Taylor & Francis
Taylor & Francis Group

A model for liver homeostasis using modified mean-reverting Ornstein–Uhlenbeck process

D.C. Trost[ab]*, E.A. Overman II[c], J.H. Ostroff [a], W. Xiong[c] and P. March[c]

[a]*Translational and Molecular Medicine, Pfizer Global Research and Development, Groton, CT, USA;* [b]*Division of Nephrology, Department of Internal Medicine, The Ohio State University, Columbus, OH, USA;* [c]*Department of Mathematics, The Ohio State University, Columbus, USA*

Short of a liver biopsy, hepatic disease and drug-induced liver injury are diagnosed and classified from clinical findings, especially laboratory results. It was hypothesized that a healthy hepatic dynamic equilibrium might be modelled by an Ornstein–Uhlenbeck (OU) stochastic process, which might lead to more sensitive and specific diagnostic criteria. Using pooled data from healthy volunteers in pharmaceutical clinical trials, this model was applied using maximum likelihood (ML) methods. It was found that the exponent of the autocorrelation function was proportional to the square root of time rather than time itself, as predicted by the OU model. This finding suggests a stronger autocorrelation than expected and may have important implications regarding the use of laboratory testing in clinical diagnosis, in clinical trial design, and in monitoring drug safety. Besides rejecting the OU hypothesis for liver test homeostasis, this paper presents ML estimates for the multivariate Gaussian distribution for healthy adult males. This work forms the basis for a new approach to mathematical modelling to improve both the sensitivity and specificity of clinical measurements over time.

1. Introduction

From a clinical standpoint, computational biology has had little impact. As others build systems biology from the bottom up, it seems that systemic models viewed from the top down are also necessary, or, at least, significantly contributory. Quantitative clinical research is needed to make important advances in medical science. The methods presented below are an attempt to stimulate such research.

For about a century [5], clinical chemistry methods [7] have been applied to blood samples obtained routinely in the course of medical care. Originally the chemical analytes, chemical components of the blood sample, were assayed one at a time in a very manual, time-consuming manner. In the 1950s, the autoanalyser was invented [9], which would perform the chemical analysis of many samples serially, reducing the labour requirements by an order of magnitude. The second generation of analysers connected these autoanalysers in parallel with each channel testing different components of the same blood sample, returning a fixed-

---

* Corresponding author. Email: craig.trost@AnalyticDynamics.com

length vector of values. As a result, in the early 1970s, physicians were obtaining as many as 20 analyte results for each blood sample drawn. The analysers were further extended to select a variable number of tests per sample in real time. Now handheld, point-of-care analysers are becoming more popular. These technologies have produced a plethora of clinical data without the corresponding quantitative science needed for efficient utilization of the output.

In order for the physician to determine the presence of disease, a statistical definition of normal was adopted [15]. This definition is essentially the boundaries of the central 95% of a Gaussian distribution determined by analysing blood samples from a reference population, usually healthy normal people, with occasional stratification by gender and age as needed. Although it is hoped that the biological distribution of the reference population is somewhat invariant, the different chemical methods and analyser characteristics add noise and possibly bias components that are specific to the analyser used. In addition, environmental conditions as well as sample collection, handling, and storage conditions affect variability. Some analytes also have biological diurnal, seasonal, and environmental variation. The non-biological sources of variation will not be addressed in this paper because of the complexity. It is reasonable to assume during the early stages of this method development that the primary component of the biological variation is much larger than the analytical and other sources of variation for most analytes. A significant problem that remains in clinical practice is that reference regions are still treated as univariate, static objects. The rectangular intersection of those reference intervals does not correspond to a probability space, at least not in any kind of regular manner that makes sense biologically. This problem as well as methods for constructing static multivariate reference regions was reviewed elsewhere [26].

The liver is an organ that synthesizes many important chemicals for the body and that metabolizes toxic chemicals to be excreted in the urine and faeces. It is an organ that regenerates itself when part of it is removed or damaged. Nutrients and other ingested substances pass directly from the intestine through the portal vein to the liver before the rest of the body has access to them. From a protective viewpoint, this is good because toxins can be removed before they circulate and damage the other organs. It is bad in at least two ways. If too much toxin is ingested, it can irreversibly destroy the liver resulting in a liver transplant or death. It also neutralizes many drugs, which necessitates the use of other routes of administration, usually less acceptable to the patient, and possibly more expensive, than swallowing a pill.

The state of the liver can be determined by sampling it directly with a biopsy needle or during abdominal surgery. Both are unpleasant for the patient and have major risks, including death. Therefore, the state of the liver is usually indirectly determined by the molecules it leaks into the blood. For many chemicals, this leakage is in a steady state, or homeostasis, which is really a dynamic steady state, i.e., fluctuating regularly in time. There are numerous chemicals that are used as surrogate measures of the liver state; many are enzymes, which are protein catalysts in biochemical reactions. It has become customary in drug development, but probably not optimal in any mathematical sense, to measure the serum enzyme activities of alanine aminotransferase (ALT), aspartate aminotransferase (AST), γ-glutamyltransferase (GGT), lactate dehydrogenase (LD), and alkaline phosphatase (ALP) and the serum concentrations of total bilirubin (TB), total protein (TP), and albumin (Alb). These are also typical analytes used by practising clinicians for assessing liver and other diseases. The distribution of these chemicals in other tissues is known to some extent [7] and none are specific to the liver, contrary to popular belief among clinicians.

Previous work described the importance of finding better methods for drug-induced liver injury (DILI) and suggested that the dynamic patterns of liver-test vectors may

convey additional diagnostic information beyond the current clinical practice [27]. This report describes an approach to constructing a scientific and quantitative framework to support those findings using these blood analytes and their dynamics as measured from clinical blood samples to assess the homeostatic state of the liver. The primary hypothesis is that the entire 8-tuple of these measurements, or a subset thereof, can be represented in a vector space as a Brownian particle [8,11,22] on a spring and that a disease or drug effect on the liver will be reflected by departures from a multivariate Ornstein–Uhlenbeck (OU) process [12,28] used to define homeostasis. This report addresses the homeostasis part of the problem but not the patterns of deviation from homeostasis (pathodynamics).

In the clinic, blood tests are invasive, somewhat expensive, and time consuming. Management of the data is also expensive. Therefore, the physician only wants to obtain samples as often as necessary. This constraint also applies to drug development. How does this translate mathematically? Samples may be obtained only at random time intervals or widely spaced, scheduled times. Only a subset of analyte measurements may be acquired at a given time because a measurement problem occurs or because some analytes may be deemed by the physician or the study protocol as unnecessary and therefore a waste of money. It is rare to have more than a few time points during a given year. This makes it difficult to use the existing computational methods of regularly spaced stochastic processes (time series) measured frequently or continuously and in large numbers on a single sample path such as those in geophysics [20], economics [6], signal processing [4], and mathematical finance [13,18]. The problem is more analogous to studying the distribution of an ensemble of sample paths of non-interacting particles [22]. Clinical medicine also ignores the formal cross-correlation and auto-correlation structures of the data but uses these haphazardly through intensive training and memorization of patterns. The approach in this paper aims to address all of these issues in a single maximum-likelihood OU model.

