



Max-Planck-Institut für demografische Forschung  
Max Planck Institute for Demographic Research  
Konrad-Zuse-Strasse 1 · D-18057 Rostock · GERMANY  
Tel +49 (0) 3 81 20 81 - 0; Fax +49 (0) 3 81 20 81 - 202;  
<http://www.demogr.mpg.de>

---

MPIDR WORKING PAPER WP 2005-011  
APRIL 2005

## **Subjective Well-Being and Mortality in Chinese Oldest Old**

Li, Qiang ([li@demogr.mpg.de](mailto:li@demogr.mpg.de))

---

This working paper has been approved for release by: James W. Vaupel ([jwv@demogr.mpg.de](mailto:jwv@demogr.mpg.de))  
Head of the Laboratory of Survival and Longevity.

© Copyright is held by the authors.

Working papers of the Max Planck Institute for Demographic Research receive only limited review.  
Views or opinions expressed in working papers are attributable to the authors and do not necessarily  
reflect those of the Institute.

## **Subjective Well-Being and Mortality in Chinese Oldest Old**

Li, Qiang

Max Planck Institute for Demographic Research, Rostock, Germany

Address for correspondence

Li, Qiang

Max Planck Institute for Demographic Research

Konrad-Zuse-Str. 1, 18057, Rostock, Germany

Tel: 0049-0381-2081258

Fax: 0049-0381-2081558

Email: [Li@demogr.mpg.de](mailto:Li@demogr.mpg.de)

## **Abstract**

The present study investigates the relationship between subjective well-being (SWB) and mortality risk, using a large sample ( $N=7852$ ) from the Chinese Longitudinal Healthy Longevity Study (age range 80-105) conducted in 2000 and 2002. Initially, we intended to contribute to the understanding of system relations between SWB, mortality risk, and unobserved heterogeneity by treating SWB as an endogenous variable, using a multi-process model. However, failure to identify unobserved heterogeneity in the mortality equation prevents us from employing this model. Given this limitation, the study examines three issues. First, we argue that the mortality model with duration dependency on the age of the study subjects is specified and that the model with duration dependency on time since the interview is misspecified. Second, we address problems associated with the identification of unobserved heterogeneity in the mortality equation. Third, we examine the association between SWB and mortality risk in the Chinese oldest old as well as the risk pattern by gender, without considering unobserved heterogeneity. We find that SWB is not a significant predictor of mortality risk when we control for socio-demographic characteristics and health status. Health plays a very important role in the relationship between SWB and mortality risk in the oldest old. Gender differences in the predictive pattern of SWB on this risk are negligible in the sample.

## 1. Introduction

Prior studies have shown that subjective well-being (SWB) is associated with mortality risk at old age (Danner, Snowdon, and Friesen 2001; Deeg et al. 1989; Levy Slade, and Kunkel 2002; Maier and Smith 1999; Smith et al. 2002). Two possible explanations have been put forward: a) SWB may have a direct impact on this risk (Levy, Slade, and Kunkel 2002); b) negative evaluations of SWB may not be in themselves causing an increase in mortality risk but rather reflect potential causes from other, unobserved domains of functioning (for example, intellectual, health) (Maier and Smith 1999; Smith et al. 2002). Older people with poor health, for example, possibly have an elevated mortality risk but tend to report lower levels of SWB. Which of the two explanations applies to the association of SWB with old age mortality risk?

Some scholars (e.g., Doblhammer and Oeppen 2003; Lillard and Panis 1996; Vaupel, Manton, and Stallard 1979) have pointed out that ignoring unobserved heterogeneity, as done in many previous studies, can lead to biased estimates of mortality risk associated with explanatory covariates. The debate on the two explanations above may be due to unadjusted unobserved (health) heterogeneity in past studies. With respect to the extent to which unobserved functioning affects the relationship between SWB and mortality risk, one possible pathway is that unobserved (health) functioning affects both SWB and this risk. This study attempts to examine the system of relations between SWB, mortality risk, and unobserved heterogeneity by specifying a multi-process (or joint) model that treats SWB as an endogenous variable. The statistical approach is developed by Lillard and Panis (2003).

We investigate such system relations in the oldest old, defined here as aged 80 and older. Some researchers have concluded that health is strongly correlated with SWB (e.g., George and Landerman 1984; Kunzmann, Little, and Smith 2000; Larson 1978; Okun et al. 1984; Smith et al. 1999, 2002; Wilson 1967). Moreover, the magnitude of health effects on SWB increases with age (George and Landerman 1984; Smith et al. 2002), that is, health is more critical to SWB at very old age. In other words, the role of health in the correlation between SWB and mortality risk in the oldest old may become increasingly evident. The small sample size of the oldest old in previous

studies impedes a meaningful analysis of this subpopulation, however. We therefore know relatively little about the relationship between SWB and mortality risk in the oldest old. Many studies have proposed that the oldest old (Fourth Age) are very different from the young old (Third Age: e.g. Baltes and Smith 2003; Maier and Smith 1999; Smith et al 2002; Suzman, Willis, and Manton 1992; Zeng et al. 2002). In general, oldest old populations typically have a higher proportion of women than men, higher levels of co-morbidity, and a greater consumption of medical and care services. Especially individuals born in China between 1895 and 1920 - they comprise the oldest old observed in 2000 - exemplify a set of characteristics that are cohort-specific. They have much lower levels of education and a much higher likelihood of widowhood. They have experienced the greatest changes in Chinese history, for example, civil wars until 1950, the Second War, the founding of the People's Republic of China, the Cultural Revolution, and economic and social reforms since 1978. These specific cohort characteristics may play an important role in the potential of maintaining positive well-being, and subsequently in the relationship between SWB and mortality risk. This study addresses system relations such as these by using a large sample of oldest old in China that yields precise estimations of the association between SWB and mortality risk.

It is a well-established fact that life expectancy is higher for women than for men. The age-adjusted risk of mortality thus is lower for women than for the opposite sex (Idler 2003). Women furthermore report more negative affect and emotional upset than do men (e.g., Smith et al. 1999). Does men's higher SWB have a different relationship with higher mortality rates compared to the lower SWB and lower mortality rates of women? Gender reflects differences of resources and social relations in the Chinese oldest old. In view of these considerations, this paper examines the system relations between SWB, mortality, and unobserved heterogeneity for men and women separately.

## **2. The method**

### **2.1 The sample**

The sample in this research is taken from the Chinese Longitudinal Survey on Healthy Longevity (CLSHL). Detailed information on the CLSHL sample, design, and assessment battery is published elsewhere (Zeng et al. 2001, 2002). The CLSHL comprises three waves of data collection: a baseline in 1998 and a second and third wave in 2000 and 2002, respectively. The base line survey was conducted in 631 randomly selected counties and cities of the 22 provinces in China. All centenarians living in these locations were interviewed on a voluntary basis. For each centenarian, one octogenarian (aged 80-89) who lived nearby and one nonagenarian (aged 90-99) who lived nearby, with a pre-designated age and sex, were matched and interviewed. "Nearby" is loosely defined to denote the same village or street if such an individual was available, or in the same town or sampled county or city (Zeng et al. 2002). Pre-designated age and sex were employed to have approximately equal numbers of male and female nonagenarians and octogenarians. For a centenarian aged 102, for example, a nearby octogenarian aged 82 and a nearby nonagenarian aged 92 were matched and interviewed. The sex of the octogenarian and nonagenarian interviewees was randomly determined with a view to have approximately equal numbers of males and females at each age from 80 to 99. If such individuals were not found, an alternative individual of the same sex and in the same five-year age group (80-84, 85-89, 90-94, 95-99) was selected (Zeng et al. 2002). The total valid sample size of the baseline data is 8805 elderly persons aged 80 to 105.

