Am J Physiol Regul Integr Comp Physiol 283: R789–R797, 2002;
10.1152/ajpregu.00069.2002.

report

Sample entropy analysis of neonatal heart rate variability

DOUGLAS E. LAKE, JOSHUA S. RICHMAN,
M. PAMELA GRIFFIN, AND J. RANDALL MOORMAN
Departments of Internal Medicine (Cardiovascular Division) and Pediatrics, University of Virginia, Charlottesville, Virginia 22908

Received 4 February 2002; accepted in final form 26 May 2002

Lake, Douglas E., Joshua S. Richman, M. Pamela Griffin, and J. Randall Moorman. Sample entropy analysis of neonatal heart rate variability. *Am J Physiol Regul Integr Comp Physiol* 283: R789–R797, 2002; 10.1152/ajpregu.00069.2002.—Abnormal heart rate characteristics of reduced variability and transient decelerations are present early in the course of neonatal sepsis. To investigate the dynamics, we calculated sample entropy, a similar but less biased measure than the popular approximate entropy. Both calculate the probability that epochs of window length m that are similar within a tolerance r remain similar at the next point. We studied 89 consecutive admissions to a tertiary care neonatal intensive care unit, among whom there were 21 episodes of sepsis, and we performed numerical simulations. We addressed the fundamental issues of optimal selection of m and r and the impact of missing data. The major findings are that entropy falls before clinical signs of neonatal sepsis and that missing points are well tolerated. The major mechanism, surprisingly, is unrelated to the regularity of the data: entropy estimates inevitably fall in any record with spikes. We propose more informed selection of parameters and reexamination of studies where approximate entropy was interpreted solely as a regularity measure.

approximate entropy; newborn infant; sepsis

PREMATURE INFANTS in the neonatal intensive care unit (NICU) are at high risk for developing bacterial sepsis. Diagnostic tests for neonatal sepsis are imperfect, especially early in the course of the illness. A possible approach to improved diagnosis of neonatal sepsis is continuous monitoring of heart rate (HR) characteristics; early neonatal sepsis is marked by reduced baseline variability and transient decelerations of HR, similar to the findings of fetal distress (6).

In their study of HR records from fetuses and newborn infants, Pincus and co-workers (19, 21) interpreted these changes as representative of a change in the complexity of the physiological processes underlying control of HR and quantified them using approximate entropy (ApEn). Pincus developed this statistical measure from theory developed in the field of nonlinear dynamic analysis and chaos (16, 20). Distressed fetuses and sick newborns sometimes showed reduced ApEn, interpreted as an increased regularity of cardiac rhythm.

In this context, entropy is the rate of generation of new information. ApEn(m,r,N) is approximately the negative natural logarithm of the conditional probability (CP) that a dataset of length N, having repeated itself within a tolerance r for m points, will also repeat itself for $m + 1$ points. An important point to keep in mind about the parameter r is that it is commonly expressed as a fraction of the SD of the data and in this way makes ApEn a scale-invariant measure. A low value arises from a high probability of repeated template sequences in the data. We define B to be the number of matches of length m, and A to be the subset of B that also matches for length $m + 1$. Thus CP = A/B. For ApEn, one calculates $-\log$ CP for each template and averages these values for all the templates. Since neither A nor B can be 0, CP must be redefined to $(1 + A)/(1 + B)$, a correction that can be rationalized as allowing templates to match themselves. This is obviously inconsistent with the idea of new information, however, and is a strong source of bias toward CP = 1 and ApEn = 0 when there are few matches and A and B are small (16, 18, 20, 22).

We developed a new related measure of time series regularity that we have called sample entropy (SampEn) (22). SampEn(m,r,N) is precisely the negative natural logarithm of the CP that a dataset of length N, having repeated itself within a tolerance r for m points, will also repeat itself for $m + 1$ points, without allowing self-matches. SampEn does not use a templatewise approach, and A and B accrue for all the templates. SampEn was designed to reduce the bias of ApEn and has closer agreement with theory for data-

Address for reprint requests and other correspondence: J. R. Moorman, Box 6012, MR4 Bldg., UVAHSC, Charlottesville, VA 22908 (E-mail: rmoorman@virginia.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

sets with known probabilistic content. Moreover, SampEn displays the property of relative consistency in situations where ApEn does not. That is, if one record shows lower SampEn than another with one set of values of m and r, it also shows lower SampEn with different values. These measures have an indirect interpretation, i.e., the extent to which the data did not arise from a random process. The terms order, regularity, complexity, and ensemble orderliness have all been used to reflect this idea, and we will use the term regularity.

Although m and r are critical in determining the outcome of either method for entropy estimation, no guidelines exist for optimizing their values. We view this as a severe shortcoming in the current art. The various existing rules generally lead to the use of values of r between 0.1 and 0.25 and values of m of 1 or 2 for data records of length N ranging from 100 to 5,000 data points (7, 15, 20). In principle, the accuracy and confidence of the entropy estimate improve as the numbers of matches of length m and $m + 1$ increase. Intuitively, the advantages of SampEn include larger values of A and B and hence more confident estimation of CP. In either method, the number of matches can be increased by choosing small m (short templates) and large r (wide tolerance). There are penalties, however, for criteria that are too relaxed (16). First, there is a theoretical concern. While these calculations only aim to estimate, entropy is defined in the limit as m approaches ∞ and as r approaches 0. Second, there are practical concerns. As r increases, the probability of matches tends toward 1 and SampEn tends to 0 for all processes, thereby reducing the ability to distinguish any salient features in the dataset. As m decreases, underlying physical processes that are not optimally apparent at smaller values of m may be obscured.

The results of Pincus and co-workers (19, 21) motivated us to examine entropy as a possibly useful indicator of early stages of neonatal sepsis. We thus have implemented SampEn analysis of a relevant clinical dataset, and we have systematically addressed several practical and general questions inherent in estimating entropy. *1*) What are the confidence intervals (CIs) of SampEn estimates? *2*) How do we pick m and r? *3*) Does entropy fall before the clinical diagnosis of neonatal sepsis? *4*) Why does entropy diminish when reduced variability and transient decelerations are present? *5*) Do missing points matter?

Our results extend those of Pincus and co-workers (19) as we detect lower SampEn before the clinical diagnosis of neonatal sepsis. In addition to their clinical relevance, the findings shed light on several general issues in using entropy estimates. In particular, the findings argue against unguided use of the parameters m and r and against an unquestioned acceptance of the idea that differences in entropy estimates are always the result of differences in time series regularity.