2. Methods

2.1 The OU process

The multivariate mean-reverting OU process can be represented by the stochastic differential equation (SDE) using Itô calculus [8,11,19]

$$d\mathbf{X}_t = -\boldsymbol{\zeta}(\mathbf{X}_t - \boldsymbol{\mu})\,dt + \boldsymbol{\sigma}\,d\mathbf{B}_t, \tag{1}$$

where $\mathbf{X}_t$ is a $p \times 1$ continuous stochastic process vector, $\mathbf{B}_t$ is a q-dimensional Brownian motion, or Wiener, process, $\boldsymbol{\zeta}$ is a $p \times p$ positive definite symmetric drift matrix, $\boldsymbol{\mu}$ is a $p \times 1$ constant location vector, and $\boldsymbol{\sigma}$ is a $p \times q$ constant diffusion matrix. The solution to this equation is

$$\mathbf{X}_t = e^{-\boldsymbol{\zeta}t}\mathbf{X}_0 + (\mathbf{I}_p - e^{-\boldsymbol{\zeta}t})\boldsymbol{\mu} + e^{-\boldsymbol{\zeta}t}\int_0^t e^{-\boldsymbol{\zeta}u}\boldsymbol{\sigma}\,d\mathbf{B}_u, \tag{2}$$

where the initial state $\mathbf{X}_0$ is Gaussian with mean $\boldsymbol{\mu}$ and cross-covariance $\mathbf{a}$ and

$$e^{-\boldsymbol{\zeta}t} = \sum_{k=0}^{\infty}\frac{1}{k!}(-t)^k\boldsymbol{\zeta}^k \tag{3}$$

is a $p \times p$ matrix. From this it can be shown that $\mathbf{X}_t$ is a stationary Gaussian stochastic process of bounded variation with mean $E[\mathbf{X}_t] = \boldsymbol{\mu}$ and, for large $s \leq t$, auto-covariance

$$\mathrm{Cov}[\mathbf{X}_s, \mathbf{X}_t] = \mathbf{a}e^{-\boldsymbol{\zeta}\tau} \tag{4}$$

for $\tau = t - s$. This implies that the joint probability density function for two time points separated by τ time units is

$$f(\mathbf{X}, \tau) = (2\pi)^{-p}|\Psi(\tau)|^{-1/2}e^{-(1/2)(\mathbf{X}-\mathbf{M})'\Psi(\tau)^{-1}(\mathbf{X}-\mathbf{M})}, \tag{5}$$

where

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_s \\ \mathbf{X}_t \end{bmatrix}$$

are the measured random variables and

$$\mathbf{M} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{bmatrix}$$

and

$$\Psi(\tau) = \begin{bmatrix} \mathbf{a} & \mathbf{a}e^{-\zeta\tau} \\ e^{-\zeta\tau}\mathbf{a} & \mathbf{a} \end{bmatrix}$$

are unknown parameters of the distribution. For the SDE, $\boldsymbol{\mu}$, $\mathbf{a}$, and $\boldsymbol{\zeta}$ are the unknown time-invariant parameters of interest where $\boldsymbol{\sigma}\boldsymbol{\sigma}' = \boldsymbol{\zeta}\mathbf{a} + \mathbf{a}\boldsymbol{\zeta}$, a positive definite symmetric matrix, is a measure of the Brownian motion scaling. Since the cross-covariances were not the focus of interest of this paper, no attempt was made to estimate a unique, biologically meaningful $\boldsymbol{\sigma}$. If $p = q$ and the orthonormal rotation matrix that diagonalizes $\boldsymbol{\zeta}$ is identical to the one that diagonalizes $\mathbf{a}$, then it can be shown that $\boldsymbol{\sigma} = (2\boldsymbol{\zeta}\mathbf{a})^{1/2} = (2\mathbf{a}\boldsymbol{\zeta})^{1/2}$.

Statistical sampling schemes for $\mathbf{X}_t$ on the interval $0 \leq t \leq T < \infty$ are well behaved in most applications. The stationary stochastic processes allow the estimation of unknown parameters from a single sample path by observing continuous data, by observing discrete samples with fixed time spacing from the sample path for large T, or by observing a large number of time points within the fixed interval. None of these are practical for clinical data sample paths, which are always discretely sampled, usually at a few widely spaced irregular times and for small or fixed T. Until the parameters are sufficiently well estimated, the number of patients observed (n), which are stochastically independent sample paths, must be large enough to achieve good convergence of the estimation sequences and the optimality properties that result.

2.2   Modelling serum analyte dynamics

The most straightforward and commonly accepted method for estimating the unknown parameters of a multivariate Gaussian probability density is the maximum-likelihood estimation (MLE) method [25]. It also produces estimators that are asymptotically Gaussian, unbiased, and minimum variance for large n. However, it may not have these properties in small samples. For simplicity, we assumed that $\boldsymbol{\zeta}$ is symmetric positive definite and $p = q$. This was done because of the sparseness of the data and to simplify the computations. Therefore each time pair was assumed to be independent, which may artificially deflate the variance. The likelihood function for a random sample of n

individuals measured at time s and then at time $s + \tau$ is

$$L(\mathbf{a}, \boldsymbol{\zeta}, \boldsymbol{\mu}; \tau) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} f(\mathbf{X}_{ij}, \tau), \tag{6}$$

where $\mathbf{X}_{ij}$ is a complete p-vector of measurements from the ith individual at the jth time interval τ. For $\tau = 1, 2, \ldots, 30$ estimation using this objective function can be simplified by maximizing the logarithm of L such that the derivatives of

$$\log L = - \left(\sum_{i=1}^{n} m_i \right) p \log(2\pi) - \frac{1}{2} \left(\sum_{i=1}^{n} m_i \right) \log |\Psi(\tau)|$$
$$- \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m_i} (\mathbf{X}_{ij} - \mathbf{M})' \Psi(\tau)^{-1} (\mathbf{X}_{ij} - \mathbf{M}) \tag{7}$$

are zero. Since the objective function is nonlinear in the parameter space, Newton's method was applied. The stratification by τ allowed the examination of the assumption that the parameters are invariant with respect to τ. Unfortunately, there were not enough observations to check if the parameters are invariant to observation time t. This last constraint was not considered to be too serious since the OU process in steady state only depends on τ. Programming for the MLE models was done using Matlab [16] code (Supplemental Files MATLAB).

2.3 Verification of the MLE code using sample statistics

Because of the complexity of the MLE code, sample statistics that converge in large samples to the same constants as the MLE estimates were used. Estimates for the means and covariances were computed on a component-wise and pair-wise basis, respectively, because of incomplete observation vectors. The overall means $\boldsymbol{\mu}$ and cross-covariances $\mathbf{a}$ were estimated using

$$\bar{X}_k = \frac{1}{N_k} \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ki}(t_j) X_i(t_j) \tag{8}$$

and

$$a_{k\ell} = \frac{1}{N_{k\ell} - 1} \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ki}(t_j) \delta_{\ell i}(t_j) \left[X_{ki}(t_j) - \bar{X}_k \right] \left[X_{\ell i}(t_j) - \bar{X}_\ell \right] \tag{9}$$

respectively, where m is the number of distinct times in the union for all measurement times, the indicator function $\delta_{ki}(t_j) = 1$ if the kth component was measured at the jth time point t_j for the ith individual or $\delta_{ki}(t_j) = 0$ if not measured,

$$N_k = \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ki}(t_j) \tag{10}$$

and

$$N_{k\ell} = \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ki}(t_j) \delta_{\ell i}(t_j). \tag{11}$$

The auto-covariances were estimated for each τ to match the MLE's using

$$b_{k\ell}(\tau) = \frac{1}{N_{k\ell}} \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ki}(t_j)\delta_{\ell i}(t_j + \tau)[X_{ki}(t_j) - \bar{X}_k(t_j)][X_{\ell i}(t_j + \tau) - \bar{X}_\ell(t_j + \tau)], \quad (12)$$

modifying $N_{k\ell}$ accordingly, and then computing the matrix equation

$$\boldsymbol{\zeta}(\tau) = \frac{-\log(\mathbf{a}^{-1}\mathbf{b}(\tau))}{\tau}, \quad (13)$$

where the log function here is the inverse of the function in Equation (3).

The standard errors for these estimators were calculated using the bootstrap method [25]. Random samples of size 1000 were chosen with replacement. Each sample was replicated 1000 times.

2.4   Analysis of clinical trial data

The clinical laboratory data used for modelling were obtained by pooling results from hundreds of drug development clinical trials (see Supplemental Files DATA). Each laboratory test was converted to have a standard unit and name. Untreated and placebo-treated healthy normal subjects' data were extracted and obvious outliers were removed, especially those that were incompatible with life and those that were inconsistent within a subject over time. After the data were prepared, a subset of males between the ages of 20 and 40 years was extracted for analysis. This subset was chosen to remove the effects of biological development and aging and because there were not enough females to do a separate analysis. It is expected that these other subpopulations will have different parameter values.