Due to the fact that the data related to SWB is of higher quality in the second wave than in the baseline survey (for detailed information, see SWB measures), we use data of respondents in Waves 2 and 3 only. The mortality information is obtained from Wave 3. As to the second wave, 4690 participants of the baseline survey had survived, 3264 had died and 850 were lost to follow up. We added 6274 new participants. The total valid sample was, therefore, 10964 oldest old aged 80 to 105 in 2000. Regarding the third wave, 6219 individuals were alive, 3240 had died between the two waves

and 1505 were lost to follow-up<sup>1</sup>. Because we have no vital information on 865 of the dead and 742 survivors, the final sample reduces to 7852 (5477 for survival and 2375 for death, respectively)<sup>2</sup>. Detailed information on sample sizes is given in Figure 1.

---Figure 1 insert here---

## 2.2 The model

A joint or multi-process model is specified to evaluate the simultaneous relationships between SWB, mortality risk, and unobserved heterogeneity. The model contains two sets of sub-models for mortality risk and SWB. The hazard model for the force of mortality in the oldest old is:

$$\ln u(t) = y(t) + \beta_1 X + \beta_2 H + \beta_3 S + \delta , \quad [1]$$

where

$u(t)$  denotes the mortality risk of the oldest old,

$y(t)$  the piecewise linear baseline log-hazard of mortality,

$X$  the socio-demographic characteristics of the oldest old,

$H$  the health status of the oldest old (self-reported ADL, self-rated health, and cognitive functioning),

$S$  subjective well-being, and

$\delta$  denotes unobserved heterogeneity for the oldest old in mortality.

The ordered probit model for the subjective well-being of the oldest old is:

$$y = \begin{cases} 0 & \text{if } y^* < \tau_1 \\ 1 & \text{if } \tau_1 < y^* < \tau_2 \\ 2 & \text{if } \tau_2 < y^* < \tau_3 \\ \vdots & \\ n & \text{if } \tau_n < y^* \end{cases} , \text{ where } y^* = \alpha_0 + \alpha_1 X + \alpha_2 H + \varepsilon + u , \quad [2]$$

<sup>1</sup> Because we do not know whether or not the 1505 individuals died during the study period, we delete them from the data.

<sup>2</sup> Among the 865 dead, the following information is not available: date of death: 10, SWB: 802, years of schooling: 26, Mini-Mental State Examination: 469, self-reported ADL: 9, self-rated health: 460. Among the 742 survivors, information is lacking on the date of information: 1, SWB: 678, years of schooling: 40, marital status at the time of the 2000 interview: 1, Mini-Mental State Examination: 275, self-reported ADL: 17, and self-rated health: 294.

where

$y$  denotes the subjective well-being of the oldest old,

$y^*$  a latent propensity dependent on independent variables,

$X$  the socio-demographic characteristics of the oldest old,

$H$  the health status of the oldest old (self-reported ADL, self-rated health, and cognitive functioning),

$\varepsilon$  unobserved heterogeneity for the oldest old in the propensity to SWB,

$u$  is a random variable, and

$\tau_i$  represents the thresholds of the ordered probit model that needs to be estimated.

The heterogeneity components  $\delta$  and  $\varepsilon$  are assumed to follow a bivariate normal distribution, represented by

$$\begin{pmatrix} \delta \\ \varepsilon \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\delta}^2 & \sigma_{\delta\varepsilon} \\ \sigma_{\delta\varepsilon} & \sigma_{\varepsilon}^2 \end{pmatrix} \right) \text{ where } \sigma_{\delta\varepsilon} = \rho_{\delta\varepsilon} \sigma_{\delta} \sigma_{\varepsilon},$$

where  $\rho_{\delta\varepsilon}$  is the correlation coefficient of heterogeneity components  $\delta$  and  $\varepsilon$ . The critical test of the simultaneous relationship between SWB, mortality, and unobserved heterogeneity is whether  $\rho_{\delta\varepsilon} = 0$ . We adopted the following decision strategy: First, if  $\rho_{\delta\varepsilon} = 0$ , then there is no individual unobserved functioning affecting both SWB and mortality. If a significant effect of SWB on mortality exists, then we accept the first explanation: that SWB has a direct impact on mortality risk. Second, if  $\rho_{\delta\varepsilon} < 0$ , which indicates that individual unobserved functioning affects both SWB and mortality, then this negative correlation suggests that the oldest old with unobserved heterogeneity tend to report lower levels of subjective well-being and have higher mortality risks. If at the same time the effect of SWB on mortality risk disappears, then we accept the second explanation: that negative evaluations of SWB are not the cause for increasing mortality risks but rather reflect potential causes from other, unobserved domains of functioning. Third, if  $\rho_{\delta\varepsilon} < 0$  and if there is a significant effect of SWB on mortality risk, then we accept both explanations, suggesting that, on the one hand, SWB has a direct impact on mortality risk, but on the other, that it also reflects the unobserved effect of functioning on mortality.

We use the software package aML2.0 (Lillard and Panis 2003). The estimation of the model is based on the maximization of the joint likelihood function for the mortality and SWB model.

For the mortality hazard model, previous studies have suggested two different duration dependencies: (a) on the age of the study subjects (e.g., Doblhammer and Oeppen 2003; Lillard and Panis 1996; Vaupel, Manton, and Stallard 1979), and (b) on the time since interview (e.g. Bath 2003; Spiers, et al. 2003; Levy, Slade, and Kunkel 2002). These dependencies have different logics underlying the specification<sup>3</sup>. This is investigated in the results section. To apply the former duration dependency, we set age of the oldest old at 2000 interview as the origin of the log-baseline intensity. To get the latter, we use zero as its origin. Following our empirical analyses, we will provide a further discussion on this.

### **2.3 Measures of variables**

Above, we simply provided a list of block variables. Below, we provide detailed information on these variables. Note that all our explanatory variables in the hazard model are time-fixed. SWB is our central variable; it is not only the explanatory variable in the hazard model but also the outcome in the ordered probit model. The other explanatory variables in the hazard model are the same as those in the ordered probit model. We introduce them together, therefore.

#### **Subjective well-being (SWB)**

The CLSHL includes six items designed to measure five aspects of SWB: (1) life satisfaction, (2) a positive attitude to life, (3) the absence of agitation, (4) the absence of loneliness and (5) a positive attitude toward one's aging. For all measures, participants are asked to indicate on a 5-point scale how well items describe their subjective outlook. Psychometric work indicates that all of the inter-item correlations are positive. The evaluation of the 1998 baseline survey data shows that SWB-related questions are not adequate mainly because some illiterate oldest old, especially

---

<sup>3</sup> I would like to thank Prof. Jan Hoem for his insightful and valuable interpretation of the two different duration dependencies.

centenarians, do not understand the questions on SWB (Zeng et al. 2002). We revised these questions in the 2000 and 2002 follow-up surveys. This measure significantly improves the reliability of the items to such a degree that the internal consistency of the items determined by Cronbach's alpha increases from 0.63 in the baseline data to 0.72 in the second wave data. Mainly for this reason, the present study uses data on participants in Waves 2 and 3. Additional tests to determine reliability for three different age groups in 2000 yield Cronbach's alphas of 0.72 for age group 80-89, 0.71 for 90-99 and 0.71 for 100-105 respectively.

A factor analysis of the six items yields one factor with an eigenvalue greater than 1.0, reflecting the latent variable SWB. Table 1 show that all six items load on this factor. Loading ranges in magnitude from 0.46 to 0.63,  $M=0.53$ . Factor scores for SWB are calculated using all six items. These items show a convergent validity (presented in Table 1) when correlated with SWB factor scores. Correlation coefficients range from 0.48 to 0.72.

---Table 1 insert here---

For the purpose of the present analysis, the scores of the composite measure are divided into four levels according to quartiles (quartile 0-25%, quartile 26-50%, quartile 51-75%, quartile 76-100%), with a higher quartile indicating a higher level of SWB.

