MODELING NONLINEAR EFFECTS IN LONGITUDINAL SURVIVAL DATA: IMPLICATIONS FOR THE PHYSIOLOGICAL DYNAMICS OF BIOLOGICAL SYSTEMS

A. Kulminski, I. Akushevich, and K. Manton

Duke University, Center for Demographic Studies, 2117 Campus Drive, Box 90408, Durham, NC, 27708-0408, USA

TABLE OF CONTENTS

1. ABSTRACT

Despite the wealth of longitudinal data on the health dynamics of human populations, information on covariates (risk factors) changes in those studies has not been systematically and fully exploited. In this work we use the 46-year follow-up of the Framingham Heart Study to analyze dynamics of these risk factors in survival models that go far beyond the standard linear dynamic formulation. We focus on improving the inferences about the physiology of human aging processes and its plasticity and on modeling state trajectories for individuals considering the effect of nonlinear interactions among covariates. We find that using standard statistical methods to construct models describing the age dependence of health status might give rise to surprising results with highly "diluted" dynamics, but with significantly improved statistical criteria. It is found that problems with the dynamics are a consequence of the intrinsic nonlinear nature of these models. We show that evolution of the risk factors measured in the Framingham study is more complicated for females than for males (i.e., female health status is more sensitive to nonlinear interactions among risk factors). We suggest that this is due to the rapid rate of decline of estrogen production after menopause.

2. INTRODUCTION

It is important to understand how the operation of human physiological systems changes with age. To examine these processes in a human population, one must develop statistical models of those physiological dynamics appropriate to available data and longitudinal observational plans.

Large amounts of epidemiologic and demographic data have been collected to link health changes over time. For example, the original Framingham the study of a cohort of 5,204 persons (aged 29 to 62) began in 1949-1950, with follow-up exams done every two years. The initial purpose of the study was to determine the relation, over time, of disease risk, especially of cardio vascular disease, to several potential (in 1950) risk factors (such as serum cholesterol, blood pressure, smoking, body mass index (BMI), etc). In addition, longitudinal risk factor studies of long standing have been conducted in community studies (in Charleston, South Carolina and Honolulu, Hawaii), in national studies (the MRFIT program and the CHS and ARIC studies), and in international studies (the seven-country study (1) among others). In Framingham new studies were initiated (in 1972) on the offspring of the original cohort members.

Despite this wealth of longitudinal data on the health dynamics of human populations, the information on state variable changes in those studies has not been systematically and fully exploited. As a consequence, some of the conclusions based on those studies proved either to be incorrect or to require serious qualification. For example, the linkage of total cholesterol to cardiovascular morbidity and overall health was found to be too crude. It was determined that at least three components of cholesterol needed to be identified (i.e., high density lipoproteins, the good type of cholesterol that transports lipids away from atherosclerotic plaques; low density lipoproteins, the "bad" type of cholesterol that enters plaques and becomes oxidized that should be reduced; and triglycerides, the component of total cholesterol that is most responsive to dietary factors (carbohydrate intake).

Although diets emphasizing low fat consumption were strongly promoted for a while, it became clear that there were important differences in the types of fats consumed, how they were metabolized, and thus how they impacted the risk of specific diseases. High carbohydrate diets ran the risk of elevating triglycerides and, in recent meta analyses, may have been implicated in the current obesity pandemic in the U.S. (2).

Indeed, the lipid profile is more complex than even this three-part decomposition. Other factors, such as

APOE allelic variation and Lipoprotein A levels, may play significant roles, as may C-reactive protein, inflammatory mechanisms, and immunological responses. Unfortunately, these details about lipid profiles and inflammatory proteins have only been recently translated into better measurements in studies. The cruder measures, however, followed over long time periods and across a wide range of ages, may, by using appropriately specified structural models, provide additional and different insights into the age variable biological effects of factors, such as cholesterol, on total health (3).

In addition many experimental and clinical studies have been made of human disease processes – again without systematically exploiting the evidence on the dynamics of the phenomena being studied. Perhaps more importantly, there have been few attempts to link the various types of information in more comprehensive (multi-level) dynamic models of human physiology in natural human populations. Furthermore, in the models that have been considered a linear paradigm was usually employed. A linear paradigm is unlikely to be a satisfactory description of a well-integrated complex physiological system where the effects of some parameters of the systems are dependent on the values of others (i.e., "field" effects).

The models used to analyze the various types of data were often incorrectly specified in terms of describing state dynamics, and thus could not correctly characterize disease risks. This is, in part, a conceptual problem because factors in both types of studies were sought that increased risk and that could be "avoided". Far less emphasis was placed on finding factors that actively reduced risk by increasing the vitality of the organism. Even less attention was paid to the fact that one variable (because of field effects, nonlinearity, or interactions with other variables) could both enhance and reduce the same disease risk simultaneously by operating through different mechanisms or structures. Thus, the conceptual model employed was predominantly one of "independent" risk factor avoidance, rather than of systematic health improvement or global functional regeneration.

Below we use the 46-year follow-up of the Framingham Heart Study to analyze state dynamics in models that go far beyond the standard linear dynamic formulation (4) and to improve the inferences about the physiology and plasticity of human aging processes that can be made from longitudinal studies of human populations. In this paper we concentrate on improving our modeling of state trajectories for individuals. In subsequent papers we will focus on modifying the mortality component of the model to make better predictions for entire populations.

3. DATA

The Framingham Heart Study began with a cohort of 5,209 persons (2,336 males and 2,873 females), aged 29 to 62, recruited in 1950. People were assessed biennially. Our database consists of up to 23 records for each person (i.e., 46 years of follow-up). For each wave of measurement, the biological measures we had available for each record were: gender, age, sex, diastolic blood pressure, systolic blood pressure, serum cholesterol, vital capacity index, hemoglobin (or hematocrit), cigarette consumption, body mass index, blood glucose, ventricular (heart) rate, and left ventricular hypertrophy.

We used this longitudinal data to analyze the dynamics of state variables and determine how they relate to the health status of persons as it changed over 46 years, from ages 30-62 (at study start) to ages 76-108 (at our "end" of follow-up). The individual risk factors are:

Age (x_1). Statistically, age (number of years) is one of the most important risk factors describing physiological activity and reflecting the process of senescence; it is, however, not directly informative about specific biological mechanisms. Some researchers have attempted to interpret age as a process by making it a nonlinear function of chronological time (a Weibull or Gompertz curve) that had an interpretation as a specific type of human "failure" process (5). Sometimes such functions were interpreted in terms of macro-molecular thermo dynamics (e.g., denaturation or unfolding of protein molecules under thermal stress (6)). We assume that age and calendar time are equivalent for making the coefficients of our model time-dependent.

Pulse pressure (x_2). Rather than dealing directly with systolic blood pressure, we consider the difference between diastolic and systolic blood pressures, which is called pulse pressure and is measured in millimeters of mercury (mmHg). Pulse pressure is less strongly correlated to diastolic blood pressure than systolic blood pressure. Its increase is a major risk factor for stroke and its decrease may reflect loss of heart pump capacity with age.

Diastolic blood pressure (x_3). Diastolic blood pressure (mmHg) has a tendency to increase with age— possibly due to the degradation of arterial elasticity and hemostatic changes. It increases the risk of stroke, atherogenesis, and renal damage.

Body mass index (x_4). Body mass index ($QI = 10 \times weight / height$; kg/m) accounts for the health risk of obesity. Low BMI may be an indicator of caloric restriction and has been found (in literature) to be an indicator of enhanced longevity and to lower disease risks in animal models (7). In humans, the relation of BMI to body mass is likely complex and multi-dimensional. Its effects may be better identified by examining the change of its covariance (or other higher order cross-moments) with other risk factors (e.g., high blood glucose, cholesterol, blood pressure).

Serum cholesterol (x_5). Increased plasma, insulin, and obesity accelerate lipolysis, increasing circulating free fatty acids and triglycerides that cause serum cholesterol (milligram/100 milliliters; $mg/100 ml$) to increase until late middle age, after which it declines. The lipid components of cholesterol have conflicting functions (HDL vs. LDL), and their levels are controlled by hepatic metabolism (e.g.,

HMCOGA enzyme). Thus we must, in analyzing the 46 year follow-up, concentrate on examining the temporal change in the covariances of cholesterol with other risk factors.

