The International Journal of Biostatistics

Volume 3, Issue 1

2007

Article 7

Survival Instantaneous Log-odds Ratio from Empirical Functions

JungAh Jung, *Novartis Pharmaceutical Corporation*

J. Wanzer Drane, *Eastern Tennessee State University*

Recommended Citation:

Jung, JungAh and Drane, J. Wanzer (2007) "Survival Instantaneous Log-odds Ratio from Empirical Functions," *The International Journal of Biostatistics*: Vol. 3: Iss. 1, Article 7.

DOI: 10.2202/1557-4679.1017

Survival Instantaneous Log-odds Ratio from Empirical Functions

JungAh Jung and J. Wanzer Drane

Abstract

The objective of this work is to introduce a new method called the Survivorship Instantaneous Log-odds Ratios (SIJOR); to illustrate the creation of SIJOR from empirical bivariate survival functions; to also derive standard errors of estimation; to compare results with those derived from logistic regression. Hip fracture, AGE and BMI from the Third National Health and Nutritional Examination Survey (NHANES III) were used to calculate empirical survival functions for the adverse health outcome (AHO) and non-AHO. A stable copula was used to create a parametric bivariate survival function, that was fitted to the empirical bivariate survival function. The bivariate survival function had SIJOR contours which are not constant. The proposed method has better advantages than logistic regression by following two reasons. The comparison deals with (i) the shapes of the survival surfaces, $S(X_1, X_2)$, and (ii) the isobols of the log-odds ratios. When using logistic regression the survival surface is either a hyper plane or at most a conic section. Our approach preserves the shape of the survival surface in two dimensions, and the isobols are observed in every detail instead of being overly smoothed by a regression with no more than a second degree polynomial. The present method is straightforward, and it captures all but random variability of the data.

KEYWORDS: copula, instantaneous log-odds ratio, log-odds, log-odds ratio, odds ratio, survival instantaneous log-odds ratio, survival instantaneous odds ratio

1 Introduction

The logistic models are commonly used to estimate odds ratios; however, they can result in misleading information on the true distributions. Linear logistic regression assumes a constant log-odds ratio for a continuous risk variable. The uses of quadratic or more complex polynomial may lead to unexplainable results such as decreasing log-odds where it should be increasing. For these reasons, survival instantaneous log-odds ratio (SIJOR) was proposed in this work.

In the paper, SIJOR is illustrated with subset of data from the Third National Health and Nutrition Examination Survey (NHANES III). It was conducted on a nationwide probability sample of approximately 33,994 persons two months and over in age (NHANES III, 2005). The hip fracture of white women was chosen as the adverse health outcome (AHO), and age by month (AGE) and body mass index (BMI) were the risk factors where both hip fracture and free of hip fracture occurred. Hip fracture is a public concern, because it is associated with increased mortality, morbidity, reduced quality of life, and it incurs significant economic and social costs (Center et al., 1999). Body weight is strongly related to bone mineral density (BMD), such that higher weight is associated with both higher mineral density and reduced fracture risk (Cummings et al., 2002). It presented that 20% of hip fractures occurred in women between 50 to 64 years old, factors in young women associated with a fracture were low bone mass (Brenneman et al., 2003). It showed that the risk of hip fracture for women with age between 50 to 64 and with low (less than 25) BMI (Cauley et al., 2003). Furthermore, women experience more falls than men (Cooper, 1999). There is much literature regarding this issue; see for example (Felson et al., 1993; Marcus et al., 1994; Ravn et al., 1999; and, Reid et al., 1992).

2 Method

In the functions of this work the variables are described as follows.

1. $Y = 0$ or 1 . $Y = 0$ indicates “free of the AHO”, and $Y = 1$ indicates “AHO present”. Y is a random variable.
2. X is a variable measured on a continuous scale on the interval $(0, B)$ where B is either finite or infinite. X can be treated as a random variable.

In many observational studies, it is very difficult to control risk factors so that dose response functions (DRF) have slightly different form which contains the cumulative distribution of the risk factors. That is $P(Y = 1|X \leq x)$. Subjects are, by and large, observed to have a measurable condition X , on a continuous scale, and they are determined to have or to be free of the AHO, but it is not possible to know *when* or at *what level* of X that the AHO actually developed. When more than one

risk factor is considered, the survival response function (SRF), $P(Y = 0|X > x)$, is more convenient to use because of its mathematical simplicity.

The survival function is simply defined as

$$P(X > x) = \int_x^B f(x')dx'$$

whatever X is, as long as it is continuous with support $I(x|0 \leq x < B)$, $0 < B < \infty$. Thus, individual's surviving or dying is not necessary in this function. It is not necessary for X to be time, although it is risk factor in this work. Non-negative support reflects reality of direct measurements on human subjects. If one uses support on the entire reals, then there must be a transformation on at least one direct observation. For example, BMI and pH are both derived.