METHODS

Clinical data. We studied 89 infants admitted consecutively to the University of Virginia NICU over a period of 9 mo using a previously described data collection and analysis procedure (6). The data collection protocol was approved by the University of Virginia Human Investigations Committee. Data records consisting of 4,096 R-R intervals were collected over ~25 min. There were 73,097 of these data records collected during the study. To remove trends (20), each was high-pass filtered by subtracting a low-pass filtered version of the record. Sepsis and sepsislike illness were defined as an acute clinical deterioration that prompted a physician to obtain a blood culture and to administer antibiotics (6). To evaluate the diagnostic potential of SampEn, we excluded data from the first 7 days after birth and for 14 days after each of the sepsis events.

For comparison of SampEn values on the day of illness to other days, records were parsed into 6-h epochs beginning at midnight and labeled according to whether an episode of sepsis and sepsislike illness occurred in the next 24 h. As a robust marker of reduced SampEn, the 10th percentile value of SampEn was determined for each 6-h epoch. Using this value as a test statistic, the receiver operating characteristic (ROC) area for distinguishing days containing events from nonevent days was calculated. The CIs and significance level for the ROC area estimates were obtained via bootstrapping. The statistical significance was assessed for the coefficient of a logistic discrimination model whose variance was robustly determined by taking into account the repeated measures on individual infants. The added diagnostic information of SampEn over traditional methods of predicting sepsis was evaluated using the Wald chi square test on models developed with the demographic variables of gestational age and birth weight.

Choosing the parameters m and r. To aid in selecting the parameters for SampEn, a random sample of 200 records of 4,096 R-R intervals was selected from the clinical dataset. The numbers of matches of lengths m and $m + 1$, CP, SampEn, and CI estimates were calculated for 16 values of r between 0.01 and 0.8 and for values of m from 1 to 10. Estimated results for r not explicitly calculated were obtained by linear interpolation. To further evaluate the parameter m, autoregressive (AR) models of various orders were fit to the data using methods discussed in APPENDIX A.

RESULTS

CIs of entropy estimates. One of our major goals is to devise a general strategy for optimal selection of m and r. Part of this strategy is to ensure that the length of the CI around the SampEn estimate is acceptable.

The statistic $CP = A/B$ estimates the CP of a match of length $m + 1$ given there is a match of length m. The accuracy of the estimate can be judged by the length of its CI, which is proportional to its SE. If B were fixed and all B of the matches of length m were independent of each other, then the random variable A would be binomially distributed and the variance of CP would simply be $CP(1 - CP)/B$. The situation is, however, more complicated because A and B are correlated. For example, there might be many pairs of matches that are dependent because of naturally occurring dependencies in the data or, more directly, because of overlapping pairs of matches with points in common.

It is shown in APPENDIX B that an estimate of the variance is

$$\sigma^2_{CP} = \frac{CP(1-CP)}{B} + \frac{1}{B^2}[K_A - K_B(CP)^2]$$

where K_A is the number of pairs of matching templates of length $m + 1$ that overlap and K_B is the number of pairs of matching templates of length m that overlap. Using the standard approximation $\sigma_{g(CP)} \cong |g'(CP)|\sigma_{CP}$ with $g(CP) = -\log(CP)$ and first derivative $g'(CP) = 1/\sigma_{CP}$, the SE of SampEn can be estimated by σ_{CP}/CP. Thus the SE of SampEn is exactly the relative error of CP. For m small enough and r large enough to ensure a sufficient number of matches, SampEn can be assumed to be normally distributed, and we define the 95% CI for each SampEn calculation to be $-\log(CP) \pm 1.96(\sigma_{CP}/CP)$.

Picking m and r for SampEn analysis of neonatal R-R interval data records. We first determined a range of m that was likely to capture essential elements of the data structure. The AR process order of each record was estimated, and we found 90% of the estimates to be between 3 and 9, inclusive. Based on this, we choose m to be no less than 3.

We now seek a value r that is neither so stringent that the number of matches is too near 0 (low confidence) nor so relaxed that CP is too near 1 (low discrimination) (16). We propose selecting r to minimize the quantity

$$\max\left(\frac{\sigma_{CP}}{CP}, \frac{\sigma_{CP}}{-\log(CP)CP}\right)$$

which is the maximum of the relative error of SampEn and of the CP estimate, respectively. This metric favors estimates with low variance and thus reflects the efficiency of the entropy estimate. In addition, this criterion represents a tradeoff of accuracy and discrimination capability, as it simultaneously penalizes CP near 0 and near 1. Figure 1 shows a color map of the median value of this criterion for 200 randomly selected records from our clinical database, calculated for a range of values of m (rows) and r (columns). The maximum relative error of either SampEn or CP ranges from 0, a deep blue, to 0.15, a deep red, and is black where no matches of length m are found.

Inspection reveals major features of the map. A window length of 1 allows confident estimation of entropy for a wide range of r. For $m \geq 2$, however, optimum values of r are clearly evident and lie generally between ~0.2 and ~0.5. The dependence of the error on r is very steep for low r, and intolerably large error precludes many combinations of m and r even in these relatively large data records of 4,096 points. We aim for a maximum relative error no higher than ~0.05, so that the 95% CI of the entropy estimate is ~10% of its value. For our data, we select $m = 3$ and $r = 0.2$ based on the findings that *1*) $m = 3$ is acceptable because of the AR analysis, *2*) $r = 0.2$ is optimum of $m = 3$, based on inspection of the color map, and *3*) the 95% CI of the estimate is ~10% of its value. We proceeded to analyze the clinical database with $m = 3$ and $r = 0.2$ based on this analysis.