Blood glucose (x_6). Elevated blood glucose (milligram/100 milliliters; *mg/100 ml*) is indicative of diabetes, which increases the risk of death because of multiple diseases. Elevated blood glucose is an indicator of insulin resistance and can cause damage to many different tissue types (e.g., retina and kidneys).

Hematocrit (x_7). Hematocrit (%) affects hemostatic and rheological factors in thrombosis, elevates cholesterol, and might be related to the inhibition of the relaxation of coronary artery endothelium by stimulating free radical production. It might be an indicator of diseases related to oxidation processes such as arterial plaque formation (8).

Vital capacity index (x_8). The vital capacity index is calculated as *10 × vital capacity in deciliters / height²* ($10 \times deciL / m^2$). This measure of pulmonary capacity is believed to be one of the biomarker's most directly related to senescence (9) possibly because of the direct exposure of pulmonary tissue to the environment.

Smoking (x_9). Smoking risk is exacerbated by the smoking rate, which is described as the mean number of cigarettes consumed per day. Similar to elevated blood glucose and hematocrit, smoking accelerates multiple age-related processes. In females it may also affect the fertility of both the mother and the female fetus by reducing the number of viable oocytes.

Left Ventricular Hypertrophy (x_{10}) (LVH). LVH prevalence is a consequence of hypertension, loss of cardiac catecholamine receptors, and obesity. With appropriate blood pressure (Ace inhibitor), hormonal control, and physical activity, LVH may now be reversible.

Pulse rate (x_{11}). Pulse rate (beats/minute) reflects physical fitness. Resting heart rate predicts cancer risk independent of physical activity.

These variables define a *J=11* dimensional risk factor space, which is conditioned on male and female status -- also on potent state variables in that many of the above variables are either controlled by, or interact with, various gender-dependent hormonal factors. Gender may also have a genetic effect (e.g., because of gender difference in the cytochrome P450 enzyme system in the liver).

4. MODELS

We constructed a stochastic process from the *J* variables, $x \equiv x_j$, measured on *I* individuals up to *N* times. As the number of observations over time increases, the discrete time stochastic process can be better approximated by a continuous-time, continuous-state

process (10). This is relevant because, in the stochastic process to be constructed, we will need information on the slopes of state trajectories. One issue is: how many measurement points are needed to approximate those durations. The model must describe the movement of a person *i* (*i=1,2,...I*) in a *J* dimensional state space. In the process, the evolution of the continuous time, vector state process, *x(t)*, is governed by a *J* dimensional stochastic differential equation,

$$dx_i(t) = u(x_i,t)dt + \sigma(x_i,t)d\xi(t).$$ [1]

The probability of death at a point in the *J* dimensional space is,

$$dP(x_i,t) = -\mu(x_i,t)P(x_i,t)dt,$$ [2]

where $\xi(x_i,t)$ is a Gaussian vector process; $u(x_i,t)$ and $\sigma(x_i,t)$ are drift and diffusion coefficients; $P(x_i,t)$ is the probability of surviving to time *t* for a person *i* and $\mu(x_i,t)$ is the mortality rate.

Equation [1] shows that the time changes in *J=11* risk factors of each person (x_i) are a function of deterministic (first term) and random (second term) components. The second equation reflects the survival probability, decaying in time exponentially with the mortality rate μ.

The random walk and mortality equations for individuals can be combined to generate the Fokker-Planck-Kolmogorov (FPK) (11) equation for temporal changes in the multivariate state variable density function *f(x)* due to drift and diffusion. It was generalized to represent state dependent mortality (12),

$$\frac{\partial f(x)}{\partial t} = -\sum_j u_j(x,t)\frac{\partial f(x)}{\partial x_j} - f(x)\sum_j \frac{\partial u_j(x,t)}{\partial x_j} \text{(deterministic change)}$$

$$+\sum_{j,j'} \frac{\partial^2}{\partial x_j \partial x_{j'}}\left[\sigma_{jj'}(x,t)f(x)\right] \quad \text{(diffusion)} \quad [3]$$

$$-\mu(x,t)f(x). \quad \text{(mortality)}$$

Rigorous analysis requires numerical integration of equation [3]. The numerical problem can be simplified by making two assumptions: a.) risk factors are measured at fixed times and are linearly related, and b.) the data are normally distributed. Under these assumptions and following the random walk specification (13), we can write the linear dynamic equations for *J* variables as,

$$x_{t+1} = u + Rx_t + \varepsilon,$$ [4]

and for mortality as a quadratic function of *x*,

$$\mu = \mu_0 + q^T x_t + x_t^T Q x_t.$$ [5]

Under these conditions (linear dynamics, quadratic mortality), this multivariate Gaussian state variable distribution remains Gaussian over time–even with mortality selection. Scalar μ_0, vector *q*, and matrix *Q* are estimated by maximum likelihood. Equation [5] may be generalized by multiplying each coefficient by an age functional (such as the Gompertz, $e^{\theta Age}$) to reflect age-

correlated latent variables increasing mortality at θ % per year.

Equation [4] describes how changes in each of the *J* state variables are related to linear superposition on their prior values. Linear models are widely used because of their simplicity and ease of interpretation. However, they are often too simple to correctly predict an organism's future health status because they may not capture essential features of underlying, linked processes. Due to interactions among the components of a system, realistic multi-parametric biological, social, and other complex systems are often nonlinear. Their dynamics can also be affected by unobserved processes reflecting nonlinear interactions and/or correlation among the system components (i.e., by "field" effects) so that standard linear models cannot fully describe data originated from these systems. For example, it was recently shown that the Framingham data do not support a linear paradigm (14). In this case, nonlinear analysis is essential. Application of nonlinear methods to the analysis of state dynamics for individuals is the primary goal of this paper. Extensions to nonlinear models, including mortality, are beyond the scope of this paper. However, given large differences in the structural physiological dynamic models we found below, we believe that the additional effects captured as nonlinearities may be larger than mortality effects within measurement intervals. Mortality rates will, however, be essential for forecasting population health.

Nonlinear time series analysis has recently gained attention (15). Any model that is not linear with respect to system variables (i.e., when the superposition principle is violated) is considered to be nonlinear in variables (versus parameters), or simply nonlinear. Comprehensive nonlinear models consider a.) the degree of nonlinearity, and b.) memory effects (i.e., how important is the system "prehistory" to assess future health). A general nonlinear model is,

$$x_j^{t+1} = u_j + f\left(x_j^t,...,x_j^{t-m}, \ \alpha_j\right) + \varepsilon_j, \qquad [6]$$

where *m* is depth of memory, α_j is a vector of parameters, and $f(\cdot)$ is a nonlinear function of the variables and perhaps of the parameters. Such nonlinear models are widely used to model and predict nonlinear time series and belong to the class of models known as "neural networks" (16).

To be useful and interpretable, such a general system needs to be substantially restricted. Here we present and study the simplest type of nonlinear model. In addition to the assumption of the *linear* superposition of all risk factors on future health, we assume that each risk factor can "interfere" with others inducing additional changes (positive or negative, depending on goodness of fit and the substantial importance of their interaction) in future health. Interference of two variables can be represented by their scalar product. We also start by assuming that only prior values of risk factors affect health. This requires re-specifying the system of 11 dynamic equations as,

$$x_j^{t+1} = u_j + R_{jj'}x_{j'}^t + C_{jj'j''}x_{j'}^t x_{j''}^t + \varepsilon_j, \qquad [7]$$

The problem now is twofold: on one hand, any nonlinear model for dynamics of risk factors violates our assumption of the normal distribution of the data; on the other hand, we have to study whether such a generalization is necessary and which terms are essential. A question that should also be answered is whether the nonlinear terms are new state variables reflecting the so-called "field" effect (due to possible unobserved processes) in the generalized linear model prescribed below. If so, our ability to identify state variables from measurements will increase, compared to using a latent variable procedure, such as principal component, which extracts only information on latent variables from second-order moments (covariance matrix). To justify applicability of nonlinear models for estimating survival functions, we note that adjustment for non-normality can be accomplished by using Monte Carlo simulation of individual trajectories and fitting simulated distributions at each time interval (17).