In multi-dimensional space

$$S(x_1, x_2, \dots, x_h) = \int_{x_1}^{B_1} \int_{x_2}^{B_2} \dots \int_{x_h}^{B_h} f(x'_1, x'_2, \dots, x'_h) dx'_1 dx'_2 \dots dx'_h,$$

$$0 < B_i < \infty, i = 1, \dots, h.$$

Thus, the SRF results from an intersection of the risk factors while the DRF is the compliment. The SRF is the probability of surviving when $X = x$ and is defined as follows.

$$\begin{aligned} P(Y = 0|X > x) &= \frac{P(Y = 0, X > x)}{P(X > x)} \\ &= \frac{P(Y = 0, X > x)}{P(X > x, Y = 0) + P(X > x, Y = 1)} \\ &= \frac{P(X > x|Y = 0)P(Y = 0)}{P(X > x|Y = 0)P(Y = 0) + P(X > x|Y = 1)P(Y = 1)} \\ &= \frac{(1 - \theta)S_0(x)}{(1 - \theta)S_0(x) + \theta S_1(x)}, \end{aligned} \tag{1}$$

where

$\theta = P(Y = 1)$ = prevalence of AHO in the population;

$S_1(x) = P(X > x|Y = 1)$ = survival distribution of X among AHO;

$S_0(x) = P(X > x|Y = 0)$ = survival distribution of X among free of AHO;

$S(x) = \theta S_1(x) + (1 - \theta)S_0(x)$ = survival distribution of X in the population.

Similarly,

$$P(Y = 1|X > x) = \frac{\theta S_1(x)}{(1 - \theta)S_0(x) + \theta S_1(x)}. \tag{2}$$

The beauty of (1) and (2) is its form does not change or take on an increased number of terms as the dimension of X increases.

The odds of surviving an AHO when $X = x$ is defined by

$$O(x) = \frac{P(Y = 0|X > x)}{P(Y = 1|X > x)} = \frac{1 - \theta}{\theta} \frac{S_0(x)}{S_1(x)},$$

from which

$$\log[O(x)] = \log \frac{1 - \theta}{\theta} + \log \frac{S_0(x)}{S_1(x)}.$$

The instantaneous log-odds ratio (Drane, 1996 and 1997) in terms of survival functions of risk variables can be derived as follows (Drane et al., 2007). For any $\Delta x > 0$,

$$\lim_{\Delta x \rightarrow 0} \frac{\log[O(x + \Delta x)] - \log[O(x)]}{(x + \Delta x) - x} = \frac{\partial}{\partial x} \log \frac{S_0(x)}{S_1(x)}.$$

Finally,

$$SILOR(x) = \frac{\partial}{\partial x} \log \frac{S_0(x)}{S_1(x)}$$

is the definition of the *Survival Instantaneous Log-Odds Ratio (SILOR)*. It is the log-odds ratio of surviving an AHO at any continuous risk value of X . Then, an important characteristic of SILOR is developed as follows.

$$\begin{aligned} SILOR(x) &= \frac{\partial}{\partial x} \log \frac{S_0(x)}{S_1(x)} \\ &= \frac{\frac{\partial}{\partial x} S_0(x)}{S_0(x)} - \frac{\frac{\partial}{\partial x} S_1(x)}{S_1(x)} \\ &= \frac{-f_0(x)}{S_0(x)} - \frac{-f_1(x)}{S_1(x)} \\ &= \lambda_1(x) - \lambda_0(x). \end{aligned} \tag{3}$$

where $\lambda(\cdot)$ is the well-known hazard function from survival and reliability theories. λ_1 and λ_0 represent the rate of being $X = x$ among AHO and non-AHO, respectively, having survived until $X = x$. SILOR is found to be the difference of hazard functions of risk variables between AHO and free of AHO, respectively.

SILOR is used here as a statistic in its own right varying over the real numbers, while all of the risk variables, X , are non-negative. X can be a vector. Two dimensions are considered for X as the risk factors in this paper. For bivariate risk factors, X_1 and X_2 ,

$$SILOR(x_1, x_2) = \nabla \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \cdot \vec{\eta}$$

where:

$\nabla = i \frac{\partial}{\partial x_1} + j \frac{\partial}{\partial x_2}$ is the vector-valued derivative or gradient; i and j are the unit vectors lying along the positive direction of the X_1 and X_2 axis, respectively; $\vec{\eta} = i \cos(\varphi_1) + j \cos(\varphi_2)$; φ_k is the angle between $\nabla \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)}$ and the X_k axis, and $\varphi_k \in [0, \pi]$, $k = 1, 2$.