SampEn of neonatal HR falls before the clinical diagnosis of sepsis and sepsislike illness. Figure 2 shows analysis of SampEn in an infant who was diagnosed with sepsis. Figure 2, *A* and *B*, shows plots of

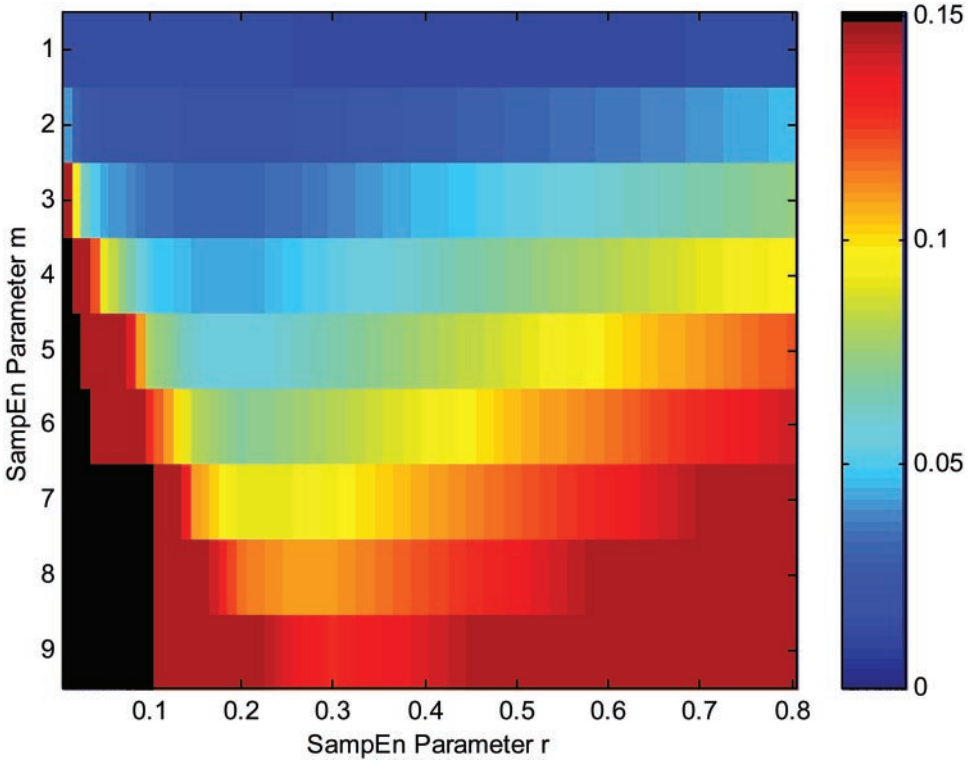

Fig. 1. A visual guide to optimal selection of window length (m) and tolerance (r) parameters for entropy estimation of neonatal heart rate time data records of length 4,096. The median value of a sample entropy (SampEn) efficiency metric for 200 randomly selected data records is plotted in pseudocolor as a function of m and r. A value of 0.05 corresponds to a 95% confidence interval (CI) that is 10% of the SampEn estimate itself. Given a value of m based on a priori reasoning, an optimal value of r can be selected to minimize the efficiency metric. In this example, we selected r to be 0.2 because it led to the best value of the efficiency metric, which was <0.05.

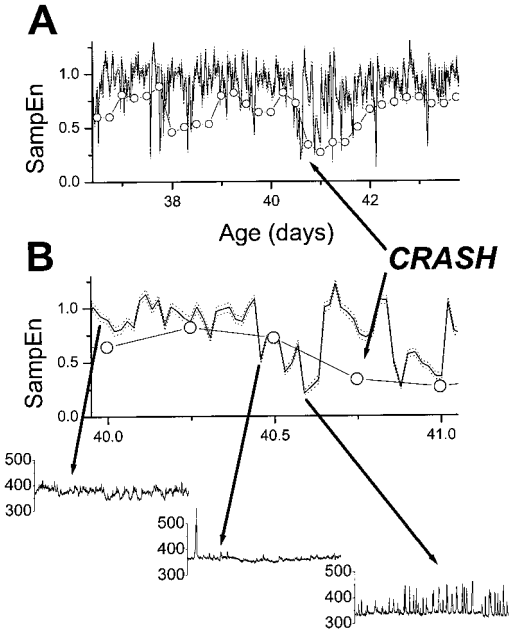

Fig. 2. SampEn is decreased before episodes of neonatal sepsis. *A*: SampEn as a function of time is shown for one infant as a solid line, with SEs shown as dotted lines. The time at which the clinical diagnosis of sepsis was suspected is labeled "CRASH" (Cultures, Resuscitation, and Antibiotics Started Here). The open circles are the 10th percentile-lowest value of the preceding 12 h and are given every 6 h. *B*: data from near the time of diagnosis on an expanded time scale. *B, insets*: time series of 4,096 R-R intervals from the designated times show the development of the characteristic abnormalities of reduced variability and transient decelerations.

SampEn(3,0.2,4,096) as a function of the infant's age, and the dotted line is the 95% CI for each estimate. The infant was diagnosed with sepsis at the point labeled CRASH but beforehand develops the previously described abnormal HR characteristics of reduced variability and transient decelerations (6). The three insets show data records of 4,096 R-R intervals and progress from normal (*left*) to reduced variability with a single large deceleration (*middle*) to repeated decelerations (*right*). None of the decelerations reach 100 beats/min (R-R interval 600 ms) and thus would not have triggered an alarm by the bedside electrocardiogram monitor. Instead, the feature that might have alerted the clinician, i.e., reduced variability and transient decelerations in the second and third records, would not have been reported. SampEn is reduced in the abnormal records, and inspection of this patient's records suggests that SampEn might be a useful tool in the early diagnosis of neonatal sepsis.

We tested this hypothesis using regression analysis in a large clinical dataset from 89 infants in whom 19 episodes of sepsis and sepsislike illness occurred. We chose the 10th percentile value of SampEn over the past 12 h as an appropriate representation. Frequency histograms of SampEn and the 10th percentile values are shown in Fig. 3. Both have near-normal distributions. We found that SampEn was significantly associated with upcoming sepsis and sepsislike illness (ROC area 0.64, 95% CI 0.56–0.74, $P = 0.001$). Moreover, SampEn significantly added diagnostic information to the variables gestational age and birth weight ($P < 0.001$).

Mechanism of reduced entropy for data with reduced variability and transient decelerations: surrogate data records. To understand the effect of reduced baseline variability and transient decelerations, or spikes, on SampEn, we performed analysis and experiments with surrogate data records. These were produced by summing pairs of records as illustrated in Fig. 4.

The first is the mean process μ (Fig. 4*A*), which consists of two modes, a constant rate and the spike. The second is the baseline process b (Fig. 4*B*), which represents the HR variability (HRV). Their sum (Fig. 4*C*) is a surrogate data record. For the spike mode of the mean process, we used either a square step for the analytic results or a clinically observed deceleration lasting 50 R-R intervals that occurred near the time of clinical diagnosis of sepsis in one patient.

