General Data Quality Assessment for the 2005 CLHLS wave[1]

Danan Gu

Center for the Study of Aging and Human Development
Medical Center, Duke University
Durham, NC 27710, USA

October 6, 2007

[1] This version is a combination of the Chinese version of the data quality assessment for the 2005 wave (Zhang, Gu, Zeng, and Liu 2006) and the general data assessment for the 2002 (Gu 2008).

Abstract: This report provides a comprehensive review of data quality of the fourth wave of the Chinese Longitudinal Healthy Longevity Survey (CLHLS) in 2005 in terms of death rate, proxy use, nonresponse rate, sample attrition, and reliability and validity of major health measures. The results show that the data quality of the 2005 wave of the CLHLS is generally good. Some recommendations in use of the dataset are provided.

Keywords: China, Chinese Longitudinal Healthy Longevity Survey, data assessment, data quality, the Chinese elderly.

INTRODUCTION

The fourth wave of the Chinese Longitudinal Healthy Longevity Survey (CLHLS) was conducted by the Center for Healthy Aging and Family Studies, Peking University and Mainland Information Group in 2005. This report provides a relatively comprehensive review of the quality of the data from the fourth wave of the Chinese Longitudinal Healthy Longevity Survey (CLHLS) in 2005 in terms of proxy use, nonresponse rate, sample attrition, and reliability and validity of major health measures. The fourth wave of the CLHLS gathered extensive questionnaire data through interviewing 15,613 elderly aged 65+. The survey was conducted in randomly selected counties and cities in 22 of China's 31 provinces.[2] Among the 15,613 sample persons in 2005, 2,797 were centenarians, 3,952 were nonagenarians, 3,909 were octogenarians, and 4,955 were aged 65-79. One thousand and forty-three elders were first interviewed in 1998 with 1,608 respondents who were first interviewed in 2000, and 5,522 respondents who were first interviewed in 2002, and 7,440 subjects who were newly recruited in 2005.

The design of the CLHLS questionnaire is based on international standards and adapted to the Chinese cultural/social context and carefully tested by pilot studies/interviews. The CLHLS emphasizes questions that might shed light on risk factors for mortality and healthy longevity. An interview and a basic health examination at each wave were performed at the interviewee's home by a team consisting of an interviewer and a medical student or a nurse. Extensive data were collected including family structure, living arrangements and proximity to children, self-rated health, self-evaluation on life satisfaction, chronic disease, healthcare needs, care costs, social activities, diet, smoking and alcohol drinking, psychological characteristics, economic resources, caregivers and family/social support, intergenerational transferring, nutrition and other health-related conditions in earlier life (childhood, adulthood, and around age 60), activities of daily living (ADL) using the Katz ADL index (Katz et al. 1963), and cognitive function measured by the Mini-Mental State Examination (MMSE) (Folstein, Folstein, and McHugh 1975). Physical performance capacity was evaluated through tests of putting a hand to the back and neck, raising hands upright, standing up from sitting in a chair without using hands, picking up a book from the floor, and turning around 360 degrees. As we did for the 2002 wave, instrumental activities of daily living (IADL) questions were added in the 2005 survey.

According to Groves (1987), there are three major potential sources of errors due to nonobservation (coverage error, nonresponse error, and sample error) and four major potential sources of errors due to observation or measurement (the interviewer, the respondents, the questionnaire, and the mode of interview). This report follows this framework in its attempt to assess the data quality of the CLHLS in 2005. All data analyses to this end are conducted using STATA 10.0 (Stata Corp 2007) and SPSS 15.01 (SPSS 2007).

[2] The 22 surveyed provinces are Liaoning, Jilin, Heilongjiang, Hebei, Beijing, Tianjing, Shanxi, Shaanxi, Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shangdong, Henan, Hubei, Hunan, Guangdong, Guangxi, Sichuan, and Chongqing. The population in the survey areas constitutes about 85 percent of the total population in China. Han Chinese people are the overwhelming majority in the 22 surveyed provinces. There were 631, 777, 866, and 943 counties or cities districts in the 1998, 2000, 2002 and 2005 surveys, respectively. The increase in numbers of survey units in 2005 was mainly due to adding an elderly comparison group aged 65-79 who were not interviewed in 1998 and 2000; and partly due to an administrative boundary change in the later waves; or some selected counties/cities that had no centenarians in an earlier wave but had centenarians in later waves.

