

Parameterizing time in electronic health record studies

RECEIVED 15 May 2014

REVISED 8 November 2014

ACCEPTED 22 December 2014

PUBLISHED ONLINE FIRST 27 February 2015

George Hripcsak^{1,2}, David J Albers¹, Adler Perotte¹

ABSTRACT

Background Fields like nonlinear physics offer methods for analyzing time series, but many methods require that the time series be stationary—no change in properties over time.

Objective Medicine is far from stationary, but the challenge may be able to be ameliorated by reparameterizing time because clinicians tend to measure patients more frequently when they are ill and are more likely to vary.

Methods We compared time parameterizations, measuring variability of rate of change and magnitude of change, and looking for homogeneity of bins of temporal separation between pairs of time points. We studied four common laboratory tests drawn from 25 years of electronic health records on 4 million patients.

Results We found that sequence time—that is, simply counting the number of measurements from some start—produced more stationary time series, better explained the variation in values, and had more homogeneous bins than either traditional clock time or a recently proposed intermediate parameterization. Sequence time produced more accurate predictions in a single Gaussian process model experiment.

Conclusions Of the three parameterizations, sequence time appeared to produce the most stationary series, possibly because clinicians adjust their sampling to the acuity of the patient. Parameterizing by sequence time may be applicable to association and clustering experiments on electronic health record data. A limitation of this study is that laboratory data were derived from only one institution. Sequence time appears to be an important potential parameterization.

Keywords: time, time series, parameterization, phenotype, electronic health record, data mining

INTRODUCTION

Most of health care is recorded with respect to the time that health events occur, and researchers who use electronic health record data must decide how to handle time. Often it is sufficient to ignore time because the important question is whether the patient ever had the condition, or to simply check if an event occurred within some relevant time window or use simple aggregation such as mean or maximum.^{1,2} Fewer studies exploit the time of clinical data in detail. Some studies use the clock time—that is, the actual time recorded for an event—to detect phenotypes and correlations^{3–8} or to detect patterns.^{9–11} Other studies use temporal abstractions that either are derived from clock time^{12–16} or intervals of clock time^{17–19} or are extracted from narrative text.^{20–23} More complex temporal processing may exploit algorithms like dynamic and irregular-time Bayesian networks, but they also usually rely ultimately on clock time.^{24,25}

When analyzing clinical data as a series of time points, one common assumption is that the distribution of the data does not change over time, and this property is called stationarity.²⁶ In theory, a system is stationary if the parameters that determine the dynamics of the system remain constant. In practice, the parameters may not be known, and a more operational definition is that over a measured time range, statistical properties of the system remain constant. This definition includes mean, variance, and differences between values. (See the first

four chapters of Kantz and Schreiber²⁶ for a particularly concise and comprehensible description of nonlinear time series analysis.) The advantage of stationarity can be seen in the following example of clustering patient types. If some patients' longitudinal records have a mix of episodes in which they are healthy with episodes in which they are severely ill, then they may get clustered not with healthy patients nor with severely ill patients, but with patients who are mildly ill—that is, their average state in some sense—even though they never actually participate in that mild state.

Clearly, health care can be nonstationary. The existence of health care is based on patients becoming ill and the health care system attempting to move them into a healthier state. Generally in research, investigators deal with this by picking patients who are similar and by picking portions of their histories that are relatively stable. For high-throughput analysis (e.g., cohorts of millions of patients), however, it may be difficult to distinguish time periods of stability.

Fortunately, some analyses are relatively robust to nonstationarity. In our previous correlation study,⁴ for example, despite lumping together over 20 years of data on millions of patients, we were able to uncover logical temporal associations that described definitional, physiological, and intention relationships. Seemingly nonstationary processes may be stationary if viewed over a long enough time period and represented with a sufficiently complex model.²⁶

Correspondence to: Dr. George Hripcsak, Vivian Beaumont Allen Professor and Chair Department of Biomedical Informatics Columbia University Medical Center 622 W 168th St, PH20, New York, NY 10032; E-mail: hripcsak@columbia.edu

©The Author 2015. Published by Oxford University Press on behalf of the American Medical Informatics Association.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work properly cited. For commercial re-use, please contact journals.permissions@oup.com

For numbered affiliations see end of article.

In addition to nonstationarity, health care has a second limitation not present in some other areas of research: sampling is highly biased. In traditional physiologic studies, experimenters use constant and relatively accurate sampling of the physiologic parameters of interest, and noise is expected to be uncorrelated with the processes and not overly large compared to the signal.²⁶ In health care, however, patients are not measured evenly, but are measured according to a clinician's judgment, usually more frequently when the patient is sicker, so the sampling rate varies with the patient state. The sampling is multiscale, with time between samples ranging from seconds to years. Furthermore, health care processes such as inpatient admission also affect the recording of events, producing a complex feed-forward loop.^{27–29}

Handling these limitations requires new methods, and an exciting new direction in nonlinear physics is to learn how to apply traditional methods to imperfect measurements. For example, in some domains, there are many short sequences that are individually insufficiently long to represent the dynamics of the system, so there are methods to aggregate those sequences, both outside of³⁰ and inside of health care.³¹ The latter study specifically addresses the irregularity of the measurements, testing whether the sequences are sufficiently homogeneous to warrant analysis.

This article addresses one piece of the puzzle. It may be possible to reduce the effects both of nonstationarity and of irregular sampling by reparameterizing time. Insofar as variables usually change more quickly when the patient is ill, and the clinician samples more frequently when the patient is ill, perhaps by considering alternatives to clock time, we can make the data more homogeneous.

Background

Sequence time is an integer index of the number of measurements starting with some first measurement as 1. Our earlier work³² showed that sequence time may be better correlated with variables than clock time. We investigated the stationarity and sampling of clinical data by calculating the mutual information between pairs of glucose measurements within patients.³² Mutual information can be used to quantify the predictability of glucose, and we studied how the predictability varied with clock time and with sequence time. Figure 1 is adapted from that work. It shows that, as expected, the predictability of glucose measurements decreases with increasing clock time between measurements. It also shows that predictability drops even more dramatically with increasing number of measurements between measurements (sequence time). This finding led us to posit that there may be better parameterizations of time than clock time. Further evidence comes from the observation that while one might expect a strong correlation between test values and time between tests, in fact, minimal correlation has been found.³³

Figure 2 shows a possible explanation. We called differences in clock time Δt and differences in sequence time τ . In general, patients' data become more variable when the patient is ill. There are many specific physiological exceptions, such as heart rate and fetal monitoring, but on the whole, measured parameters move out of range during illness (e.g., glucose out of control, drop in hemoglobin during bleed, rise in white cell count, rise in temperature, and drop in blood pressure). Clinicians tend to sample patients more frequently when they are ill so that while the patient's data may be changing more rapidly than usual, the increased sampling may make up for the increase. The overall effect is illustrated in the bottom of figure 2: variability with respect to the sampling rate (sequence time, τ) does not shift as much as variability with respect to clock time. Thus medicine is a self-regulating feedback loop that drives the timing of measurements.²⁸

We subsequently studied how to aggregate time sequences whether time was parameterized by clock time or sequence time.³¹ In another study, we used clock time but constrained the measurements to specific ranges of sequence time.³⁴ This can be important if one is applying a mechanistic model, which is usually driven by clock time (e.g., reaction rates).^{35,36} That study³⁴ demonstrated the strong reliance of the univariate predictability of a variable on the number of measurements between two values. Another approach is to warp time through a transformation such as a power function⁵; this is effectively an intermediate parameterization between clock time and sequence time.