We generalize the problem by treating products of original state variables as new variables, $y \equiv y_j = x_j x_{j'}$ to produce the system of dynamic equations,

$$x_{t+1} = u^x + R^x x_t + C^x y_t + \varepsilon^x, \qquad [8]$$
$$y_{t+1} = u^y + R^y y_t + \varepsilon^y.$$

This can be written in a more compact form by defining a new vector *X*:

$$X^T = \begin{bmatrix} x & y \end{bmatrix}, \qquad [9]$$

which consist of 11 original state variables and 66 new ones:

$$X_{t+1} = \tilde{u} + \tilde{R}X_t + \tilde{\varepsilon}, \qquad [10]$$

Unlike system [7], system [10] is linear with respect to the expanded set of variables. Systems [4], [7] and [10] are studied below.

5. DYNAMICS OF LINEAR AND NONLINEAR MODELS

In matrix form a linear system of ordinary differential equations (ODE) is written,

$$\frac{dy}{dt} = Ay, \qquad [11]$$

where *y* is a vector of state variables and *A* is a matrix of coefficients, independent of *y*. This system has the solution

$$\bar{y} = y_0 \exp(\Lambda t), \qquad [12]$$

where y_0 is a vector of initial conditions and Λ is a diagonal matrix of eigenvalues λ of matrix *A*. Depending on whether eigenvalues are real, complex, or purely imaginary, there are three different types of system evolution: exponential, exponential with oscillations, and oscillatory. The sign of the real part of system eigenvalues determines whether there is exponential growth or decay. If the real part of at least one eigenvalue is positive, the stationary solution of the system (i.e., when $A\bar{y} = 0$) is

unstable. That is, any small perturbation of this solution gives rise to exponential divergence. In contrast, when all the real parts in Λ are negative, the solution is stable and phase trajectories eventually converge to this solution – irrespective of the starting point in the state space (i.e., irrespective of initial conditions y_0). Classically, the loss of function with age in humans has been viewed as monotonic and as leading increasingly to an exponential increase in mortality. In the system described above, short term changes (during transient) need not be monotonic (i.e., age reversals of the trajectories of the measured and state variables are possible).

An example of such behavior in biology is given by the well-known logistic equation:

$$\frac{dN}{dt} = (b - \mu)N,$$

where N is a number of members in a population; b is a birth rate, and μ is a mortality rate. Note that b and μ are constants and do not depend on N.

In nonlinear models, things are more complex. In general, a nonlinear system of ODE is,

$$\frac{dy}{dt} = f(y, \alpha), \qquad [13]$$

where α is a vector of parameters and $f(\cdot)$ is a vector of nonlinear functions of state variables and parameters. Equation [13] can be rewritten in the same form as [11], but with coefficients dependent on y.

Since $f(\cdot)$ is nonlinear, system [13] can have more than one solution. Hence, the phase space can be composed of several basins of attraction or repulsion. This could correspond to multiple modes of failure or to different disease processes as causes of death. Starting from different initial conditions, the phase trajectory can settle onto one or another solution, depending on which basin of attraction/repulsion a particular set of initial conditions belongs to (e.g., a person genetically predisposed to diabetes is likely to die of circulatory disease). This gives rise to qualitatively different results during the transient (and after the transient has died out) constituting the first qualitative difference between linear and nonlinear models. A second difference is that the rate of convergence to (or divergence from) a certain solution can no longer be ascribed to the eigenvalues of any matrix composed from the coefficients of the original nonlinear system [13]. System behavior can, however, be predicted locally (i.e., in the vicinity of a solution).

Suppose we have a stationary solution that is a fixed point of [13], for example., $dy/dt = 0$. Vectors of solutions are given by $f(\bar{y}, \alpha) = 0$. For each fixed point in phase space $\bar{y}$, using only the first two terms, we can carry out a Taylor series expansion of $f(y, \alpha)$ in its neighborhood.

$$f(y, \alpha) = f(\bar{y}, \alpha) + \left. \frac{df}{dy} \right|_{y=\bar{y}} (y - \bar{y}). \qquad [14]$$

Since by definition $f(\bar{y}, \alpha) = 0$ at each fixed point, we obtain,

$$\frac{dz}{dt} = Bz, \qquad [15]$$

where $z = y - \bar{y}$ and $B = \left. \frac{df}{dy} \right|_{y=\bar{y}}$ is a Jacobian matrix that is independent of the variables. Hence, [15] is a system of linear ODE, like [11], that is obtained from the original nonlinear system [13] linearizing it in the neighborhood of solution $\bar{y}$. The behavior of the phase trajectories in the vicinity of this solution is governed by the eigenvalues of B, which are called Lyapunov exponents λ_j (where j is number of equations, or degrees of freedom). They show whether the trajectory converges to (or diverges from) this solution, as was described for eigenvalues of the linear problem. The difference between this and the linear case is that [15] predicts only short term behavior, while eigenvalues in Λ in equation [12] govern long term evolution.

The time of settlement of the trajectory onto the solution, the transient, can be "large." This is the case for systems with small "dissipation" and, as a limiting case, for Hamiltonian systems (i.e., systems attempting to conserve, for instance, energy). In the latter systems, the sum of all the eigenvalues (i.e., trace of matrix A or B), ideally, is zero, which conserves the volume of the state space over time. A particular case is a system with neutral stability when all $\mathrm{Re}\,\lambda = 0$. Then, if $\mathrm{Im}\,\lambda = 0$, the system remains on the initial conditions irrespective of time. When $\mathrm{Im}\,\lambda \neq 0$, the system evolves periodically, with amplitude of oscillations given by the initial conditions. In the former case, the trace is close to zero, but is negative. Hence, each set of initial conditions in Hamiltonian systems has its own trajectory that is not attracted to another set. In systems with small dissipation all trajectories are attracted to each other, converging to a solution, but the transient time to reach it can be large; so the final state cannot be reached during a reasonable time of integration.

A famous example of such systems in population biology is the so-called predator-prey model constructed by Lotka and Volterra:

$$\frac{dN_1}{dt} = N_1(r_1 - b_1 N_2),$$

$$\frac{dN_2}{dt} = N_2(-r_2 + b_2 N_1).$$

The constant b_1 describes the death rate of prey eaten by predators; constant b_2 measures the skill of the predator to catch prey; r_1 denotes the rate of prey increase in the absence of predators; and r_2 denotes the natural (without prey) mortality rate of the predators.

Apart from the trivial solution ($N_1 = N_2 = 0$), this system has a nontrivial ($N_{1,2} = r_{2,1}/b_{2,1}$) solution that has neutral stability. For this solution, all $\mathrm{Re}\,\lambda = 0$, while $\mathrm{Im}\,\lambda$ is purely imaginary, and $\mathrm{Im}\,\lambda = \pm i\sqrt{r_1 r_2}$. Hence, this model

shows oscillatory behavior being an image of a family of closed curves in phase space in which each curve corresponds to a different initial condition.

This discussion is also valid for a system of difference equations (i.e., a system similar to equation [7]) for which corresponding critical eigenvalues are equal–not to 0–but to 1 because $dy/dt \approx (y_{t+\Delta t} - y_t)/\Delta t$ and, assuming $\Delta t = 1$, the matrix of eigenvalues Λ' for the system of difference equations is $\Lambda' = \Lambda + I$, where I is a unit matrix.

If there is no stationary solution, we can still characterize divergence (or convergence) between two different trajectories, which start from initially near points, estimating Lyapunov exponents numerically – even if the trajectories do not settle onto any attractive solution (attractor) (18). The set (spectrum) of Lyapunov exponents can be defined following Wolf et al., (19) in terms of the length of the ellipsoidal principal axis $p_j(t)$ as a limit in the long term evolution of the "spherical" initial volume in J-dimensional phase space as,

$$\lambda_j = \lim_{t \to \infty} \frac{1}{t} \log_2 \frac{p_j(t)}{p_j(0)}, \qquad [16]$$

where $p_j(0)$ is initial length.