Finally, $\text{SIOR}(x_1, x_2)$ is defined as follows: Its magnitude is

$$\sqrt{\left[\frac{\partial}{\partial x_1} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \right]^2 + \left[\frac{\partial}{\partial x_2} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \right]^2}, \quad (4)$$

and its direction is

$$\varphi_k(x_1, x_2) = \cos^{-1} \left[\frac{\frac{\partial}{\partial x_k} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)}}{\sqrt{\left[\frac{\partial}{\partial x_1} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \right]^2 + \left[\frac{\partial}{\partial x_2} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \right]^2}} \right], \quad (5)$$

where $\varphi_k(x_1, x_2)$ is the angle between $\nabla \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)}$ and its respective axis, $0 \leq \varphi_k(x_1, x_2) \leq \pi$, $k = 1, 2$. If $\frac{\partial}{\partial x_k} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \geq 0$, then $0 \leq \varphi_k(x_1, x_2) \leq \frac{\pi}{2}$. Conversely, if $\frac{\partial}{\partial x_k} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} < 0$, then $\frac{\pi}{2} < \varphi_k(x_1, x_2) \leq \pi$. Thus, the direction must be given in terms of all of the angles. In two dimensions the quadrants are identified according to the $\cos(\varphi_k(x_1, x_2))$, $k = 1, 2$. This does not affect $\text{SIOR}(x_1, x_2)$ because it is the square root of a sum of squares. These directions translate into harmful or helpful effect with respect to the each risk factor. As $\frac{\partial}{\partial x_k} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)}$ closes to zero, the influence on SIOR is shifted from x_k to the other, $k = 1, 2$. $\cos^2(\varphi_k(x_1, x_2))$ gives the fraction of the square of the partial derivative with respect to X_k . Thus, larger $\cos^2(\varphi_k(x_1, x_2))$, the greater is the effect X_k has on $\text{SIOR}(x_1, x_2)$. Keep always in mind that $\sum_k \cos^2(\varphi_k) = 1$ where k is greater than equal to 2. The survival instantaneous odds ratio (SIOR) can simply be derived by exponentiating SIOR (4).

3 A modeling approach for bivariate survival distributions

One of the objectives of this paper is to illustrate SIOR of two dimensions of risk factors with real data. There is a great deal of flexibility in modeling multivariate survival data. Fully parametric approaches are attractive when they are identifiable and produce smooth survival functions. Fully nonparametric approaches, while providing flexibility, can be inconsistent or inefficient in the presence of

partial derivatives. Semi-parametric models combine the best features of both approaches (Shih and Louis, 1995), so this approach will be investigated. Copula models of bivariate survival functions will be discussed in the following section.

3.1 Copula models of bivariate survival functions

Copulas offer a tool for understanding relationships in multivariate data. The basic concept is that a multivariate distribution can be “organized” into its marginals plus a dependence function called copula. The stable copula (Gumbel-Hougaard) family was used for the bivariate survival models because of its mathematical convenience when combined with Weibull marginals. There have been several applications of stable copulas (Oakes, 1989, Shih and Louis, 1995, Shemyakin and Youn, 2006). The choice of the bivariate survival model is as follows (Shemyakin and Youn, 2006).

$$S(x_1, x_2) = \exp \left[- \left\{ \left(\frac{x_1}{\beta_1} \right)^{\alpha\gamma_1} + \left(\frac{x_2}{\beta_2} \right)^{\alpha\gamma_2} \right\}^{1/\alpha} \right]. \quad (6)$$

where $\alpha \geq 1$, $\beta_j > 0$ and $\gamma_j > 0$, $j = 1, 2$.

Frees and Valdez (1998) discussed the measures of association in copula models. In stable copula family, the role of the association parameter α can be understood through its relationship to Kendall’s correlation coefficient (τ). It is known that $\alpha = \frac{1}{1-\tau}$. Large values of α provide large correlations, and x_1 and x_2 are independent when α is equal to 1.

3.2 Empirical survival functions

The empirical survival distributions of risk factors (X_1, X_2) among AHO and among free of AHO respectively. That is $S_i(x_1, x_2) = \frac{m_i(x_1, x_2)}{n_i(x_1, x_2)}$, $i = 0, 1$, where $m_i(x_1, x_2)$ is the count of subjects with $X_1 > x_1$ and $X_2 > x_2$, and n_i is the total number of the sample among each stratum: $i = 0$ represents free of AHO and $i = 1$ represents the AHO.

4 Example

4.1 Data

There are 93 subjects with hip fracture and 2276 without hip fracture in the NHANES III data. Thus, the estimated prevalence (θ) of hip fracture is 0.039 ($\frac{93}{93+2276}$). The selected limits are from 648 to 1080 months (54 to 90 years) for AGE and from 15.4 to 43.8 for BMI. The limits of AGE and BMI were selected

Figure 1: Plot of age by month versus body mass index. Top panel is the plot among those with hip fracture. Bottom panel is the plot among those without hip fracture.

where both fracture and free of hip fracture occurred. The plots of AGE versus BMI with hip fracture and without hip fracture are displayed in Figure 1.

Because of the limited dependence parameter space for the stable copular family, $\alpha \geq 1$, this family permits only non-negative correlations in order to ensure a proper probability density function. AGE and BMI have negative correlations for both hip fracture and free of hip fracture so that a transformation of 43.8 - BMI was used. In all the mathematical functions and their analyses in this work, BMI represents the reverse coding of BMI. However, all the graphic illustrations and their interpretations will use the original values of BMI. The transformation of 1080 - AGE could also have used. However, the former fitted the data better than the latter.