The goal of this article is to study different parameterizations of time. Specifically, we seek a parameterization that maximizes stationarity and that makes measurements that are separated by similar times most homogeneous across a patient's record.

Our general approach is to study time in health records with the intent that our findings could then be used in association and prediction algorithms. We recognize that alternative approaches are possible. One approach is to directly incorporate the irregularity of time into the algorithms. Examples include subset time series of increasing lengths,⁶ coarsened granularity with linear interpolation,⁷ or a hierarchical model that uses a linear dynamical system with a window-based temporal model nested inside of it to handle the irregularity of time.⁸ All three examples are set in the intensive care unit, where data are measured irregularly but fairly frequently. In the general electronic health record, measurements occur on multiple scales: minutes or hours for intensive care units and years for outpatient care. No single window size can adequately cover the breadth. A reparameterization, on the other hand, may be able to accommodate multiple scales. Several types of Bayesian networks have been used in the context of irregular time series,^{24,25} but it remains unclear how nonstationarity of the time series affects the results. Furthermore, using time reparameterization is not limited to a single algorithm or set of algorithms and may potentially be applied in many situations.

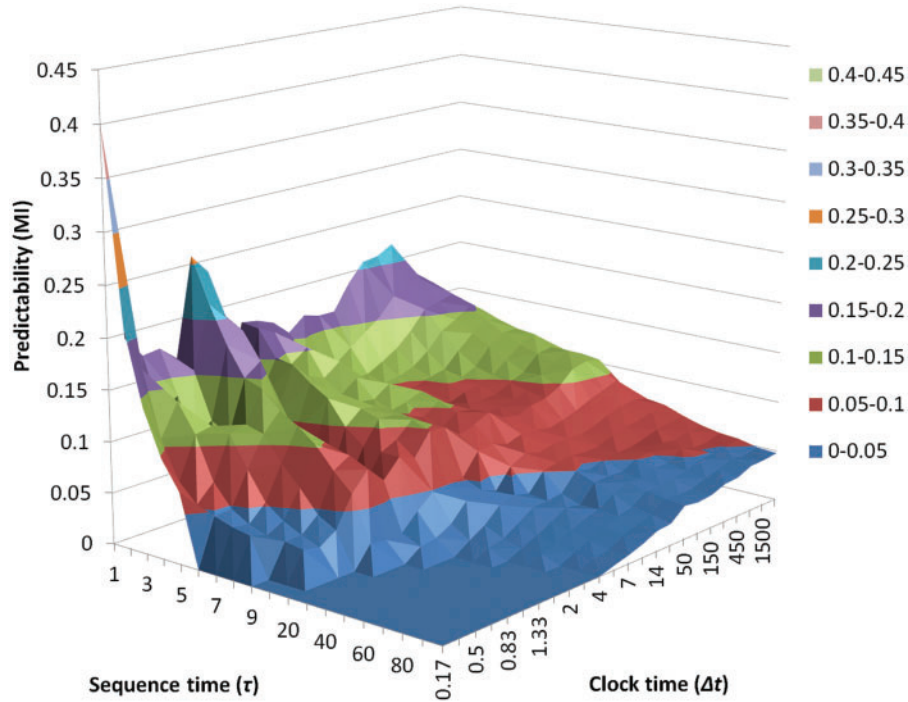
A different, more explicit approach is to model the care process as a set of temporal contexts so that temporal concept abstraction can be tailored to the context.^{14–16} In other words, rather than pick one time parameterization that hopes to accommodate all time scales, divide the health record into contexts (e.g., intensive care unit stay, outpatient care, or presence of a particular diagnosis or medication) and allow the algorithm (in this case, temporal abstraction) to treat each context differently. In theory, this would offer the most flexibility, but it is not applicable to all types of numeric algorithms that one might seek to employ. Nevertheless, the approaches may complement each other. For example, one could reparameterize time to regularize the time series as much as possible and then perform temporal abstraction.

METHODS

We desire a time parameterization that maximizes stationarity of the time series. There are effectively two primary sources of temporal information in the series: the clock time of measurements and the number of measurements between the members of any pair, which we refer to as sequence time. We expect measurement values to change over time, and we expect that values can drift further given greater amounts of time. If a time series is stationary, however, we expect that values will change about the same amount on average given equal separations in time.

We can derive an intermediate warped time by using an exponent to transform the difference in clock time between successive measurements. This is illustrated in figure 3 using 1/3 as the exponent (labeled $\Delta t^{1/3}$). Lasko *et al*⁵ arrived at this warping parameter (1/3)

Figure 1: Predictability of glucose plotted versus sequence time and clock time. Predictability is quantified as mutual information between pairs of points separated by the given clock time and sequence time. At the shortest clock and sequence time (far left), predictability is the highest. It drops off with increasing clock time, but it drops off more precipitously with increasing sequence time, raising the possibility that sequence time is an important potential parameterization for time. (Adapted from Albers and Hripcsak³²).



heuristically using domain knowledge and a human-guided search.⁵ Because clock time can be seen as warped time with an exponent of 1 and sequence time can be seen as warped time with an exponent of 0, warped time with an exponent of 1/3 can be seen as an intermediate between the two.

Experiment one

In our first experiment, we assessed the effect of time parameterization on stationarity. We focused on two components of stationarity that are likely to be affected by time parameterization: the rate of change of a variable, which is defined in terms of time, and the magnitude of change in a variable. While the latter is not defined in terms of time, it may be affected by how measurements are grouped.

Assume M patients with N_m measurements each, recorded in time. We will notate this time series by $x_i(t_{i,j})$, where i identifies the index of the patient, j identifies the index of the measurement for patient i , t is the real time when the measurement was taken, and x_t is the measurement function; that is, $x_i(t_{i,j})$ is the j 'th measured value for patient i . Therefore, in the clock time parameterization, a patient time series can be represented as pairs of times and measurements as follows: $((t_{i,1}, x_i(t_{i,1})), (t_{i,2}, x_i(t_{i,2})), \dots, (t_{i,N_i}, x_i(t_{i,N_i})))$, and in the sequence time parameterization, patient time series can be similarly represented as pairs of indices and measurements as follows: $((1, x_i(t_{i,1})), (2, x_i(t_{i,2})), \dots, (N_i, x_i(t_{i,N_i})))$. Warped time raises the difference in clock time between sequential measurements to a power, with time between more distant measurements simply being

equal to the sum of the sequential jumps. For a warp factor of 1/3, warped time is given as follows:

$$\left((t_{i,1}^{\frac{1}{3}}, x_i(t_{i,1})), \left(\sum_{j=1}^2 (t_{i,j} - t_{i,j-1})^{\frac{1}{3}}, x_i(t_{i,2}) \right), \dots, \left(\sum_{j=1}^{N_i} (t_{i,j} - t_{i,j-1})^{\frac{1}{3}}, x_i(t_{i,N_i}) \right) \right),$$

where $t_{i,0}$ is defined as 0.