While the existence of limit [16] has been questioned (20), the orbital divergence of any data set can be quantified. This divergence defines how the range of physiological diversity can expand with age (without the selection of the mortality effect). Even if this limit does not exist for the underlying system or if it cannot be approached due to predicting only short term behavior or to having only finite amounts of data, the Lyapunov exponent estimates still provide a useful characterization of data (19).

Unlike the linear problem, to find Lyapunov exponents for a nonlinear model, we need the derivatives of the nonlinear functions $f(y, \alpha)$. Unfortunately, difference equations do not provide this information. This is the case of our prediction nonlinear model in the form of equation [7]. This system alone is not sufficient to find Lyapunov exponents. It gives us a function connecting consecutive data points, but says nothing about the derivatives of this function at each point. To avoid this difficulty, we can either use different fitting techniques (18) or convert [7] to a system of ODEs,

$$\frac{dx_j}{dt} \approx \frac{x_j^{t+1} - x_j^t}{\Delta t} = \frac{1}{\Delta t}\left[u_j + (R_{jj'} - I)x_{j'}^t + C_{jj'j''}x_{j'}^t x_{j''}^t + \varepsilon_j \right], \quad [17]$$

where $\Delta t = t_{t+1} - t_t$ and I is a unit matrix.

6. LINEAR RISK FACTOR DYNAMIC RESULTS

One way to estimate coefficients of the drift vector (u) and regression matrices (R, C) is to use ordinary least squares (OLS). While there are more elaborate techniques (21, 22), employment of OLS in our studies is justified because we are using the discrete time

approximation of the continuous-time stochastic process. The improved efficiency of more sophisticated estimation procedures can give unsatisfactory or misleading results in our approximate model.

We find (23, 24) linear dynamics crucially sensitive to methods for filling in missing data points. The most satisfactory results are obtained when filling in missing data by the Monte Carlo method (MC). Such procedures are consistent with the Missing Information Principle or Paradigm (25). A comparison of the dynamics using the MC method with different initial seeds, which randomize the magnitudes of the filled data, gives essentially the same result in that the dynamics do not significantly change. Hence, the percentage of missing data in the Framingham Study approaches a critical level at which special caution should be paid when choosing the method of imputation.

Coefficients of equation [4] for the 46-year follow-up are in Table 1 for males and females. They are estimated over all ages, starting from ages 28 to 62 at the beginning of the study. Persistence for Framingham risk factors varies little by gender. Drift and regression coefficients correlate well both for this model and for the model for 11 risk factors estimated for the 34-year follow-up (4). They are also consistent with the corresponding coefficients for an 8-risk factor model for the 20-year data (26).

Figure 1 depicts the predicted linear dynamics of 10 risk factors with age for males and females. Pulse pressure first declines and then, after age 35, rises (until age 100) for males and females. The male mean is higher before age 47. It crosses the female mean at this age, decreases, and then rises slightly, diverging more slowly at advanced ages. Diastolic blood pressure increases by age 40 (48 in observed data) and then declines - regardless of gender. The age dependence of BMI reach a maximum at age 55 for males and 60 for females (the maximum for males is 6% higher than for females), after which they both decline. Cholesterol rises until age 42 for both males and females, and then declines. The rate of decline is larger for males. Male blood glucose first decreases (by age 35) and then increases (until age 100). At all ages it remains higher than for females. Postmenopausal female hematocrit is flat, but slightly decreasing. Male hematocrit reaches a maximum of 48% at age 40 (declining to 43% by age 95). The vital capacity index declines to age 100 because of a negative drift coefficient (Table 1). Smoking is projected to end at age 98 for both males and females. Male LVH prevalence is higher at ages 30 through 80. Pulse rates first increase (until age 40) and then decline. The male pulse rate declines slightly faster. The female pulse rate is higher, which compensates for lower hematocrit. Since these dynamics do not take into account mortality, they deviate from the model for the 34-year follow-up at advanced ages when mortality grows rapidly.

7. NONLINEAR MODELS, EXTENDED LINEAR MODELS AND DYNAMIC ANALYSIS RESULTS

In Equation [7] $C \equiv C_{jj'j''}$ is a tensor of rank 3 (11x11x11), with 1331 coefficients. Due to the symmetry
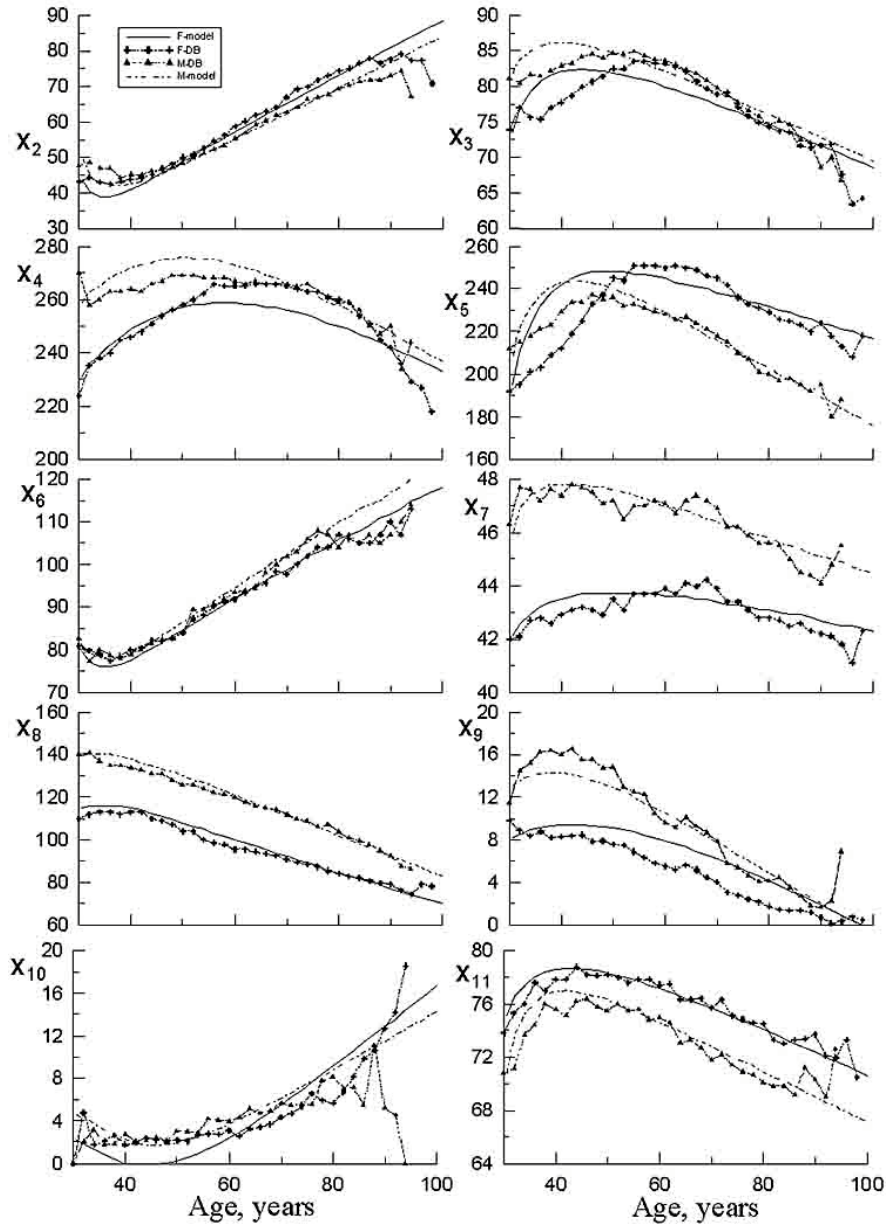
Figure 1. Predicted dynamics (model) and mean values over individuals for given ages in the Framingham data base (DB) of 10 risk factors for basic linear model for females (F) and males (M).

of the problem, with respect to the permutation of the last two indices (j' and j''), the number of coefficients in Equation [7] can be reduced to $J^2((J-1)/2+1)=726$. Hence, even a simple quadratic restriction on the nonlinear terms significantly extends the number of the parameters to be estimated. Fortunately, the Framingham database has enough records for OLS to be used.