### **Socio-demographic characteristics**

Five measures of socio-demographic characteristics are employed. Age is collapsed into five categories: 80-84, 85-89, 90-94, 95-99, and 100-105. Sex has two levels: female and male. The type of residence is defined as urban and rural. Years of schooling is divided into three groups: no schooling, 1-6 years of schooling and 7 + years of schooling. Marital status at the time of the 2000 interview is grouped into two groups: married and unmarried. The group "unmarried" includes separated, divorced, widowed, and never married persons.

## Health status

Three measures of health status are used. Functional condition is assessed by self-reported ADL, which is grouped into three levels: no, one, and two or more functioning limitations. Self-rated health is measured by a single-item question, namely “How do you rate your health at present?” Participant answers are placed into four categories: very good, good, fair, and bad<sup>4</sup>.

The Mini-Mental State Examination (MMSE) is used to assess the cognitive functioning of the Chinese oldest old under study. MMSE is culturally translated into the Chinese language, based on established international standards for the MMSE questionnaire, and carefully tested in pilot survey interviews (Zeng et al. 2003). It includes brief measures of orientation, registration, attention, and calculation, recall, and language, with scores ranging from 0 to 23. Scores of MMSE are graded into two levels, using a median split (0-50%, 51-100%), with “51-100%” indicating a higher level of cognitive functioning.

Table 2 shows a breakdown of the sample according to the levels of the factors (see the second column). It also displays the number and percentage of people who died (see the third and fourth column).

---Table 2 insert here---

## 3. Results

We first estimate the hazard model and ordered probit model separately, and then run the joint model. Using the hazard model/ordered probit model, we begin with the model without unobserved heterogeneity. We then proceed to the model that includes unobserved heterogeneity in order to compare both models.

### 3.1 Mortality risk in the oldest old (*hazard model*)

We first run the model with duration dependency on the age of the oldest old, followed by the model with duration dependency on time since the 2000 interview. Upon presenting the results of the hazard model, a graph is employed for the log-

---

<sup>4</sup> Because of the small number of respondents who evaluated their SRH as “very bad” (n=54), we added them to the group of respondents who answered “bad”.

baseline spline, and a table for the estimations of the categorical covariates. With the intercept (log-hazard at the beginning of the spell if all categorical covariates are zero), nodes<sup>5</sup>, and slope parameters, we can plot the baseline spline. In fact, such a graph is the only sensible way of displaying the results of the log-baseline intensity for interpretation (Beise and Volland 2002). The tables display the relative risk for the reference group (anti-log of the estimated coefficients in aML), for the estimation of categorical covariates as well as the log-baseline intensity.

### 3.1.1 The model with duration dependency on the age of the oldest old

Equation 1 in the model describes the general format of the hazard model. When we adopt duration dependency on the age of the oldest old, the mortality equation is:

$$\ln u_i(x) = y_1(x) + \beta_1 X + \beta_2 H + \beta_3 S + \delta, \quad (3)$$

where  $\ln u_i(x)$  is the logarithm of the force of mortality at attained age  $x$ , and  $y_1(x)$  is the log-baseline spline with nodes at 85, 90, 95, and 100 years of age. These nodes are the same for all individuals. The log-baseline hazard duration is always a piecewise-linear spline (also known as generalized Gompertz or piecewise-linear Gompertz) in aML. The remaining terms are the same as in Equation 1.

Age group<sup>6</sup> is not included in this model, mainly because the model already contains age. The joint estimation of age and age group fails to converge. Generally, age group and age can be used in the same model, but in our case<sup>7</sup>, there is little variation in the period effects. Consequently, putting age group and age together may be problematic.

The graph in Figure 2 shows the shape of the mortality hazard for the oldest old. The hazard of mortality increases typically with age in the two years of observation, though the slopes are very flat during the whole spell.

---Figure 2 insert here---

Table 3 shows that women, the higher educated and married oldest old have lower mortality risks. A better health status is associated with a lower mortality risk, that is,

---

<sup>5</sup> Nodes are sometimes called knots or bend points. Nodes are used to cut the spell duration into several pieces. For each piece, the log-baseline intensity is assumed to be linear.

<sup>6</sup> A time-fixed variable, we therefore considered it as the indirect definition of birth cohort.

<sup>7</sup> There are only two wave surveys and the interval between two waves is relatively small.

higher levels of MMSE, no functioning limitation and better self-rated health are related to lower mortality risk.

---Table 3 insert here---

The findings on the relationship between SWB, our central important variable, and mortality risk are interesting. As we already know, health plays an important role in the relationship between SWB and mortality risk. We, therefore, first run the model without adjustment with three health indicators (Model 1), and then the model with adjustment with the same three health indicators (Model 2). Model 1 shows that without controlling for the health status, SWB is a significant predictor of mortality risk. However, when we do control for the health status, SWB is not significantly related to mortality risk in the oldest old. This means that measured health factors may explain the effect of SWB on mortality risk in our sample. Recalling the question we asked earlier (Which of the two explanations on the relationship between SWB and mortality risk applies?) our current result seems to support the second explanation: SWB may have no direct effect on mortality risk but rather reflect the effects of other health factors. Estimations of the mortality risk associated with risk factors may be biased because we omit unobserved heterogeneity. Thus we next introduce unobserved heterogeneity into the hazard model and use a multi-process model to further investigate this matter.

### **3.1.2 Failure to identify unobserved heterogeneity in mortality equation**

Unfortunately, the estimation with unobserved heterogeneity fails to converge. We suspect that it is close to being unidentified in our sample. Many reasons may have led to the failure, two of which we consider in this paper: One explanation is related to aML *per se*, that is, to some extent its assumption of piecewise linear log-baseline intensity is so flexible that we were not able to capture unobserved heterogeneity. In other words, the effect of unobserved heterogeneity is attributed to the piecewise-linear log-baseline intensity. The other possible reason is that the risk factors in our model help to determine unobserved heterogeneity rather than do so conversely (Hoem 1989). In view of these reasons, we take further explorative steps to identify unobserved heterogeneity.

To eliminate the influence of the assumption of piecewise linear log-baseline intensity, we suppose that the log-baseline intensity is constant. This is feasible

because the slopes of the log-baseline intensity are so flat in the two years of observation that they could be negligible. We arrive at the constant log-baseline intensity in aML by defining a spline with intercept and without nodes, and by fixing the slope coefficient to zero. Then we can take into account the age group as covariate in our model. The mortality equation is now:

$$\ln u_i(x) = C + \beta_1 X + \beta_2 H + \beta_3 S + \delta, \quad [4]$$

where  $C$  is the constant log-baseline intensity and age group is included in the socio-demographic block  $X$ . The remaining terms are the same as in Equation 1.

The model without unobserved heterogeneity runs well. Table 4 shows that the older the participants are, the higher is the mortality risk. The relative risks of other covariates are similar to those of the earlier model. Again, the additional inclusion of unobserved heterogeneity in the model makes the estimation fail to converge.

---Table 4 insert here---

To adjust our model to the consideration that risk factors help determine unobserved heterogeneity rather than conversely, we estimate the model with only one covariate: sex. We choose sex as the only covariate mainly because the number of females and males is almost the same in our study. Our central variable, SWB, is not considered here because the aim is to see whether unobserved heterogeneity can be identified. Again, even in a model that has sex as the only covariate, the inclusion of unobserved heterogeneity makes the estimation impossible.

None of the two steps has led to the identification of unobserved heterogeneity in the mortality equation in our data set<sup>8</sup>.