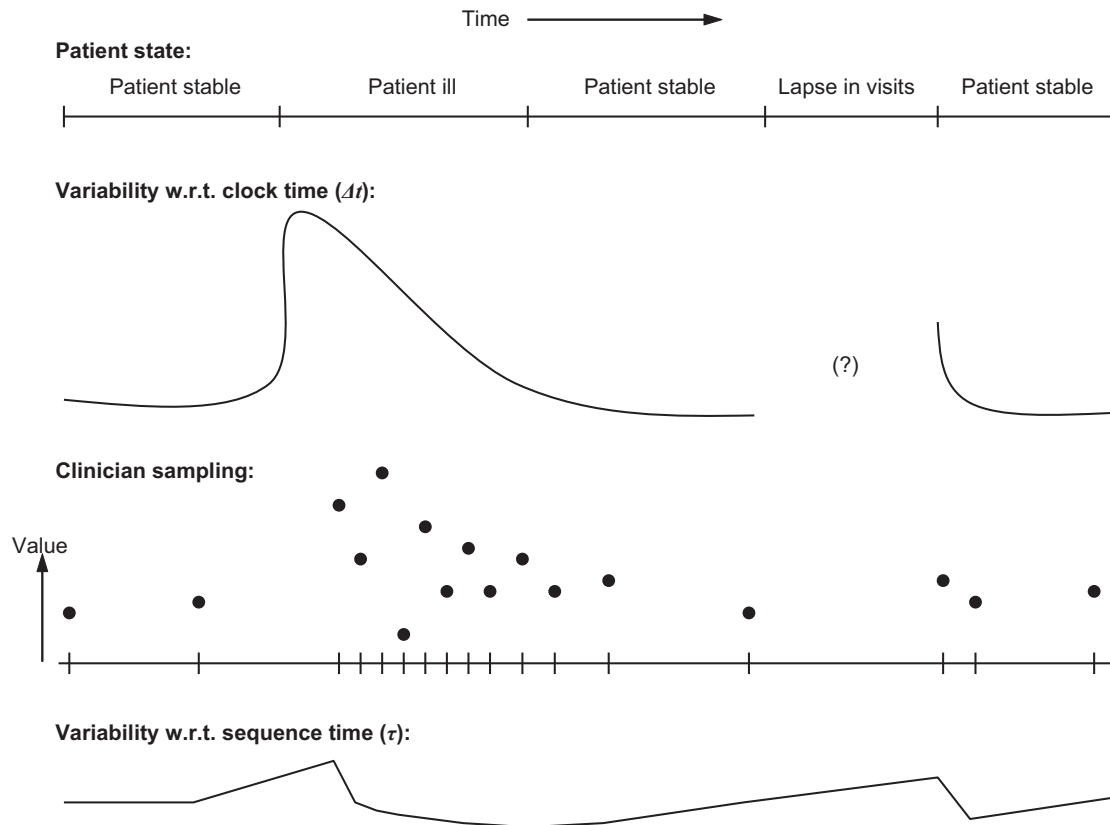
To assess stationarity, we quantified the degree to which rate of change and magnitude of change of a variable varied over time for patients. Both quantities should remain stable over time for each patient for a stationary process. We created time windows of size w and broke each patient's time series into nonoverlapping, adjacent temporal bins. The units of w depend upon the time parameterization. We use k to index the B_j bins per patient. We estimated the variability across bins of the patient's median absolute value of rate of change as follows: we use median rather than mean because the distributions may be highly skewed. For clock time, sequence time, and warped time, the median rate of change for bin k in patient i is given as $r_{i,k}$, $r'_{i,k}$, and $r''_{i,k}$, respectively:

$$r_{i,k} = \text{median}_{j,j-1 \in \text{bin}_k} \left(\left| \frac{x_i(t_{i,j}) - x_i(t_{i,j-1})}{t_{i,j} - t_{i,j-1}} \right| \right),$$

$$r'_{i,k} = \text{median}_{j,j-1 \in \text{bin}_k} \left(\left| \frac{x_i(t_{i,j}) - x_i(t_{i,j-1})}{1} \right| \right),$$

$$r''_{i,k} = \text{median}_{j,j-1 \in \text{bin}_k} \left(\left| \frac{x_i(t_{i,j}) - x_i(t_{i,j-1})}{(t_{i,j} - t_{i,j-1})^{\frac{1}{3}}} \right| \right).$$

Figure 2: The relation between patient illness, sampling, and variability. The patient is stable, then becomes ill and recovers slowly; the patient is then lost to follow-up, and returns after unknown health. If the patient were observed very frequently and regularly, one could calculate a detailed “variability with respect to clock time (Δt)” (e.g., glucose in the setting of diabetes with an episode of poor control). In this example (but not necessarily in all cases), the variability increases with illness. The clinician orders tests on the patient on the basis of the patient’s known level of illness (“Clinician sampling”), missing the beginning of the illness, and over-sampling as the illness resolves until the clinician is comfortable with the patient’s status. After being lost to follow-up, the patient is somewhat off from baseline (either status post another illness or due to drift over the missing time) and returns to stability. The “variability with respect to sequence time (τ)” accounts for both the variability with respect to time and the sampling rate, and it illustrates the degree to which the clinician successfully compensates for the change in variability.



The coefficient of variation (cv) quantifies the standard deviation of the median rate of change, normalized by the mean of the median rate of change. A stationary time series should minimize cv, although it will remain nonzero owing to variation produced by chance. We estimated cv as follows for clock time:

$$cv = \frac{1}{M} \sum_{i=1}^M \sqrt{\frac{\frac{1}{B_i-1} \sum_{k=1}^{B_i} (r_{i,k} - \bar{r}_{i,k})^2}{\frac{1}{B_i} \sum_{k=1}^{B_i} r_{i,k}}},$$

with cv for sequence and warped time given analogously, replacing $r_{i,k}$ with $r'_{i,k}$ and $r''_{i,k}$.

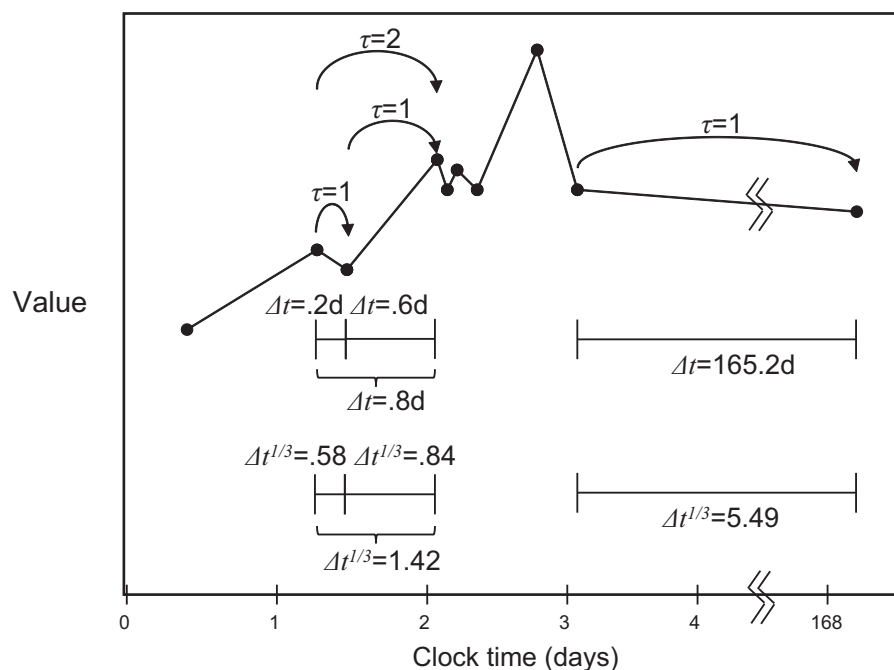
We compared the cv for the three time parameterizations for four common laboratory variables: blood glucose, creatinine, potassium, and sodium. We used data for all patients who had the laboratory test performed from 1989 to 2011.

