Journal Articles   Biomechanics Research Building

2-2013

The Appropriate Use of Approximate Entropy and Sample Entropy with Short Data Sets

Jenna M. Yentes
University of Nebraska at Omaha, jyentes@unomaha.edu

Nathaniel Hunt
University of Nebraska at Omaha

Kendra K. Schmid
University of Nebraska Medical Center

Jeffrey P. Kaipust
University of Nebraska at Omaha, jkaipust@unomaha.edu

Denise McGrath
University of Nebraska at Omaha

See next page for additional authors

Follow this and additional works at: http://digitalcommons.unomaha.edu/biomechanicsarticles

Part of the Biomechanics Commons

Authors

Jenna M. Yentes, Nathaniel Hunt, Kendra K. Schmid, Jeffrey P. Kaipust, Denise McGrath, and Nicholas Stergiou

The appropriate use of approximate entropy and sample entropy with short data sets

Jennifer M Yentes[1], Nathaniel Hunt[1], Kendra K Schmid[2], Jeffrey P Kaipust[1], Denise McGrath[1], Nicholas Stergiou[1,3]

[1] Nebraska Biomechanics Core Facility

Department of Health, Physical Education, and Recreation

University of Nebraska at Omaha

6001 Dodge Street

[2] College of Public Health

Department of Biostatistics

University of Nebraska Medical Center

984355 Nebraska Medical Center

Omaha, NE 68198

[3] College of Public Health

Department of Environmental, Agricultural & Occupational Health

University of Nebraska Medical Center

984355 Nebraska Medical Center

Omaha, NE 68198

Abbreviated title: Entropy with short data sets

Corresponding Author:

Nicholas Stergiou, PhD

University of Nebraska at Omaha

6001 Dodge Street, HPER 207

Omaha, NE 68134

Phone: 402-554-3247

Fax: 402-554-3693

nstergiou@unomaha.edu

Abstract

Approximate entropy (ApEn) and sample entropy (SampEn) are mathematical algorithms created to measure the repeatability or predictability within a time series. Both algorithms are extremely sensitive to their input parameters: m (length of the data segment being compared), r (similarity criterion) and N (length of data). There is no established consensus on parameter selection in short data sets, especially for biological data. Therefore, the purpose of this research was to examine the robustness of these two entropy algorithms by exploring the effect of changing parameter values on short data sets. Data with known theoretical entropy qualities as well as experimental data from both healthy young and older adults was utilized. Our results demonstrate that both ApEn and SampEn are extremely sensitive to parameter choices, especially for very short data sets, $N \leq 200$. We suggest using N larger than 200, an m of 2 and examine several r values before selecting your parameters. Extreme caution should be used when choosing parameters for experimental studies with both algorithms. Based on our current findings, it appears that SampEn is more reliable for short data sets. SampEn was less sensitive to changes in data length and demonstrated fewer problems with relative consistency.

Keywords: step length, step width, step time, nonlinear analysis, entropy, locomotion

Abbreviations

ApEn – approximate entropy

SampEn – sample entropy

SE_{diff} – Standard error of the difference

Introduction

Entropy is defined as the loss of information in a time series or signal. Within the past twenty years, the use of entropy methods to define periodicity or regularity in human data has become quite popular (Figure 1). Presently, the two most commonly used methods for biological data are approximate entropy (ApEn) and sample entropy (SampEn). Entropy has been used to describe changes in postural control [8, 33, 40], physical activity measures [3, 38], as well as other movement tasks [13, 35, 37]. ApEn and SampEn have also been utilized with human walking data [11, 14-16, 18, 20, 21, 28, 32]. The two algorithms are similar yet the validity or accuracy of these algorithms along with the selection of correct parameters used with short data sets has yet to be determined.

ApEn was developed by Steven Pincus [29-31] as a measure of regularity to quantify levels of complexity within a time series [30]. The ability to discern levels of complexity within biological data sets has become increasingly important. In terms of human movement, newer theories of movement control do not view variability in movements as error (i.e. dynamical systems theory). In fact, these theories consider complexity of movement patterns to be associated with system stability (i.e. rich behavioral state) [39]. In contrast to older theories of movement control, variability is not defined only through the amount of variance (e.g. standard deviation) but rather through examination of the temporal variations in the movement output. Thus, Lipsitz and colleagues further defined complexity as signifying the presence of chaotic temporal variations in the steady state output of a healthy biological system representing the underlying physiologic capability to make flexible adaptations to everyday stresses placed on the human body [22, 23]. Importantly, there are certain benefits for the nervous system for adopting chaotic regimes that allow a wide range of

potential behaviors. This leads to healthy biological systems that are adaptable and flexible in an unpredictable and ever-changing environment [9, 17]. We should mention here that this definition of complexity arises from the clinical domain [22, 23]. However, other definitions of complexity also exist mostly from the mathematical domain. For example, Bar-Yam defined complex as "consisting of interconnected or interwoven parts" [2]. Gell-Mann suggests that the degree of connectedness is a measure for complexity [10]. Thus, complex systems are systems composed of connected subsystems. Lastly, Mayer-Kress stated that the characteristics of complex systems are a) many interacting subsystems, b) multiple interactions within and between levels of analysis, c) emergence of movement coordination modes, and d) exhibition of varying levels of the system output that continually evolve with learning and development [26]. Our research group has adopted the definition by Lipsitz and colleagues and propose that mature motor skills and healthy states are associated with optimal movement variability where the temporal structure of the movement output reveals information about the control processes [39]. Furthermore, a movement with high regularity (low entropy) or a very random movement (high entropy) would reveal low complexity, inferring the motor control process as either unyielding or too erratic, respectively [39]. Thus, entropy provides researchers the ability to quantify complexity within relatively short data sets based on meaningful experimental comparisons to control groups.