4.2 Analysis

The initial step in the model fitting is to determine the appropriate marginals. The quality of the fit of a marginal distribution can be examined visually comparing the fitted function against the empirical version, as displayed in Figure 2.

SAS PROC NLMIXED (SAS Institute Inc., 2004) was used to estimate the parameters. Though there is information loss such as the survival function of BMI among hip fractures, $S_1(bmi)$, when BMI is higher than 30 and the survival function of AGE among free of hip fractures, $S_0(age)$, with young AGE, 648 -750 months, these marginal fits are considered to be acceptable.

Empirical survival distributions of AGE and BMI as the bivariate risk variables are calculated and displayed in Figure 3 for both hip fracture and free of hip fracture. A 433×285 AGE - BMI grid was created (123405 points) to illustrate the empirical survival distributions. The choices of the numbers of each grid are arbitrary as long as the surfaces of the empirical survival distributions are smooth enough. We should keep in mind that the empirical distributions may differ from true values because they depend on how we select the grid of AGE and BMI to estimate those values. SAS PROC NLMIXED (SAS Institute Inc., 2004) was used to estimate the parameters. The NLMIXED procedure fits nonlinear mixed models, that is, models in which both fixed and random effects enter nonlinearly. The starting values are from the respective marginal Weibull distributions and the Kendall's correlations ($\alpha_1 = 1.2171$, $\alpha_0 = 1.1212$). The model (6) was fitted with the empirical distributions among the hip fracture data and among free of hip fracture data, respectively. The estimation of parameters are shown in Table 1, and their approximate standard errors of the estimates are very small as can be seen. The retained parameters of both bivariate models are significantly different from their null values at $\alpha=0.05$ ($p < .0001$). Therefore the fitted survival distribution functions are

Figure 2: Upper row: Marginal survival distribution function among hip fractures; left panel is $S_1(\text{age})$ and right panel is $S_1(\text{bmi})$. Lower row: Marginal survival distribution function among free of hip fractures; left panel is $S_0(\text{age})$ and right panel is $S_0(\text{bmi})$. Broken line represents the empirical distribution and solid line represents the fitted distribution in all graphs.

Parameter	Estimate	SE	95% LCL	95% UCL
$S_1(age, bmi)$				
α_1	1.24	0.00210	1.2335	1.2417
a_1	993.7	0.0517	993.6	993.8
b_1	9.18	0.00556	9.1693	9.1911
c_1	21.45	0.00318	21.4437	21.4562
d_1	5.03	0.00423	5.0221	5.0387
$S_0(age, bmi)$				
α_0	1.02	0.00150	1.0187	1.0246
a_0	887.4	0.0403	887.4	887.5
b_0	6.94	0.00277	6.9337	6.9446
c_0	19.20	0.00516	19.1942	19.2144
d_0	4.44	0.00437	4.4324	4.4495

Table 1: Estimates of parameters for the bivariate survival distribution functions of age by month and body mass index among hip fractures and among free of hip fractures. S_j : bivariate survival functions; $j = 0$ represents free of hip fracture and $j = 1$ represents hip fracture. SE: approximate standard error. 95% LCL and 95% UCL: approximate 95% lower and upper limits.

$$S_1(age, bmi) = \exp \left\{ - \left[\left(\frac{age}{993.7} \right)^{(1.24)(9.18)} + \left(\frac{bmi}{21.45} \right)^{(1.24)(5.03)} \right]^{1/1.24} \right\} \quad (7)$$

and

$$S_0(age, bmi) = \exp \left\{ - \left[\left(\frac{age}{887.4} \right)^{(1.02)(6.94)} + \left(\frac{bmi}{19.20} \right)^{(1.02)(4.44)} \right]^{1/1.02} \right\}. \quad (8)$$

These fitted survival distributions are shown in Figure 3 with their empirical distributions. Those graphs appear very close to the empirical graphs in both hip fracture and free of hip fracture.

5 Comparison with a logistic regression

How can SIJOR be shown to perform better than the log-odds ratio calculated from a logistic regression? Drane et al. (2007) showed that the fundamental equation resulting from in the SIJOR regardless of the dimension of the vector of

Figure 3: Upper row: The survival distribution of AGE and BMI as the bivariate risk variables among hip fractures; left panel is the empirical distribution and right panel is the fitted distribution. Lower row: The survival distribution of AGE and BMI as the bivariate risk variables among free of hip fractures; left panel is the empirical distribution and right panel is the fitted distribution. AGE represents age by month and BMI represents body mass index in all graphs.

risk factors and confounders is as follows.

$$SILOR(x) = \sqrt{\sum_i \left[\frac{\partial}{\partial x_i} \log[O(x)] \right]^2},$$

where $O(x)$ is odds.