### 3.1.3 Model with duration dependency on time since the 2000 interview

In the next model, we adopt duration dependency on the time since the 2000 interview. The mortality equation is:

---

<sup>8</sup> Besides the above two explorative steps, we estimate the model by fixing empirical values to the unobserved heterogeneity. If the model fit improved, the unobserved heterogeneity in the model would be identified. However, the model fit worsened and the estimation of the coefficients of variables slightly increased when compared to those in the model without fixed unobserved heterogeneity. Furthermore, the higher the value of fixed unobserved heterogeneity, the lower the log-likelihood and the higher the estimation of the variables' coefficients. This means that it is hard to identify unobserved heterogeneity.

$$\ln u_i(x_{i_0}, t) = y_2(t) + \sum_{j=1}^5 \alpha_j g_j(x_{i_0}) + \beta_1 X + \beta_2 H + \beta_3 S + \delta, \quad [5],$$

Where  $x_{i_0}$  denotes the age at the 2000 interview,  $t$  represents the time since the 2000 interview, and  $g_j(x_{i_0}) = 1$  if this age is in group  $j$ ,  $g_j(x_{i_0}) = 0$  otherwise ( for  $j = 1, 2, 3, 4, 5$ ). The linear spline  $y_2(t)$  has no intercept and has nodes at ages duration  $t = 0.5, 1$ , and  $1.33^9$ , which corresponds to ages  $x_{i_0} + 0.5, x_{i_0} + 1$ , and  $x_{i_0} + 1.33$ . Equation 5 means that an individual who is in age group  $j$  at the 2000 interview will have the force of mortality  $\ln u_{ij}(x_{i_0}, t) = y_2(t) + a_j + \text{more}$ . That is to say, the forces of mortality for individuals in the various initial age groups<sup>10</sup> differ from each other only through different implicit intercepts  $\{a_j\}$ .

The correspondence between  $a_1, a_2, a_3, a_4$  and  $a_5$  on the one hand, and the parameters we have estimated on the other, is as follows: “Constant” is the maximum likelihood estimate for  $a_1$ . The parameters for “age group 85-89”, “age group 90-94”, “age group 95-99”, and “age group 100-105” are the antilogarithm of the maximum likelihood estimators for  $a_2 - a_1, a_3 - a_1, a_4 - a_1$  and  $a_5 - a_1$ , respectively.

Due to a lacking intercept, we use zero as the beginning of the linear spline when we plot the log-baseline intensity. As can be seen from Figure 3, the mortality hazard fluctuates randomly during the two years of observation.

---Figure 3 insert here---

Table 5 shows that the relative risks associated with the covariate levels without controlling for unobserved heterogeneity are very similar to those in the former model. Also, SWB is not significantly associated with mortality risk when we control for other risk factors. Interestingly, unobserved heterogeneity is identified in this model specification. A model with unobserved heterogeneity greatly improves the model. Unobserved heterogeneity is significantly different from zero and thus significantly associated with mortality risk. Failure to account for unobserved

---

<sup>9</sup> In the analyses, month is the time unit. In the text, we change the time unit from month to year in order to facilitate understanding.

<sup>10</sup> Or equivalently, in the various cohorts, for the age groups can be seen to represent different birth cohorts.

heterogeneity results in the bias estimation of the effects of covariates and log-baseline intensity.

---Table 5 insert here---

However, this model is logically misspecified on the following grounds: It indicates that mortality risk at attained age  $x_{i_0} + t$  should depend on  $x_{i_0}$  and  $t$  separately when instead it ought to depend on their sum. Further, the model shows that the log-baseline intensities for different respondents have different nodes and therefore are different from one individual to another. This contradicts the very idea of a log-baseline hazard.

Evidently, these two problems have not prevented the estimation process from converging, nor have they prevented the ensuing estimates for other parameters, as indicated in Table 5. The logic underlying the specification in this model remains unsatisfactory.

The model with duration dependency on the age of the oldest logically is a much more satisfactory specification of the forces of mortality. The mortality risk for the oldest old in this model depends on the sum of the age at interview and time since interview. The nodes of the log-baseline intensity are the same for all individuals.

Failure to identify unobserved heterogeneity in the mortality equation prevents further work from being undertaken, that is, with the multi-process model that is specified by the relationship between the unobserved heterogeneity in the mortality model and the one in the ordered probit model.

### **3.2 Gender difference in the relationship between SWB and mortality**

Before running the hazard model for men and women separately, we run descriptive analyses comparing the survival status, SWB, and the covariates in the two gender groups (see Table 6). There are significant gender differences in the survival status when we control for age<sup>11</sup>. Women are less likely to be married and

---

<sup>11</sup> There are no significant gender differences when we do not control for age.

have fewer years of education. They also report a lower level of SWB, self-rated health, have more functional disability, and score less well in MMSE than do men.

---Table 6 insert here---

It is in the context of significant gender differences that we perform hazard analyses separately for men and women (see table 7). Similarly to the findings in the whole sample, SWB is a significant predictor of mortality risk for both sexes if no adjustment in the health indicators is made. Again, when we control for health status, SWB is significantly related to mortality risk for neither men nor women. The effect of SWB on mortality risk may be attributed to the health status of the two sexes. Differences between the two gender groups are found on the log-baseline intensity and as regards the marital status at the time of the 2000 interview. The significant increase of log-baseline intensity emerges in less advanced ages for men, while such increase applies to women at more advanced ages, although the increase is small in both groups. The marital status at the time of the 2000 interview is a significant predictor of mortality risk only for men. Unmarried men have a mortality risk that is higher by 49% compared to married oldest old men.

---Table 7 insert here---

#### **4. Discussion**

In this study, we aimed at contributing to the understanding of the relationship between SWB and mortality in the oldest old by treating SWB as an endogenous variable, using a multi-process model in a sample of individuals aged 80 to 105 years. However, failure to identify unobserved heterogeneity in the mortality equation prevented us from employing the multi-process model. Given this limitation, we proceeded along the following lines: First, we argued that the mortality model with the duration dependency on the age of the study subjects was specified and that the model with duration dependency on the time since interview was misspecified. Second, we described the problems associated with the identification of unobserved heterogeneity in the mortality equation. Third, we examined the association between SWB and mortality risk in Chinese oldest old as well as the risk pattern by gender, without considering unobserved heterogeneity.

#### **4.1 The specification of duration dependency on the age of the study subjects / the misspecification of duration dependency on the time since interview in the mortality hazard model**

We argued that the mortality hazard model with duration dependency on the age of the oldest old was correctly specified on the following grounds: The force of mortality in the oldest old depended on the sum of the age at and time since interview. Further, the nodes of the log-baseline intensity were the same for all individuals. Another two reasons are noteworthy here. First, the hazard model has an intercept and this is consistent with the specification of the piecewise linear log-baseline intensity model. As we know, the piecewise linear intensity model has an intercept, and the intercept corresponds to the log-hazard at the beginning of the spell if all covariates are zero. Second, the plot of the log-baseline intensity provides us with a clear and empirical interpretation of the baseline intensity. The force of mortality in the oldest old typically increases with age, although the slope is flat. Some scholars have pointed out that the estimation of the baseline intensity may provide some information on the type of time-dependence in a set of episode data and that it may give us some insights as to whether a fully parametrical model would be appropriate or not (Blossfeld et al., 2002). Obviously, the model with duration dependency on the time since the 2000 interview could not meet the four requirements. Although in our case the misspecification does not prevent an estimation of covariate coefficients, a further investigation of the harm such misspecification does in the long run and the presence of time-varying covariates is needed due to the lack of time-varying covariates and the short observation time in our study.

#### **4.2. Problems associated with the identification of unobserved heterogeneity in the mortality equation**

Unobserved heterogeneity in the mortality equation was hard to identify. To arrive at the identified estimation, we took two further explorative steps with aML. First, we employed constant log-baseline intensity. Second, we simplified the model by using one covariate only. None of this led to an improvement. It seems that something

beyond aML techniques results in failure to identify unobserved heterogeneity in the mortality equation. We proceeded to use a gamma distribution with unobserved heterogeneity, as suggested by Vaupel, Manton, and Stallard (1979). Although we get the converging results with sex as the only covariate and unobserved heterogeneity with Stata (we do not list the results here), the unobserved heterogeneity is not significantly associated with the mortality risk. Clearly, unobserved heterogeneity should be significant in such a case.