ApEn measures the "likelihood that runs of patterns that are close remain close on next incremental comparisons" [31]. Upon development of this algorithm, Pincus demonstrated that as compared to other nonlinear algorithms [41], ApEn could differentiate between noisy and chaotic time series with a relatively short number of data points, i.e. 1,000 [31]. It has been suggested that ApEn could be used on data sets as short as 75 to 100 data points [29, 30]. This proved to be essential to

researchers working with human subjects, as typically it is difficult to acquire long, continuous, biological data sets, particularly when pathological or older adults are of interest. It can be difficult for these individuals to complete tasks and fatigue is generally an issue.

The ApEn algorithm has not gone without scrutiny. Richman and Lake (2000) introduced SampEn as an algorithm to counteract the following shortcomings of ApEn [34]. First, ApEn inherently includes a bias towards regularity, as it will count a self-match of vectors. SampEn does not count a self-match, thus eliminating the bias towards regularity [34]. Second, ApEn lacks relative consistency [29], as the input parameters are changed, the value of ApEn may "flip". As an example, white noise may demonstrate a much smaller ApEn value than a known periodic signal when one parameter is set very small. Eventually this will "flip" and the ApEn value will become greater in the white noise signal as the input parameters are changed. Third, the parameters of ApEn must be fixed and comparing data should only be done when the input parameters are the same for both datasets [29]. This is due not only to the issue of relative consistency, but also due to the overall sensitivity of the algorithm to the parameters of choice and to data length [19, 34]. SampEn has been suggested to be independent of data length and demonstrates relative consistency [34]. However, Richman and Lake did state that when using data sets of less than 100, SampEn diverged from their predictions [34]. Thus, the choice of parameters has become important due to the inherent sensitivity of the algorithms [4]. It is imperative that parameters be carefully chosen as findings reported in the literature should not be an artifact of parameter choice, meaning that if the parameters are changed, even slightly, the conclusions reached through the experimental process could be vastly different. Clearly this is a precarious situation that needs to be addressed.

The input parameters are 1) *m*, the length of data that will be compared, 2) *r*, the similarity criterion and 3) *N*, the length of the data. Typically, for clinical data, *m* is to be set at 2, *r* to be set between 0.1 and 0.25 times the standard deviation of the data and *N* as 1,000 [30], even as short as 75 data points [29, 30]. Data length does not have as great of an impact on the calculation of SampEn [34]. Yet, caution has been advised when using data sets of less than 200 points for either ApEn or SampEn [4]; however, this study utilized data with different *m* and *N* values for three different animals (cat: *m*=4, *N*=2000; rat: *m*=3, *N*=1000; mouse: *m*=4, *N*=180). Non-stationary (or drift) is frequently present in biological data and this could also affect the calculation of entropy. As *N* increases, it is likely that the drift will increase as well [6].

Amount of variance in the time series may also affect the calculation of entropy, as well as the presence of spikes and outliers within the data [6, 27]. A greater standard deviation will increase the tolerance for consideration of a match and vice versa with a smaller standard deviation. Hence, the selection of *r* appears to be the most difficult to choose. As originally suggested, *r* is to be chosen between 0.1 and 0.25 times the standard deviation of the entire time series. These suggestions do not always demonstrate the best results for all data sets and therefore, elaborate methods to choose *r* have been developed. The parameter *r* may be chosen based on the minimization of the maximum SampEn relative error and conditional probability [19] or to provide the maximum value of ApEn [5, 25]. While recently, it was suggested that the maximum value method may not be appropriate for analyzing nonlinear signals and is only appropriate for known random time series [24]. Yet another method in which a fixed tolerance value is used [36], the *r* value does not depend on the standard deviation of each vector during comparison. In this calculation of ApEn, the time series is

differenced and the standard deviation in moving windows of a certain length is calculated. The maximum and minimum standard deviation values found serve as the bounds of *r*.

Clearly, no consensus has been established to properly select the parameters needed to calculate both ApEn and SampEn. While the sensitivity of ApEn and SampEn to changing parameters in biological data from three different animals has been investigated [4], there has yet to be a study that investigates this issue in human movement data and specifically with gait data. Considering that it is difficult to acquire large time series from human walking, the choice of parameters used is critical for these short data sets. This is especially true for studies with pathological populations that are limited in their walking capacity and can fatigue easily. In addition, it has not been agreed upon which algorithm performs consistently independent of parameter choice. SampEn has demonstrated more consistent results than ApEn [4]; yet, it has been suggested that SampEn is highly dependent on the relationship between sampling frequency and Nyquist rate as well as the signal-to-noise ratio [1, 6]. In fact, a study investigating body temperature in patients with multiple organ failure concluded that ApEn provided better discrimination between two groups as compared to SampEn [7]. Thus, there is a need to determine 1) the correct parameters to use, 2) which algorithm clearly demonstrates relative consistency and 3) which algorithm provides the best discrimination between groups. The latter is important due to the fact that entropy values are relative. It is not meaningful to report an entropy value for a pathological group without reporting the "healthy" entropy value. Thus, if true differences exist, the algorithm should be able to discriminate between groups and/or conditions and maintain relative consistency as the parameter is changed slightly.

The purpose of this research was to examine 1) the effect of changing the parameter values m, r and N needed for the calculations of both ApEn and SampEn on short data sets with respect to gait; 2) describe the relative consistency of the algorithms; and 3) determine the ability of each algorithm to discriminate between groups. To do this, we first performed ApEn and SampEn calculations for different combinations of parameter values on known theoretical data sets. Theoretical data represent a situation in which accurate entropy values can be determined based on data of unlimited length. Yet, the choice of parameters or consistency of an algorithm may not be similar when applied to biological data, which are constrained by short data sets. Hence, using both theoretical and experimental data we can examine the algorithms performance in both ideal and actual situations. Next, we utilized these same combinations of parameters on spatio-temporal gait data from healthy young and older adults. By using both young and older adults, this allowed for the investigation of different parameter combinations with ApEn and SampEn to discriminate between two different groups.