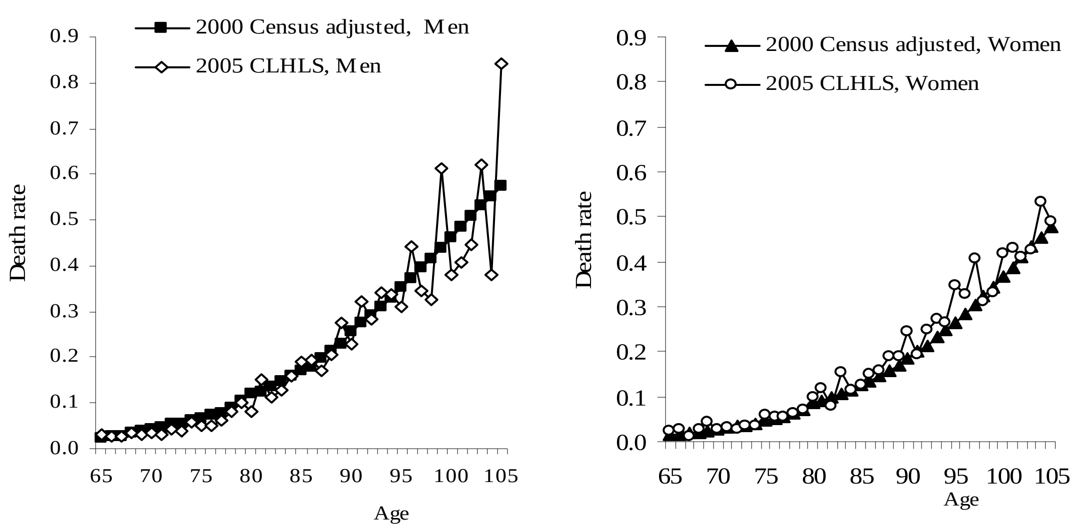
1. ALL-CAUSE MORTALITY ASSESSMENT

The accuracy of mortality rates at old ages is critical for an understanding of human aging and longevity. Numerous studies have documented distinct patterns of mortality at older ages (Horiuchi and Wilmoth 1998; Kannisto 1994; Thatcher, Kannisto, and Vaupel 1998; Vaupel et al. 1998; Zeng and Vaupel 2003). For example, countries with high-quality data frequently show that age-specific mortality rates after age 80 tend to follow logistic, Kannisto, or quadratic patterns (Kannisto 1994; Thatcher et al. 1998; Vaupel et al. 1998). Departure from these patterns may signify age misreporting and produce inaccurate mortality estimates that will bias predicted outcomes and lead to erroneous conclusions.

Unlike respondent deaths in most longitudinal surveys in Western nations, which can be linked to a national death database with relatively accurate mortality reports, the survival status of respondents in longitudinal surveys in most developing countries are often collected through interviews. Consequently, mortality data from studies in developing countries may introduce inaccuracies due to age exaggeration and age-at-death misreporting by proxies, which are further confounded by sample attrition and selection. Such inaccuracies may well introduce biases in the estimation of associations between the study variables and mortality and lead researchers to draw incorrect conclusions.

Figure 1 compares age-sex-specific average annual death rates observed in the CLHLS with corresponding death rates in the 2000 census. The 2000 census rates after age 80 are adjusted by Kannisto method to reflect the true force of mortality for these ages (see Gu and Dupre 2008). The comparisons clearly show that there is no mortality underestimation in the period of 2000-2005 in the CLHLS except for a slightly lower mortality rates found among the elderly men aged 65 to 79.