In considering a new (nonlinear) model, we need a criterion to compare it with our base (linear) model. One possibility is the Akaike Information Criterion (27),

$$AIC = N\left(\log\frac{2\pi}{N}\sum_{n=1}^{N}(x_n-\hat{x})^2+1\right)+2M \qquad [18]$$

where N is the number of observations and M is the number of free parameters. $\hat{x}$ is the mean of the given x_j over all samples.

Estimating all the coefficients in [7], we obtain a model called a full nonlinear (FN) model. We can also estimate, as in the previous section, coefficients of an extended linear (EL) model [10].

Applying OLS to both systems and averaging results over all age intervals, we find $u=u^x$, $R=R^x$ and $C=C^x$. This can be understood by recognizing that

Table 1. Coefficients u_j and $R_{jj'}$ of the linear regression model [4]: Estimated for males and females followed for 46 years with biennial exams in the Framingham Heart Study

State at time t+1	Gender	u	State at time t $X \times 100$										
			X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
X_2	M	-0.3	24.7	65.9	8.8	-0.2	0.2	1.5	-4.7	-1	1.8	3.2	0.3
	F	-2.3	29.1	62.5	8.1	1	0.5	1.8	-6.5	-1.8	1.4	2.8	0.5
X_3	M	20.5	-10.1	3	66.5	1.1	0.3	-0.1	8.5	0.2	-1.3	-0.7	4
	F	20.7	-10.7	2.8	65.9	1.4	0.7	-0.5	9	-1.2	-1.7	0	4.5
X_4	M	29.8	-9.1	-3	0.7	89.6	0.2	-0.4	12.2	0.2	-2	-1.1	-2.4
	F	21.4	-6.9	-1.9	3.1	92.2	-0.4	-0.4	9.5	0.1	-5.2	-2.8	-1.6
X_5	M	80.7	-33.1	-0.6	-1.1	-0.9	72.1	-3.1	6.5	-0.3	-0.3	-4	6
	F	67.2	-15.8	0.7	11	0.4	71.9	-1.9	2.9	-3	-1.9	-6.3	4.4
X_6	M	17.2	28.6	14	-2.2	6.4	-1.8	49	-20.3	-1	-3.4	-1	9.9
	F	22	28	9.5	1	3.8	-0.8	49.7	-24	-0.2	6.9	3	6.1
X_7	M	21.6	-1.9	-0.5	1.2	0.2	0	0	53.3	-0.4	1.2	-0.5	0.9
	F	19.1	0.1	-0.6	1.8	0.4	0	0.1	48.9	-0.4	3.4	-0.4	1.1
X_8	M	47.1	-21.5	-2.2	2.1	-0.1	0	-0.4	-7.4	77.4	-3.7	-1	-6.5
	F	46.4	-18.5	-3.2	-0.8	-1	-0.2	0.2	-6	74.7	-2.8	-1	-4.3
X_9	M	3.6	-5.3	0	-2.6	-0.4	-0.1	0	6.7	-0.7	86	0.5	1.9
	F	1.4	-2.4	0	-0.6	-0.2	0.1	0	2.8	-0.2	92.2	0.2	0.3
X_{10}	M	-2.2	1	7	2.6	-0.3	-0.3	0.1	-2.1	0.1	0	77.8	-1.3
	F	2.4	1.9	4.4	1.8	-0.4	-0.5	0.6	-6.7	-0.9	2.2	83	-1.3
X_{11}	M	28.4	-6.9	0.4	10.9	-0.7	0.7	0.8	7.9	-3.4	5.1	-0.6	54.2
	F	32.4	-6.5	0.1	8.3	-0.3	0.5	0.5	4.8	-2.9	4	-1.1	53.6

equation [10] is composed of two subsystems (see [8]). Residuals for all x_j in both models will be the same. Since the number of parameters estimated and the number of observations are identical for both models, AICs will be the same. The dynamics of x_j, however, will be different because the EL model captures dynamics of *y*, and the calculated value of *y* is used in the equation for *x* in the EL model. In other words, the EL model takes into account possible "field" effects that originated from unobserved processes, interactions, and/or correlations among the risk factors.

Figure 2 shows the dynamics of 10 risk factors with age for males and females using both FN and EL models. Comparing dynamics described by the linear (Figure 1) and EL (Figure 2) models, we conclude that the "field" effects might be more important at "young" and advanced ages when the homeostasis of the organism is tightly controlled.

While AICs for 9 risk factors (excluding x_8) improved for the FN and EL models, in comparison with the base linear model (see Table 2), "propagation" is more reasonable for the EL model-for all risk factors (Figure 2). This has two consequences: First, analyzing the standard statistical criteria of nonlinear models may not be sufficient to determine which of them is the "best" model. To choose a "best fitting" model analysis of "predicted" behavior is important. Second, due to inherent features of nonlinear models, averaging drift and regression coefficients over all ages can lead to unsatisfactory predictions of risk factor behavior (see Section 5). To show this, we estimated the Lyapunov exponents. Use of equation [17] for our problem can be justified because the dynamics predicted by [7] and [17] are nearly identical numerically.