We hypothesized that the value of ApEn and SampEn would change as a function of m, r and N; however, SampEn would maintain relative consistency and that this would be true for both the theoretical data and the gait data. We hypothesized that each algorithm would be able to discriminate between the theoretical data with periodic data providing the lowest entropy values, random data providing the highest entropy values and chaotic data to provide a value in between. For the experimental data, we hypothesized that the older adults would demonstrate a loss of complexity due to the aging process, either an increased or decreased entropy value, as compared to the young adults.

Materials and Methods

Theoretical Procedures

In order to perform our investigation, it was essential to understand the effect of parameter choice in time series with known entropy qualities. To investigate this aspect, known periodic, chaotic and white noise signals were generated using custom written MatLab code (Mathworks Inc., Natick, Massachusetts). These time series were generated in lengths of 200 data points. This length was chosen, as this was the same data length as the minimum data length for the experimental data.

To increase the applicability of the theoretical data analysis to the experimental data, a known discrete theoretical time series, rather than continuous time series was generated. The periodic time series was generated using the logistic map equation (Equation 1) with *a* set at 3.4 [5] . This choice in *a* allowed the equation to bifurcate into a two-point attractor, perfectly repeating itself.

$$x_{n+1} = a(x_n)(1 - x_n) \hspace{3cm} \text{Equation 1.}$$

Where x_n is a real number and *a* represents the rate of growth or decay. Initial conditions were chosen at random. Since any initial condition within the basin of attraction (zero to one) would bifurcate to this same two-point attractor when *a* was equal to 3.4, only one periodic time series was generated.

The chaotic time series were generated using the same logistic map (Equation 1) but using an *a* of 4.0. When *a* was set at 4.0, nearly all initial conditions within the basin of attraction (zero to one) iterated to a chaotic pattern. White noise time series were generated using the random number function (rand) in MatLab. Twenty chaotic and white noise time series were generated with the

initial conditions being chosen at random for each time series. For all theoretical time series, the first 100 iterations were discarded in order for the dynamics to stabilize. Representative theoretical time series generated are presented in Figure 2.

Experimental Procedures

Twenty-six healthy young adult (25.9 ± 3.0 years; 175.7 ± 8.8 cm; 72.9 ± 11.9kg; 12 male) and 24 healthy older adult (70.9 ± 4.1 years; 173.2 ± 11.2 cm; 76.2 ± 13.2 kg; 13 male) subjects were recruited and provided informed consent for this study. All subjects were independently residing in the community, were able to ambulate independently for a distance of 60 meters without an assistive device and were not diagnosed with a progressive neurologic condition. All subjects were free of any pathological condition that directly affects the musculoskeletal system, leading to an abnormal walking pattern. The University's Institutional Review Board approved all study procedures.

Prior to data collection, subjects were asked to walk on a treadmill for a maximum of eight minutes. This eight-minute warm-up has been considered sufficient for individuals to achieve a proficient treadmill movement pattern [12] . During the eight-minute warm-up a self-selected speed was found. If a subject indicated that a speed was comfortable, they continued to walk at that speed for one minute and then asked again if the speed was too fast or too slow. If they indicated it was too fast, the treadmill was slowed or vice versa. This continued until a comfortable speed was found. After the warm-up period, subjects were asked to walk on the treadmill at their selected speed for a total of three minutes while three-dimensional marker trajectories were recorded. Active rigid body markers were placed on the lateral sides of the foot and six position sensors recorded at 100 Hz (Optotrak Certus system;

Northern Digital Inc., Waterloo, Canada). In addition, virtual markers were identified prior to data collection through the use of wand marking. These markers included the location of the toe, heel and the first and fifth metatarsal heads. The position data of the virtual markers was tracked in real-time (First Principles software; Northern Digital Inc., Waterloo, Canada) with reference to the corresponding rigid bodies. The unfiltered position data for the x, y, and z coordinates of each virtual marker were exported and processed using custom computer code (MatLab; Mathworks Inc., Natick, Massachusetts). This software calculated step length, step width and step time for each subject. Step length was defined as the distance between heel contact and subsequent heel contact of the contralateral foot. Step width was defined as the mediolateral distance between heel markers at successive heel strikes. Step time was defined as the amount of time from heel strike of one foot to the subsequent heel strike of the contralateral foot. All subjects walked a minimum of 200 steps during the data collection period. Therefore, the time series of step length, step width and step time were cut to 200 data points.

Data Analysis

Each individual time series was subjected to calculation of ApEn(m, r, N) and SampEn(m, r, N). The calculation of ApEn is described in detail elsewhere [4, 30]. Given the time series $f(n) = f(1), f(2), ...,$ $f(N)$, where N was the total number of data points, a sequence of m-length vectors (a data segment of length m) was formed. Comparisons were then made against each m-length vector within the time series. Vectors were considered alike if the tail and head of the vector fall within a tolerance level, $\pm r*$standard deviation [31]. The sum of the logarithm of the total number of like vectors was divided by N-m+1. The process was then repeated after m was increased by 1 (m+1). ApEn was calculated by subtracting the conditional probabilities of m+1 from m. A time series with similar distances between data points resulted in lower ApEn values, while large differences in distances

between data points resulted in higher ApEn values. Theoretically, a perfectly repeatable time series would elicit an ApEn value ~0 and a perfectly random time series would elicit an ApEn value ~2. For human movement, a consistent or periodic gait pattern (i.e. robotic gait) would elicit a low ApEn value and a disorderly gait pattern would elicit a higher ApEn value.