Lillard and Panis (1996) have pointed out that “conceptually, mortality risk also might be heterogeneous. Because ‘you only live once,’ heterogeneity in the mortality equation is hard to identify from a single occurrence.” A further step, that of a linear combination of heterogeneity in marriage and divorce, also failed in their research. Hoem (2004) stated that “we know in principal, identification is much more difficult when you only have at most one event per individual than when you can have several events.”<sup>12</sup> Another way of looking at our finding is that our failure to identify unobserved heterogeneity in the mortality equation is in line with Hoem’s assertion in a paper (1989) where he discussed the limitations of current heterogeneity techniques.

### **4.3. The effect of SWB on mortality risk in the Chinese oldest old**

The present study examined the predictability of SWB’s effect on mortality in the oldest old in China. Although we failed to treat SWB as an endogenous variable in a multi-process model used to investigate the relationship between SWB and mortality, our study is meaningful in four respects.

First, although SWB is not a statistically significant predictor of mortality in the oldest old in China after controlling for socio-demographic characteristics and health status, our study points into the expected direction: the higher the level of SWB, the lower the level of mortality risk. To some degree, our finding is not consistent with earlier findings that SWB is a significant predictor of mortality risk. This may be the result of three reasons. First, the difference may stem from the use of different confounder covariates in the model. In our model, the measurement of health status

---

<sup>12</sup> The quote is from a manuscript by Prof. Jan Hoem arising from a discussion between us on the two different duration dependencies and the identification of unobserved heterogeneity.

includes self-rated health, MMSE, and self-reported ADL. By contrast, Levy, Slade, and Kunkel (2002) did not control for MMSE, and Maier and Smith (1999) did not do so for functional health, for example. Furthermore, when we exclude MMSE from the model, SWB is significantly related to mortality risk<sup>13</sup>. Second, the different measurement of SWB in our study may produce inconsistent results. A significant effect of SWB on mortality risk is only found in the self-perceptions on aging and loneliness, which have been two SWB sub-components in past research (Danner, Snowdon, and Friesen 2001; Deeg et al. 1989; Levy, Slade, and Kunkel 2002; Maier and Smith 1999; Smith et al. 2002). The measurements of SWB in our study, however, include three additional constructs: life satisfaction, a positive attitude to life, and agitation. These three components have not predicted survival status in previous studies. A composite of these five constructs may reduce the effect of self-perception on aging and loneliness on mortality risk to such an extent that SWB is not significantly related to mortality. Finally, sample selection may be another reason. We delete the missing values of SWB, that is, the sample includes only the participants who have complete information on all six items that make up SWB. In other words, SWB is more homogeneous in the current sample than in the original one. Recently, Li (2003) found that an additional level of missing SWB values was significantly predictive of mortality risk in the oldest old.

Second, health status plays a critical role in the relationship between SWB and mortality risk in the oldest old. Without control for health indicators, SWB is significantly associated with this risk. However, when controlling for health, SWB loses its significant effect. This means we do find that observed health indicators are very important in explaining the association between SWB and mortality risk in the oldest old. Although failure to identify unobserved heterogeneity prevents us from understanding this issue to greater detail, our finding so far supports Maier and Smith's explanation that negative evaluations of SWB are not a cause for increased mortality risk but rather may reflect potential causes from other domains of functioning.

Third, in contrast to previous studies that investigate the relationship between SWB and mortality risk at old age or mainly in the young old, our study is focussing

---

<sup>13</sup> The exclusion of self-reported ADL does not result in a significant effect of SWB on mortality risk.

exclusively on the oldest old. Moreover, our large sample of the oldest old allows for the precise estimation of the relationship between SWB and mortality risk. It is of interest that we do not find a significant effect of SWB on mortality risk when we control for socio-demographic characteristics and health status. With regard to the increasing magnitude of the effects of health on SWB with age, as mentioned earlier, we suggest that, in the oldest old, health may play a far more important role in the relationship between SWB and mortality risk than in the young old.

Fourth, although there exists an age-adjusted gender difference of SWB in the oldest old and a lower mortality risk for women than for men, we do not find different predictive patterns of SWB on mortality risk between men and women. SWB is not a significant predictor of mortality risk for both men and women.

Admittedly, this study has its limitations in several aspects. First, our measurement of SWB does not cover all domains of SWB, suffice it to mention positive affect, happiness, personal growth, satisfaction with social relationships, and autonomy (e.g., Diener 1984; Lawton 1975, 1991; Ryff 1989). Consequently, we were not able to investigate the full relationship between SWB and mortality risk. Second, due to the very low education level of the Chinese oldest old under study, especially the centenarians, they did not at times fully understand the questions on SWB. Although we revised the questions in Wave 2, the factor analysis is just acceptable. Third, our observation covers a two-year period only. A deeper understanding of the relationship between SWB and mortality risk in the long run is needed.

To sum up, the present study yields precise estimations of the relationship between SWB and mortality risk in the oldest old, using a large sample from CLHLS. Statistical controls for health factors reveal that SWB is not related to this. Moreover, our discussion on the two different duration dependencies and the problems of identifying unobserved heterogeneity in mortality equation is relevant to the specification of mortality models.

## **Acknowledgements**

I am deeply grateful to my mentor, Dr. Heiner Maier, for his help during the whole spell of this study, beginning with choosing a topic, subsequent discussions on the sample and variables, his suggestions on the identification of unobserved heterogeneity and the structure of the paper, and finally his comments on the paper. I would like to extend my thanks to Prof. Jan Hoem, too, for his insightful and valuable interpretation of two different duration dependencies in mortality equation, and his suggestions as to the identification of unobserved heterogeneity and the investigation of risk patterns by gender.

This study was carried out as part of Prof. Jan Hoem's course "Regression analysis for duration data (event-history analysis) incorporating selectivity/unobserved heterogeneity" during my stay at the Max Planck Institute for Demographic Research (Winter Semester of 2003/2004) for which I received financial support. For language editing, I would like to thank Susann Backer.

## References

- Baltes, P. B., Smith, J. (2003). "New frontiers in the future of aging: From successful aging of the young old to the dilemmas of the fourth age." *Gerontology*, 49. 123-135.
- Bath, P. A. (2003). "Differences between older men and women in the self-rated health-mortality relationship." *The Gerontologist* 43(3). 387-395.
- Beise, J., Volland, E. (2002). "A multilevel event history analysis of the effects of grandmothers on child mortality in a historical German population (Krummhörn, Ostfriesland, 1720-1874)." *Demographic Research*, 7(13). 467-497.
- Blossfeld, H. P., Rohwer, G. (2002). *Techniques of event history modeling: New approaches to causal analysis* (2nd ed.). USA: Lawrence Erlbaum Associates, Inc. Publishers.
- Danner, D. D., Snowdon, D. A., Friesen, W. V. (2001). "Positive emotions in early life and longevity: findings from the nun study." *Journal of Personality and Social Psychology*, 80. 804-813.
- Deeg, D.J.H., Zonneveld R. J. V., Maas P.J. van der, Habbema J. D. F.. (1989). "Medical and social predictors of longevity in the elderly: Total predictive values and interdependence." *Social Science and Medicine*, 29. 1271-1280.
- Diener, E. (1984). "Subjective well-being." *Psychology Bulletin*, 95. 542-575.
- Doblhammer, G., Oeppen, J. (2003). "Reproduction and longevity among te British peerage: the effect of frailty and health selection." *Proceedings of the Royal Society of London/B*, 270:1524. 1541-1547.
- George, L. K., Landerman, R. (1984). "Health and subjective well-being: A replicated secondary data analysis." *International Journal of Aging and Human Development*, 19. 133-156.
- Hoem, J. M. (1989). *Limitation of heterogeneity technique: Selectivity issues in conjugal union disruption at parity zero in contemporary Sweden*. Stockholm Research Reports in Demography. No. 56. The essence of this report was presented to the Conference on Convergent Questions in Genetics and Demography at the Univeristy of Michigan, Ann Arbir, 7-8 October 1988.
- Idler, E. L. (2003). "Discussion: Gender differences in self-rated health, in mortality, and in the relationship between two." *The Gerontologist*, 43(3). 372-375.
- Kunzmann, U., Little, T. D., Smith, J. (2000). "Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin Aging Study." *Psychology and Aging*, 15(3). 511-526.