We considered three kinds of baseline processes. First, for the analytic results, we used Gaussian random numbers and refer to this as random data or white noise. Second, as demonstrated in Fig. 4, *D–F*, we constructed isospectral surrogate data records. The method is to randomize the phase of inverse Fourier transforms of the power spectral densities of observed data records. Figure 4*D* is an observed clinical data record, scaled in milliseconds on the left axis and in SD on the right axis. Figure 4*E* shows the isospectral surrogate scaled so that it has mean = 0 and SD = variance = 1. Figure 4*F* shows the sum of the record in Fig. 4*E* and the clinically observed deceleration that has been scaled so that the overall variance of the

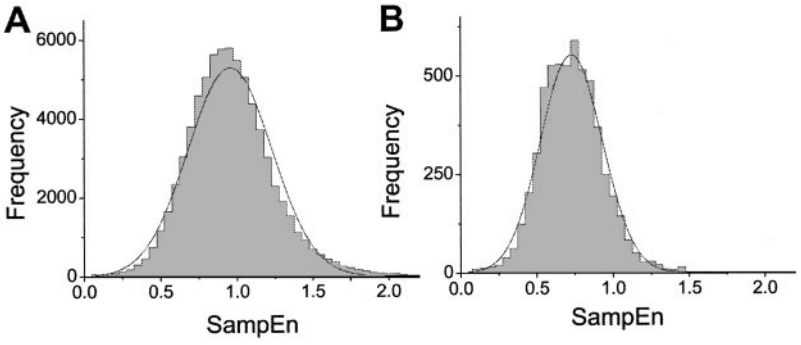

Fig. 3. Frequency histograms of SampEn ($n = 73,097$; *A*) and the 10th percentile value of SampEn from each 12-h epoch ($n = 5,626$; *B*). The smooth lines are Gaussian functions.

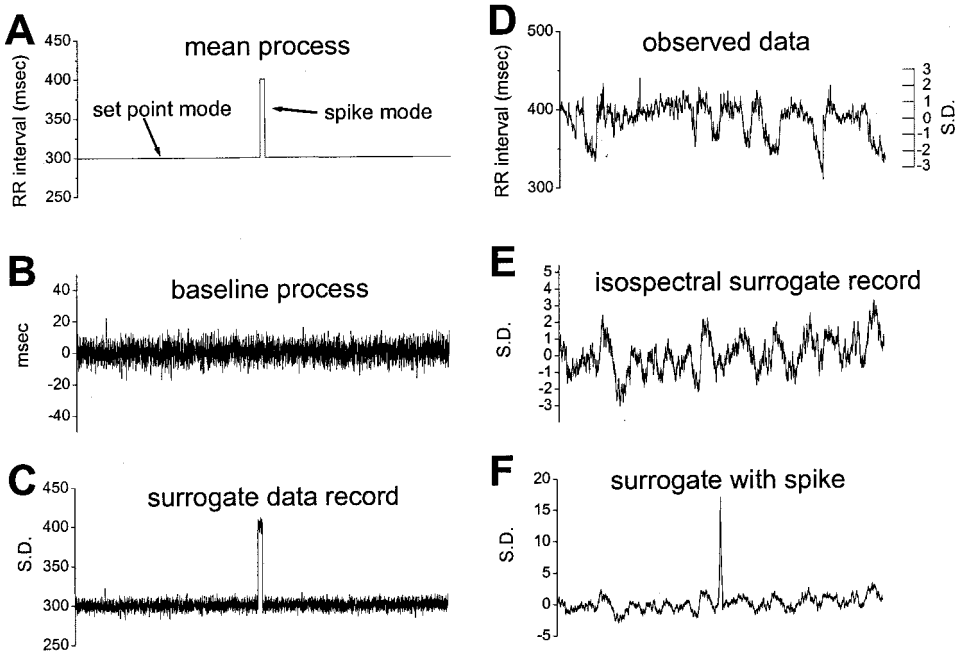

Fig. 4. Surrogate data records. *A* and *B* show the major components. *A*: the mean process, which has set point and spike modes. *B*: the baseline process, here meaning the heart rate variability, modeled as Gaussian random numbers. *C*: their sum, a surrogate data record. *D–F*: a more realistic surrogate with the same frequency content as the observed data. *D*: a clinically observed data record of 4,096 R-R intervals. The lefthand ordinate is labeled in ms and the righthand ordinate in SD. *E*: a 4,096-point isospectral surrogate dataset formed using the inverse Fourier transform of the periodogram of the data in *D*. *F*: the surrogate data after addition of a clinically observed deceleration lasting 50 points and scaled so that the variance of the record is increased from 1 to 2.

record is 2. Thus the spike contributes 50% of the combined variance in this example record.

Finally, as an example of a deterministic model, we mapped series from the logistic map $x_{i+1} = 4x_i(1 - x_i)$ onto a Gaussian random process. The resulting series has the identical marginal density and approximates the spectral properties of Gaussian white noise but, being a deterministic process after an initial random seed x_1, has entropy that does not approach infinity as r approaches 0.

Mechanism of reduced entropy for data with reduced variability and transient decelerations: analytic results. We first analyzed a mean process with a square-topped spike of height Δ and duration εN beats where ε is small. Because of the spike, the mean varies and can be considered as a random variable with two modes, one a constant value μ_0 with probability $1 - \varepsilon$ and another a constant value $\mu_0 + \Delta$ with probability ε. Thus the mean process has variance $\sigma^2_\mu = \Delta^2\varepsilon(1 - \varepsilon)$. We denote the variance of the baseline process by σ^2_b. The combined variance σ^2 of the entire record is the sum of the variances of the mean process and the baseline, or $\sigma^2 = \sigma^2_b + \sigma^2_\mu = \sigma^2_b + \Delta^2\varepsilon(1 - \varepsilon)$. In APPENDIX C it is shown that in this case, the sample entropy can be approximated by

$$\text{SampEn}(m,r,N) \approx -\log\left(\frac{r}{\sqrt{\pi}}\right) - \frac{\Delta^2\varepsilon(1-\varepsilon)}{2\sigma_b^2}$$

The critically important finding from this analysis is that CP increases and SampEn decreases when spikes that are large (increased Δ) or long lasting (increased ε) inflate the combined variance σ^2 and, with it, the tolerance $r\sigma$. As is shown in APPENDIX C, this model can be generalized to a signal that consists of a spectrum of modes. We note that the combined variance can be large relative to the baseline variance in a number of ways other than a large single spike. So, for example, SampEn will also be significantly reduced if there are a moderate number of moderate sized spikes, as we have observed clinically.