SampEn has been defined as the negative natural logarithm for conditional properties that a series of data points a certain distance apart, m, would repeat itself at $m+1$ [34]. SampEn differs from ApEn in that SampEn: 1) eliminates the counting of self-matches and 2) takes the logarithm of the sum of conditional probabilities, rather than the logarithm of each individual conditional property as ApEn does (Figure 3). Given the time series $g(n) = g(1), g(2), \ldots, g(N)$, where N was the total number of data points, a sequence of m-length vectors was formed. Comparisons were then made against each m-length vector within the time series. As with ApEn, vectors were considered alike if the tail and head of the vector fell within the set tolerance level. The sum of the total number of like vectors was divided by $N-m+1$ and defined as B. Further, SampEn defined A as the subset of B that also matched for $m+1$. SampEn was then calculated as $-\ln(A/B)$. Similar to ApEn, a time series with similar distances between data points would result in a lower SampEn value and large differences would result in greater SampEn values with no upper limit. Thus, a perfectly repeatable time series elicited a SampEn value ~0 and a perfectly random time series elicited a SampEn value converging toward infinity.

To examine the effect of the choice of parameters length of data for comparison (m), sensitivity criterion (r) and data length (N), each time series was subjected to the calculation of ApEn and SampEn under all combinations of $m = 2, 3$ and 4; $r = 0.05, 0.1, 0.15, 0.20, 0.25$ and 0.30 times the standard

deviation of the entire time series; and N = 100, 120, 140, 160, 180 and 200, for a total of 108 combinations. Simply, m was the vector or window length that is compared during runs of data and r was the sensitivity criterion in which like vectors or window lengths would be considered similar. For the experimental data, m represented the number of steps. For instance, for the interpretation of step length data, if m is equal to two, this was the length of two steps. The parameter r represented the tolerance of variance in step lengths. When r was equal to 0.2, then the tolerance level in finding like step lengths was within 20% of the standard deviation of all step lengths within the entire time series. Finally, N was the total number of step lengths within the entire time series.

In addition, the total number of self-matches within the time series was also calculated. This was the count of the number of vector comparisons (m) where only the self-match was found. The total number of comparisons for each time series when ApEn was calculated was $N-m+1$.

All time series were visually inspected for spikes and outliers. Average experimental time series for both young and older adults are presented in Figure 2. Experimental data were then examined for stationarity. Approximately 1/3 of the trials demonstrated a significant difference in mean values from the first 100 data points as compared to the second 100 data points. To eliminate stationarity, the data were differenced and ApEn and SampEn were calculated on the differenced time series. No significant differences were found between ApEn and SampEn between the original time series and the differenced time series. Thus, the original time series were used for statistical analysis.

Statistical Analysis

When the parameter value $m = 4$ was used, almost all SampEn values converged toward infinity. Therefore, m of 4 was dropped for all analyses.

The mean and standard deviation of the periodic time series was calculated, for all other combinations of parameters. Group means and standard deviations for ApEn, SampEn and the number of self-matches, from chaotic time series, white noise time series, step length, step width and step time for both the young and older adult groups were calculated for all other combinations of parameters. A repeated measures ANOVA was conducted in SAS (SAS Institute, Inc., Cary, North Carolina) to determine the effect of type (chaotic vs. white noise) and group (young vs. older adults) on ApEn and SampEn. The effect of changing m, r and N on ApEn and SampEn for the theoretical and experimental data was also investigated. If a significant 3-way interaction between m, r and N was found, this indicated that the entropy value was different depending on the values of m, r and N. If a significant 2-way interaction was found, this indicated that entropy values were different depending on the values of m, r and N yet, consistent across the levels of the third variable. For all analyses, there were two levels of m, there were six equally-spaced levels of N and r, so these were treated as continuous variables for analysis purposes. Standard error of the difference (SE_{diff}) between groups was also calculated.

Results

Theoretical Data

The periodic data set produced a value of zero for all combinations of m, r, and N. This was true for the number of self-matches, ApEn and SampEn. Therefore, only the chaotic times series and white noise time series were compared.

After accounting for the effects of m, r and N, a significant difference in ApEn (SE_{diff}: 0.021; $F_{1,19}$: 98.93; $p<0.0001$) and SampEn (SE_{diff}: 0.046; $F_{1,19}$: 1838.66; $p<0.0001$) between the chaotic and white noise time series was found. The chaotic data set provided higher values of ApEn, but lower values of SampEn. A significant interaction between m, r and N was found for ApEn ($F_{1,2852}$: 29.97; $p<0.0001$), meaning the value of ApEn is different depending on the combination of values. For SampEn, significant 2-way interactions between m, r and N were found [($r*m$: $F_{1,2754}$: 55.48; $p<0.0001$), ($N*m$: $F_{1,2754}$: 29.72; $p<0.0001$), ($r*N$: $F_{1,2754}$: 112.14; $p<0.0001$)]. SampEn is consistent across values of N, but is different depending on values of r and m. (Figures 4-7).

Experimental Data

Step Length: After accounting for the effects of m, r and N, a significant difference in ApEn (SE_{diff}: 0.013; $F_{1,48}$: 5.77; $p=0.02$) and SampEn (SE_{diff}: 0.043; $F_{1,48}$: 14.27; $p=0.0004$) between the young and older adults was found. Older adults had lower values of SampEn, but higher values of ApEn. A significant 3-way interaction between m, r and N was found for ApEn ($F_{1,3543}$: 435.36; $p<0.0001$). The slope of ApEn as a function of r and N were different from each other depending on the m chosen, and the value of the other (Figures 4-7). For SampEn, significant 2-way interactions between m, r and N were found [($r*m$: $F_{1,3312}$: 151.02; $p<0.0001$), ($N*m$: $F_{1,3312}$: 27.62; $p<0.0001$), ($r*N$: $F_{1,3312}$: 145.37; $p<0.0001$)]. Figures 4 and 5 ($m = 2$ or 3, respectively) present SampEn as a function of N. SampEn was consistent across values of N although there were

differences in SampEn as r changed. In Figures 6 and 7, SampEn is presented as a function of r. SampEn tended to decrease as r increased although SampEn was consistent across the levels of N.