To do the comparison, we had to pick a time window size w for each parameterization. The choice is a balance between having

enough points within each bin to get a stable estimate of the median and having enough bins to detect a change over time. We tested 7–180 days for clock time, picking 30 days as the primary target. For sequence time we tested 5–20 units and for warped time we tested 7–180 days raised to the one-third power. For our primary comparison, we picked the window sizes for sequence and warped time that best matched the average proportion of patients with at least two bins and the average number of measurements per bin for clock time at 30 days. To make sure that the choice of window size was not responsible for the result, we compared all combinations of window sizes (684 comparisons).

We used bootstrap estimates of variance³⁷ to calculate 95% confidence intervals for the coefficients and p values for the differences between coefficients. p Values were Bonferroni corrected for 684 comparisons. We repeated the experiment for root-mean-square error in place of the median absolute value of rate of change to ensure that the choice of metric was not responsible for the result.

Figure 3: Time parameterizations. Ten measurements for one variable for one patient are shown. Measurements are plotted according to their clock time (labeled Δt). The clock time between measurements is shown for several points: the second through fourth (0.2 d and 0.6 d for a total of 0.8 d) and the last two (165.2 d). Also shown is sequence time (τ) of 1 for sequential measurements and 2 from the second to the fourth point, and warped time ($\Delta t^{1/3}$) of 0.58 and 0.84 for a total of 1.42, and 5.49 for the last two. Note that the total warped time from the second to the fourth measurement is not the cube root of the total clock time, 0.8, but rather the sum of 0.58 and 0.84; that is, the cube root is taken only between sequential measurements.



We then repeated the above experiment using magnitude of change $c_{i,k}$ defined for all three time parameterizations as follows:

$$c_{i,k} = \text{median}_{j-j-1 \in \text{bin}_k} (|x_i(t_{i,j}) - x_i(t_{i,j-1})|),$$

The cv for $c_{i,k}$ is defined analogously to that for $r_{i,k}$, replacing $r_{i,k}$ with $c_{i,k}$. It was expected that magnitude of change would be less affected by time parameterization than rate of change because time is not directly quantified in the definition (instead, time serves merely as an index). Whereas for rate of change, the scale of the metric varies widely between time parameterizations, for magnitude of change, the scale should stay approximately the same. Even though the cv should correct the scale, magnitude of change is less affected by bias if the cv fails to correct the scale perfectly (e.g., due to chance variation, the mean value in the denominator may not be a perfect correction for differences in scale). Furthermore, at the very least, it ensures that any improvement in stability of rate of change is not at the expense of magnitude of change.

Experiment two

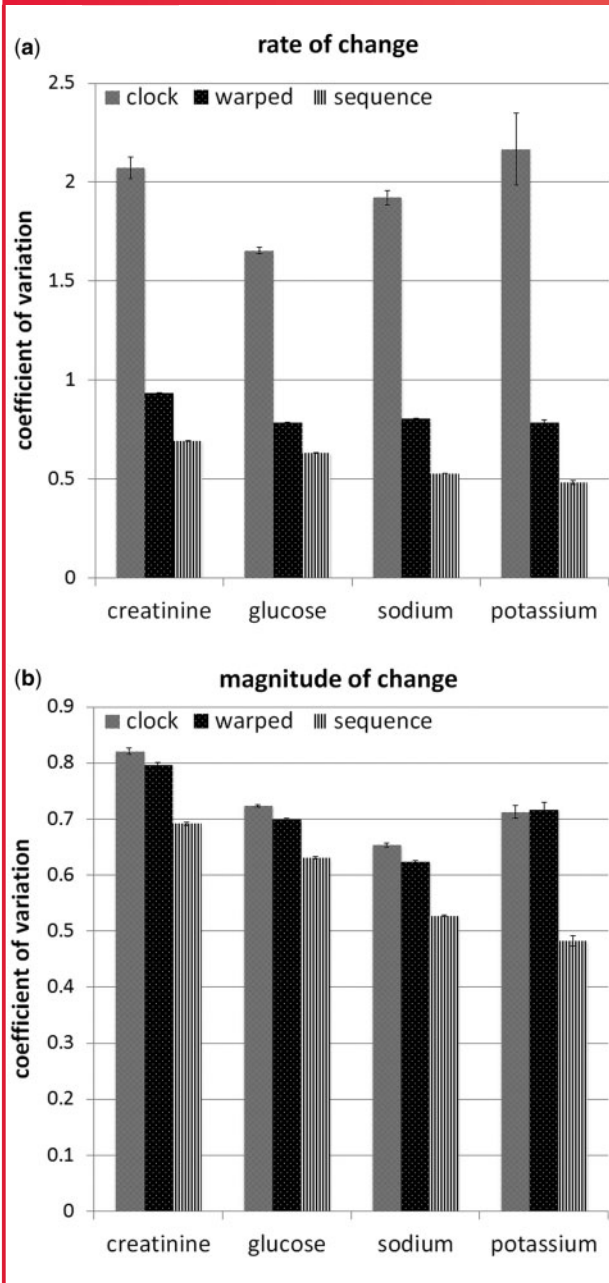
To give a reader a more concrete view of the effect of nonstationarity and the potential improvement afforded by sequence time, our second experiment compared the homogeneity of changes in values stratified by the clock time between them. For pairs of measurements j and j' within patient i , we plotted the typical change in value ($x_i(t_{i,j}) - x_i(t_{i,j'})$) for a series of clock time differences ($t_{i,j} - t_{i,j'}$) on

a logarithmic scale, ranging from about an hour to about a year. We used the same four variables: blood glucose, creatinine, potassium, and sodium. We carried out the experiment for all pairs of sequential measurements (i.e., difference in sequence time, $j - j'$, equal 1, labeled $\tau = 1$) and for all possible pairs of measurements (regardless of number of measurements in between).

Experiment three

In our third experiment, we studied the relationship between different time parameterizations and the degree to which measurements separated by a given difference in time tended to differ by the same amount. We created 10 bins of time separation for each parameterization. Specifically, we stratified all pairs of measurements into 10 deciles of logarithm-scaled time in three dimensions: clock time, sequence time, and warped time using an exponent of 1/3. We sampled about 10 million pairs randomly from all possible pairs. We used bins as predictors and median absolute value of magnitude of change as the outcome, and we created a series of linear regression models to assess dependencies among the time parameterizations. We considered a parameterization to be better if pairs within its bins are more homogeneous. In the regression results, this shows up as follows: an ideal parameterization should have relatively large positive coefficients that are statistically significant, and the other predictors should be smaller and less significant. We repeated the experiment for the four variables, and we repeated the experiment at other warp values: 1/2, 1/4, 1/7, and 1/100.