We conclude from this analysis that SampEn must fall when spikes inflate the variance of the series.

Mechanism of reduced entropy for data with reduced variability and transient decelerations: experimental results. To investigate the effects of spikes in HR data records, we calculated SampEn in the sample of 200 observed data records, their isospectral surrogates, Gaussian random data, and the deterministic model of the Gaussian-mapped logistic map. Scaled versions of the clinically observed deceleration, or spike, were added, and the results were plotted as a function of the size of the spike measured as its contribution to the combined variance.

Figure 5*A* shows the results of SampEn calculations for white noise, raw data, isospectral surrogate data, and deterministic white noise. As expected from the analytic results above, SampEn of random data falls in the presence of spikes. More relevant to the clinical problem, SampEn values of the observed HR and surrogate data both fall in the presence of spikes. For example, the median SampEn value for the observed data falls from 1.3 to 0.75 when a spike that doubles the combined variance (i.e., 50% of the combined variance is due to the spike) is added. An analogous example is shown as Fig. 4*F*.

We reject the interpretation that these changes in SampEn reflect alterations in the regularity of the points because 99% of the data are not altered. There is only one spike, so recent findings suggesting that ApEn is sensitive to pulse frequency do not apply (28). Instead, we interpret the changes to be the arithmetic

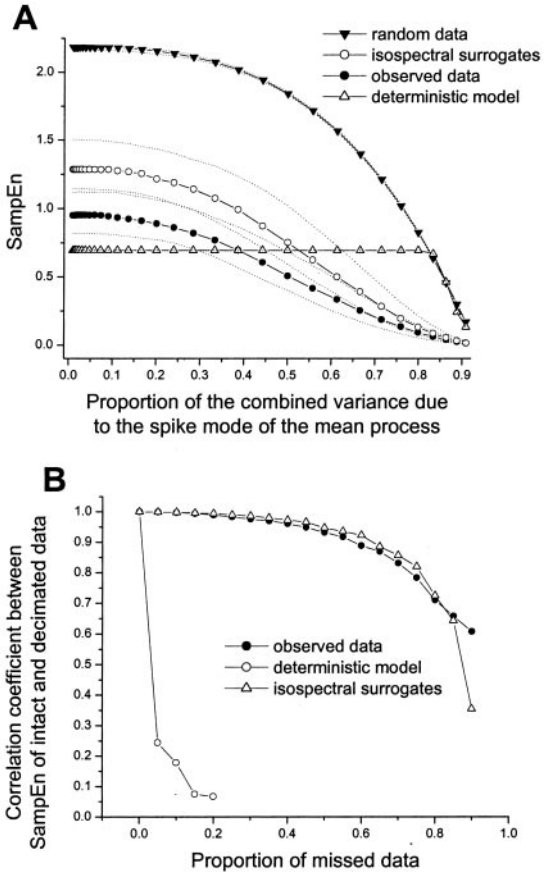

Fig. 5. Entropy estimates of heart rate and surrogate data are affected by spikes but not by missing points. *A*: SampEn falls in the presence of spikes. Data points are the median, and the dotted lines are the 95% CI for SampEn values of 200 series of Gaussian random numbers, observed data and their isospectral surrogates, and a deterministic model. *B*: entropy estimates of 200 observed data records and their isospectral surrogates are largely unchanged by removal of even large proportions of data points. The entropy estimate of a deterministic series, however, is altered by removal of a small proportion of points.

consequence of outlying points that inflate the combined variance.

SampEn values of the raw data were significantly less than the surrogate data ($P < 0.001$, Wilcoxon rank sum test). This result has been used in fetal HR analysis to demonstrate nonlinearity in the data (7). Given the presence of spikes, this is not a proper conclusion here. This result only suggests that neonatal HR data contain some combination of nonstationary and nonlinear characteristics not captured by a linear stationary process. Finding methods that can distinguish among these components is a very important future task.

A possible approach to addressing this issue would be to model nonstationarity into the surrogate data to isolate nonlinear effects. Similarly, spikes, outliers, and other nonstationary properties of the HR data could be removed from the data before the analysis. These methods rely on a sufficient understanding of the mechanism of nonstationarity in HR data and the development of a robust mathematical model. For example, the occurrence of spikes can be modeled as a Poisson process with the height and duration of the spikes following some random distribution. Neonatal HR data, however, include many nonstationary features, and more sophisticated models are likely necessary for their description.

It is noteworthy that SampEn of the deterministic model is unaffected by spikes contributing <80% of the combined variance. Thus a process in which SampEn is driven exclusively by the regularity is much more resistant to the effect of outlying points that inflate the variance.

We conclude from this analysis that reduced SampEn in neonatal HR data is strongly influenced by spikes rather than increased regularity.

Is SampEn sensitive to missing points? Long HR data records are vulnerable to missing points, as electrocardiographic waveforms are easily distorted by motion of the patient. SampEn, however, is potentially very sensitive to missing points as it is based entirely on the ordering of the data. We randomly removed varying proportions of the 200 sample clinical datasets and calculated the correlation coefficient between SampEn of the intact and the decimated datasets. Figure 5*B* shows that randomly removing as much as 40% of the R-R interval data did not reduce the correlation coefficient below 0.95. On the other hand, removal of even a small proportion of points from a series from the logistic map drastically alters SampEn. We conclude from this analysis that missing points do not greatly alter the SampEn estimate for clinical HR data records, arguing further against increased regularity as the major mechanism of low entropy before the clinical diagnosis of neonatal sepsis.

DISCUSSION

We studied SampEn, which was developed as a measure of time series regularity of R-R interval data records from newborn infants. Our most important findings are *1*) parameters for SampEn estimation can be optimized; *2*) SampEn falls early in the course of neonatal sepsis and sepsislike illness; *3*) SampEn falls in the presence of spikes in a record with reduced variability; *4*) the mechanism is not a change in regularity; and *5*) SampEn of neonatal HR records is not very sensitive to missing points. The findings corroborate the earlier work of Pincus and co-workers, who found reduced ApEn in acidotic fetuses (21) and sick newborns (19), and point to a clinical utility for SampEn monitoring of infants at risk of sepsis and sepsislike illness.