Step Width: After accounting for the effects of m, r and N, no differences in ApEn (SE$_{diff}$: 0.012; $F_{1,48}$: 0.18; p=0.67) and SampEn (SE$_{diff}$: 0.034; $F_{1,48}$: 2.48; p=0.12) between the young and older adults were found. A significant interaction between m, r and N was found for ApEn ($F_{1,3543}$: 722.95; p<0.0001). The slope of ApEn as a function of r and N were different from each other depending on the m chosen, and the value of the other (Figures 4-7). For SampEn, significant 2-way interactions between m, r and N were found [($r*m$: $F_{1,3245}$: 165.83; p<0.0001), ($N*m$: $F_{1,3245}$: 20.54; p<0.0001), ($r*N$: $F_{1,3245}$: 146.85; p<0.0001)]. Just as with step length, SampEn was consistent across values of N although there were differences in SampEn as r changed (Figures 4 and 5). When SampEn is presented as a function of r (Figures 6 and 7), SampEn tended to decrease as r increased although SampEn was consistent across the levels of N.

Step Time: After accounting for the effects of m, r and N, no differences in ApEn (SE$_{diff}$: 0.033; $F_{1,49}$: 1.15; p=0.29) and SampEn (SE$_{diff}$: 0.068; $F_{1,49}$: 0.33; p=0.57) between the young and older adults were found. However, a significant interaction between m, r and N was found for ApEn ($F_{1,6854}$: 70.08; p<0.0001). The slope of ApEn as a function of r and N were different from each other depending on the m chosen, and the value of the other (Figures 4-7). For SampEn, significant 2-way interactions between m, r and N were found [($r*m$: $F_{1,6580}$: 71.67; p<0.0001), ($N*m$: $F_{1,6580}$: 36.25; p<0.0001), ($r*N$: $F_{1,6580}$: 30.76; p<0.0001)]. Just as with step length and width, SampEn was consistent across values of N although SampEn changed as r changed (Figures 4 and 5).

SampEn tended to decrease as r increased although it was consistent across the levels of N (Figures 6 and 7).

Observational Analysis of relative consistency

The relative consistency of ApEn became problematic for all experimental data sets for $m=2$ (Figure 4) and r was set at 0.25 and 0.3. Older adults demonstrated a higher ApEn when $N=100$ but as N increased to 200, younger adults demonstrated a higher ApEn value. Figure 4 also reveals that relative consistency for step width SampEn was challenged for smaller r values. The issue with relative consistency in ApEn was clearly demonstrated in the theoretical data as well (Figure 6). At small r values, the randomized logistic data yielded lower ApEn values than the chaotic logistic map, which is known to be incorrect [4]. SampEn demonstrated a similar issue in the step time data where the young adults yielded lower SampEn values at lower r values and higher SampEn values at higher r values (Figure 6).

Bias in calculation of ApEn

The number of self-matches increased as the data length increased (Figure 8). As r increased the number of self-matches decreased (Figure 9). This was true for both $m=2$ and $m=3$. There were far less self-matches for $m=2$ and when r was between 0.2 and 0.3. When $m=2$, the number of self-matches was stable at larger r values, i.e. 0.15 and greater. Figures 8 and 9 demonstrate the magnitude of the bias associated with the ApEn algorithm; when $m=2$, the use of a larger r value appeared to limit the amount of bias associated with the calculation of ApEn.

Discussion

The use of entropy as a mathematical algorithm to describe predictability of spatio-temporal gait parameters (e.g. step length, step width, minimum toe clearance) is an emerging practice in human movement research [14-16]. While previous studies have provided new and important insights into a variety of clinical problems using these nonlinear analyses, it is essential to adopt a prudent approach to the various techniques available for the quantification of entropic properties of a signal. Previous studies have investigated the use of these techniques in various biological signals [4, 33-35]; however, the application of entropy calculations to gait data requires further exploration. The current study was aimed at the following three goals: 1) to determine the correct input parameters to use, 2) determine which algorithm clearly demonstrates relative consistency and 3) to determine which algorithm provides the best discrimination between groups. It was hypothesized that 1) the value of ApEn and SampEn would change as a function of m, r and N, 2) SampEn would demonstrate better relative consistency as compared to ApEn and that 3) both algorithms would able to discriminate between the theoretical data types and the experimental data groups, young and older adults. The results fully supported hypothesis 1 and only partially supported hypotheses 2 and 3. Overall the results demonstrate that both ApEn and SampEn are extremely sensitive to parameter choice in short data sets and extreme caution should be used when choosing parameters for gait related studies.

Significant 2-way or 3-way interactions between m, r and N were clearly shown for all data sets employed. This indicates that the entropy values are dependent on the combination choice of m, r, and N. ApEn entropy values were influenced depending on input parameter combination regardless of whether theoretical or experimental data was analyzed. On the other hand, SampEn produced

results for the experimental data sets that clearly demonstrated independence of data length regardless of the choice of *m* or *r*, consistent with previous findings [27] (Figures 4 and 5).

It is reasonable to speculate that the sensitivity to data length would plateau at a certain level. Stabilization in entropy would be expected with greater *N*. In unpublished data from our laboratory, we examined the chaotic theoretical data for *N* lengths up to 10,000 and found that ApEn and SampEn clearly stabilized around 2,000 data points (see Appendix A). It is unknown at this time if this would be true for experimental data as well. Future work should repeat this investigation in longer spatio-temporal gait parameter time series. For now we would recommend the *N* to be larger than 200, and as large as possible with respect to the practical constraints of the experiment (i.e. fatigue, pathology investigated). It must be kept in mind however, that stationarity and drift in the data may increase with an increasing *N* and that this drift would have an effect on entropy calculations [6] . Although non-stationarity was present in roughly 1/3 of our experimental trials, it did not result in significant difference in the entropy calculations. This may not be true for other data sets and stationarity should always be examined. Another consideration to keep in mind when choosing *N* is whether or not the length of the time series being analyzed is sufficient to capture the dynamics of the system. It is possible that the dynamics and therefore the related biological complexity of the movement pattern could not be captured in such short data sets thus affecting the results. On the other hand, for some pathological populations, collecting a trial of 200 data points may be difficult and therefore, if even shorter data sets are to be considered, a similar procedure to what is presented in this paper should be conducted. This will allow for the best choice in parameters.