- Larson, R. (1978). "Thirty years of research on the subjective well-being of older Americans." *Journal of Gerontology*, 33. 109-125.
- Lawton, M. P. (1975). "The Philadelphia Geriatric Center Morale Scale: A review." *Journal of Gerontology*, 30. 85-89.
- Lawton, M. P. (1991). "A multidimensional view of quality of life in frail elders." In J. E. Birren, J. E. Lubben, J. C. Rowe, D. E. Deutchman (Eds.), *The concept and measurement of quality of life in the frailty elderly* (pp. 3-27). San Diego, CA: Academic Press.
- Levy, B. R., Slade M. D., Kunkel S. R.. (2002). "Longevity increased by positive self-perception of aging." *Journal of Personality and Social Psychology*, 2. 261-270.
- Li, Qiang (2003). Subjective well-being and mortality in the oldest old in China. Mini-project for MP301 (Introduction to Event History Analysis instructed by Gunnar Andersson) in the winter semester 2002/03 in International Max Planck Research School for Demography.
- Lillard L. A., Panis. C.W.A. (2003). aML multilevel multiprocess statistical software, Version 2.0. Econ Ware, Los Angeles, California.
- Lillard L. A., Panis. C.W.A. (1996). "Marital status and mortality: the role of health." *Demography*, 33(3). 313-327.
- Maier, H., Smith, J. (1999). "Psychology predictors of mortality in old age." *Journal of Gerontology: Psychological Sciences*, 1. 44-54.
- Okun M. A., Stock, W. A., Haring, M. J., Witter, R. A. (1984). "Health and subjective well-being: A meta-analysis." *International Journal of Aging and Human Development*, 19. 111-132.
- Ryff, C. D. (1989). "Happiness is everything, or is it? Exploration on the meaning of psychological well-being." *Journal of Personality and Social Psychology*, 57. 1069-1081.
- Smith Jacqui, Fleeson W., Geiselman B., Settersten R. A. Jr., Kunzmann U.. (1999). Sources of well-being in very old age. In *The Berlin Aging Study: Aging from 70 to 100*. New York: Cambridge University Press. 450-471.
- Smith, J., Borchelt M., Maier H., Jopp D. (2002). "Health and well-being in the young old and oldest old." *Journal of Social Issues*, 4. 715-732.
- Spiers, N., Jagger, C., Clarke M., Arthur A. (2003). "Are gender differences in the relationship between self-rated health and mortality enduring? Results from three birth cohorts in Melton Mowbray, United Kingdom." *The Gerontologist*, 43(3). 406-411.
- Suzman, R. M., Willis, D. P., Manton, K. G. (1992). *The oldest old*. New York:

Oxford University Press.

Vaupel, J. W., Manton, K. G., Stallard, E. (1979). "The impact of heterogeneity in individual frailty on the dynamics of mortality." *Demography*, 16(3). 439-454.

Wilson, W. (1967). "Correlates of avowed happiness." *Psychological Bulletin*, 67. 294-306.

Zeng Yi., Vaupel J.W., Xiao Zhenyu, Zhang Chunyun, Liu Yuzhi. (2001). "The healthy longevity survey and the active life expectancy of the oldest old in China." *Population*,. (Engl Selection) 13(1), 95-116.

Zeng Y, Vaupel J.W., Xiao Zhenyu, Zhang Chunyun, Liu Yuzhi. (2002). "Sociodemographic and health profiles of the oldest old in China." *Population and Development Review*, 28(2). 251-273.

Zeng Yi, Liu Yuzhi, Linda K. G.(2003). "Gender differentials in the oldest old in China." *Research on Aging*, 1. 65-80.

**Table 1: Factor Loading Matrix**

	Item	Factor Loading	Correlation between item and SWB factor scores
1. Life satisfaction	How do you rate your life at present? 1=very bad; 2=bad; 3=so so; 4=good; 5=very good <sup>1</sup>	0.46	0.48
2. A positive attitude to life	Do you always look on the bright side of things? 1=never; 2=seldom; 3=sometimes; 4=often; 5=always <sup>1</sup>	0.53	0.56
3. Agitation	Do you often feel fearful or anxious? 1= always; 2=often;3=sometimes; 4=seldom; 5= never	0.55	0.64
4. Loneliness	Do you often feel lonely and isolated? 1= always; 2=often;3=sometimes; 4=seldom; 5= never	0.63	0.72
5. Attitude towards one's aging	Do you feel the older you get, the more useless you become? 1= always; 2=often;3=sometimes; 4=seldom; 5= never	0.50	0.57
	Are you as happy as when you were younger? 1=never; 2=seldom; 3=sometimes; 4=often; 5=always <sup>1</sup>	0.51	0.58

Note: The order of the five response levels of each item in the questionnaire is different from the order in the table, which is reversed here for factor analysis on the one hand, and for higher scores indicating higher levels of SWB on the other.

**Table 2: Sample description in 2000, and number and percentages of participants who died before the third wave in 2002, by all time-fixed factors in the analysis.**

	Total (N=7852)	Deceased (N=2375)	
	No.	No.	Percent
<b>Subjective well-being</b>			
0-25%	1,964	724	36.86
26-50%	1,961	603	30.75
51-75%	1,954	556	28.45
76-100%	1,973	492	24.94
<b>Socio-demographic characteristics</b>			
Age group			
80-84	2,144	378	17.63
85-89	1,739	430	24.73
90-94	1,732	521	30.08
95-99	943	405	42.95
100-105	1,294	641	49.54
Sex			
Male	3,504	1,057	30.17
Female	4,348	1,318	30.31
Type of residence			
Urban	4,783	1,371	28.66
Rural	3,069	1,004	32.71
Years of schooling			
0 year	4,871	1,567	32.17
1-6 years	2,257	643	28.49
7+ years	724	165	22.79
Marital status at the time of the 2000 interview			
married	1,629	328	20.14
unmarried	6,223	2,047	32.89
<b>Health status</b>			
MMSE			
0-50%	3,732	1,436	38.48
51-100%	4,120	939	22.79
Self-reported ADL			
No functional limitation	5,649	1,352	23.93
One functional limitation	1,054	398	37.76
2+ functional limitations	1,149	625	54.4
Self-rated health			
Bad	813	340	41.82
Fair	2,464	821	33.32
Good	3,298	898	27.23
Very good	1,277	316	24.75

**Table 3: Piecewise linear log-baseline intensity, and relative risks of mortality for the oldest old, in the model with duration dependency on the age of the oldest old, without unobserved heterogeneity.**

	Model 1 <sup>a</sup>	Model 2 <sup>b</sup>
<b>Piecewise linear log-baseline intensity</b>		
80-84	0.0064**	0.0058**
85-89	0.0046**	0.0035**
90-94	0.0044***	0.0036**
95-99	0.0069***	0.0051***
100+	0.0048***	0.0047***
Intercept	-10.9016***	-10.0863***
<b>Socio-demographic characteristics</b>		
Sex		
Male	1	1
Female	0.71***	0.65***
Type of residence		
Urban	1	1
Rural	1.00	1.03
Years of schooling		
0 year	1	1
1-6 years	0.95	0.96
7+ years	0.79**	0.79***
Marital status at the time of the 2000 interview		
Married	1	1
Unmarried	1.38***	1.41***
<b>Subjective well-being</b>		
0-25%	1	1
26-50%	0.83***	0.96
51-75%	0.77***	0.95
76-100%	0.70***	0.93
<b>Health status</b>		
MMSE		
0-50%		1
51-100%		0.77***
Self-reported ADL		
No functional limitation		1
One functional limitation		1.41***
2+ functional limitations		1.95***
Self-rated health		
Bad		1
Fair		0.80***
Good		0.73***
Very good		0.69***
LnL	-11851.73	-11707.87

Note: <sup>a</sup> Model 1 is the intensity regression with piecewise linear baseline intensity, gender, type of residence, years of schooling, marital status at 2000 interview, and SWB.