Four representative Lyapunov exponents are plotted for males and females in Figure 3. This gives surprisingly good results by explaining the significant deviation in the dynamics of x_9 and x_{10}. Eight Lyapunov exponents are negative, regardless of age and gender (as shown for λ_2 and λ_5). Two are positive at "young" ages. As long as they are positive, our model will be sensitive to initial conditions - even when they belong to the same basin of attraction (or repulsion, if Lyapunov exponents are positive). If the underlying model allows long term dynamics, this gives rise to complex behavior. Because our model can predict only short-term evolution, strong dependence on the initial conditions does not lead to complex, possibly chaotic, behavior, but it is the reason that our predictions become imprecise (with regard to the dynamics of smoking cigarettes and LVH risk). To understand this, recall that we do not know exact (or true) values of the initial conditions for which equation [7] was designed (i.e., drift vector and regression matrix were estimated) because this is the model of a stochastic process. Starting from even slightly different initial conditions causes exponential divergence (at the rate given by the
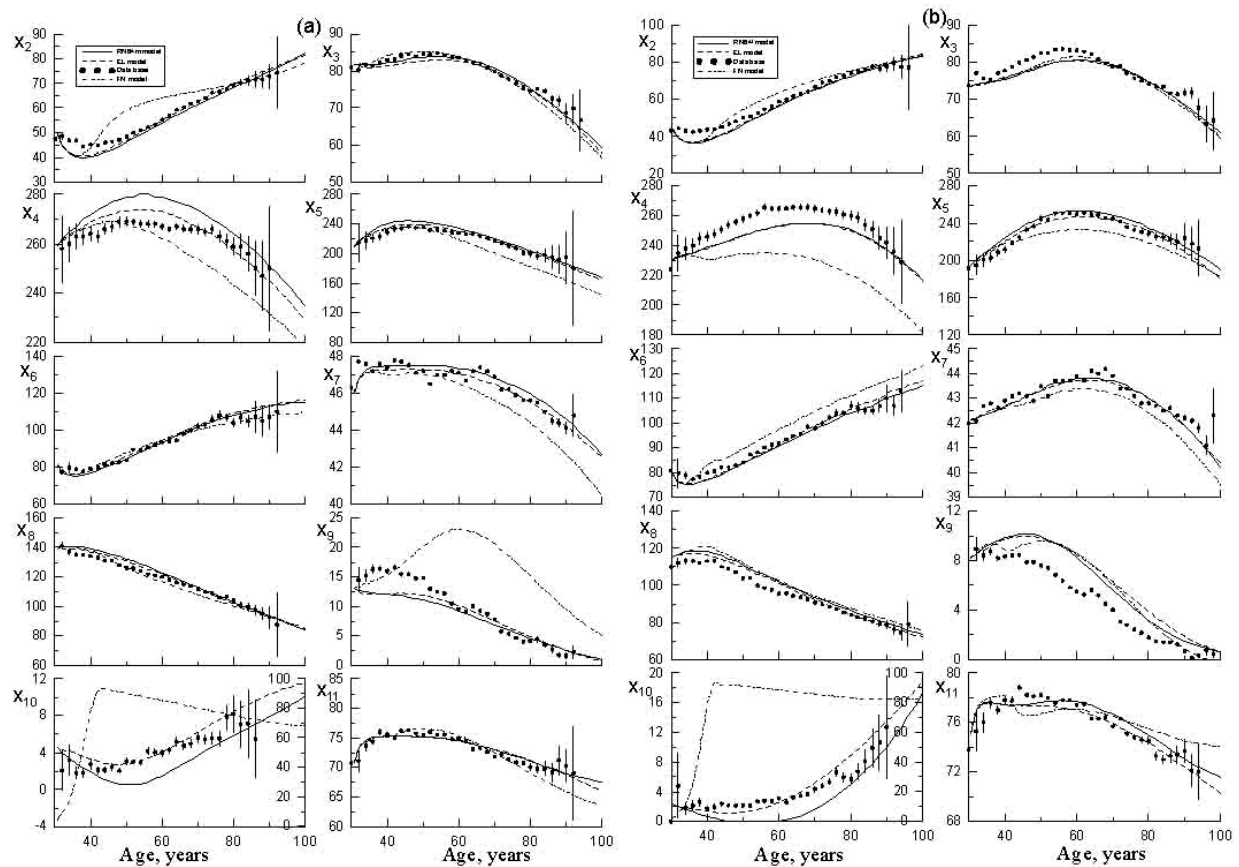
Figure 2. Dynamics of 10 risk factors for the FN, RN64 and EL models for (a) males and (b) females. Note that propagation for LVH for the FN models is linked with the right axis.
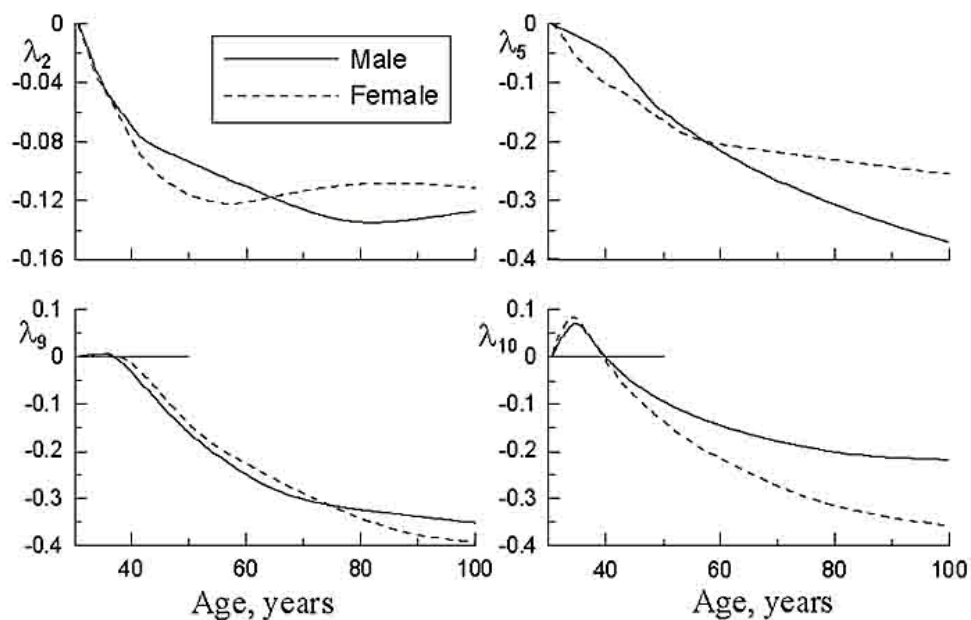

Figure 3. Estimations of the four Lyapunov exponents for males (solid line) and females (dashed line).

Table 2. The difference between AICs of the given nonlinear and the base linear models

Model	Gender	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
FN	M	-107	-402	-86.9	-136	-166	-77.7	11.9	-2130	-1100	-117
	F	-0.111	-136	-81.3	-220	1.76	-3.28	-102	-1240	-2310	-51
RN64	M	-110	-396	-72.1	-138	-156	-80.5	7.91	-975	-842	-118
	F	-194	-879	-160	-517	-346	-153	-110	-928	-1420	-151

Table 3. Coefficients u_j and $R_{jj'}$ of the full nonlinear regression model [7]: Estimated for males and females followed for 46 years with biennial exams in the Framingham Heart Study

| State at time t+1 | Gender | u | State at time t $X \times 100$ | | | | | | | | | | |
			X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
X_2	M	-19.4	52.0	85.0	19.6	1.4	-2.4	-5.2	25.5	10.2	2.0	7.7	-23.0
	F	-7.2	62.9	64.4	24.7	0.8	-1.5	8.6	-73.3	-2.2	0.8	12.1	5.1
X_3	M	-12.8	60.1	8.6	60.0	11.8	0.9	9.0	-2.4	-2.3	-6.6	1.8	7.3
	F	-28.5	73.5	26.9	61.2	6.4	0.9	1.7	29.6	6.1	-3.2	-4.8	19.0
X_4	M	-18.3	29.5	21.7	-15.9	112.3	0.3	-1.7	21.2	4.5	-27.1	-6.4	9.6
	F	-36.2	58.6	10.7	21.3	86.7	-0.4	27.9	92.7	-1.3	-8.1	-24.2	15.0
X_5	M	-138.0	160.8	-33.7	2.8	35.2	97.1	17.9	263.7	14.7	44.2	-18.9	48.9
	F	-72.1	213.7	10.4	-1.4	6.7	109.1	12.2	-10.9	16.0	6.6	1.5	41.9
X_6	M	33.5	29.8	-12.6	103.1	-13.5	6.7	49.3	-207.2	9.3	3.4	16.5	17.0
	F	67.9	68.7	-8.6	9.9	-9.8	2.6	-2.2	-148.3	7.4	-27.1	33.0	23.3
X_7	M	19.4	-3.8	1.9	4.5	2.6	2.0	-0.9	32.8	1.0	1.0	-4.2	0.8
	F	25.4	-1.1	-0.6	4.6	0.6	-0.2	0.3	13.9	-1.3	2.4	1.1	4.3
X_8	M	-0.8	38.2	-11.7	-12.4	6.8	-5.9	-4.8	60.9	115.5	-24.2	11.9	-4.1
	F	-8.2	38.7	-14.9	-15.2	5.6	-0.7	-1.7	27.1	123.9	6.4	12.5	16.5
X_9	M	-18.1	51.1	-4.0	-1.9	3.5	0.6	-1.2	-17.3	10.6	147.7	0.9	-9.1
	F	1.4	3.5	-1.0	3.0	0.6	0.0	-0.5	-8.8	-2.6	116.5	-3.7	-1.5
X_{10}	M	27.2	0.8	-30.2	-46.4	4.1	0.0	10.3	-63.9	-4.7	15.6	312.8	11.5
	F	-3.8	-1.6	23.9	-19.2	2.2	-1.3	-4.2	-10.7	11.8	-7.2	310.5	5.0
X_{11}	M	13.8	18.9	7.3	31.3	1.2	-2.8	1.1	-5.2	0.8	7.4	8.4	49.3
	F	32.3	-14.6	15.8	25.4	-4.5	-1.3	0.7	18.8	-9.6	17.3	5.4	48.7

positive Lyapunov exponent) of the current phase trajectory from the "true" one (i.e., the trajectory that started from exact or true initial conditions).