The 80 unknown parameters of the multivariate OU model in Equation (1) were estimated using MLE methods [2,25]. Since the OU model depends only on time differences (τ) between observations, the analysis was stratified by $\tau \in \{1, 2, \ldots, 30\}$ in days to evaluate how the estimates varied with τ. It was not possible to stratify by the raw time because of insufficient sample sizes, although this effect would not be expected in an unsynchronized sample. Preliminary MLE analyses suggested that replacing τ with $\sqrt{\tau}$ in Equation (5) leads to time-invariant parameters; this was termed the modified OU model.

2.5   Numerical methods

Because of the constraints of clinical data, new software had to be developed to obtain estimates of the unknown parameters in the OU model. Since this Matlab code was written only to handle complete vectors, the sample size necessarily decreased as the dimension increased. Ideally, an 8-dimensional analysis should have been sufficient to estimate all the parameters of the full model. However, because of the dimension versus sample-size problem, the analyses were stratified by dimension and the estimates for all possible combinations of analytes for the given dimension were averaged to get a mean parameter estimate. This involved running $30 \cdot (2^8 - 1) = 7{,}650$ models. Some did not converge, primarily due to insufficient sample size. The rule that the sample size had to be greater than $10p$ was implemented to improve the modelling behaviour. Sample statistics, which are model-free and distribution-free, for the means and covariances from Equations (8) and (9) based on the first two sample moments were used to check the MLE results. The final descriptive statistical analyses of the results were performed using R [21].

For the full set of models, processing time for a single processor was on the order of weeks. A grid of over 200 nodes reduced the computing time to a few days, depending on resource contention.

3. Results

3.1 Descriptive statistics

Although the sample included data from about 5000 subjects, only about 2000 had one or more follow-up visits and the number in each stratum was on the order of hundreds. The frequency distribution of the number of subjects (n) relative to day 0 is shown in Figure 1. The frequencies decline exponentially, making it difficult to get precise estimates over long periods of time and demonstrating the need for special approaches to estimating the parameters. The frequency of measurements by time interval was somewhat flatter but only had larger numbers in the intervals of 1–14 days. Figure 2 shows the frequencies for ALT, which is typical of the other analytes. However, intervals up to 30 days were kept to salvage some additional information. The histograms for all individual analytes are shown in Supplemental File Figures S3.1–3.9 online. The numbers were significantly reduced when analytes were combined.

The parameters were statistically evaluated in detail for dimensions $p = 1$ (the degenerate case used clinically), $p = 2$ (the simplest multivariate case), and $p = 6$ (the largest dimension with reasonable numbers of subjects). Again ALT is used as a typical example of the findings. Figure 3(a) shows that the mean is time-invariant relative to the first observation time and Figure 3(b) shows the same property for the interval times. Figures 4(a) and (b) show the same for the cross-covariances, respectively. The results for the all means and cross-covariances were clearly invariant with respect to τ (Supplemental Files Figures S1–S3).
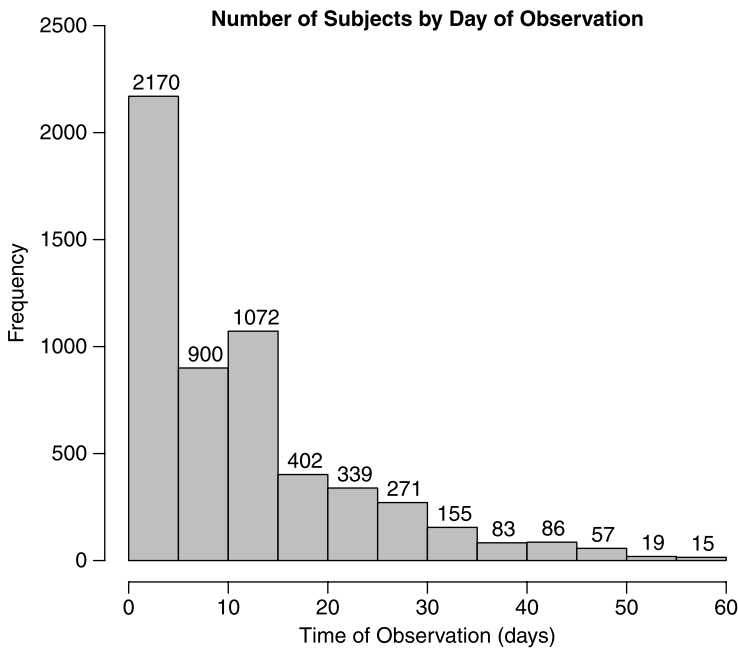

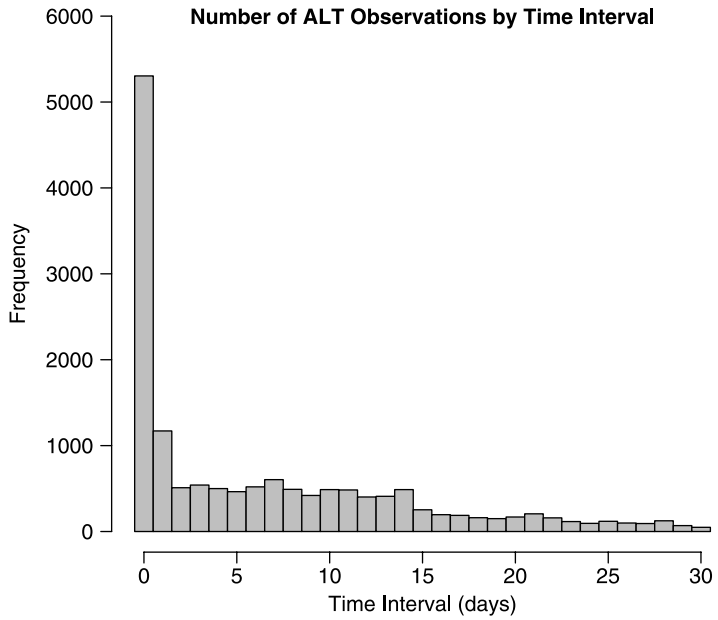Figure 1. Frequency of subjects by time of observation.

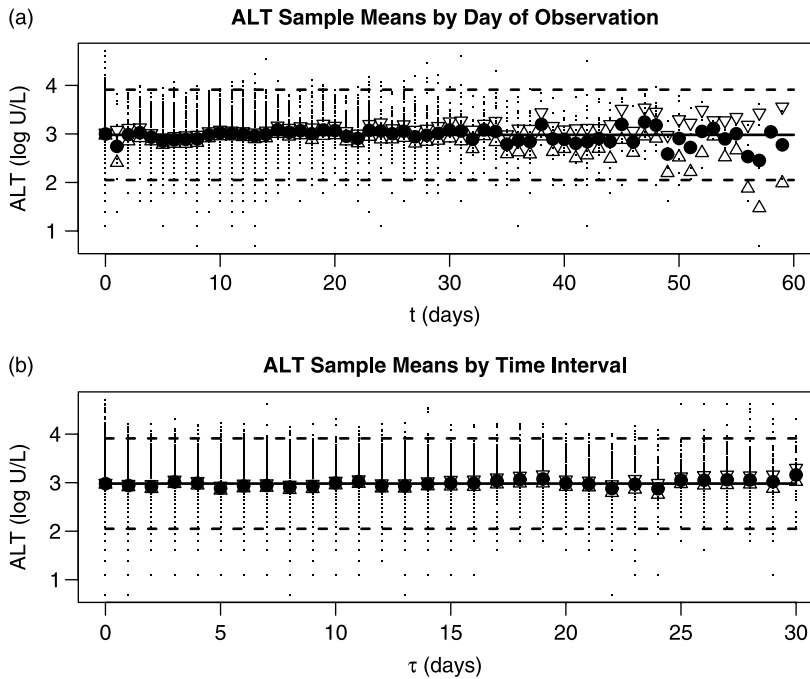Figure 2.   Frequency of measurements for ALT.

Figure 3.   Sample means for ALT by (a) time of observation and (b) time between observations. The raw data are indicated by points, the means by solid circles, the 95% confidence intervals by open triangles, the overall mean by the solid line and the overall central 95% region by the dash lines.