For the linear logistic that same relationship is

$$SILOR(x) = \sqrt{\sum_i \beta_i^2},$$

where $\log(O(x)) = x'\beta$.

A bivariate logistic regression was fitted with the same data. That is

$$\log \frac{P(Y = 0|age, bmi)}{1 - P(Y = 0|age, bmi)} = \beta_0 + \beta_1 age + \beta_2 bmi + \beta_3 age \cdot bmi + \beta_4 age^2 + \beta_5 bmi^2 \quad (9a)$$

where $P(Y = 0|age, bmi)$ is the probability of free of hip fracture with AGE and BMI. Using SAS PROC logistic (SAS Institute Inc., 2004) procedure, the fitted logistic model is

$$\log[O(age, bmi)] = 8.4895 - 0.00782age + 0.3167bmi - 0.0113bmi^2. \quad (9b)$$

From SAS analysis, the maximum likelihood estimates are shown in Table 2. The retained parameters of logistic regression model are significantly different from their null values at $\alpha=0.05$ ($p < .0001$). Neither the effect of AGE square nor their interaction were detected by the logistic model. The log-odds from the logistic model is displayed with those from empirical distribution as well as the proposed method, respectively, as can be seen the upper row in Figure 4; log-odds from the logistic regression (9b) appears to reach its maximum near BMI = 30 and minima at both extreme values where AGE = 1080 or 648 months (90 or 54 years). This may not be true but the fitted logistic regression model has this property. The log-odds from the proposed method appears closer to the empirical log-odds in character and actual values (see Figure 4). The estimated probability of free of hip fracture that corresponds to the survival response function (1) is

$$P(Y = 0|age, bmi) = \frac{\exp(\log[O(age, bmi)])}{1 + \exp(\log[O(age, bmi)])} \quad (10)$$

where $\log[O(age, bmi)]$ is found in (9b). The empirical log-odds and survival distribution of the risk factors are not estimable near the minimum value of BMI and the maximum value of AGE because of lack of observations. These grid values are treated as missing (2086 grid points out of 123405 were excluded). In Figure 4 the

Figure 4: Upper row: From the left to right, log-odds of surviving hip fracture from the empirical distribution, logistic model, and proposed method, respectively. Lower row: From the left to right, survival response functions (SRF) from the empirical distribution, logistic model, and proposed method, respectively. AGE represents age by month and BMI represents body mass index in all graphs.

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chisq
Intercept	1	8.4895	1.0960	60.0042	<0.0001
AGE	1	-0.00782	0.00110	50.8476	<0.0001
BMI	1	0.3167	0.0770	16.9242	<0.0001
BMI^2	1	-0.0113	0.00235	23.1202	<0.0001

Table 2: Estimates of parameters for the logistic regression with age by month (AGE) and body mass index (BMI) as the bivariate risk factors.

SRF from the logistic regression (10) decreases as BMI increases beyond 35. This may not be true but it is projected by the logistic regression model. The SRF from the proposed method appears closer to the empirical SRF in character and actual values (see Figure 4).

Now, a statistical procedure is needed to compare the fits of log-odds and the SRF as obtained from the logistic model to the proposed method. AIC (Akaike, 1973 and 1974), and BIC (Schwarz, 1978) were used in this study. In model selection, the values of the AIC and BIC for a given data set have no meaning. The model with the lowest AIC and BIC being the best model among all models specified for the data at hand. For the log-odds, AIC and BIC were 5436.08 and 5542.85, respectively for the proposed method, and 416854.7 and 416903.3, respectively for the logistic model. For the probability of surviving hip fracture, AIC and BIC were -337501.8 and -337395, respectively for the proposed method, and -218176.5 and -218128, respectively for the logistic model. Smaller AIC and BIC are better fit; thus, it can be concluded that the proposed method fits the data better than the logistic model.

For a comparison purpose, SIOR was computed from the logistic regression model (9b). SIOR is derived by exponentiating the SIOR and the results are illustrated in Figure 5. The directional cosines are $\cos(\varphi_1(age, bmi)) < 0$ for AGE, while $\cos(\varphi_2(age, bmi)) > 0$ for BMI < 30, and $\cos(\varphi_2(age, bmi)) < 0$ for BMI > 30. The negative $\cos(\varphi_2(age, bmi))$ beyond 30 in BMI indicates that BMI has a harmful effect on hip fracture when BMI is greater than 30. This may not be true, however it is projected by the logistic regression. The changes in SIOR appear to be almost the same magnitude as BMI increases or decreases. AGE effects on SIOR are hardly detected. The changes of both directional cosines appear to be almost the same as BMI reaches its maximum or minimum. Therefore it can be concluded that the logistic model resulted in misleading information about the true relationship.

Figure 5: Contour plots of logistic survival instantaneous odds ratio (SIOR) of hip fracture on white women with age by month and body mass index as the bivariate risk factors. Upper row is the magnitude and lower row are the directional cosine by age by month (left panel) and by body mass index (right panel), respectively.