CIs for entropy estimates. We derived an expression to estimate the variance of the CP that epochs within a series that match within tolerance r for m points will also match for $m + 1$ points, taking into account the fact that epochs might overlap. This new estimate is based on data and is independent of the mechanism underlying the process.

We note that this variance measure is fundamentally different from the estimate of Pincus (20a) of the SD of ApEn, which was calculated using replicates of

the MIX process. This is another and more empirical approach, which we also performed by determining the CI directly from analysis of multiple surrogate datasets with the same properties as HR. We calculated the 95% CI of multiple realizations of isospectral surrogate data using this bootstrap analysis and found that there was close correlation between the analytic and empirical results. The CIs from the analytic method for $r = 0.2$, however, were about twice those of the surrogate data. The discrepancy may be due to the failure of the surrogate data, which are matched for frequency content, to capture all the dynamic features of the real data. We used the more conservative analytic method to select values of m and r.

Pincus and co-workers (20, 20a) noted that analytic definition of the SE of ApEn is much more complicated. Because the ApEn calculation involves the sum of the logarithms rather than the logarithm of the sums, our methods are not applicable. Recently, it has been noted that the asymptotic distribution of ApEn is related to the chi-square distribution (23). This result, however, only applies to large sets of uniformly distributed discrete data and is not applicable to most experimental data. We know of no general methods to determine a CI for ApEn.

Parameters for entropy estimation can be optimized. Optimal selection of m and r has been an unexplored area of paramount importance. Pincus (17), for example, showed very large uncertainty in the ApEn of a 1,000-point MIX(0.4) series with the popular choices of $m = 2$ and $r = 0.05 \times$ the range of the data, or $r \approx 0.3 \times$ SD. While some of the error is the bias of ApEn (20), the rest may be an unwittingly suboptimal choice of m and r.

We developed a systematic general approach to picking m and r based on a new metric of the efficiency of the entropy estimate. The two steps are to *1*) pick a range of values of m based on an understanding of the physical process or by fitting AR models, and *2*) calculate SampEn for a range of r and select the value that optimizes an efficiency metric such as the maximum of σ_{CP}/CP and $\sigma_{CP}/[-\log(CP) \times CP]$.

Our analysis led us to select $m = 3$ and $r = 0.2$ for neonatal HR data records. The resulting CIs were acceptably low (Fig. 2). We note, however, that these values are not universally applicable to all datasets. In fact, datasets of length 4,096 are quite long compared with many for which ApEn analysis has been used. The confidence with which ApEn results can be viewed for short, clinically observed datasets for any values of m and r is not known. We suggest that calculation of CI is an essential part of entropy estimation.

SampEn falls early in the course of neonatal sepsis and sepsislike illness. We observed reduced variability and transient decelerations in HR data before neonatal sepsis, and we hypothesized that SampEn would fall before the clinical diagnosis. This was the case: multivariable logistic regression modeling showed that SampEn added independent information to birthweight, gestational age, and days of age in predicting sepsis by up to 24 h. We note, however, that low SampEn in these records does not distinguish between HR decelerations (which are of interest) and HR accelerations (which are not). To enhance the diagnostic usefulness of SampEn in this setting, we tested the effect of adding measurements of the third moment, or skewness, to the multivariable model. By its sign, skewness cleanly separates records of low SampEn with decelerations, where long intervals lead to skewness >0, from those with accelerations. A model incorporating the skewness had improved discriminating ability (ROC area 0.77, CI 0.69–0.85, $P < 0.001$). Thus skewness adds information to SampEn, justifying a multivariate approach to the clinical goal of early detection of sepsis.

Entropy falls in the presence of spikes in a record with reduced variability, but the mechanism is not a change in regularity. Our findings suggest that a reduction in SampEn in a time series might be due to two very different mechanisms: an increase in the degree of regularity, or outlying data points that inflate the SD. We demonstrate analytically and experimentally that SampEn falls in the presence of spikes, and we conclude that reduced SampEn in neonatal HR data could be due, at least in part, to the spikes rather than increased regularity. In principle, neither measure should be sensitive to outlying points: the result is due only to establishing r as a function of the SD of the data. This conclusion contradicts that of Pincus et al. (20a), who stated that ApEn is insensitive to outliers, and differs from those of previous studies that attribute reduced entropy solely to changes in regularity; some investigators go so far as to draw conclusions about nonlinear dynamics in heart beat data (3, 5, 7, 9–12, 15, 21).

Contrary to prediction, very small subsets of the data can indeed dramatically affect entropy estimates (20). In general, we suggest that a reduction in entropy estimates indicates increased regularity, the presence of spikes, or both. Finding methods that distinguish these components is a very important future task.

SampEn analysis of neonatal HRV is not sensitive to missing points. Frequency domain analysis of HRV dissects sympathetic and parasympathetic activity (1, 14, 27) but is very sensitive to missing data points (2, 24). We found SampEn little affected by loss of more than one-third of the data, the practical limit that we might encounter. This is surprising, since loss of small amounts of data significantly impaired the detection of regularity in truly deterministic data. The finding, however, is consistent with other results presented here that SampEn of HR records reports on spikes as well as regularity. In this context, loss of data points is irrelevant to the calculation.

Physiological and clinical relevance of SampEn analysis of neonatal HRV. The physiological mechanism underlying reduced variability and transient decelerations is not known but seems likely to represent dysfunction of the autonomic nervous system or of intracellular signal transduction processes perhaps by circulating cytokines (8, 13). Since the pathophysiology may not be specific for sepsis, SampEn may find use as

a general estimate of the health of the infant. Regardless of the mechanism, the findings are important in clinical medicine, as SampEn can be considered a candidate measure for monitoring at-risk infants, either alone or as part of a multivariable scheme. Given the frequency and severity of sepsis and sepsislike illness in premature newborn infants (26), any scheme that improves on the current clinical practice would be welcome. Moreover, it seems reasonable to conjecture that cumulative measures of SampEn might reflect the burden of illness in infants and be useful both in estimating prognosis and in resource utilization in the newborn intensive care unit.