(a)

ALT Sample Cross–covariance by Day of Observation

Figure 4. Sample cross-covariance for ALT by (a) time of observation and (b) time between observations. The estimates are indicated by solid circles, the 95% confidence intervals by open triangles, the overall estimate by the solid line and the overall central 95% region by the dash lines.
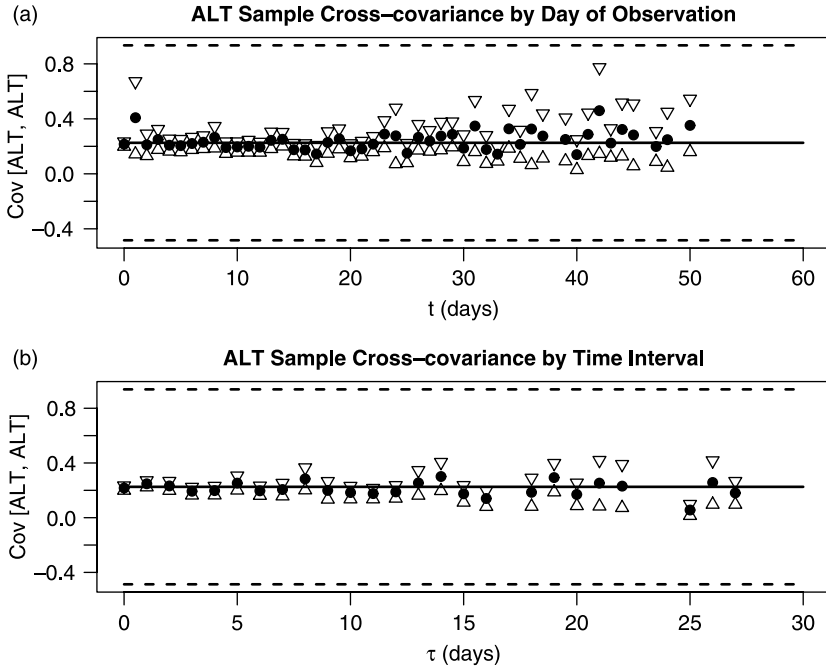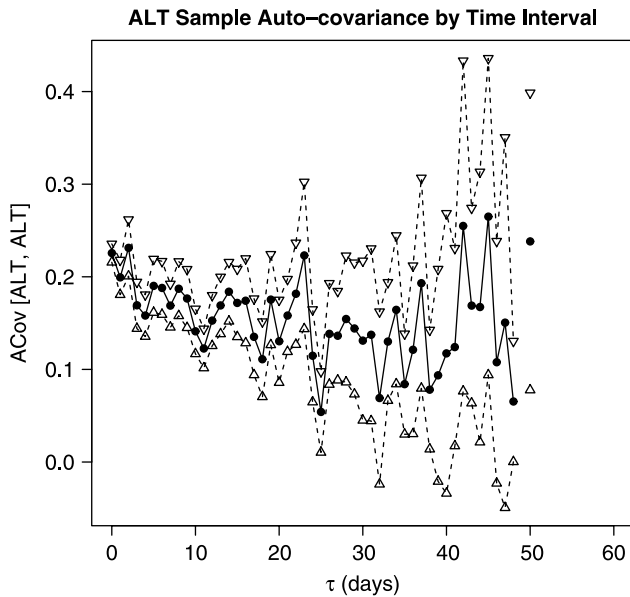
ALT Sample Auto–covariance by Time Interval

Figure 5. Sample auto-covariance for ALT by time between observations. The estimates are indicated by solid circles and the 95% confidence intervals by open triangles.

The sample estimates for the auto-covariance showed an unexpected behaviour. Although the auto-covariance was clearly not zero for ALT (Figure 5), it appeared to be flatter than an exponential decay. This time-varying behaviour for $\boldsymbol{\zeta}$ was studied in the MLE framework because the estimates were more stable, unlike Figure 5.

3.2 Maximum likelihood estimates (MLE)

The evaluation of the components of $\boldsymbol{\zeta}$ and $\boldsymbol{\sigma\sigma'}$ with respect to τ showed a relationship that was described by the regression $\zeta_{ij} = k_{ij}/\sqrt{\tau}$ as demonstrated in Figures 6 and 7, respectively, for ALT. The complete set of estimates is plotted over time in the Supplementary File Figure S1. The MLE model for Equation (7) was modified and rerun substituting $\sqrt{\tau}$ for τ in the Matlab code. After this modification, both $\boldsymbol{\zeta}$ and $\boldsymbol{\sigma\sigma'}$ were essentially constant with the exception of $\tau = 1$ (see Figures 8 and 9). It was suspected that there was something unusual about this particular time point such as time intervals significantly less than 24 h and these were dropped from further summary analyses. The complete set of plots for the case using $\sqrt{\tau}$ is shown in the Supplemental File Figure S2. All of the estimates for $\boldsymbol{\mu}$ and $\mathbf{a}$ were essentially constant over time in both situations. It was also found that this time modification did not work as well for TB (subscript 6).

The summary of the parameter estimation results using the sample statistics and modified OU models are shown in Tables 1–4. Although many off-diagonal parameters were near zero, many were not. Even if they were all zero, a multivariate model is still required to control the probability of false positives [26].

3.3 Cross-correlation

Based on the MLE's of the unknown parameters for $p = 6$, a graphical representation of the static relationship between the natural logarithms of ALT and AST demonstrates the

Figure 6. The MLE estimates of $\boldsymbol{\zeta}$ for ALT by time between observations. The estimates are indicated by solid circles, the 95% confidence intervals by open triangles, the overall mean by the dash line, and the regression line using the square root of the interval by the solid line.
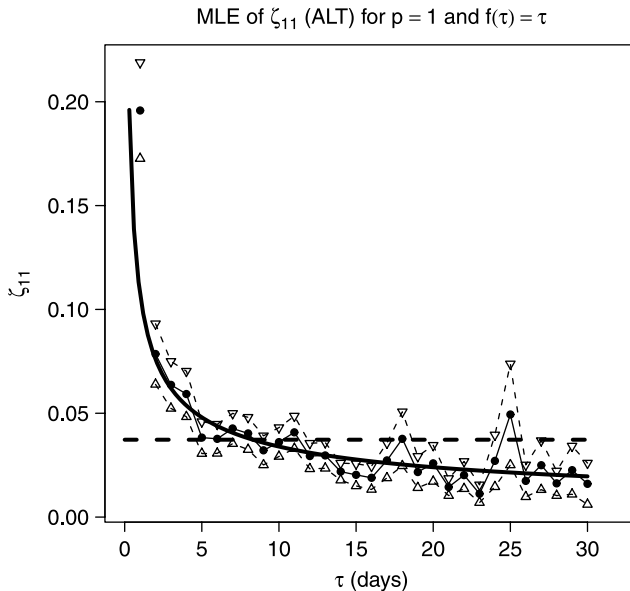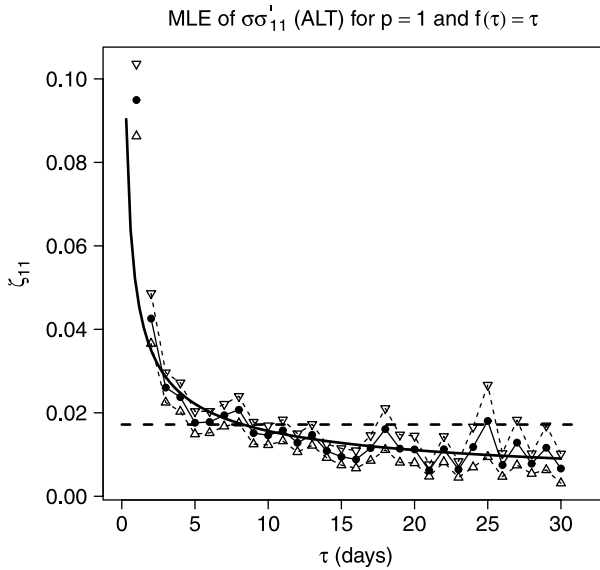
Figure 7. The MLE estimates of $\sigma\sigma'$ for ALT by time between observations. The estimates are indicated by solid circles, the 95% confidence intervals by open triangles, the overall mean by the dash line, and the regression line using the square root of the interval by the solid line.

cross-correlation between these two analytes (Figure 10). The inner rectangle is the union of the two 95% reference intervals and the outer frame is approximately a 3-fold change in those reference intervals in the raw-data space, both clinical standards. Results falling anywhere on this entire graph would be considered as not clinically significant under current practice. The ellipse is the joint 95% MLE reference region [26]. The striations occurred because of the excessive rounding that is customary for clinical data presentation. This has an unusually large effect on TB data. It can be easily observed that the data do not fit the box. In higher dimensions, this defect is exacerbated.