6 Results

$\text{SIOR}(age, bmi)$ is illustrated in this section. SIOR is derived by exponentiating SIOR and the contours of SIOR are shown in Figure 6. Since the directional cosines are found to be $\cos(\varphi_1(age, bmi)) < 0$ for AGE, and $\cos(\varphi_2(age, bmi)) > 0$ for BMI with a support of (AGE, BMI). Thus, odds of surviving hip fracture decreases as increasing AGE and decreasing BMI. The maximum value of SIOR was 1.22 and occurred at the maximum value of AGE (1080 months) and the minimum value of BMI (15.4). It is an indication that compared to the referent there are 22% decreases on the odds of surviving hip fracture or increases on the odds of hip fracture where AGE = 1080 months and BMI = 15.4. SIOR is 1 when women have the minimum AGE of 648 months (54 years), and the highest BMI of 43.8, which implies that this is the referent combination against which all the SIOR are compared. $\cos(\varphi_1(age, bmi)) < 0$ for all value of AGE implies that the odds of surviving hip fracture decreases as AGE increases so that AGE has a harmful effect on hip fracture. $\cos(\varphi_2(age, bmi)) > 0$ for all value of BMI implies that the odds of surviving hip fracture increases as BMI increases so that BMI has a protective effect on hip fracture. $\cos(\varphi_1(age, bmi)) \rightarrow \pm 1$ and $\cos(\varphi_2(age, bmi)) \rightarrow 0$ indicates a nearly complete dependence on AGE. Conversely, as $\cos(\varphi_1(age, bmi)) \rightarrow 0$ and $\cos(\varphi_2(age, bmi)) \rightarrow \pm 1$ indicates a nearly complete dependence on BMI as the cause hip fracture. When women whose BMI is lower than 35, the risk of hip fracture is more likely to be caused by their low BMI and this effect is greater for younger women whose AGE are smaller than 864 months (72 years) old.

The variances of the estimates of the SIOR can be calculated by the Delta method (Greene 2000; Agresti 2002). The delta method is a general approach for computing an approximate variance for functions of maximum likelihood. The standard errors of the estimates of the SIOR are displayed in Figure 7. Standard errors of the estimates of the SIOR have very small values for all AGE and BMI. Their range is from 0 to 0.0033. There is a variation on the estimation of the SIOR beyond 35 in BMI, and it increases as AGE increases. It can be concluded that it is caused by the sparseness of the data (see Figure 1). Except those areas, the values of the standard errors of the estimates are almost zero. It is an indication that the estimates of the SIOR converge to the true values. The estimates of the SIOR, their standard errors, 95% confidence limits and directional cosines are shown in Table 3.

Figure 6: Contour plots of survival instantaneous odds ratio (SIOR) of hip fracture on white women with age by month and body mass index as the bivariate risk factors. Upper row is the magnitude and lower row are the directional cosine by age by month (left panel) and by body mass index (right panel), respectively.

Figure 7: Standard errors of the estimates of survival instantaneous odds ratio (SIOR) of hip fracture on white women with AGE and BMI as the bivariate risk factors. AGE represents age by month and BMI represents body mass index in the graph.

Table 3: The estimates of the SIOR (SIOR), their standard errors (SE_SIOR), 95% lower (LCL_SIOR) and upper (UCL_SIOR) confidence limits, and directional cosines ($\cos(\varphi_1)$, $\cos(\varphi_2)$) for the selected values of age by month (AGE) and body mass index (BMI). The number of values is 36.