Figure 4: Stationarity with different time parameterizations. The effect of time parameterization on (a) rate of change and (b) magnitude of change is shown for four laboratory tests. Coefficient of variation (standard deviation divided by mean) of the median absolute value of rate of change and of magnitude of change is plotted for clock time, warped time, and sequence time. Sequence time shows the least variability in all cases.



Experiment four

In the fourth experiment, we illustrated the use of time reparameterization for prediction. We performed probabilistic predictions of future values given previous values in time series of glucose values. To do this, we leveraged the Bayesian formulation of a stochastic process

known as Gaussian processes.^{38,39} This probabilistic model has been used to model patient data in previous work and is a natural one for noisy irregularly sampled time series.^{5,40,41}

A total of 400 patients with at least 11 glucose values were randomly selected from the clinical data warehouse. Half of these patients were randomly assigned to the development cohort and half were randomly assigned to the validation cohort.

As in the first experiment, assume M individuals with N_m measurements each, and measurement function $x_i(t_{ij})$. A Gaussian process is fully specified given a mean function, $m(z)$, and a covariance function, $k(z, z')$.

$$f(z) \sim GP(m(z), k(z, z'))$$

where in this case we have chosen the mean function to be a constant, $m(z) = \alpha$ and the kernel to be $k(z, z') = \theta_0 e^{\{-\frac{\theta_1}{2}|z-z'|^2\}} + \theta_2 + \theta_3 z^T z'$.

In our case, either the measurement times t_{ij} or the measurement indices j represent z , and the measurement values are represented by the Gaussian process $f(z)$ with added noise: $x_i(t_{ij}) \sim \mathcal{N}(f(t_{ij}), \beta)$ in the case of clock time parameterization and $x_i(t_{ij}) \sim \mathcal{N}(f(j), \beta)$ in the case of sequence time parameterization. Performing a prospective prediction given previous observations requires evaluating the conditional distribution of the measurement at either the next measurement time or the next measurement index. Detailed derivations can be found in Rasmussen.³⁸

In the case of clock time, the base time unit was defined to be 71 days (the expected time between glucose measurements) because this placed the clock time and sequence time parameterizations on equal footing with respect to average spacing. Maximum a posteriori values for the set of parameters $\{\alpha, \beta, \theta_0, \theta_1, \theta_2, \theta_3\}$ were learned based on the development set of patients using the L-BFGS algorithm.

To quantify the difference in performance between these models, the macro-averaged difference in predictive log-likelihood between clock time and sequence time parameterizations was evaluated. From the 11th point forward, we evaluated $p(t_i | t_{1:i-1})$. The bootstrap was used to evaluate the standard error of the differences.

RESULTS

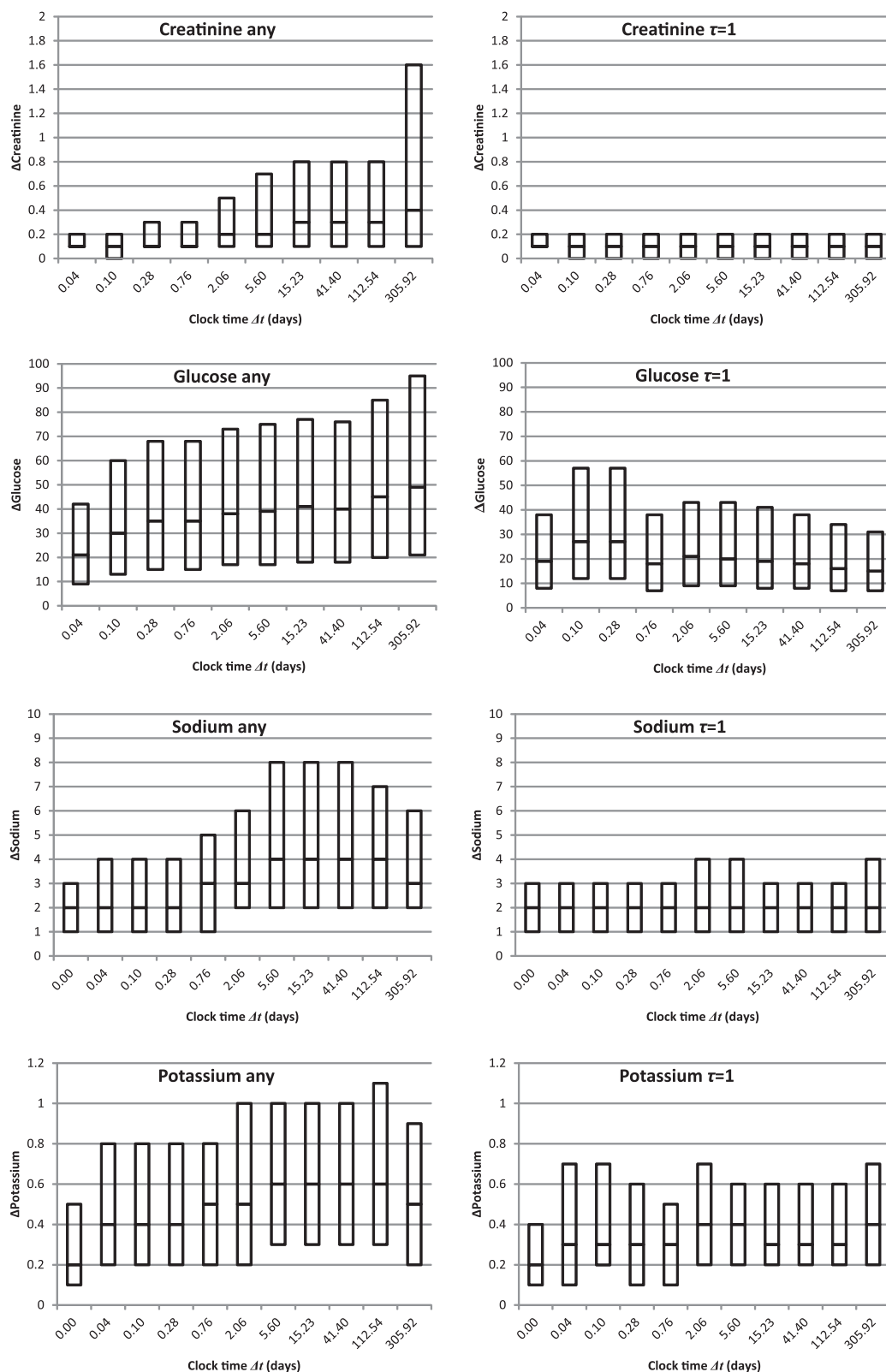
Experiment one

Figure 4a shows the results of comparing rate of change for the three parameterizations for the four variables at a window size of 30 days for clock time, 5–10 units for sequence time (depending on the variable), and 5.6 days¹ for warped time. Sequence and warped time were both significantly less variable (more stationary) than clock time, and sequence time was statistically significantly less variable than warped time. Figure 4b shows the comparison for magnitude of change, and again sequence time was less than warped time, and warped time less than clock time. Of all the comparisons to sequence time (2 metrics, 4 variables, 19 total bin sizes), there was only one match where another parameterization beat sequence time: for potassium at a window size of 180 days in clock time and five units in sequence time, and this comparison was highly mismatched for window size. One other comparison was not statistically significant, and in all other comparisons for all laboratory variables, sequence time was statistically significantly more stable than the other parameterizations. The root-mean-square results were similar to median absolute value.