3.4 Auto-correlation

A major advantage of using dynamic models is that they capture the time information between measurements, usually reflected in the auto-covariance (auto-correlation) structure or parameters, in this case ζ. In the OU model, the auto-covariance decays at an exponential rate. This is a common assumption embedded in many software packages for longitudinal, or

Table 1. Sample and MLE estimates of μ.

| Analyte | Sample | MLE | | |
		$p = 1$	$p = 2$	$p = 6$
ALT	2.98	2.98	2.98	2.93
AST	3.07	3.03	3.03	3.03
GGT	2.82	2.98	2.96	2.79
LD	5.07	5.14	5.12	5.03
ALP	4.32	4.31	4.31	4.32
TB	-0.51	-0.60	-0.61	-0.55
TP	7.26	7.31	7.31	7.26
Alb	4.44	4.41	4.41	4.43

D.C. Trost et al.

Table 2. Sample and MLE estimates of **a**.

Analyte	ALT	AST	GGT	LD	ALP	TB	TP	Alb
ALT								
Sample	0.2256	0.1030	0.0725	0.0349	−0.0068	0.0009	0.0167	0.0106
MLE p = 1	0.2314							
MLE p = 2	0.2253	0.1215	0.0690	0.0532	−0.0033	−0.0170	0.0089	0.0189
MLE p = 6	0.2312	0.1057	0.0656	0.0306	0.0015	−0.0145	0.0174	0.0120
AST								
Sample		0.1277	0.0193	0.0458	−0.0080	0.0100	0.0208	0.0191
MLE p = 1		0.1543						
MLE p = 2		0.1484	0.0186	0.0963	−0.0153	0.0080	0.0176	0.0293
MLE p = 6		0.1316	0.0170	0.0477	−0.0048	0.0068	0.0145	0.0239
GGT								
Sample			0.3195	0.0046	0.0103	−0.0528	0.0309	−0.0109
MLE p = 1			0.2471					
MLE p = 2			0.2633	−0.0006	0.0149	−0.1125	0.0320	−0.0116
MLE p = 6			0.3537	0.0017	0.0183	−0.0500	0.0425	−0.0084
LD								
Sample				0.1309	0.0015	0.0139	0.0047	0.0006
MLE p = 1				0.2043				
MLE p = 2				0.1852	−0.0260	0.0517	0.0135	0.0174
MLE p = 6				0.1023	−0.0022	0.0123	0.0040	0.0077
ALP								
Sample					0.0987	−0.0136	0.0130	0.0163
MLE p = 1					0.1019			
MLE p = 2					0.0979	−0.0157	0.0177	0.0169
MLE p = 6					0.0890	−0.0077	0.0135	0.0166
TB								
Sample						0.2484	0.0339	0.0258
MLE p = 1						0.4033		
MLE p = 2						0.3885	0.0295	0.0239
MLE p = 6						0.2915	0.0263	0.0291
TP								
Sample							0.2570	0.0870
MLE p = 1							0.2364	
MLE p = 2							0.2338	0.0876
MLE p = 6							0.2336	0.0752
Alb								
Sample								0.1091
MLE p = 1								0.1274
MLE p = 2								0.1274
MLE p = 6								0.1113

time-series, data. In our modified OU model, the auto-covariance decays exponentially with respect to the square root of time. This rate of decay is much slower. Based on a $p = 6$ model of simultaneous parameter estimates, the size of the 95% central marginal probability ellipse in two dimensions grows slowly (Figure 11), although it moves to the asymptotic state along the minor axis faster than along the major axis. Conversely, the ability to detect time changes or deviations from the expected randomness is greatly enhanced because the region that represents no statistical change is much smaller. The outer ellipse is the 95% reference

Table 3. Sample and MLE estimates of ζ.

Analyte	ALT	AST	GGT	LD	ALP	TB	TP	Alb
ALT								
Sample	0.2054	−0.1777	−0.0914	0.0220	−0.0305	−0.0373	0.0425	−0.0171
MLE p = 1	0.1108							
MLE p = 2	0.1192	−0.0443	−0.0206	−0.0248	0.0066	−0.0040	0.0192	0.0003
MLE p = 6	0.1640	−0.0731	−0.0318	−0.0104	−0.0002	−0.0073	0.0079	0.0072
AST								
Sample		0.3398	0.0453	−0.1090	0.0757	0.0456	−0.0033	−0.1044
MLE p = 1		0.1142						
MLE p = 2		0.1479	0.0013	−0.0195	0.0222	0.0051	0.0230	−0.0144
MLE p = 6		0.2415	0.0045	0.0355	0.0124	0.0101	0.0163	−0.0045
GGT								
Sample			0.1754	−0.0069	−0.0167	0.0709	−0.0487	0.0621
MLE p = 1			0.1039					
MLE p = 2			0.1185	−0.0296	0.0024	0.0268	0.0130	0.0474
MLE p = 6			0.1411	−0.0140	0.0031	0.0317	−0.0063	0.0220
LD								
Sample				0.0303	0.0478	−0.0344	0.0069	−0.0038
MLE p = 1				0.1119				
MLE p = 2				0.1615	0.0153	−0.0059	−0.0368	−0.0097
MLE p = 6				0.1921	−0.0105	0.0141	0.0259	0.0229
ALP								
Sample					0.0510	−0.0116	0.0077	−0.0319
MLE p = 1					0.0409			
MLE p = 2					0.0441	0.0035	0.0158	0.0091
MLE p = 6					0.0464	0.0032	0.0167	0.0056
TB								
Sample						0.0740	−0.0189	0.0408
MLE p = 1						0.1145		
MLE p = 2						0.1219	0.0194	0.0248
MLE p = 6						0.1952	0.0105	0.0066
TP								
Sample							0.1409	0.0328
MLE p = 1							0.2021	
MLE p = 2							0.2183	0.0798
MLE p = 6							0.1986	0.0761
Alb								
Sample								0.1004
MLE p = 1								0.1645
MLE p = 2								0.1661
MLE p = 6								0.1708

boundary for a single measurement (Figure 11) and it is also the boundary for a second measurement taken infinitely long after the first one is observed at the centre. For this analyte pair, the ellipse effectively reaches the asymptotic boundary after about 180 days. It may be faster or slower for other analyte combinations but without the homeostasis model, this information cannot be used efficiently.

The numbers of days for the specified marginal ($p = 6$) decay are shown in Table 5 for the 8 analytes. Somewhere in the range of 90% decay or higher, the observation vectors

Table 4. Sample and MLE estimates of $\boldsymbol{\sigma\sigma'}$.