AGE	BMI	SIOR	SE_SIOR	LCL_SIOR	UCL_SIOR	$\cos(\varphi_1)$	$\cos(\varphi_2)$
648	15.4	1.17	8.28E-07	1.17	1.17	-0.0066	1
648	21	1.1099	2.05E-05	1.1098	1.1099	-0.01	1
648	27	1.0486	9E-05	1.0484	1.0487	-0.0215	0.9998
648	33	1.0126	0.000251	1.0121	1.013	-0.0768	0.997
648	39	1.0011	0.001366	0.9984	1.0038	-0.8654	0.5011
648	43.8	1.0009	0.002881	0.9953	1.0066	-1	0
700	15.4	1.1703	1.27E-06	1.1703	1.1703	-0.0148	0.9999
700	21	1.1106	2.15E-05	1.1106	1.1107	-0.0215	0.9998
700	27	1.0504	6.7E-05	1.0502	1.0505	-0.0431	0.9991
700	33	1.0139	0.000239	1.0134	1.0143	-0.1423	0.9898
700	39	1.002	0.001377	0.9993	1.0047	-0.9632	0.2688
700	43.8	1.0019	0.002893	0.9963	1.0076	-1	0
800	15.4	1.1726	2.87E-06	1.1726	1.1726	-0.0277	0.9996
800	21	1.1147	2.34E-05	1.1147	1.1148	-0.0378	0.9993
800	27	1.0556	8.61E-05	1.0555	1.0558	-0.068	0.9977
800	33	1.0151	0.000249	1.0147	1.0156	-0.2311	0.9729
800	39	1.0035	0.001406	1.0008	1.0063	-0.9881	0.1538
800	43.8	1.0035	0.002927	0.9977	1.0092	-1	0
900	15.4	1.1818	8.98E-06	1.1818	1.1819	-0.0427	0.9991
900	21	1.1271	2.45E-05	1.1271	1.1272	-0.0521	0.9986
900	27	1.0633	0.000139	1.063	1.0636	-0.0896	0.996
900	33	1.0163	0.000344	1.0157	1.017	-0.3281	0.9446
900	39	1.0054	0.001484	1.0024	1.0083	-0.995	0.1003
900	43.8	1.0053	0.003011	0.9994	1.0112	-1	0
1000	15.4	1.2074	2.57E-05	1.2074	1.2075	-0.0499	0.9988
1000	21	1.1497	3.02E-05	1.1497	1.1498	-0.0552	0.9985
1000	27	1.0705	0.000193	1.0701	1.0709	-0.1009	0.9949
1000	33	1.0173	0.000535	1.0162	1.0183	-0.3934	0.9194
1000	39	1.0068	0.001676	1.0035	1.0101	-0.9969	0.0786
1000	43.8	1.0068	0.003212	1.0005	1.0131	-1	0
1080	15.4	1.2202	3.46E-05	1.2201	1.2203	-0.0488	0.9988
1080	21	1.1581	3.52E-05	1.158	1.1581	-0.0531	0.9986
1080	27	1.0724	0.000208	1.072	1.0728	-0.0998	0.995
1080	33	1.0174	0.000602	1.0162	1.0186	-0.397	0.9178
1080	39	1.0069	0.001772	1.0034	1.0104	-0.997	0.0771
1080	43.8	1.0069	0.003315	1.0004	1.0134	-1	0

7 Discussion

Logistic models are widely used to estimate odds ratios in statistical analyses. However, their use for continuous risk factors is not ideal because logistic regression has a constant odds ratio for a unit change in a continuous risk factor. It may mislead or cause one to ignore some information about the true relationship. Quadratic or more complex polynomial logistic regressions could also lead to unexplainable results such as decreasing risk rate where it should be increasing. For these reasons, we proposed a new statistic called survival instantaneous log-odds ratio (SIOR). SIOR is an extension of log-odds ratio with continuous risk factors and is derived by taking partial derivatives of log-odds of fitted empirical survival distribution functions. It is explained as the log-odds of surviving an AHO with continuous risk variables, and it is found to be the difference of hazard functions of risk variables between AHO and free of AHO.

In many observational studies, it is very difficult to control risk factors so that dose response functions have slightly different form which contains the cumulative distribution of the risk factors. When more than one risk factor is considered, the survival response function (SRF) is more convenient to use because of its mathematical simplicity. The survival distribution function is simply defined as $P(X > x) = \int_x^B f(x')dx'$ whatever X is, as long as it is a continuous with support $I(x|0 \leq x < B)$, $0 < B < \infty$. Thus, individual's surviving or dying is not necessary in this function. It is not necessary for X to be time. They are risk factors expressed as AGE and BMI in this paper.

The uniqueness of this approach is two-fold. First of all the empirical distributions of the risk factors among AHO and free of AHO contain the total of information of the data to connect the risk variables with the outcomes. Secondly, the SIOR is a vector product of partial derivatives of the survival functions of the risk variables and is independent of prevalence and incidence of AHO. It has been shown here that with multiple risk factors and their empirical survival functions one can immediately use the theory to a greater advantage than other approaches, especially that of logistic regression. Logistic regressions with continuous multiple risk variables can result in misleading information on the true distributions. However, SIOR is based on their empirical distributions of risk variables so that it has the strength such as more unbiased and informative against logistic models. The standard errors of the estimates of SIOR are very small except in those areas where the data are sparse. It indicates that the estimates of the SIOR converge to the true values. SIOR permits assessment of the interaction as well as overall risk factor effects. It characterizes the rate of change of surviving an AHO in log-odds at any continuous risk factor. It also helps to determine the range of risk factors at which the risk of AHO increases or decreases.

Important findings in this work are described in the following paragraph. For two risk factors SIOR has a vector of partial derivatives, which are con-

verted until converted to a scalar by multiplying by a unit directional vector, $i \cos(\varphi_1(x_1, x_2)) + j \cos(\varphi_2(x_1, x_2))$, where

$$\cos(\varphi_k(x_1, x_2)) = \frac{\frac{\partial}{\partial x_k} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)}}{\sqrt{\left[\frac{\partial}{\partial x_1} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \right]^2 + \left[\frac{\partial}{\partial x_2} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \right]^2}}, \quad k = 1, 2.$$

Thus, direction is needed in order to fully understand and interpret $\text{SILOr}(x_1, x_2)$ with the magnitude of

$$\sqrt{\left[\frac{\partial}{\partial x_1} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \right]^2 + \left[\frac{\partial}{\partial x_2} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \right]^2}.$$