Summary. We propose new and general methods for optimal selection of m and r for SampEn and ApEn. Our most important finding, however, applies to all studies using either measure. SampEn and ApEn are modulated by outlying points that are irrelevant to the intuitive idea of the regularity of the process. Thus increased regularity is not always the mechanism of lower entropy. This observation allows a simpler interpretation of why reduced variability and transient decelerations in fetal and neonatal HR records lead to lower ApEn and SampEn. Instead of regularity that might or might not be visually apparent (17, 21), we suggest that these time series features inevitably lead to lower entropy estimates for reasons unrelated to regularity.

APPENDIX A

AR Models

AR models of various orders were fit to the data. The motivation for this approach was that if data come from an AR(p) process then $m \geq p$. For a series $[u(i)\colon 1 \leq i \leq N]$ of length N, the parameters of an AR(p) process $a_1, a_2, \ldots, a_p$ were estimated to minimize the least-squares fit to the data

$$SS(a_1, a_2, \ldots, a_p) = \sum_{i=p+1}^{N} [u(i) - a_1 u(i-1) - a_2 u(i-2) - a_p u(i-p)]^2$$

which is accurately determined by solving the first p Yule-Walker equations with correlations estimated using the sample autocorrelation coefficients (4). The order of the process was then estimated to minimize the Schwarz's Bayesian criterion (SBC) (Ref. 25)

$$\mathrm{SBC}(p) = \log [SS(a_1, a_2, \ldots, a_p)/N] + \frac{p}{N} \log (N)$$

which represents a tradeoff between the fit of the AR model and the number of parameters estimated.

APPENDIX B

Estimation of the Variance of SampEn

Let $x_m(i)$ denote the template $[u(i + k)\colon 0 \leq k \leq m - 1]$ of length m from a time series $[u(i)\colon 1 \leq i \leq N]$ of length N assumed to be independent and identically distributed. The number of matches of length $m + 1$ can be expressed as $A = \Sigma U_{ij}$, where $U_{ij} = 1$ if $x_{m+1}(i)$ matches $x_{m+1}(j)$, and 0 otherwise. The summation can be restricted to the B pairs (i,j) of matches of length m, where $x_m(i)$ matches $x_m(j)$. The variance of CP can thus be written

$$\sigma^2_{\mathrm{CP}} = \frac{1}{B} \mathrm{Var}\,(A) = \frac{1}{B^2} \sum_{i,j} \sum_{k,l} \mathrm{Cov}\,(U_{ij}, U_{kl})$$

For the B pairs where $i = k$ and $j = l$ in the above sum, $\mathrm{Cov}(U_{ij}, U_{kl}) = \mathrm{Var}(U_{ij}) = \mathrm{CP}(1 - \mathrm{CP})$. If the templates involved for U_{ij} and U_{kl} have no points in common, they are independent and thus uncorrelated so that $\mathrm{Cov}(U_{ij}, U_{kl}) = 0$. If the templates overlap, the covariance can be estimated by $U_{ij}U_{kl} - \mathrm{CP}^2$, which is $1 - \mathrm{CP}^2$ when both pairs of $m + 1$ templates match and $-\mathrm{CP}^2$ otherwise. So the variance can be estimated as

$$\sigma^2_{\mathrm{CP}} = \frac{\mathrm{CP}(1-\mathrm{CP})}{B} + \frac{1}{B^2} [K_A - K_B(\mathrm{CP})^2]$$

where K_A is the number of pairs of matching templates of length $m + 1$ that overlap and K_B is the number of pairs of matching templates of length m that overlap. Calculating K_A and K_B is not trivial from either the computational or algorithmic point of view, especially for smaller values of m. The calculation involves consideration of all B^2 pairs of m matches. Because B is of order N^2, this is of order N^4 and could be very large for smaller m and larger r. Note that the condition for the template pair (i,j) to overlap the template pair (k,l) is equivalent to $\min(i - k, i - l, j - k, j - l) \leq m$, and care needs to be taken to avoid double-counting overlapping pairs.

APPENDIX C

Analysis of the Effect of Spikes on SampEn

A spike of height Δ and duration εN has variance $\sigma^2_{\mu} = \Delta^2\varepsilon(1 - \varepsilon)$, and the combined variance of the entire record is $\sigma^2 = \sigma^2_{\mathrm{b}} + \sigma^2_{\mu} = \sigma^2_{\mathrm{b}} + \Delta^2\varepsilon(1 - \varepsilon)$. Recall that $r\sigma$ is the tolerance for finding matches. Assume Δ is large enough relative to r and σ that there are few if any matches between points inside the spike and points outside the spike. That is to say, matches occur only within the same mode of the mean process. When the baseline process is Gaussian white noise, the difference D between two potential matching points has a Gaussian distribution with mean 0 and variance $2\sigma^2_{\mathrm{b}}$. Because of independence, the CP of a match is independent of m and equal to

$$\mathrm{CP} = P[|D| < r\sigma] = \int_{-r\sigma}^{r\sigma} \frac{1}{2\sqrt{\pi}\sigma_{\mathrm{b}}} e^{-\frac{x^2}{4\sigma_b^2}}\, dx \approx \frac{r\sigma}{\sqrt{\pi}\sigma_{\mathrm{b}}} = \frac{r}{\sqrt{\pi}} \sqrt{1 + \frac{\Delta^2\varepsilon(1-\varepsilon)}{\sigma_b^2}}$$

where the approximation is valid for small r. We convert CP to SampEn by taking its negative logarithm and make two approximations

$$\mathrm{SampEn}(m,r,N) \approx -\log\left(\frac{r}{\sqrt{\pi}}\right) - \log\left(\sqrt{1 + \frac{\Delta^2\varepsilon(1-\varepsilon)}{\sigma^2_{\mathrm{b}}}}\right) \approx -\log\left(\frac{r}{\sqrt{\pi}}\right) - \frac{\Delta^2\varepsilon(1-\varepsilon)}{2\sigma_b^2}$$

where the second approximation is valid for $\Delta^2\varepsilon(1 - \varepsilon) < \sigma^2_{\mathrm{b}}$.