Analyte	ALT	AST	GGT	LD	ALP	TB	TP	Alb
ALT								
Sample	0.0458	−0.0091	−0.0179	−0.0049	−0.0023	−0.0038	0.0134	−0.0082
MLE p = 1	0.0505							
MLE p = 2	0.0504	0.0178	0.0087	0.0009	0.0015	−0.0045	0.0117	0.0049
MLE p = 6	0.0630	0.0144	0.0015	0.0042	0.0017	−0.0032	0.0098	0.0060
AST								
Sample		0.0375	0.0075	−0.0151	0.0113	0.0098	0.0042	−0.0160
MLE p = 1		0.0346						
MLE p = 2		0.0376	0.0087	0.0128	0.0030	0.0043	0.0142	0.0044
MLE p = 6		0.0461	0.0024	0.0169	0.0026	0.0081	0.0138	0.0078
GGT								
Sample			0.0883	−0.0005	−0.0043	0.0244	−0.0075	0.0161
MLE p = 1			0.0541					
MLE p = 2			0.0622	−0.0080	0.0043	0.0024	0.0143	0.0099
MLE p = 6			0.0962	−0.0072	0.0049	0.0071	0.0095	0.0082
LD								
Sample				−0.0013	0.0144	−0.0118	0.0014	−0.0065
MLE p = 1				0.0528				
MLE p = 2				0.0589	0.0001	0.0036	−0.0262	−0.0032
MLE p = 6				0.0316	0.0002	0.0075	0.0119	0.0062
ALP								
Sample					0.0086	−0.0030	0.0029	−0.0010
MLE p = 1					0.0082			
MLE p = 2					0.0079	−0.0006	0.0093	0.0052
MLE p = 6					0.0072	0.0011	0.0094	0.0057
TB								
Sample						0.0303	0.0069	0.0141
MLE p = 1						0.0658		
MLE p = 2						0.0644	0.0198	0.0147
MLE p = 6						0.0814	0.0212	0.0130
TP								
Sample							0.0754	0.0340
MLE p = 1							0.0945	
MLE p = 2							0.1013	0.0550
MLE p = 6							0.0969	0.0526
Alb								
Sample								0.0230
MLE p = 1								0.0415
MLE p = 2								0.0426
MLE p = 6								0.0419

might be considered as stochastically independent, from a practical standpoint. Otherwise, a dynamic, homeostasis model appears to be needed to make proper clinical decisions.

4. Discussion

These findings suggest that dynamic models of liver tests might improve the ability to detect when the liver is transitioning away from homeostasis in the presence of variation that is currently considered clinically insignificant. This can be achieved by using the time
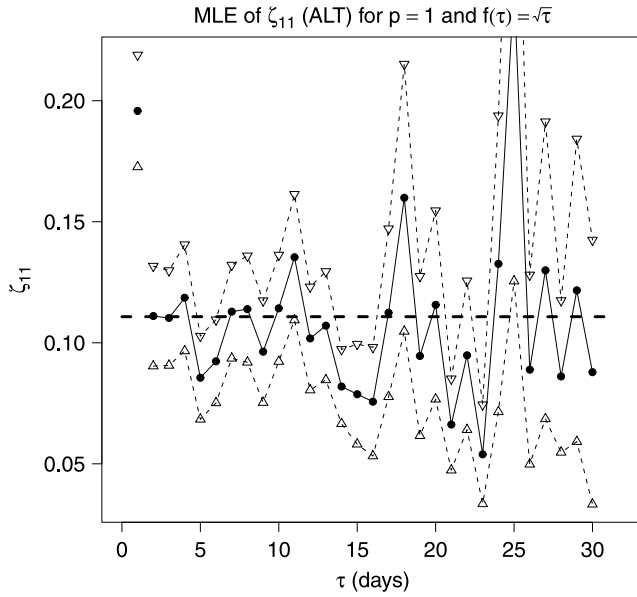
Figure 8. The MLE estimates of $\boldsymbol{\zeta}$ for ALT by time between observations using $\sqrt{\tau}$. The estimates are indicated by solid circles, the 95% confidence intervals by open triangles, and the overall mean by the dash line.
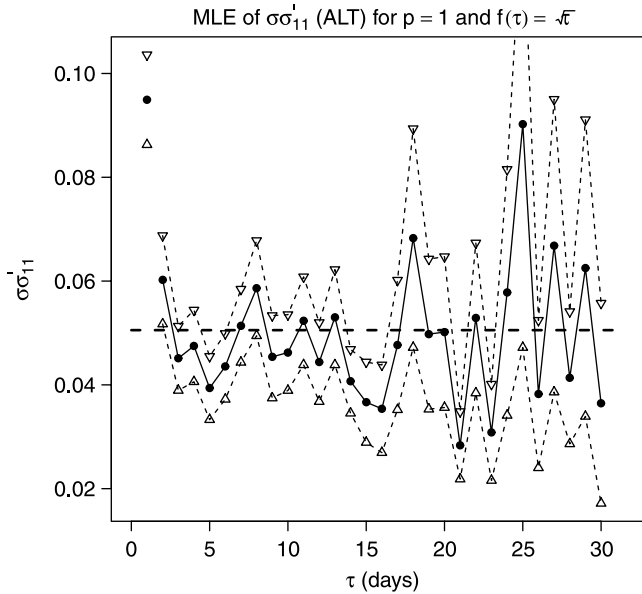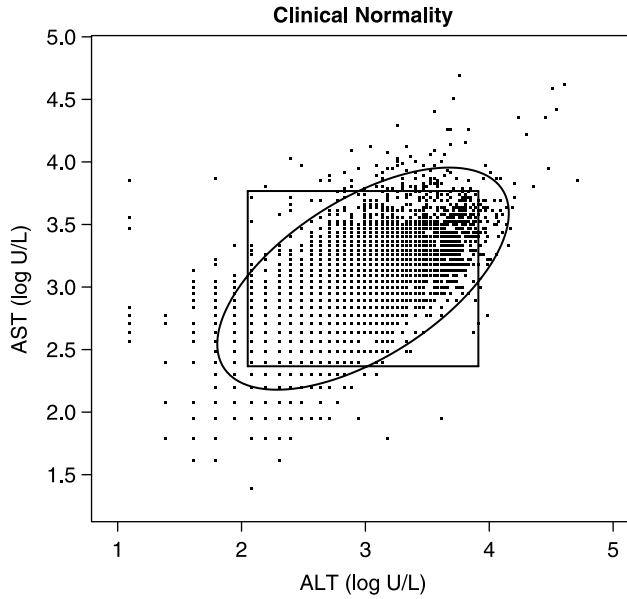

Figure 9. The MLE estimates of $\boldsymbol{\sigma\sigma}'$ for ALT by time between observations using $\sqrt{\tau}$. The estimates are indicated by solid circles, the 95% confidence intervals by open triangles, and the overall mean by the dash line.

Figure 10. Clinical normality for the natural logarithms of AST and ALT. The box is the intersection of the two individual 95% reference intervals. The ellipse is the joint 95% reference region using the cross-correlation information. The graph frame is sized to enclose the region that is considered clinically insignificant for these two analytes. Actual data are plotted using dots to demonstrate the lack of fit when using the box for clinical decisions. The discretization of the data is an artefact of clinical rounding error in data reporting.
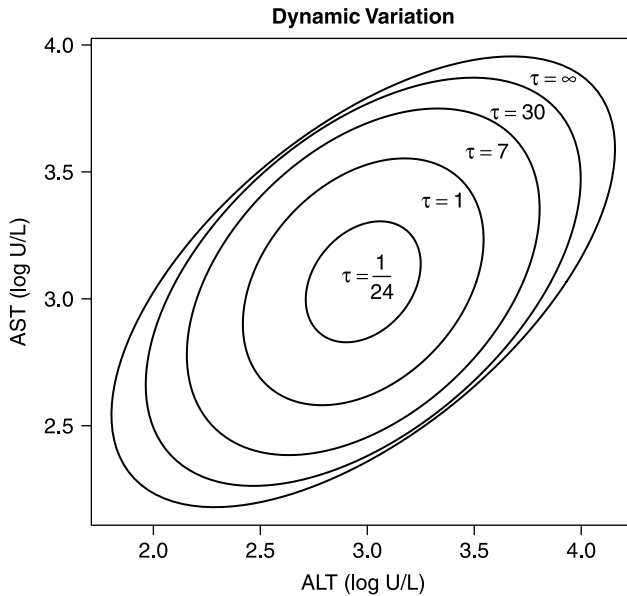
Figure 11. Dynamic variation of the joint 95% reference region for the natural logarithms of AST and ALT for two measurement times. It is assumed that the first measurement is at the centre of the ellipse. The time interval τ is in days.

Table 5. Elapsed time in days to the indicated percentage decay in the marginal auto-correlation.

Analyte	Auto-correlation decay (%)			
	50	75	90	95
ALT	18	71	197	334
AST	8	33	91	154
GGT	24	97	266	451
LD	13	52	144	243
ALP	223	893	2463	4168
TB	13	50	139	236
TP	12	49	134	228
Alb	16	66	182	308

information (memory) between measurements in a modified OU model to detect nonrandom movement or changes in stochastic behaviour. A candidate for the SDE for liver homeostasis is

$$d\mathbf{X}_t = -t^{-1/2}\boldsymbol{\zeta}(\mathbf{X}_t - \boldsymbol{\mu})\,dt + t^{-1/4}\boldsymbol{\sigma}\,d\mathbf{B}_t. \tag{14}$$

Unfortunately, this equation is not physically intuitive because the homeostatic force appears to be relative to the amount of time between observations. The properties of this equation are given in the Appendix. However, it will require further study to see if there is a formulation that leads to a solution that is only a function of τ.