Experiment two

Figure 5 shows the result for all possible pairs (labeled any) and for sequential measurements (labeled $\tau=1$) for the four tests. As

Figure 5: Variability of laboratory tests versus clock time. The median and quartiles of the average difference between measurements that are separated by a given clock time, shown in logarithmic deciles in days. “Any” implies all pairs were included. “ $\tau = 1$ ” implies that only pairs of sequential points were included. The magnitude of the differences increases with increasing clock time, as expected, but when sequence time is held to 1, the association essentially disappears.



expected, for all possible pairs (any), the median change and variability of the change (interquartile range) increased with longer time scales. The effect is most noticeable for creatinine and for sodium, and less so for glucose and for potassium. Creatinine and sodium are slower changing than glucose and potassium, so we may simply not have gone down to a short enough time scale to effectively freeze the glucose and potassium measurements, whereas creatinine and sodium are expected to change little on the shorter times scales (hours). For example, glucose may change within minutes³⁵ yet measurements usually occur over hours or more.

The results for sequential measurements ($\tau = 1$) are striking. There is no increase in variability of changes as time scale increases (if anything, that for glucose drops). Furthermore, for creatinine and sodium, the magnitude of the change for the all possible pairs (any) is bigger than that for sequential measurements ($\tau = 1$) by a factor of three to five. This goes along with the hypothesis presented in figure 2. Over time and on average, clinicians tend to order tests when they are deemed necessary. Healthy patients are naturally sampled less often. When they become ill, laboratory tests relevant to the illness are sampled more frequently, and as they recover, the frequency drops. Thus clinicians appear to naturally but perhaps unconsciously sample to aim for a constant rate of change.

Experiment three

The result of the third experiment is shown in table 1. The table shows regression coefficients and statistical significance for the three time parameterizations to predict variability of values (i.e., median magnitude of change between values separated by a given time separation). Several models are shown for each laboratory test, each model using different combinations of predictors (blank cells in the table imply that the predictor was not included in the model).

Surprisingly, as a sole predictor (models G5, P5, etc.) clock time is a relatively poor predictor of variation. Sequence time and warped time are both predictive of variation on their own for all the laboratory tests (G6, G7, P6, P7, etc.). For multivariable models, once sequence time is included in a model, the other predictors are negative or not significant (G2, G4, P1, P2, P4, S1, S2, S4, C1, C2, C4), with one exception (G1) in which clock time has a smaller positive correlation. Therefore, sequence time, which has the most positive and most significant set of regression coefficients (see the sequence time column in table 1), explains most of the variation in the bins, and therefore should have the most homogeneous bins, especially for creatinine. We repeated the analysis on a subset with other powers such as 1/2, 1/4, 1/7, and 1/100, and the results were similar to those in table 1 (which uses 1/3).

Experiment four

The macro-averaged predictive log-likelihood was 0.071 (95% CI, 0.034–0.117) higher (more predictive) for sequence time parameterized data than clock time. Figure 6a and b demonstrate the posterior distributions for the two parameterizations on two patients. The posterior distribution for sequence time (figure 6a.2 and b.2) follows the trends in the observations well with most lying within the posterior one standard deviation area, whereas clock time (a.1 and b.1) do not track well, possibly due to the varying granularity of the measurements.

DISCUSSION

Of the three parameterizations, we found that sequence time made the time series statistically significantly more stable in terms of rate of change and magnitude of change over time and appeared to create

Table 1: Dependence of variability of values on transformed time

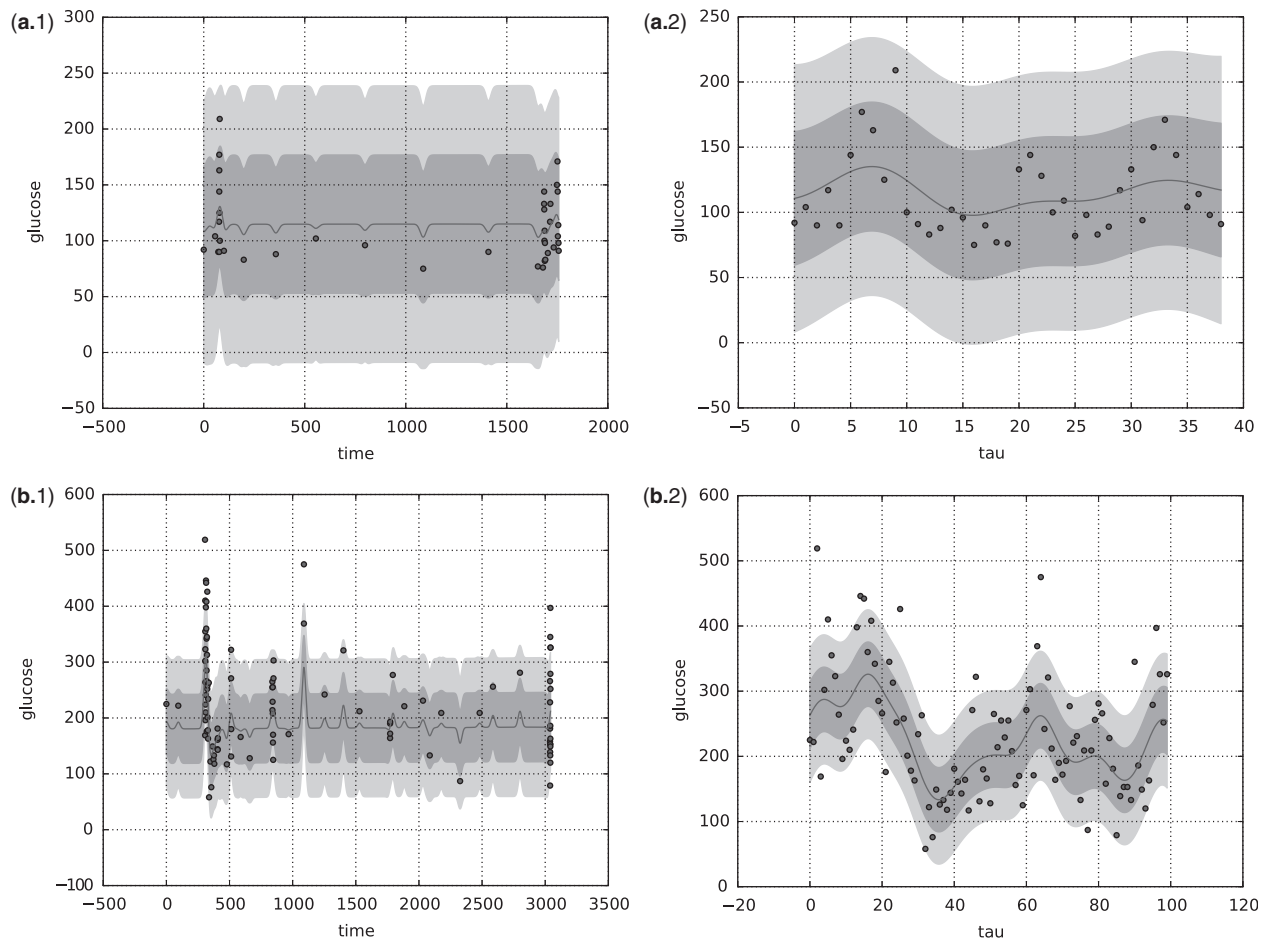
Model ^a	Predictor		
	Clock time	Sequence time	$\Delta t^{1/3}$ warped time
Glucose			
G1	2.8***	6.2***	−5.1***
G2	−0.6**	2.8***	
G3	−1.7***		3.2***
G4		4.6***	−2.8***
G5	0.6NS		
G6		2.2***	
G7			1.5***
Potassium			
P1	−0.008 NS	0.045***	−0.040***
P2	−0.009**	0.028***	
P3	−0.015***		0.026***
P4		0.040***	−0.019***
P5	0.005**		
P6		0.028***	
P7			0.021**
Sodium			
S1	−0.01 NS	0.50***	−0.28***
S2	−0.27***	0.42***	
S3	−0.36***		0.45***
S4		0.61***	−0.32***
S5	−0.08 NS		
S6		0.39***	
S7			0.22**
Creatinine			
C1	−0.006 NS	0.117***	−0.030 NS
C2	−0.026**	0.086***	
C3	−0.046***		0.072***
C4		0.126***	−0.057***
C5	0.003 NS		
C6		0.065***	
C7			0.042***