<sup>b</sup> Model 2 is the intensity regression with piecewise linear baseline intensity, gender, type of residence, years of schooling, marital status at 2000 interview, SWB, MMSE, self-reported ADL, and self-rated health.

\* p<0.1; \*\* p<0.05, \*\*\* p<0.01

**Table 4: Constant log-baseline intensity and relative risks of mortality for the oldest old, without unobserved heterogeneity.**

Variable	Relative risks
Constant	-4.36 <sup>***</sup>
<b>Socio-demographic characteristics</b>	
Age group	
80-84	1
85-89	1.28 <sup>***</sup>
90-94	1.48 <sup>***</sup>
95-99	2.21 <sup>***</sup>
100+	2.51 <sup>***</sup>
Sex	
Male	1
Female	0.66 <sup>***</sup>
Type of residence	
Urban	1
Rural	1.04
Years of schooling	
0 year	1
1-6 years	0.95
7+ years	0.77 <sup>***</sup>
Marital status at the time of the 2000 interview	
Married	1
Unmarried	1.43 <sup>***</sup>
<b>Subjective well-being</b>	
0-25%	1
26-50%	0.96
51-75%	0.96
76-100%	0.93
<b>Health Status</b>	
MMSE	
0-50%	1
51-100%	0.76 <sup>***</sup>
Self-reported ADL	
No functional limitation	1
One functional limitation	1.43 <sup>***</sup>
2+ functional limitations	1.97 <sup>***</sup>
Self-rated health	
Bad	1
Fair	0.80 <sup>***</sup>
Good	0.73 <sup>***</sup>
Very good	0.69 <sup>***</sup>
LnL	-11724.37

Note: . \* p<0.1; \*\* p<0.05, \*\*\* p<0.01

**Table 5: Piecewise linear log-baseline intensity, and relative risks of mortality for the oldest old, in the model with duration dependency on the time since the 2000 interview.**

Variables	Without heterogeneity	With heterogeneity
<b>Piecewise linear log-baseline intensity</b>		
0-6 months	0.0929 <sup>***</sup>	0.2383 <sup>***</sup>
7-12 months	-0.0143	0.0634 <sup>***</sup>
13-16 months	0.1563 <sup>***</sup>	0.2355 <sup>***</sup>
16+ months	-0.0482 <sup>***</sup>	0.02
<b>Socio-demographic characteristics</b>		
Age group		
constant	-4.8814 <sup>***</sup>	-7.0577 <sup>***</sup>
80-84	1	1
85-89	1.29 <sup>***</sup>	1.35 <sup>***</sup>
90-94	1.49 <sup>***</sup>	1.67 <sup>***</sup>
95-99	2.26 <sup>***</sup>	3.06 <sup>***</sup>
100-105	2.60 <sup>***</sup>	3.74 <sup>***</sup>
Sex		
Male	1	1
Female	0.65 <sup>***</sup>	0.54 <sup>***</sup>
Type of residence		
Urban	1	1
Rural	1.04	1.04
Years of schooling		
0 year	1	1
1-6 years	0.95	0.94
7+ years	0.79 <sup>***</sup>	0.69 <sup>***</sup>
Marital status at the time of the 2000 interview		
Married	1	1
Unmarried	1.50 <sup>***</sup>	1.63 <sup>***</sup>
<b>Subjective well-being</b>		
0-25%	1	1
26-50%	0.96	0.95
51-75%	0.95	0.92
76-100%	0.92	0.89
<b>Health status</b>		
MMSE		
0-50%	1	1
51-100%	0.76 <sup>***</sup>	0.68 <sup>***</sup>
Self-reported ADL		
No functional limitation	1	1
One functional limitation	1.44 <sup>***</sup>	1.67 <sup>***</sup>
2+ functional limitations	2.02 <sup>***</sup>	2.90 <sup>***</sup>
Self-rated health		
Bad	1	1
Fair	0.79 <sup>***</sup>	0.72 <sup>***</sup>
Good	0.73 <sup>***</sup>	0.63 <sup>***</sup>
Very good	0.68 <sup>***</sup>	0.55 <sup>***</sup>
Sigma		1.412 <sup>***</sup>
LnL	-11624.09	-11611.22

Note: \* p<0.1; \*\* p<0.05, \*\*\* p<0.01

**Table 6: Sample characteristics by gender, and comparison between men and women**

Variables	Male (N=3504)		Female (N=4348)		Significant Gender Difference
	Frequency	Percent	Frequency	Percent	
<b>Censor</b>					$F(1,7846)=19.65, p<0.001^1$
Death	1,057	30.17	1,318	30.31	
Censor	2,447	69.83	3,030	69.69	
<b>Subjective well-being</b>					$\chi^2(3) = 106.14, p<0.001$
0-25%	737	21.03	1,227	28.22	
26-50%	808	23.06	1,153	26.52	
51-75%	919	26.23	1,035	23.8	
76-100%	1,040	29.68	933	21.46	
<b>Socio-demographic characteristics</b>					
Age group					$\chi^2(4) = 342.53, p<0.001$
80-84	1,142	32.59	1,002	23.05	
85-89	850	24.26	889	20.45	
90-94	843	24.06	889	20.45	
95-99	370	10.56	573	13.18	
100-105	299	8.53	995	22.88	
Type of residence					<i>ns</i>
Urban	2,169	61.9	2,614	60.12	
Rural	1,335	38.1	1,734	39.88	
Years of schooling					$\chi^2(2) = 2200, p<0.001$
0 year	1,180	33.68	3,691	84.89	
1-6 years	1,729	49.34	528	12.14	
7+ years	595	16.98	129	2.97	
Marital status at the time of the 2000 interview					$\chi^2(2) = 1000, p<0.001$
Married	1,293	36.9	336	7.73	
Unmarried	2,211	63.1	4,012	92.27	
<b>Health status</b>					
MMSE					$\chi^2(1) = 395.42, p<0.001$
0-50%	1,228	35.05	2,504	57.59	
51-100%	2,276	64.95	1,844	42.41	
Self-reported ADL					$\chi^2(2) = 90.58, p<0.001$
No functional limitation	2,700	77.05	2,949	67.82	
One functional limitation	419	11.96	635	14.6	
2+ functional limitations	385	10.99	764	17.57	
Self-rated health					$\chi^2(3) = 33.48, p<0.001$
Bad	324	9.25	489	11.25	
Fair	1,028	29.34	1,436	33.03	
Good	1,512	43.15	1,786	41.08	
Very good	640	18.26	637	14.65	

Note: Age-adjusted comparison. There is no significant gender difference without control for age. We did not adjust for age when comparing other variables.