Because DILI is so difficult to detect in clinical trials, it is important to find more sensitive and specific ways to detect these signals [27]. This model provides a mathematical framework for enhanced signal detection and classification and is consistent with clinical observations in untreated healthy normal subjects. Even though stochastic calculus is ubiquitous in finance and economics, it is largely untapped in the biological sciences, especially in clinical medicine [1,3,10,14,17,23,24,29,30]. The use of a q-dimensional ($q \neq p$) Brownian motion process to describe the driving force of the biological system could be very useful. The value of q is unknown and does not enter into any of the calculations presented here. Many aspects of the whole biological organism may not be measurable and thus are unknowable, especially when interactions with the environment are included. It is conceivable that the summation of large numbers of stochastic processes should lead to an observable Gaussian process, for which some function of Brownian motion may be a reasonable representation. Because of insufficient data, the goodness-of-fit of the Brownian motion assumption was not tested statistically but the results from the model appear to be consistent with diffusion process that is slower than Brownian motion.

The next step in modelling homeostasis is to develop an MLE model that uses all the data simultaneously for all time intervals and for all combinations of incomplete vectors. This work is currently underway but is much more difficult from a numerical analysis standpoint. After this new software is ready, besides finding an appropriate diffusion process, additional questions need to be explored. These include determining if the population value of $\boldsymbol{\mu}$ is appropriate or if it needs to be individualized [23], if the parameters vary by gender, age, and other demographic aspects, and if the variation due to the analytical variation in laboratory methods can be separated from the biological variation. The applicability of this approach may be useful in modelling relative liver

homeostasis in chronic diseases such as diabetes so that changes due to DILI or another disease can be detected efficiently even though the patient is abnormal at baseline. There is nothing here that makes the approach restricted to liver biology. Similar homeostasis models can be developed for other combinations of clinical measurements.

A clinician may wonder why so much attention is being paid to constructing a more sensitive method for detecting 'ordinary' DILI when the medical problem is really finding methods for detecting 'idiosyncratic' DILI with greater specificity. For the nonclinician readers, the medical definition of *idiosyncratic* is something characterised as being peculiar to the given individual, i.e., particular genetic and environmental factors make a person peculiarly susceptible. For instance, idiosyncratic drug reactions are a problem for drug development because they tend to be very rare with an incidence on the order of 1 in 10,000 or lower, and the drug is removed from the market or the development is stopped if the reaction is serious and occurs in only a few patients. For truly idiosyncratic reactions, this problem may be unsolvable both clinically and mathematically because the combinations of factors can be very large and there may never be enough data to verify a model or a laboratory test. The homeostasis model was pursued first for the following reasons:

- As described in this paper and elsewhere [26], current methods of detecting abnormalities are crude and do not use clinical measurements efficiently. A better or more mathematical definition of normal is needed to form a foundation for modelling specific pathological states.
- Better models of normal should lead to smaller regions, increasing sensitivity, and therefore increase the size of the complement set, indirectly improving specificity when the region for a particular abnormality is close to or intermingled with the normal region.
- Medical statisticians tend to focus on the null hypothesis (normal) but clinicians are interested in the various alternative hypotheses (pathological states), which are generally not modelled by statisticians. Although the latter is the real aim of our line of research, proper mathematical definitions of normal have not been used previously, therefore detecting deviations from normal requires better definitions.
- The space between the normal region and the 'clinically significant' region, commonly defined univariately as a three-fold change from the normal boundary (Figure 10), is largely ignored in drug development and clinical practice. By using the appropriate models of homeostasis and disease, we may find many more specific patterns in this space that would otherwise be ignored. It is hypothesised that some types of so-called idiosyncratic DILI are not really idiosyncratic but are relatively common and undetected because of the current insensitive approach. Better homeostasis models may lead to the discovery of these types of DILI and may lead to predicting which patients are susceptible to idiosyncratic reactions, even though the actual rare event may not be amenable to such modelling.
- A dynamic model of homeostasis whose variation can be considered largely as biological randomness provides a standard for measuring changes in or deviations from the normal random fluctuations even though the patient's sample path may never cross the normal boundary. It is hypothesised that changes in random behaviour can be detected and modelled to predict specific types of DILI.
- The homeostasis model should be the simplest dynamic case for describing the behaviour of clinical measurements. If models for this case are unknown or unknowable, it is not likely that specific dynamic disease models can be

constructed. This paper is meant to show that such a model exists for liver tests. It may be possible now to generalise it to the various liver lesions.

- Multivariate properties of existing common laboratory tests may be less expensive, more sensitive, and more specific than new biomarkers.

To some extent, the concept of clinical significance is used by clinicians as a device to avoid unnecessary costs and procedures. It most likely came about when clinicians discovered that statisticians could make the variance of a statistic arbitrarily small by increasing the sample size in a study. Common statistical approaches not only filter out measurement errors but also tend to remove much of the biological variability. Part of the intent of this research is to understand the dynamic biological variability in clinical measurements to determine if the regions of normality and clinical insignificance contain any diagnostic information.

Once we have a good quantitative definition of homeostasis, the specificity problem can be attacked using pathodynamic (nonequilibrium) models of disease. Such phenotypic models may then map to specific genotypes or genotype-environment combinations and may provide a good transitional link between systems biology and clinical medicine.

References

[1] O.O. Aalen and H. Gjessing, *Survival models based on the Ornstein–Uhlenbeck process*, Lifetime Data Anal. 10 (2004), pp. 407–423.
[2] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley & Sons, New York, 1984.
[3] P.K. Andersen and K. Liestøl, *Attenuation caused by infrequently updated covariates in survival analysis*, Biostatistics 4 (2003), pp. 633–649.
[4] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software*, John Wiley & Sons, New York, 2001.
[5] D. Berger, *A brief history of medical diagnosis and the birth of the clinical laboratory: Part 2 – Laboratory science and professional certification in the 20th century*, Med. Lab. Obs. (1999), p. 32, August.
[6] P.J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, 2nd ed., Springer-Verlag, New York, 1991.
[7] C.A. Burtis and E.R. Ashwood (eds), *Tietz Textbook of Clinical Chemistry*, 3rd ed., W.B. Saunders Company, Philadelphia, 1999.
[8] R. Durrett, *Stochastic Calculus: A Practical Introduction*, CRC Press, Boca Raton, FL, 1996.
[9] K.B. Ekelman (ed.), *New Medical Devices: Invention, Development, and Use*, National Academy Press, Washington, DC, 1988.
[10] R.A. Heath, *The Ornstein–Uhlenbeck model for decision time in cognitive tasks: an example of control of nonlinear network dynamics*, Psychol. Res. 63 (2000), pp. 183–191.
[11] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer-Verlag, New York, 1991.
[12] P. Langevin, *On the theory of Brownian motion*, C.R. Acad. Sci. Paris 146 (1908), pp. 530–533.
[13] X.S. Lin, *Introductory Stochastic Analysis for Finance and Insurance*, John Wiley & Sons, New York, 2006.
[14] Y. Madec and C. Japhet, *First passage time problem for a drifted Ornstein–Uhlenbeck process*, Math. Biosci. 189 (2004), pp. 131–140.

[15] H.F. Martin, B.J. Gudzinowicz, and H. Fanger, *Normal Values in Clinical Chemistry: A Guide to Statistical Analysis of Laboratory Data*, Marcel Dekker, Inc., New York, 1975.

[16] Matlab Development Team, *Matlab: The Language of Technical Computing*, The MathWorks Inc, Natick MA, 2006.

[17] J.S. Metcalfe, L.-C. Chen, T.-Y. Chang, K. McDowell, J.J. Jeka, and J.E. Clark, *The temporal organization of posture changes during the first year of independent walking*, Exp. Brain Res. 161 (2005), pp. 405–416.