Our results highlighted issues with relative consistency in the current data sets for both ApEn and SampEn calculations. Relative consistency relates to the stability of the measure. When using an entropy algorithm, the investigator should consider if it is providing a consistent value across different r values. Inconsistent values for ApEn and SampEn in the current data are problematic because the true relative direction of differences between the young and older adults is unknown. When utilizing any nonlinear mathematical tool (e.g. entropy) it is critically important that the relative differences between groups are not an artifact of parameter choice. Slight changes in the input parameter choices should be investigated and their effect on the results should be reported. This ensures that the relative difference, not necessarily the magnitude of difference but rather the direction of differences, between groups remains stable.

In the current study, the relative magnitude of the numbers switched when r = 0.15 and when r = 0.20, suggesting that ApEn with the typical r choice of 0.2 may not be the best. However, our results do not clearly reveal a better alternative. Choosing a higher r value of 0.25 or 0.3 then the relationship becomes unstable with respect to changing data length, consistent with previous reports [6]. Conversely, choosing a smaller r can lead to an increased number of self-matches. This truly illustrates the fact that investigators should be meticulous and careful when choosing parameter values, especially r. Data sets are unique and thus, r values may differ. Karmakar et al (2007) investigated the entropy of minimum toe clearance from 500 consecutive steps using $m = 3$ and a range of r values from 0.1 to 0.9 [15]. In their study, they tested the group differences in their data across a range of r values, which provides important information regarding the consistency of their results. This approach may be the most appropriate in dealing with the issue of choosing the correct r value. Certainly, we recommend that a range of r values should be tested

in a piloting phase. Going a step further and reporting these results may lead to greater transparency and understanding of these methods, as they are applied to gait data.

Typically it is suggested that for clinical data, m is to be set at 2 when utilizing the ApEn algorithm [30]. An m of 2 has also been utilized in even the earliest papers reporting SampEn [34]. Yet, an m of 3 has been found to be acceptable for analysis when using SampEn [19]. For the current study, m would represent the distance between consecutive steps that are compared to each other. By choosing an m of 3, one would compare the first and last step, of three consecutive steps, to the first and last step of another set of three consecutive steps. By choosing an m of 2, the algorithm compares one set of consecutive steps to another set of consecutive steps. We found that an m of 2 provides reasonable results for the theoretical data and a lower number of self-matches for the experimental data. In addition, SampEn does not contain an upper limit in values and this made it difficult to compare results when m was increased to 3. For many of the smaller r values, SampEn converged toward infinity. Therefore, thoughtful consideration is required when choosing the m parameter value, bearing in mind what the choice of m represents in biological terms with respect to individual data sets (in our case, the steps being compared). Choosing a value, like 2, that has been utilized frequently in literature, will allow comparison of study results to previously published findings. Yet, a choice of the right parameter combinations for individual data sets may yield an m value that is different from 2.

Based on our current findings, it appears that SampEn is more reliable for short gait data sets, as proposed in previous studies [19, 34]. SampEn was less sensitive to changes in data length, demonstrated fewer problems with relative consistency and does not contain the inherent bias

associated with the ApEn algorithm. How exactly should one go about choosing parameter values? Based on the current findings, as well as past studies, it appears that there is not a set combination of parameters that will work every time. Time and care will have to be devoted to parameter choice when piloting a project and based on the current findings, it is suggested that a range of parameter combinations be examined before data collection. It will be important to examine $m = 2$ and 3, r values ranging from 0.1 to 0.3 and N as appropriate to capture the dynamics of the system. Further, one should avoid a parameter combination that demonstrates an abrupt change from the parameter combinations nearby. For example, using data from the current study, one would not want to choose a parameter combination of $m = 2$: $r = 0.25$, $N = 200$ for step length ApEn, as the parameter combinations immediately surrounding it yielded an opposite finding, young adults demonstrating a greater ApEn.

This study used spatio-temporal gait data from young and elderly subjects to investigate the performance of ApEn and SampEn with changing parameters. ApEn and SampEn of the step length time series were able to discriminate between groups. Step width and step time were not sensitive enough to the differences in age groups. However, we do not know if healthy young do in fact demonstrate more (or less) regularity in step length as compared to older adults, e.g. do the older adults exhibit consistently higher values across the range of r values or does the trend invert? This is a critical question that needs to be answered. Entropy quantifies the degree in which complexity is present in movement. A healthy gait pattern is flexible and adaptable, demonstrating complexity. A loss of complexity, either too rigid (low entropy) or erratic (high entropy) may reveal important information about motor control processes associated with aging. Thus, future studies should investigate the ApEn, SampEn and other measures of entropy (e.g. Multiscale

Entropy [6]) of multiple gait parameters in healthy young and healthy older adults to determine if age is in fact associated with a loss in regularity.

Future studies may also evaluate the sensitivity of entropy analysis through the generation of periodic signals with increasing levels of noise. This could be done for known signals with increasing levels of periodicities as well. These generated signals may reflect the nested periodicities that are found within human movement patterns and would be useful for a theoretical perspective.

The purpose of this study was to investigate short gait data sets. It is feasible that a data length of 200 is simply too short and that a longer data set would yield more consistent findings. Future work should determine what data lengths are required so that ApEn and SampEn provide comparable insights (e.g. both ApEn and SampEn demonstrate that young adults walk with more regular patterns than older adults) and exhibit a consistent pattern with changing parameters.

Appendix A

In order to understand the stabilization of entropy values as N increased in theoretical data, the chaotic logistic map was subjected to entropy analysis up to an N of 10,000 data points. For this particular analysis, an m of 2 and r of 0.20 times the standard deviation of the time series were chosen, as they are the most popular choice in past and current literature. Using the same procedures as outlined in the methods above, SampEn and ApEn were quantified on generated

data from 100 to 10,000 data points, in increments of 100. As can be seen in Figure 10, the entropy values for both SampEn and ApEn stabilize around an *N* of 2,000.