In two dimensions the signs of the partial derivatives together indicate the gradient of the vector $\text{SILOr}(x_1, x_2) = \nabla \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)}$. As $\frac{\partial}{\partial x_k} \log \frac{S_0(x_1, x_2)}{S_1(x_1, x_2)} \rightarrow 0$, the influence on SILOr is shifted from x_k to the other, $k = 1, 2$. This will also be reflected in $(\varphi_1(x_1, x_2), \varphi_2(x_1, x_2))$ and the differences in hazard ratio. These directions translate into harmful or protective effect on the AHO with respect to single risk variables. The direction of SILOr is not always the same for all combinations of the risk factors because it depends on the correlation between risk factors. It is found to be different since there is a negative correlation between AGE and BMI.

In this paper a stable copula is used for the construction of models of bivariate survival functions. The models fitted very well with the empirical survival distributions. The stable copular model has an advantage of its mathematical convenience when combined with Weibull marginals. One limitation of the stable copula model is to permit only nonnegative correlation between two variables. However it can be worked out by taking a transformation which makes one of the variables the reverse order. To estimate the parameters of the models SAS PROC NLMIXED procedure was used and found to be easy as well as producing good estimates. The method of copulas has a wide range of applications tying together seemingly disparate distributions. Thus, it can allow one to connect multivariate risk variables of more than two dimensions as well to an AHO simultaneously, which are worth further investigations.

References

- Agresti A. (2002) *Categorical Data analysis*, 2nd edn. New York: Wiley.
- Akaike, H. (1973) Information theory and an extension of the maximum likelihood principal. In B.N. Petrov and F. Csak (eds.), *Second International Symposium on Information Theory. Budapest: Akademiai Kiado.*, 267-281.
- Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control.*, AC-19, 716-723.

- Brenneman S. K. et al. (2003) Evaluation of decision rules to identify postmenopausal women for intervention related to osteoporosis. *Disease Management*, 6, 159-68.
- Cauley, J. A. et al. (2003) Effects of Estrogen Plus Progestin on Risk of Fracture and Bone Mineral Density: The Women's Health Initiative Randomized Trial. *JAMA*, 290, 1729-1738.
- Center, J. R. et al. (1999) Mortality after all major types of osteoporotic fracture in men and women: an observational study. *Lancet*, 353(9156), 878-882.
- Cummings, S. R. et al. (2002) Clinical use of bone densitometry: scientific review. *JAMA*, 288(15), 1889-1897.
- Cooper C. (1999) Epidemiology of osteoporosis. *Osteoporos Int.*, 9(2), S2-8.
- Drane, J. W. (1996) Instantaneous Log Odds Ratio. *Int. Biometric Soc., ENAR: Richmond, VA, USA*.
- Drane, J. W. (1997) The odds ratio for continuous risk factors, some new results. *Int. Biometric Soc., ENAR: Memphis, TN, USA*.
- Drane, J. W. et al. (2007) Two theorems and thier application in public health. *Journal of modern applied statistical methods* will appear in May 2007.
- Felson, D. T. et al. (1993) Effects of weight and body mass index on bone mineral density in men and women: the Framingham study. *J. Bone Miner. Res.*, 8(5), 567-573.
- Frees, E. and Valdez, E. (1998) Understanding relationships using copulas. *North Amer. Actuarial J.*, 2(1), 1-25.
- Greene, W. H. (2000) *Econometric Analysis*, 4th edn. New York: Prentice Hall.
- Marcus, R. et al. (1994) Correlates of bone mineral density in the postmenopausal estrogen/progestin interventions trial. *J. Bone Miner. Res.*, 9(9), 1467-1476.
- NHANES III data. (2005) National Center for Health Statistics. Hyattsville, MD. Available at <http://www.cdc.gov/nchs/products/elec-prods/subject/nhanes3.htm>. [18 March 2007].
- Oakes, D. (1989) Bivariate survival models induced by frailties. *J. of the Amer. Statist. Assoc.*, 84, 487-493.
- Ravn, P. et al. (1999) Low body mass index is an important risk factor for low bone mass and increased bone loss in early postmenopausal women. Early Postmenopausal Intervention Cohort (EPIC) study group. *J. Bone Miner. Res.*, 14(9), 1622-1627.
- Reid, I. R. et al. (1992) Determinants of total body and regional bone mineral density in normal postmenopausal women—a key role for fat mass. *J. Clin. Endocrinol. Metab.*, 75(1), 45-51.
- SAS Institute, Inc. (2004) SAS System V9 for Windows [Computer program]. Carey, NC: SAS Institute, Inc.
- Schwarz G. (1978) Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.

- Shemyakin, A. E. and Youn, H. (2006) Copula models of joint survival analysis
Applied Stochastic Models in Business and Industry., 22(2), 211-224.
- Shih, J. H. and Louis, T. A. (1995) Inferences on the association parameter in
copula models for bivariate survival data. *Biometrics*, 51, 1384-1399.