[18] M. Musiela and M. Rutkowski, *Martingale Methods in Financial Modeling: Theory and Applications*, Springer-Verlag, Berlin, 1997.

[19] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 5th ed., Springer-Verlag, Berlin, 1998.

[20] J.D. Pelletier and D.L. Turcotte, *Self-affine time series: II. Applications and models*, Adv. Geophys. 40 (1999), pp. 91–166.

[21] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2007. ISBN 3-900051-07-0, URL http://www.R-project.org.

[22] L.E. Reichl, *A Modern Course in Statistical Physics*, 2nd ed., John Wiley & Sons, New York, 1998.

[23] G.K. Rosenkranz, *Modeling laboratory data from clinical trials*, Comp. Stat. Data Anal. 53 (2009), pp. 812–819.

[24] J.W. Sleigh, D.A. Steyn-Ross, M.L. Steyn-Ross, C. Grant, and G. Ludbrook, *Cortical entropy changes with general anaethesia: Theory and experiment*, Physiol. Meas. 25 (2004), pp. 921–934.

[25] A. Stuart and J.K. Ord, *Kendall's Advanced Theory of Statistics: Volume 2 Classical Inference and Relationship*, Oxford University Press, New York, 1991.

[26] D.C. Trost, *Multivariate probability-based detection of drug-induced hepatic signals*, Toxicol. Rev. 25 (2006), pp. 37–54.

[27] D.C. Trost and J.W. Freston, *Vector analysis to detect hepatotoxicity signals in drug development*, Drug Info. J. 42 (2008), pp. 27–34.

[28] G.E. Uhlenbeck and L.S. Ornstein, *On the theory of Brownian motion*, Phys. Rev. 36 (1930), pp. 823–841.

[29] S.Y. van der Werf, K.I. Kaptein, P. de Jonge, J. Spijker, R. de Graaf, and J. Korf, *Major depressive episodes and random mood*, Arch. Gen. Psychiatry 63 (2006), pp. 509–518.

[30] X. Wu and D.Q. Naiman, *P-value simulation for affected sib pair multiple testing*, Hum. Hered. 59 (2005), pp. 190–200.

A. Appendix

We seek a modification of the stationary, mean-reverting OU process (Equation (1)) that has constant mean vector $\boldsymbol{\mu}$, constant covariance matrix $\mathbf{a}$, and time-dependent drift $\boldsymbol{\zeta}(t) = t^{-1/2}\boldsymbol{\zeta}$. This can be achieved at the expense of a time-dependent diffusion matrix $\boldsymbol{\sigma}(t) = t^{-1/4}\boldsymbol{\sigma}$. Specifically, we show that the equation

$$d\mathbf{X}_t = -\boldsymbol{\zeta}(t)(\mathbf{X}_t - \boldsymbol{\mu})\,dt + \boldsymbol{\sigma}(t)\,d\mathbf{B}_t \tag{15}$$

has a stationary solution such that $E[\mathbf{X}_t] = \boldsymbol{\mu}$ and

$$E[(\mathbf{X}_s - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'] = \mathbf{a}e^{-2|\sqrt{t}-\sqrt{s}|\boldsymbol{\zeta}}.$$

Applying Itô's formula [8,11,19] to $e^{2\sqrt{t}\boldsymbol{\zeta}}(\mathbf{X}_t - \boldsymbol{\mu})$ and using Equation (15), we find

$$e^{2\sqrt{t}\boldsymbol{\zeta}}(\mathbf{X}_t - \boldsymbol{\mu}) - e^{2\sqrt{t}\boldsymbol{\zeta}}(\mathbf{X}_s - \boldsymbol{\mu}) = \int_s^t e^{2\sqrt{u}\boldsymbol{\zeta}}\boldsymbol{\sigma}\frac{1}{u^{1/4}}\,d\mathbf{B}_u \tag{16}$$

and it follows that

$$\mathbf{X}_t = \boldsymbol{\mu} + e^{-2(\sqrt{t}-\sqrt{s})}\left[\mathbf{X}_t - \boldsymbol{\mu} + \int_s^t u^{-1/4}e^{2\sqrt{u}\zeta}\boldsymbol{\sigma}\,d\mathbf{B}_u\right]. \tag{17}$$

Note that the stochastic integral in Equation (17) can be extended down to 0, because the integral is square integrable on the interval $[0, t]$. It follows immediately that

$$E[\mathbf{X}_t] = \boldsymbol{\mu} + e^{-2\sqrt{t}\zeta}(E[\mathbf{X}_0] - \boldsymbol{\mu})$$

so that $E[\mathbf{X}_t] = \boldsymbol{\mu} \quad \forall t > 0$ if $E[\mathbf{X}_0] = \boldsymbol{\mu}$.

To compute the covariance matrix, note that

$$e^{2\sqrt{s}\zeta}E\left[(\mathbf{X}_s - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'\right]e^{2\sqrt{t}\zeta} = E\left[(e^{2\sqrt{s}\zeta}(\mathbf{X}_s - \boldsymbol{\mu}))(e^{2\sqrt{t}\zeta}(\mathbf{X}_t - \boldsymbol{\mu}))'\right]$$

$$= E\left[(e^{2\sqrt{s}\zeta}(\mathbf{X}_s - \boldsymbol{\mu}))\left(e^{2\sqrt{s}\zeta}(\mathbf{X}_t - \boldsymbol{\mu}) + \int_s^t u^{-1/4}e^{2\sqrt{u}\zeta}\boldsymbol{\sigma}\,d\mathbf{B}_u\right)'\right]$$

$$= e^{2\sqrt{s}\zeta}E[(\mathbf{X}_s - \boldsymbol{\mu})(\mathbf{X}_s - \boldsymbol{\mu})']e^{2\sqrt{s}\zeta}.$$

Therefore, if $0 \le s < t$,

$$E[(\mathbf{X}_s - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'] = E[(\mathbf{X}_s - \boldsymbol{\mu})(\mathbf{X}_s - \boldsymbol{\mu})']e^{-2(\sqrt{t}-\sqrt{s})\zeta}.$$

It remains to compute the one-time covariance, and this hangs on the fact that the standard, OU process (Equation (1)) is stationary with covariance matrix $\mathbf{a}$ if and only if $\mathbf{a}\zeta + \zeta\mathbf{a} = \boldsymbol{\sigma}\boldsymbol{\sigma}'$. Using Equation (16) with $s = 0$, we find

$$E[(e^{2\sqrt{t}\zeta}(\mathbf{X}_t - \boldsymbol{\mu}))(e^{2\sqrt{t}\zeta}(\mathbf{X}_t - \boldsymbol{\mu}))'] = E[(\mathbf{X}_0 - \boldsymbol{\mu})(\mathbf{X}_0 - \boldsymbol{\mu})'] + \int_s^t u^{-1/2}e^{2\sqrt{u}\zeta}\boldsymbol{\sigma}\boldsymbol{\sigma}'e^{2\sqrt{u}\zeta}\,du.$$

Letting

$$\mathbf{R}(t) = E[(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'],$$

we find

$$\mathbf{R}(t) = e^{-2\sqrt{t}\zeta}\left(\mathbf{R}(0) + \int_s^t u^{-1/2}e^{2\sqrt{u}\zeta}\boldsymbol{\sigma}\boldsymbol{\sigma}'e^{2\sqrt{u}\zeta}\,du\right)e^{-2\sqrt{t}\zeta}$$

and it follows that $\mathbf{R}$ satisfies the differential equation

$$\dot{\mathbf{R}}(t) = -\frac{\zeta}{\sqrt{t}}\mathbf{R}(t) - \mathbf{R}(t)\frac{\zeta}{\sqrt{t}} + \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sqrt{t}}.$$

This has a stationary solution $\mathbf{R}(t) \equiv \mathbf{a}$ if and only if

$$0 = -\frac{\zeta\mathbf{a}}{\sqrt{t}} - \frac{\mathbf{a}\zeta}{\sqrt{t}} + \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sqrt{t}},$$

that is, if and only if $\zeta\mathbf{a} + \mathbf{a}\zeta = \boldsymbol{\sigma}\boldsymbol{\sigma}'$.