**Table 7: Piecewise linear log-baseline intensity, and relative risks of mortality for the oldest old by gender, in the model with duration dependency on the age of the oldest old, without unobserved heterogeneity.**

Variables	Male(N=3504)		Female(N=4348)	
	Model 1 <sup>a</sup>	Model 2 <sup>b</sup>	Model 1 <sup>a</sup>	Model 2 <sup>b</sup>
<b>Piecewise linear log-baseline intensity</b>				
80-84	0.0076 <sup>**</sup>	0.0071 <sup>*</sup>	0.0052	0.0046
85-89	0.0051 <sup>**</sup>	0.0044 <sup>*</sup>	0.004	0.0028
90-94	0.0009	-0.0006	0.0084 <sup>***</sup>	0.0073 <sup>***</sup>
95-99	0.0084 <sup>***</sup>	0.0081 <sup>***</sup>	0.0054 <sup>***</sup>	0.0034
100+	0.0026	0.001	0.0055 <sup>***</sup>	0.0054 <sup>***</sup>
Constant	-12.1822 <sup>***</sup>	-11.3706 <sup>***</sup>	-9.9715 <sup>**</sup>	-9.3222 <sup>**</sup>
<b>Socio-demographic characteristics</b>				
Type of residence				
Urban	1	1	1	1
Rural	1.06	1.09	0.94	0.97
Years of schooling				
0 year	1	1	1	1
1-6 years	0.92	0.93	1.03	1.02
7+ years	0.81 <sup>**</sup>	0.81 <sup>**</sup>	0.66 <sup>**</sup>	0.66 <sup>*</sup>
Marital status at the time of the 2000 interview				
Married	1	1	1	1
Not married	1.46 <sup>***</sup>	1.49 <sup>***</sup>	1.18	1.15
<b>Subjective well-being</b>				
0-25%	1	1	1	1
26-50%	0.86 <sup>*</sup>	1.00	0.81 <sup>***</sup>	0.94
51-75%	0.77 <sup>***</sup>	0.96	0.78 <sup>***</sup>	0.95
76-100%	0.68 <sup>***</sup>	0.93	0.72 <sup>***</sup>	0.92
<b>Health Status</b>				
MMSE				
0-50%		1		1
51-100%		0.77 <sup>***</sup>		0.76 <sup>***</sup>
Self-reported ADL				
No functional limitation		1		1
One functional limitation		1.41 <sup>***</sup>		1.41 <sup>***</sup>
2+ functional limitations		1.89 <sup>***</sup>		1.97 <sup>***</sup>
Self-rated health				
Bad		1		1
Fair		0.76 <sup>***</sup>		0.83 <sup>***</sup>
Good		0.62 <sup>***</sup>		0.83 <sup>***</sup>
Very good		0.64 <sup>***</sup>		0.71 <sup>***</sup>
LnL	-11843.67	-11697.91	-11843.67	-11697.91

Note: <sup>a</sup> Model 1 is the intensity regression with piecewise linear log-baseline intensity, type of residence, years of schooling, marital status at 2000 interview, and SWB.

<sup>b</sup> Model 2 is the intensity regression with piecewise linear log-baseline intensity, type of residence, years of schooling, marital status at 2000 interview, SWB, MMSE, self-reported ADL, and self-rated health.

\* p<0.1; \*\* p<0.05, \*\*\* p<0.01

# CLHLS

Age 80+

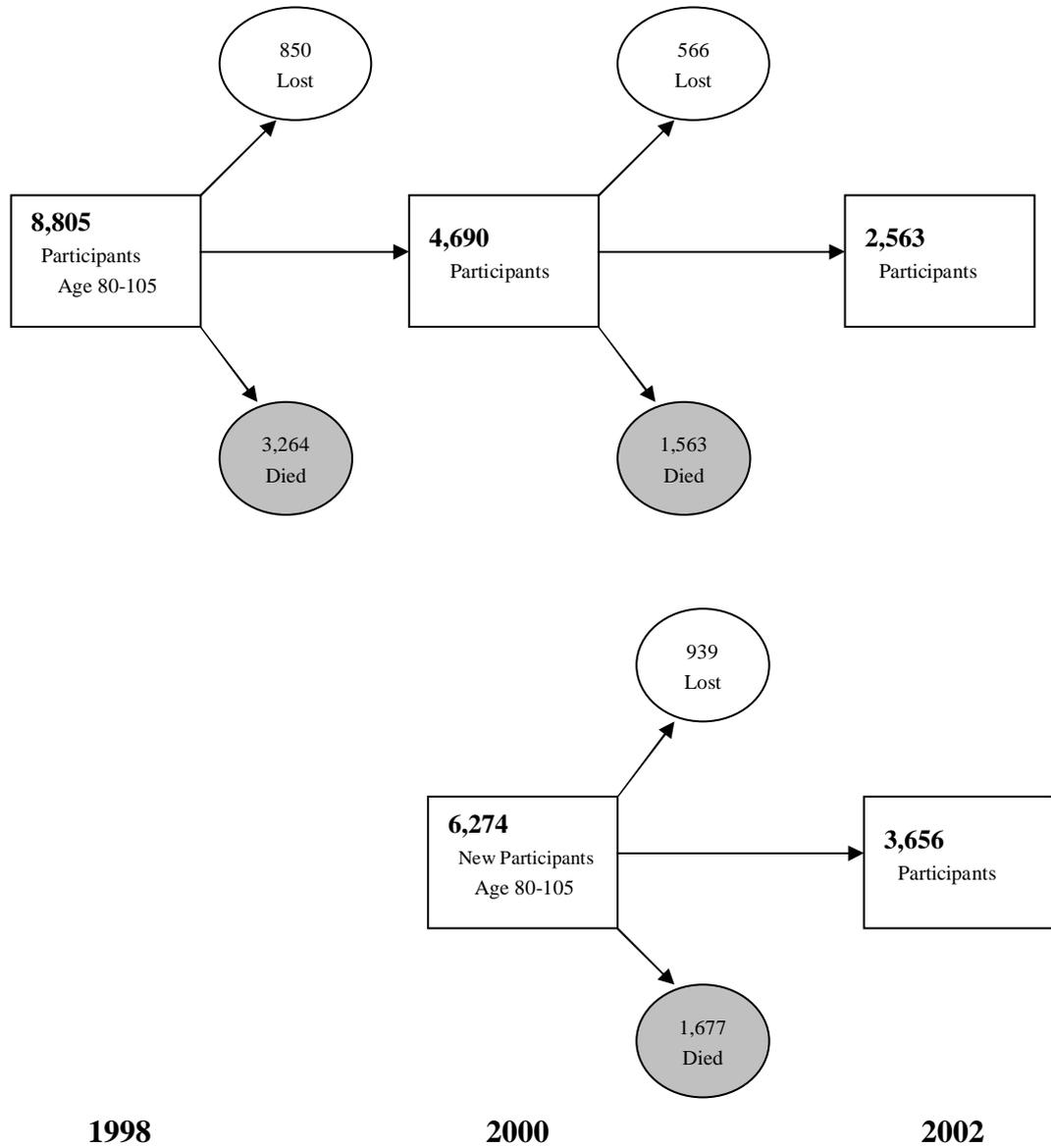
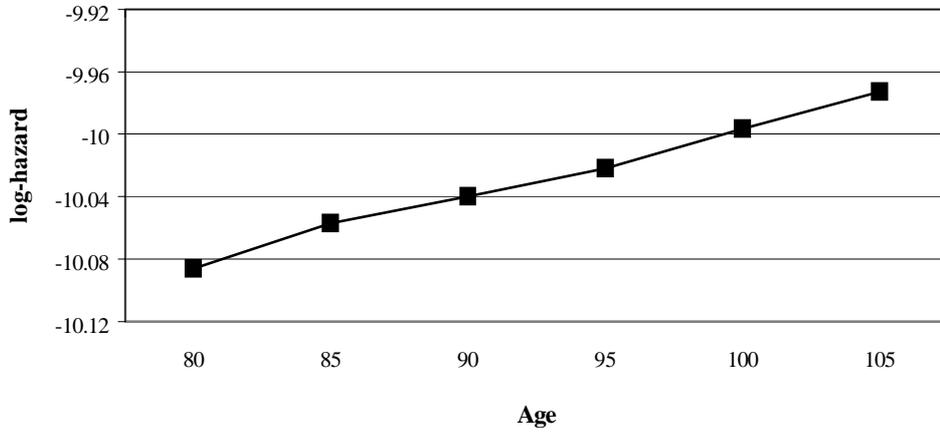


Figure 1: Overview of the Chinese Longitudinal Healthy Longevity Study (CLHLS).

**Figure 2: Piecewise-Linear Baseline Log-Hazard of Mortality in the Oldest Old**



**Figure 3: Piecewise linear baseline log-hazard of mortality in the oldest old**