The Lyapunov exponents do not necessarily govern the dynamics of original state variables, but rather their superposition, since they are eigenvalues of a *linearized* system near a solution. However, we found a good correlation of certain Lyapunov exponents with risk factors, as we see in Figure 3. We concluded that Lyapunov exponents are negative for all risk factors except smoking cigarettes and LVH. For x_9 and x_{10} they are positive in the 30-to-37 year age range (males) and the 30-to-38 year range (females) for smokersand in the 30-to-40 age range (both males and females) for LVH. Moreover, the range in which λ_{10} is positive is correlated with domains of fast divergence of the phase trajectories for x_{10} from the "true" one (compare Figure 2 and Figure 3). Such findings can be understood if quadratic (i.e., diagonal) coefficients in R dominate an interaction (off diagonal) and nonlinear coefficients. For the FN model, the clearest

dominance is seen for diagonal elements for LVH (males and females [Table 3]). For that reason λ_{10} is better correlated with LVH than with λ_9 for x_9. The rate of divergence of the dynamics of LVH and smoking cigarettes is related to the magnitude of positive Lyapunov exponents λ_9 and λ_{10}. Since λ_9 is small and positive and the diagonal element $R_{9,9}$ does not clearly dominate, we observe sharper growth for the risk of LVH than for smoking.

As we add variables (or their combinations) into a model, we want to assess the importance of each variable in equations for both prediction and propagation. We want to know whether certain variables (or combinations of them) can be ignored, little affecting the model; whether the contribution of a variable is real and not random variation; and whether the correlation of variables can result in the contribution of a variable being enhanced or diluted by including other terms, etc. To do this, there are several techniques. Most frequently used are forward selection and backward elimination procedures. In forward

Table 4. Standard deviation sigma

		X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
F	Linear	11.59	8.12	18.75	31.91	23.27	2.964	11.66	3.362	11.75	10.47
	FN	11.55	8.032	18.69	31.69	23.15	2.955	11.63	3.305	11.43	10.44
	RN64	11.55	8.033	18.69	31.69	23.15	2.955	11.63	3.325	11.55	10.44
M	Linear	10.43	7.943	16.87	28.59	25.79	3.054	12.72	6.444	13.07	10.43
	FN	10.4	7.879	16.82	28.48	25.67	3.044	12.69	6.226	12.83	10.39
	RN64	10.4	7.88	16.82	28.48	25.68	3.044	12.69	6.337	12.88	10.39

selection, variables are successively added depending on a criterion and are retained if the fit is significantly improved. In backward elimination, variables are removed if their combination is not significant. The aim is to find a "best fit" model.

A more comprehensive approach is to examine all possible models for a given set of variables. This is feasible only for small models. This is not our case because, if we consider only nonlinear terms in equation [7], there are $2^{66 \times J} = 2^{726}$ models. While the latter method seems unreasonable due to computational restrictions, the former two strategies must be used with caution going from linear to nonlinear models.

We combined nonlinear analysis with forward selection or backward elimination techniques. The nonlinear analysis provides substantial background to understand the effects of the interactions.

From our nonlinear analysis, we determined that the most crucial variables in our FN model are x_9 and x_{10}, which dilute predicted dynamics. To simplify the problem, we considered the integral effect of smoking cigarettes and having LVH in combination with each health state risk factor in J-dimensional state space. We did not analyze which risk factor is most significantly enhanced or diluted by combining the effect of x_9 or x_{10} with each of the other variables. Since LVH associated with Lyapunov exponent λ_{10}, the most significant coefficient giving rise to divergence for LVH, is $C_{j,10,10}$. It is reasonable to assume that $C_{j,9,9}$ gives a considerable contribution to the dynamics of cigarette smoking. To check this, we chose the FN model as a reference and used backward elimination of the contribution of each nonlinear term $x_9 x_j$ and $x_{10} x_j$. This confirmed our assumption that a significant improvement of predicted dynamics for LVH is achieved when we set $C_{j,10,10} = 0$. For smoking, none of the individual contributions from $C_{j,j',9}$ is so dramatic. Most significantly, the dynamics of smoking cigarettes changes when $C_{j,9,9} = 0$. Further contribution to the dynamics is achieved when $C_{j,9,1} = 0$. By setting $C_{j,9,9}$ and $C_{j,10,10}$ to zero and evaluating the remaining coefficients, we obtain reduced nonlinear models that are free of divergent dynamics (RN64m for males and RN64f for females [Figure 2]). Dropping these coefficients improves the dynamics of other risk factors.

Comparing linear and nonlinear models, the distribution of the residuals is most changed for cigarettes

and LVH. Figure 4 shows this distribution for all risk factors (for the RN64f model). We also show, by a dashed line, residuals for cigarettes and LVH for the linear model. For all nonlinear models, the standard deviation is smaller (Table 4). Thus, the result of our nonlinear analysis suggests how systems can be separated into dynamic sub systems.

Models RN64m and RN64f give surprising results. For males AICs for x_2, x_5, x_7, x_8 and x_{11} are better than corresponding AICs for the EL model; AICs for x_3, x_4 and x_6 are slightly worse. Elimination of $C_{j,9,9}$ and $C_{j,10,10}$, makes AICs for x_9 and x_{10} worse (Table 2). For females, AICs for x_9 and x_{10} are also worse than for the EL model. For the remaining risk factors, AICs are significantly better (Table 2). Because the number of observations for these models and for the EL model is the same, while the number of the estimated parameters is 76 instead of 78 (only slightly less), we conclude that nonlinear models better fit the mean distribution of risk factors over age for females. Hence, we speculate that females' health status is more sensitive to nonlinear interactions than that of males. It appears that the different age trajectories of hormonal status for females (here unmeasured) might make their dynamics more complex because of an interaction of estrogens and other sex hormones with a number of physiological systems (e.g., oxidation processes in the arterial endothelium). It may be that these are intrinsic genetic gender differences as well. For example, cigarette smoking may have a greater effect on females because of a certain Cytochrome P450 enzyme that operates to produce more cancer promoting substances in blood from inhaled combustion products of tobacco (28).

8. CONCLUSION

We have shown that considerable additional information on human aging processes can be extracted if one assesses nonlinear effects. However, if we go beyond linear models to describe the age dependence of health status, standard statistical methods might not be sufficient to determine the best nonlinear model. Such models can give rise to surprising behavior as a consequence of their intrinsic nonlinear nature. It is necessary to choose a "best fit" model in nonlinear analysis. Such an analysis of the nonlinear models describing evolution of the risk factors measured in the Framingham study reveals that age-dependent risk factor dynamics for females are more
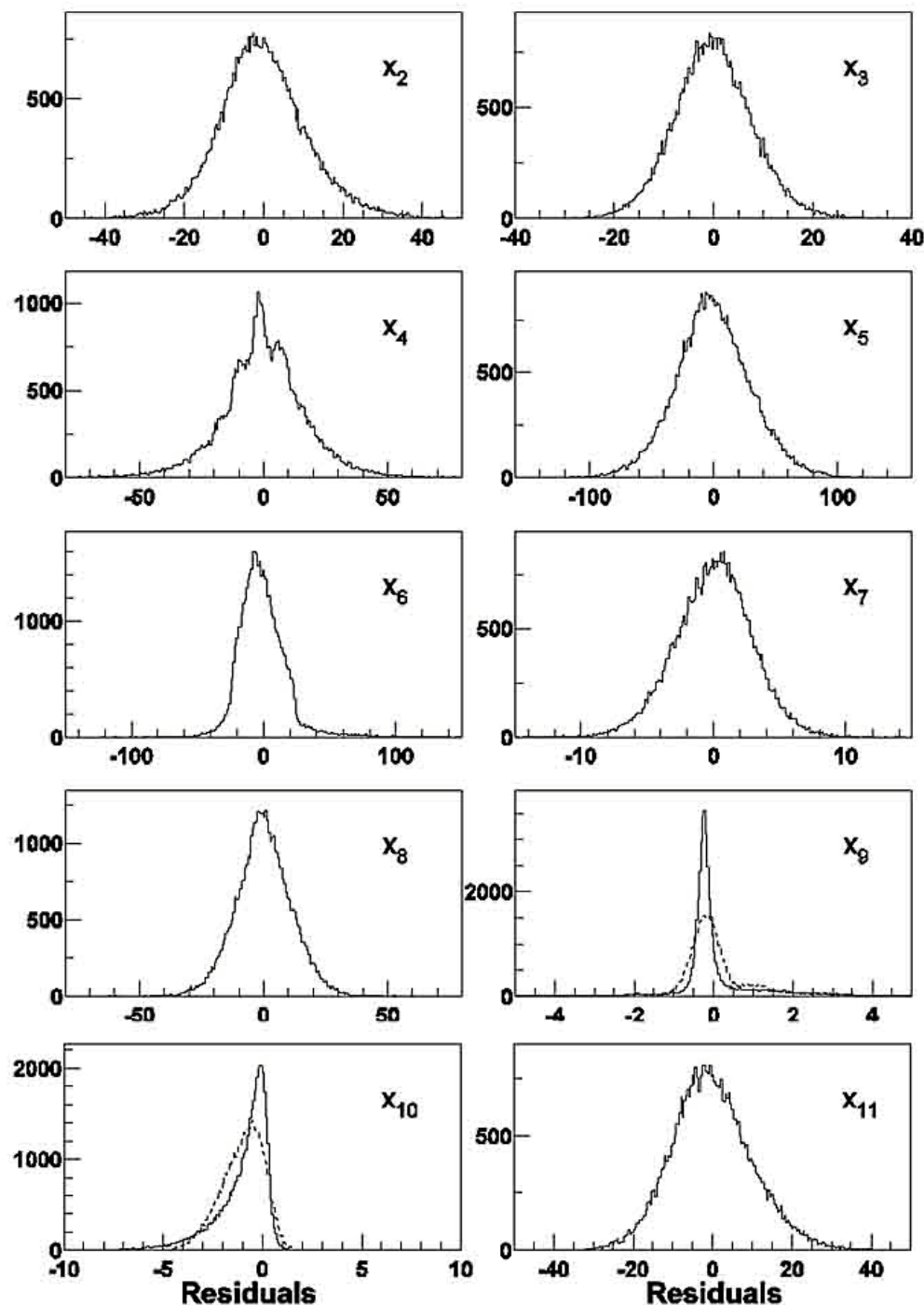
Figure 4. Distribution of residuals for the RN64f model (solid line) and for cigarettes and LVH for females for the base linear model.