For a signal that consists of a spectrum of modes, CP depends in a complex way on how the process makes transitions from one mode to another. If matches predominantly

occur only within the same mode, the analysis holds, and SampEn can be approximated for small r by

$$\text{SampEn}(m,r,N) \approx -\log\left(\frac{r}{\sqrt{\pi}}\right) - \log\left(\sqrt{1+\frac{\sigma_\mu^2}{\sigma_b^2}}\right)$$

$$\approx -\log\left(\frac{r}{\sqrt{\pi}}\right) - \frac{\sigma_\mu^2}{2\sigma_b^2}$$

in agreement with our result above where $\sigma^2{}_\mu = \Delta^2\varepsilon(1 - \varepsilon)$.

REFERENCES

1. **Baldzer K, Dykes FD, Jones SA, Brogan M, Carrigan TA, and Giddens DP.** Heart rate variability analysis in full-term infants: spectral indices for study of neonatal cardiorespiratory control. *Pediatr Res* 26: 188–195, 1989.
2. **Berntson GG and Stowell JR.** ECG artifacts and heart period variability: don't miss a beat. *Psychophysiology* 35: 127–132, 1998.
3. **Chaffin DG Jr, Barnard JM, Phernetton T, and Reed KL.** Decreased approximate entropy of heart rate variability in the hypoxic ovine fetus. *J Maternal-Fetal Med* 8: 23–26, 1998.
4. **Chatfield C.** *The Analysis of Time Series*. New York: Chapman and Hall, 1985.
5. **Garde S, Regalado MG, Schechtman VL, and Khoo MC.** Nonlinear dynamics of heart rate variability in cocaine-exposed neonates during sleep. *Am J Physiol Heart Circ Physiol* 280: H2920–H2928, 2001.
6. **Griffin MP and Moorman JR.** Toward the early diagnosis of neonatal sepsis and sepsis-like illness using novel heart rate analysis. *Pediatrics* 107: 97–104, 2001.
7. **Groome LJ, Mooney DM, Holland SB, Smith LA, Atterbury JL, and Loizou PC.** Human fetuses have nonlinear cardiac dynamics. *J Appl Physiol* 87: 530–537, 1999.
8. **Kuster H, Weiss M, Willeitner AE, Detlefsen S, Jeremias I, Zbojan J, Geiger R, Lipowsky G, and Simbruner G.** Interleukin-1 receptor antagonist and interleukin-6 for early diagnosis of neonatal sepsis 2 days before clinical manifestation. *Lancet* 352: 1271–1277, 1998.
9. **Lipsitz LA, Pincus SM, Morin RJ, Tong S, Eberle LP, and Gootman PM.** Preliminary evidence for the evolution in complexity of heart rate dynamics during autonomic maturation in neonatal swine. *J Auton Nerv Syst* 65: 1–9, 1997.
10. **Makikallio TH, Ristimae T, Airaksinen KE, Peng CK, Goldberger AL, and Huikuri HV.** Heart rate dynamics in patients with stable angina pectoris and utility of fractal and complexity measures. *Am J Cardiol* 81: 27–31, 1998.
11. **Makikallio TH, Seppanen T, Niemela M, Airaksinen KE, Tulppo M, and Huikuri HV.** Abnormalities in beat to beat complexity of heart rate dynamics in patients with a previous myocardial infarction. *J Am Coll Cardiol* 28: 1005–1011, 1996.
12. **Newlin DB, Wong CJ, Stapleton JM, and London ED.** Intravenous cocaine decreases cardiac vagal tone, vagal index (derived in Lorenz space), and heart period complexity (approximate entropy) in cocaine abusers. *Neuropsychopharmacology* 23: 560–568, 2000.
13. **Oddis CV and Finkel MS.** Cytokines and nitric oxide synthase inhibitor as mediators of adrenergic refractoriness in cardiac myocytes. *Eur J Pharmacol* 320: 167–174, 1997.
14. **Ori Z, Monir G, Weiss J, Sayhouni X, and Singer DH.** Heart rate variability: frequency domain analysis. *Cardiol Clin* 10: 499–533, 1992.
15. **Palazzolo JA, Estafanous FG, and Murray PA.** Entropy measures of heart rate variation in conscious dogs. *Am J Physiol Heart Circ Physiol* 274: H1099–H1105, 1998.
16. **Pincus SM.** Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA* 88: 2297–2301, 1991.
17. **Pincus SM.** Approximate entropy (ApEn) as a complexity measure. *Chaos* 5: 110–117, 1995.
18. **Pincus SM.** Quantifying complexity and regularity of neurobiological systems. *Methods Neurosci* 28: 336–363, 1995.
19. **Pincus SM, Gladstone IM, and Ehrenkranz RA.** A regularity statistic for medical data analysis. *J Clin Monit* 7: 335–345, 1991.
20. **Pincus SM and Goldberger AL.** Physiological time-series analysis: what does regularity quantify? *Am J Physiol Heart Circ Physiol* 266: H1643–H1656, 1994.

20a. **Pincus SM and Huang W-M.** Approximate entropy: statistical properties and applications. *Commun Statist-Theory Meth* 21: 3061–3077, 1992.

21. **Pincus SM and Viscarello RR.** Approximate entropy: a regularity measure for fetal heart rate analysis. *Obstet Gynecol* 79: 249–255, 1992.
22. **Richman JS and Moorman JR.** Physiological time series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* 278: H2039–H2049, 2000.
23. **Rukhin AL.** Approximate entropy for testing randomness. *J Appl Probability* 37: 88–100, 2000.
24. **Schechtman VL, Kluge KA, and Harper RM.** Time domain system for assessing variations in heart rate. *Med Biol Eng Comput* 26: 367–373, 1988.
25. **Schwarz G.** Estimating the dimension of a model. *Ann Stat* 6: 461–464, 1978.
26. **Stoll BJ, Gordon T, Korones SB, Shankaran S, Tyson JE, Bauer CR, Fanaroff AA, Lemons JA, Donovan EF, Oh W, Stevenson DK, Ehrenkranz RA, Papile LA, Verter J, and Wright LL.** Late-onset sepsis in very low birth weight neonates: a report from the National Institute of Child Health and Human Development Neonatal Research Network. *J Pediatr* 129: 63–71, 1996.
27. **Task Force of the European Society of Cardiology, and the North American Society of Pacing and Electrophysiology.** Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation* 93: 1043–1065, 1996.
28. **Veldhuis JD, Johnson ML, Veldhuis OL, Straume M, and Pincus SM.** Impact of pulsatility on the ensemble orderliness (approximate entropy) of neurohormone secretion. *Am J Physiol Regul Integr Comp Physiol* 281: R1975–R1985, 2001.