Acknowledgments

References

1. Aboy M., Cuesta-Frau D., Austin D. and Mico-Tormos P. Characterization of sample entropy in the context of biomedical signal analysis. Conf.Proc.IEEE Eng.Med.Biol.Soc. 2007:5943-5946, 2007.

2. Bar-Yam Y. Dynamics of complex systems. Boulder, Colorado: Westview Press. 2003.

3. Cavanaugh J.T., Kochi N. and Stergiou N. Nonlinear analysis of ambulatory activity patterns in community-dwelling older adults. J.Gerontol.A Biol.Sci.Med.Sci. 65:197-203, 2010.

4. Chen X., Solomon I. and Chon K. Comparison of the use of approximate entropy and sample entropy: applications to neural respiratory signal. Conf.Proc.IEEE Eng.Med.Biol.Soc. 4:4212-4215, 2005.

5. Chon K., Scully C.G. and Lu S. Approximate entropy for all signals. IEEE Eng.Med.Biol.Mag. 28:18-23, 2009.

6. Costa M., Goldberger A.L. and Peng C.K. Multiscale entropy analysis of biological signals. Phys.Rev.E 71:021906, 2005.

7. Cuesta-Frau D., Miro-Martinez P., Oltra-Crespo S., Varela-Entrecanales M., Aboy M., Novak D. and Austin D. Measuring body temperature time series regularity using Approximate Entropy and Sample Entropy. Conf.Proc.IEEE Eng.Med.Biol.Soc. 2009:3461-3464, 2009.

8. Deffeyes J.E., Harbourne R.T., Stuberg W.A. and Stergiou N. Approximate entropy used to assess sitting postural sway of infants with developmental delay. Infant Behav.Dev. 34:81-99, 2011.

9. Faure P. and Korn H. Is there chaos in the brain? I. Concepts of nonlinear dynamics and methods of investigation. C.R.Acad.Sci.III 324:773-793, 2011.

10. Gell-Mann M. The quark and the jaguar: Adventures in the simple and the complex. New York, New York: Owl Books. 1994.

11. Georgoulis A.D., Moraiti C., Ristanis S. and Stergiou N. A novel approach to measure variability in the anterior cruciate ligament deficient knee during walking: the use of the approximate entropy in orthopaedics. J.Clin.Monit.Comput. 20:11-18, 2006.

12. Grabiner M.D. and Troy K.L. Attention demanding tasks during treadmill walking reduce step width variability in young adults. J.Neuroeng.Rehabil. 2:25, 2005.

13. Hong S.L. and Newell K.M. Motor entropy in response to task demands and environmental information. Chaos 18: 033131, 2008.

14. Kaipust J.P., Huisinga J.M., Filipi M. and Stergiou N. Gait Variability Measures Reveal Differences Between Multiple Sclerosis Patients and Healthy Controls. Motor Control 16:229-244, 2012.

15. Karmakar C.K., Khandoker A.H., Begg R.K., Palaniswami M. and Taylor S. Understanding ageing effects by approximate entropy analysis of gait variability. Conf.Proc.IEEE Eng.Med.Biol.Soc. 2007:1965-1968, 2007.

16. Khandoker A.H., Palaniswami M. and Begg R.K. A comparative study on approximate entropy measure and poincaré plot indexes of minimum foot clearance variability in the elderly during walking. J.Neuroeng.Rehabil. 5:4, 2008.

17. Korn H. and Faure P. Is there chaos in the brain? II. Experimental evidence and related models. C.R.Biol 326:787-840, 2003.

18. Kurz M.J. and Hou J.G. Levodopa influences the regularity of the ankle joint kinematics in individuals with Parkinson's disease. J.Comput.Neurosci. 28:131-136, 2010.

19. Lake D.E., Richman J.S., Griffin M.P. and Moorman J.R. Sample entropy analysis of neonatal heart rate variability. Am.J.Physiol.Regul.Integr.Comp.Physiol. 283:R789-R797, 2002.

20. Lamoth C.J., van Deudekom F.,J., van Campen J.P., Appels B.A., de Vries O.J. and Pijnappels M. Gait stability and variability measures show effects of impaired cognition and dual tasking in frail people. J.Neuroeng.Rehabil. 8:2, 2011.

21. Lamoth C.J.C., Ainsworth E., Polomski W. and Houdijk H. Variability and stability analysis of walking of transfemoral amputees. Med.Eng.Phys. 32:1009-1014, 2010.

22. Lipsitz L.A. Dynamics of stability: The physiologic basis of functional health and frailty. J.Gerontol.A Biol.Sci.Med.Sci. 57:115-125, 2002.

23. Lipsitz L.A. and Goldberger A.L. Loss of 'complexity' and aging. Potential applications of fractals and chaos theory to senescence. JAMA 267:1806-1809, 1992.

24. Liu C., Liu C., Shao P., Li L., Sun X., Wang X. and Liu F. Comparison of different threshold values r for approximate entropy: application to investigate the heart rate variability between heart failure and healthy control groups. Physiol.Meas. 32:167-180, 2011.

25. Lu S., Chen X., Kanters J., Solomon I.C. and Chon KH. Automatic selection of the threshold value R for approximate entropy. IEEE Trans.Biomed.Eng. 55:1966-1972, 2008.

26. Mayer-Kress G., Liu Y. and Newell K.M. Complex systems and human movement. Complexity 12:40-51, 2006.

27. Molina-Picó A., Cuesta-Frau D., Aboy M., Crespo C., Miró-Martínez P. and Oltra-Crespo S. Comparative study of approximate entropy and sample entropy robustness to spikes. Artif.Intell.Med. 53:97-106, 2011.

28. Myers S.A., Stergiou N., Pipinos I.I. and Johanning J.M. Gait variability patterns are altered in healthy young individuals during the acute reperfusion phase of ischemia-reperfusion. J.Surg.Res. 164:6-12, 2010.