complicated than for males (i.e., female health status is more sensitive to nonlinear interactions among risk factors). We suggest that this is due to the rapid rate of decline of estrogen production after menopause.

We can suggest several ways to improve the nonlinear models: 1.) account for the selective effect of a

risk factor, or its product, on dynamics; 2.) account for the age-dependence of the coefficients in our regression by estimating them at each 2-year step; 3.) account for higher order nonlinear terms. We recommend the third approach. The weakness of the first approach is that it requires a significant increase of numerical power. That, however, is still reasonable if we combine statistical approaches with

nonlinear analysis. The weakness of the second approach concerns lack of data at advanced ages. This can be avoided by averaging the regression coefficients over longer than 2 year-age intervals.

9. REFERENCES

1. Keys A.: Wine, garlic, and CHD in seven countries. *Lancet* 1, 145-146 (1980)

2. Willett W. C., P. J. Skerrett; E. L. Giovannucci: Eat, drink and be healthy: The Harvard Medical School Guide to Healthy Eating, Simon & Schuster (2001)

3. Koopmans T. C. and O. Reiersol: The identification of structural characteristics. *Annals of Mathematical Statistics* 21, 165-181 (1950)

4. Manton K. G., E. Stallard; M. A. Woodbury; J. E. Dowd: Time-varying covariates in models of human mortality and aging: Multidimensional generalization of the Gompertz. *J Gerontol A Biol Sci* 49, B169-B190 (1994)

5. Strehler B.L. and A.S Mildvan: A general theory of mortality and aging. *Science* 132, 14-21 (1960)

6. Rosenberg B., G. Kemeny, R.C. Switzer, T.C. Hamilton: The kinetics and thermodynamics of death in multicellular organisms. *Mech Ageing Dev* 2, 275-293 (1973)

7. Masoro, EJ: Dietary restriction and aging. *J Am Geriatr Soc* 41 (9), 994-999 (1993)

8. Sullivan D.M., R. Vander Zwaag, J.P. Hughes, V. Maddock, F.W. Kroetz, K.B. Ramanathan, D.M. Mirvis: Estrogen replacement and coronary artery disease: Effect on survival in postmenopausal women. *Arch Intern Med* 150, 2557-2566, (1990)

9. Strehler, B.L: Time, Cells and Aging. Academic Press, New York (1977)

10. Gillespie, D. T: The mathematics of simple random walks. *Naval Research News* 35, 46-52 (1983)

11. Risken, H.: The Fokker-Planck Equation. Springer-Verlag, New York (1996)

12. Woodbury, M.A.; K.G. Manton: A random-walk model of human mortality and aging, *Theoretical Population Biology* 11, 37-48 (1977)

13. Woodbury M.A. and K.G. Manton: A Theoretical Model of the Physiological Dynamics of Circulatory Disease in Human Populations. *Hum Biol* 55 (2), 417-441 (1983)

14. Port S., A. Garfinkel, N. Boyle: There is a non-linear relationship between mortality and blood pressure. *Eur Heart J* 21, 1635-1638 (2000)

15. Peña, D., G.C. Tiao, R.S. Tsay: A Course in Time Series Analysis. Wiley, New York (2001)

16. Lai T. L., S. P. Won: Stochastic Neural Networks with Applications to Nonlinear Time Series. *J Am Stat Assoc: Theory and Methods* 96, 968-981 (2001)

17. Akushevich I., A. Kulminski, K. Manton: Time-Varying Covariates in Models of Aging: Nonlinear Analysis of Multivariate Biological Systems. Nonlinear Analysis: *Real World Applications* to be published (2003)

18. Kaplan D and L Glass: Understanding nonlinear dynamics. Springer-Verlag, New York (1995)

19. Wolf A., J. B. Swift, H. L. Swinney, J. A. Vastano: Determining Lyapunov exponents from a time series. *Physica D* 285-317 (1985)

20. Haken, H.: At least one Lyapunov exponent vanishes if the trajectory of an attractor does not contain a fixed point. *Phys Lett* 94A, 71 (1983)

21. Ware, J. H.: Linear models for the analysis of longitudinal studies. *Amer. Statistician* 39, 95-101 (1985)

22. Tuma N.B., M.T. Hannan: Social Dynamics. Academic Press, New York (1984)

23. Akushevich I., L. Akushevich, K. Manton, A.Yashin: Stochastic Process Model of Mortality and Aging: Application to Longitudinal Data. *Nonlinear phenomena and complex systems*, in press (2003)

24. Kulminski A., I. Akushevich, K.Manton: Time-Varying Covariates in Models of Aging: Nonlinear Analysis of Multivariate Biological Systems. *Nonlinear analysis: Real world applications*, to be published (2003)

25. Orchard G. and M.A. Woodbury: A missing information principle: Theory and application. Sixth Berkley Symposium on Mathematical Statistics and Probability (eds: L. LeCam, J. Heyman, E. Scott), University of California Press, Berkley, CA, 697-715 (1971)

26. Manton KG, E. Stallard: Chronic Disease Modeling. Oxford University Press, New York (1988)

27. Sakamoto Y, M. Ishiguro; G. Kitagawa: Akaike Information Criterion Statistics. D. Reidel Publishing Company, Dordrecht, Holland (1986)

28. Anderson, G. D.: Sex Differences in Drug Metabolism: Cytochrome P-450 and Uridine Diphosphate Glucuronosyltransferase. *J Gend Specif Med* 5, 25-33 (2002)

29. Manton K.G. and E. Stallard: Chronic Disease Risk Modeling: Measurement and Evaluation of the Risks of Chronic Disease Processes. In: Griffin Series of the Biomathematics of Diseases. Charles Griffin Limited, London, England (1988)

30. Willett, W: Dietary fat and breast cancer. *J Natl Cancer Inst* 83, 1035-1036 (1991)

Send correspondence to: A. Kulminski, Duke University, Center for Demographic Studies, 2117 Campus Drive, Box 90408, Durham, NC, 27708-0408, USA, Tel: (919) 684-6126, Fax: (919) 684-3861, E-mail: kam@cds.duke.edu