^aThe models are the different combinations of predictors (clock, sequence, and warped time): all three, all pairs, and each one.

*Significant at $p < 0.05$; **significant at $p < 0.01$; ***significant at $p < 0.001$; blank cell implies that the predictor was not included in that model.

NS, not significant.

Figure 6: Effect of time parameterization on Gaussian process modeling. Posterior distributions for the time series of two patients, (a) and (b), for clock time parameterization (a.1 and b.1) and sequence time parameterization (a.2 and b.2). The dark gray areas indicate one standard deviation around the posterior mean and light gray areas indicate two standard deviations. The clock time parameterization appears to miss important series features because of the inconsistency of the granularity of the measurements.



the most stationary process for the laboratory variables. Once pairs of measurements were stratified by sequence time, further division by clock time or warped time had less effect, especially for creatinine. This is in contrast to clock time, for example, where stratifying by clock time resulted in bins of pairs that could be further separated by sequence time. Using other warp exponents, such as $1/2$, $1/4$, $1/7$, and $1/100$, produced a result similar to $1/3$; sequence time was still superior. We note that this does not eliminate nonstationarity; it only removes or reduces the temporal component. These results imply that clinicians do appear to correct for changes in variability fairly well, similar to what is illustrated in figure 2, and that parameterizing by sequence time should maximize stationarity.

We illustrated the use of sequence time in a Gaussian process model and demonstrated statistically significantly improved prediction. Graphs of the posterior distributions suggest an explanation: the model that uses sequence time appears to track features of the time series better.

That clock time and warped time had smaller negative coefficients when sequence time is included may have an interesting

interpretation. It implies that given the same sequence time (e.g., all pairs of points separated by five measurements), increasing clock time is associated with *less* variability. This implies that clinicians are not quite correcting for illness enough and that, in fact, a *time inversion* might be useful. Measurements closer together in clock time, on average, are really further apart in terms of their variability. One can imagine a different warped time that uses a small negative exponent, such as $-1/5$, which does in fact invert time. While a negative-exponent warping is possible, the simplicity of sequence time offers greater understanding and, because time is expressed as integers, it may lead to more efficient algorithms.

We recommend that researchers consider parameterizing by sequence time in their mining and clustering experiments that do not, by their nature, require clock time. While they could measure the improvement in stationarity as we did in the first experiment, we have not yet correlated size of improvement in stationarity with improvement in performance. If the experiment requires clock time, for example because the data are linked to a mechanistic model set in clock time or because resulting predictions must be interpreted in terms of

clock time, then we suggest that the researcher parameterize by clock time but consider stratifying by sequence time or incorporating sequence time as an additional input to the model.

Parameterizing by sequence time may have an important added advantage in the context of patient privacy. Because all dates can be removed, retaining just the sequence of events, such a data set may qualify for Safe Harbor⁴² rendering it nonhuman-subjects research.

This work builds on previous work to bring methods from time series analysis, including nonlinear physics, into biomedical informatics and specifically, health record analysis. We loosely refer to this as the “physics of the medical record.”²⁸ Our article in *Chaos*³¹ details an approach to aggregating short sequences of data across patients, with a subsequent article on assessing its bias.⁴³ The approach was designed to accommodate both clock time and sequence time, and we provide a method to quantify heterogeneity in the data set. We applied this and related approaches to creatinine data,³⁴ glucose data,^{35,36} and seizures in the neurological intensive care unit.⁴⁴ Lasko *et al*⁵ also used time series analysis methods in using unsupervised learning to address noisy, sparse, irregular health record data, applying it to uric acid laboratory data to distinguish clinical context. This article presents a complementary way to assess heterogeneity specifically related to time parameterization.

As noted in the Background section, a number of alternative approaches exist. A number of researchers have directly accommodated the irregularity of time into their algorithms.^{6–9,24,25} Whereas this handles the temporal irregularity, it remains unclear what effect the non-stationary of the series has on the results, and each approach is relatively specific to its algorithm. Our current approach to measuring stationarity might be employed to test the effect on some of these alternative methods. The explicit modeling of temporal context^{14–16} more directly handles nonstationarity in the sense that the time series may be stationary within each context. Once again, our approach to measuring stationarity might be employed to verify the explicit modeling. In addition, it might be beneficial to apply the explicit modeling to a reparameterized time series or to intervals^{17–19} derived from reparameterized time (e.g., use 5-measurement intervals instead of 6-month intervals).

Our main limitations are the use of data from a single medical center and the restriction to laboratory tests. We believe that given the breadth of the medical center, which includes both inpatient and outpatient data, there are unlikely to be major differences in overall time structure from other large health systems. The use of laboratory data provides frequently sampled, continuous data. The analysis could not be done in this way on categorical data. Nevertheless, further validation is warranted.

CONCLUSIONS

Medicine is nonstationary and irregularly sampled, and it may benefit from alternate time parameterizations. We developed an approach to compare time parameterizations. We found that of clock time, sequence time, and warped time, sequence time produced the most stability in rate of change and magnitude of change over time, and it best predicted the degree of changes between pairs of measurements for four laboratory variables. This finding implies that clinicians do in fact adjust their sampling rate to accommodate increased variability of the patient state during illness. We recommend that researchers consider parameterizing by sequence time, and if clock time is required, consider accounting for sequence time in other ways.

CONTRIBUTORS

All authors made substantial contributions to the conception and design of the work or the acquisition, analysis, or interpretation of data for the work; drafted the work or revised it critically for important intellectual content; had final approval of the version to be published; and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This work was supported by National Library of Medicine grant number R01 LM006910.

COMPETING INTERESTS

None.

