LONGITUDINAL SURVIVAL DATA AND PROBLEMS OF NONLINEAR DYNAMICS

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

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

OVERVIEW OF THE TOPIC

Large amounts of epidemiologic and demographic data have been collected to link health changes over time. For example, the original Framingham 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.

In addition 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.

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 [2] 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 presentation we concentrate on improving our modeling of state trajectories for individuals.

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. 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. 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 has a tendency to increase with age. It increases the risk of stroke, atherogenesis, and renal damage.

Body mass index (x_4). Body mass index 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 [3]. In humans, the relation of BMI to body mass is likely complex and multi-dimensional.

Serum cholesterol (x_5). Increased plasma, insulin, and obesity accelerate lipolysis, increasing circulating free fatty acids and triglycerides that cause serum cholesterol 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 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.

Vital capacity index (x_8). The vital capacity index is calculated as $10 \times$ *vital capacity in deciliters* / *height*^{2}. This measure of pulmonary capacity is believed to be one of the biomarker's most directly related to senescence [4] 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.

THEORETICAL BACKGROUND AND ANTICIPATED RESULTS

We constructed a stochastic process from the J variables measured on I individuals up to N times. The model to be constructed must describe the movement of a person i ($i=1,2,\ldots I$) in a J dimensional state space. To construct the model we use the random walk specification [5] which leads to the linear dynamic equations for J variables.

Such a model 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. 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. Dynamics in such systems can also be affected by unobserved processes reflecting nonlinear interactions and/or correlation among the system components so that standard linear models cannot fully describe data originated from these systems. Nonlinear time series analysis has recently gained attention [6]. For example, it was recently shown that the Framingham data do not support a linear paradigm [7]. In this case, nonlinear analysis is essential. However, employing nonlinear models one can face a problem which might be related with specific features originating from nonlinear interactions. This can require application of special methods of nonlinear dynamics to the analysis of state dynamics for individuals. In this presentation we specifically deal with such kind of problems.

At the beginning, we focus on the study of the simplest type of nonlinear model, when nonlinear terms are assumed to be proportional to the product of two risk factors. This is in addition to the assumption of the linear superposition of all risk factors on future health. Hence, we assume that each risk factor can "interfere" with each other inducing additional changes (positive or negative, depending on goodness of fit and the substantial importance of their interaction) in future health. Then, we extend our model assuming that interference of three and four variables can be also important. Respectively, this can be represented by scalar product of three and four variables.

We expect that considerable additional information on human aging processes can be extracted if one assesses nonlinear effects. However, such models can give rise to surprising behavior as a consequence of their intrinsic nonlinear nature. Hence, we elaborate a way how standard statistical methods used to describe the age dependence of health status in traditional linear models can be enhanced to determine the best nonlinear model.

REFERENCES

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

2. 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)

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

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

5. 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)

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

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