of sleep apnea-hypopnea syndrome based on heart rate variability. In *Computing in Cardiology Conference (CinC)*, 2013 (pp. 959-962). IEEE.
- [30] Ravelo-Garca, et.al, An approach to the enhancement of sleep apnea detection by means of detrended fluctuation analysis of RR intervals. In *Computing in Cardiology Conference (CinC)*, 2014 (pp. 905-908). IEEE.

Address for correspondence:

Miguel Vizcardo Cornejo  
Calle Santa Catalina N 117 Arequipa, Peru  
mvizcardoc@unsa.edu.pe