29. Pincus S. Approximate entropy (Apen) as a complexity measure. Chaos 5:110-117, 1995.

30. Pincus S. and Huang W. Approximate entropy - Statistical properties and applications. Commun.Stat.Theory Methods 21:3061-3077, 1992.

31. Pincus S. Approximate entropy as a measure of aystem-complexity. Proc.Natl.Acad.Sci.U.S.A. 88:2297-2301, 1991.

32. Rathleff M.S., Samani A., Olesen C.G., Kersting U.G. and Madeleine P. Inverse relationship between the complexity of midfoot kinematics and muscle activation in patients with medial tibial stress syndrome. J.Electromyogr.Kinesiol. 21:638-644, 2011.

33. Rhea C.K., Silver T.A., Hong S.L., Ryu J.H., Studenka B.E., Hughes C.M.L. and Haddad J.M. Noise and complexity in human postural control: interpreting the different estimations of entropy. PLoS One 6:e17696, 2011.

34. Richman J.S. and Moorman J.R. Physiological time-series analysis using approximate entropy and sample entropy. Am.J.Physiol.Heart Circ.Physiol. 278:H2039-H2049, 2000.

35. Rose M.H., Bandholm T. and Jensen B.R. Approximate entropy based on attempted steady isometric contractions with the ankle dorsal- and plantarflexors: reliability and optimal sampling frequency. J.Neurosci.Methods 177:212-216, 2009.

36. Sarlabous L., Torres A., Fiz J.A., Gea J., Martinez-Llorens J., Morera J. and Jane R. Interpretation of the approximate entropy using fixed tolerance values as a measure of amplitude variations in biomedical signals. Conf.Proc.IEEE Eng.Med.Biol.Soc. 2010:5967-5970, 2010.

37. Smith B.A., Teulier C., Sansom J., Stergiou N. and Ulrich B.D. Approximate entropy values demonstrate impaired neuromotor control of spontaneous leg activity in infants with myelomeningocele. Pediatr.Phys.Ther. 23:241-247, 2011.

38. Sosnoff J.J., Goldman M.D. and Motl R.W. Real-life walking impairment in multiple sclerosis: preliminary comparison of four methods for processing accelerometry data. Mult.Scler. 16:868-877, 2010.

39. Stergiou N., Harbourne R. and Cavanaugh J. Optimal movement variability: a new theoretical perspective for neurologic physical therapy. J.Neurol.Phys.Ther. 30:120-129, 2006.

40. Turnock M.J.E. and Layne C.S. Variations in linear and nonlinear postural measurements under achilles tendon vibration and unstable support-surface conditions. J.Mot.Behav. 42:61-69, 2010.

41. Wolf A., Swift J.B., Swinney H.L. and Vastano J.A. Determining Lyapunov exponents from a time-series. Physica D 16:285-317, 1985.

Figure Captions

Figure 1. Total number of publications using Approximate Entropy or Sample Entropy listed on PubMed from 1991- July 2012.

Figure 2. Representative time series for theoretical and average time series for experimental data sets. Theoretical data includes: periodic logistic map (A), chaotic logistic map (B) and random white noise (C). Experimental data shown are from step length (D), step time (E) and step width (F). For experimental data black represents the older adults and gray are the young adults. The mean and standard deviation of the entire time series is also shown in the figure.

Figure 3. Visual representation of the calculation of ApEn and SampEn.

Figure 4. ApEn (top row) and SampEn (bottom row) as a function of data length when m=2. For theoretical data, the filled shapes represent the various levels of r for chaotic logistic maps, while the open shapes represent he various levels of r for randomized logistic maps. For all experimental data, the filled shapes represent young adults and the open shapes represent older adult data.

Figure 5. ApEn (top row) and SampEn (bottom row) as a function of data length when m=3. For theoretical data, the filled shapes represent the various levels of r for chaotic logistic maps, while the open shapes represent he various levels of r for randomized logistic maps. For all experimental data, the filled shapes represent young adults and the open shapes represent older adult data. All values in which SampEn is equal to 4 are not true data. SampEn does not contain an upper limit

for calculation. In order to make graphs more legible at lower resolutions (between 0-3), extremely large SampEn values were replaced with the value of 4 for illustration purposes only. These numbers were typically larger than 1,000.

Figure 6. ApEn (top row) and SampEn (bottom row) as a function of tolerance (r) when m=2. For theoretical data, the filled shapes represent the various levels of N for chaotic logistic maps, while the open shapes represent he various levels of N for randomized logistic maps. For all experimental data, the filled shapes represent young adults and the open shapes represent older adult data.

Figure 7. ApEn (top row) and SampEn (bottom row) as a function of tolerance (r) when m=3. For theoretical data, the filled shapes represent the various levels of N for chaotic logistic maps, while the open shapes represent he various levels of N for randomized logistic maps. For all experimental data, the filled shapes represent young adults and the open shapes represent older adult data.

Figure 8. The number of self-matches (bias in ApEn) for when m=2 (top row) and m=3 (bottom row) as a function of data length (N). For theoretical data, the filled shapes represent the various levels of r for chaotic logistic maps, while the open shapes represent he various levels of r for randomized logistic maps. For all experimental data, the filled shapes represent young adults and the open shapes represent older adult data.

Figure 9. The number of self-matches (bias in ApEn) for when m=2 (first row) and m=3 (second row) as a function of tolerance (r). For theoretical data, the filled shapes represent the various levels of N for chaotic logistic maps, while the open shapes represent he various levels of N for

randomized logistic maps. For all experimental data, the filled shapes represent young adults and the open shapes represent older adult data.

Figure 10. The Apen and SampEn values are plotted for the chaotic logistic map using an *m* of 2 and *r* of 0.2. The entropy values were calculated for data of lengths 100 to 10,000 increasing in increments of 100. Based on this figure, it appears that the entropy values stabilize around an *N* of 2,000.