REFERENCES

- Chen Y, Carroll RJ, McPeck Hinz ER, *et al*. Applying active learning to high-throughput phenotyping algorithms for electronic health records data. *J Am Med Inform Assoc*. 2013;20:e253–e259.
- Saria S, Rajani AK, Gould J, *et al*. Integration of early physiological responses predicts later illness severity in preterm infants. *Sci Transl Med*. 2010;2(48):48ra65.
- Warner JL, Zollanvari A, Ding Q, *et al*. Temporal phenome analysis of a large electronic health record cohort enables identification of hospital-acquired complications. *J Am Med Inform Assoc*. 2013;20:e281–e287.
- Hripcsak G, Albers DJ, Perotte A. Exploiting time in electronic health record correlations. *J Am Med Inform Assoc*. 2011;18(Suppl 1):i109–i115.
- Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PLoS ONE* 2013;8:e66341.
- Hauskrecht M, Visweswaran S, Cooper G, *et al*. Data-driven identification of unusual clinical actions in the ICU. Annual American Medical Informatics Association Symposium, November 16–20, 2013, Washington, DC, 2013.
- Liu Z, Hauskrecht M. Sparse linear dynamical system with its application in multivariate clinical time series. NIPS 2013 Workshop on Machine Learning for Clinical Data Analysis and Healthcare, Lake Tahoe, Nevada on December 5–8, 2010, December 2013.
- Liu Z, Hauskrecht M. Clinical time series prediction with a hierarchical dynamical system. In: The 14th Conference on Artificial Intelligence in Medicine; May 2013; Murcia, Spain.
- Wang F, Lee N, Hu J, *et al*. Towards heterogeneous temporal clinical event pattern discovery: a convolutional approach. In: KDD’12; August 12–16, 2012, Beijing, China. 2012:453–461.
- Batal I, Valizadegan H, Cooper GF, *et al*. A pattern mining approach for classifying multivariate temporal data. In: Proceedings IEEE Int Conf Bioinformatics Biomed. Atlanta, Georgia on 11 Nov 12 to 15, 2011, 2011:358–365.
- Noren GN, Hopstadius J, Bate A, *et al*. Temporal pattern discovery in longitudinal electronic patient records. *Data Min Knowl Discov*. 2010;20:361–387.
- Klimov D, Shahar Y, Taieb-Maimon M. Intelligent interactive visual exploration of temporal associations among multiple time-oriented patient records. *Methods Inf Med*. 2009;48(3):254–262.
- Stacey M, McGregor C. Temporal abstraction in intelligent clinical data analysis: a survey. *Artif Intell Med*. 2007;39:1–24.
- Shahar Y. A framework for knowledge-based temporal abstraction. *Artif Intell*. 1997;90(1–2):79–133.

15. Shahar Y. Dynamic temporal interpretation contexts for temporal abstraction. *Ann Math Artif Intell*. 1998;22(1–2):159–192.
16. Shahar Y. Knowledge-based temporal interpolation. *J Exp Theor Artif Intell*. 1999;11:123–144.
17. Moskovitch R, Shahar Y. Medical temporal-knowledge discovery via temporal abstraction, San Francisco, CA on November 14 to 18, 2009. *AMIA Annu Symp Proc*. 2009:452–456.
18. Moerchen F, Fradkin D. Robust mining of time intervals with semi-interval partial order patterns. In: Proceedings of SIAM Data Mining. Columbus, OH on April 29 to May 1, 2010, 2010.
19. Moskovitch R, Peek N, Shahar Y. Classification of ICU patients via temporal abstraction and temporal patterns mining. Notes of the Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP 2009) Workshop, Verona, Italy; 2009:35–40.
20. Sun W, Rumshisky A, Uzuner O. Temporal reasoning over clinical text: the state of the art. *J Am Med Inform Assoc*. 2013;20:814–819.
21. Sohn S, Savova GK. Mayo clinic smoking status classification system: extensions and improvements. *AMIA Annu Symp Proc*. 2009;2009:619–623.
22. Zhou L, Hripcsak G. Temporal reasoning with medical data—A review with emphasis on medical natural language processing. *J Biomed Inform*. 2007;40:183–202.
23. Hripcsak G, Elhadad N, Chen C, et al. Using empirical semantic correlation to interpret temporal assertions in clinical texts. *J Am Med Inform Assoc*. 2009;16:220–227.
24. Sebastiani P, Mandl KD, Szolovits P, et al. A Bayesian dynamic model for influenza surveillance. *Stat Med*. 2006;25(11):1803–1816.
25. Ramati M, Shahar Y. Irregular-time Bayesian networks. In: Proceedings of the 26th July 8–11, 2010 Conference on Uncertainty in Artificial Intelligence (UAI-2010), Catalina Island, CA, USA, 2010.
26. Kantz H, Schreiber T. Nonlinear time series analysis, 2nd ed. Cambridge: Cambridge University Press; 2004.
27. Hripcsak G, Albers DJ. Correlating electronic health record concepts with healthcare process events. *J Am Med Inform Assoc*. 2013;20:e311–e318.
28. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc*. 2013;20:117–121.
29. Pathak J1, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. *J Am Med Inform Assoc*. 2013;20(e2):e206–e211.
30. Komalapriya C, Thiel M, Ramano MC, et al. Reconstruction of a system's dynamics from short trajectories. *Phys Rev E*. 2008;78:066217.
31. Albers DJ, Hripcsak G. Using time-delayed mutual information to discover and interpret temporal correlation structure in complex populations. *Chaos*. 2012;22:013111.
32. Albers DJ, Hripcsak G. An information-theoretic approach to the phenome (abstract). AMIA Summit on Translational Bioinformatics; 2009 March 15–17; San Francisco, CA. 2009.
33. Pivovarov R, Albers DJ, Sepulveda JL, Elhadad N. Identifying and mitigating biases in EHR laboratory tests. *J Biomed Inform*. 2014;51:24–34.
34. Albers DJ, Hripcsak G. A statistical dynamics approach to the study of human health data: resolving population scale diurnal variation in laboratory data. *Phys Lett A*. 2010;374:1159–1164.
35. Albers DJ, Hripcsak G, Schmidt M. Population physiology: leveraging electronic health record data to understand human endocrine dynamics. *PLOS One* 2012;7(12):e48058.
36. Albers DJ, Elhadad N, Tabak E, Perotte A, Hripcsak G. Dynamical phenotyping: using temporal analysis of clinically collected physiologic data to stratify populations. *PLOS One* 2014;9:e96443.
37. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall; 1993:45–57.
38. Rasmussen E, Williams KI. *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press; 2006:248.
39. Bishop CM. *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006:303–320.
40. Pimentel M, Clifton D, Clifton L, et al. Modelling patient time-series data from electronic health records using Gaussian processes. In: Advances in Neural Information Processing Systems: Workshop on Machine Learning for Clinical Data Analysis; 2013:1–4.
41. Hauskrecht M, Liu Z, Wu L. Modeling clinical time series using Gaussian process sequences. In: SDM. SIAM 2013:623–631.
42. Office for Civil Rights, Department of Health and Human Services. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html> Accessed January 11, 2014.
43. Albers DJ, Hripcsak G. Estimation of time-delayed mutual information and bias for irregularly and sparsely sampled time-series. *Chaos, Solitons, Fract*. 2012;45:853–60.
44. Claassen J, Albers D, Schmidt JM, et al. Nonconvulsive seizures in sub-arachnoid hemorrhage link inflammation and outcome. *Ann Neurol*. 2014;75:771–781.

AUTHOR AFFILIATIONS

¹Department of Biomedical Informatics, Columbia University Medical Center, New York, USA

²Medical Informatics Services, NewYork-Presbyterian Hospital, New York, USA