Figure 1. Comparison of all-cause mortality observed in the 2005 CLHLS and the 2000 census



2. PROXY USE

As frequently reported in most empirical studies, it is normal for a survey of the elderly to have more than 20 percent of respondents unable to complete the questionnaire due to cognitive or linguistic impairments (Coroni-Huntley et al. 1986; DeHaan et al. 1993; Magaziner et al. 1988). Studies of the elderly that fail to use proxies often have a higher rate of nonresponse or missing data (Blazer, Hughes, and George 1987). Therefore, proxies are frequently used as an alternative so to reduce elderly nonresponse, especially for the oldest-old because substantial proportions are usually incapable of providing accurate responses or even participating due to impaired hearing/vision, frail health or recall problems (Rodgers and Herzog 1992). Although it is not known for certain whether proxy information is similar to that provided by the subjects themselves (Pierre et al. 1998), there is a general consensus among investigators that proxy respondents should be used in research focusing on the oldest-old in order to avoid biasing the data in favor of healthy older persons (Rodgers and Herzog 1992).

Given that a proxy reporter is likely to be used, who then should be the proxy reporter? The existing literature in epidemiological studies suggests that validity varies considerably, depending on the relationship of the proxy to the respondent, the type of information sought, and the time period involved (Tang and McCorkle 2002). Caregivers may be more knowledgeable than personal friends and family members about physical health and functional symptoms of institutional respondents, although some studies have shown that caregivers tend to overrate the respondent's disability (Rothman et al. 1991); whereas family members may be more knowledgeable about personal, familial, and economic situations, and the like. Among family members, wives have been shown to be particularly reliable proxy reporters (Kolonel, Hirohata, and Nomura 1977).

However, the use of proxy reporters rests to a large extent on the tradeoff between nonresponse and inaccurate reporting. Errors due to unit and item nonresponse may be reduced by seeking information from proxies, but in such a case errors due to inaccurate responding may increase. To date, both of these assumptions remain unsubstantiated (Rodgers and Herzog 1992). However, it is widely understood that proxies can be used to report about factual issues, and produce fairly accurate information. Sometimes with a good questionnaire design, good quality data can be collected even on subjective questions (Basset and Magaziner 1988; Rodgers 1988; Rodgers and Herzog 1992). The consistent finding across studies is that the accuracy of proxy ratings is high when the information sought is concrete and observable (Klinkenberg et al. 2003; Tang and McCorkle 2002).

In the CLHLS, questions such as self-rated health, life satisfaction, and MMSE tests on cognitive functioning are answered by the interviewees only. Other questions are answered by the interviewees themselves, as much as possible. For those who are not able to answer these questions, a close family member or another knowledgeable proxy (i.e., significant other) provides answers as indicated earlier. An indicator question is marked by the interviewer to signify whether the answer is provided by the interviewee or the proxy.

Table 1 shows the proportion of proxy use in the 2005 wave. Consistent with the first three waves, proxy use increases with age. Table 2 suggests that about 90 percent of proxies are close relatives such as a spouse, children, and grandchildren. Given the fact that proxies are used mainly to answer objective and factual questions in the CLHLS, the higher proportion of close relative proxies suggests that any potential bias is not substantial. Previous studies have shown that the level of agreement between respondents and proxies is influenced by a number of factors such as education, age, and living arrangement (Rothman et al. 1991; Tang and McCorkle 2002; Zsembik 1994). Our multivariable analysis indicates that respondents with an older age, lower education, rural residence, lower cognitive functioning, and higher disability are more likely to use a proxy. Data from previous waves show that the respondents with a proxy have a 20 percent more relative risk of death compared to those without a proxy (data not shown here).

Table 1. Comparison of proportion of proxy in the CLHLS (%)

	Age group			
	65-79	80-89	90-99	100-105
1998 Wave				
Without proxy		61.50	36.57	16.43
Mix		38.02	62.23	81.18
Full proxy*		0.48	1.19	2.39
2000 Wave				
Without proxy		62.05	37.41	15.92
Mix		37.39	61.07	80.64
Full proxy*		0.57	1.52	3.43
2002 Wave				
Without proxy	88.40	64.35	36.91	19.04
Mix	11.39	35.43	61.57	76.41
Full proxy*	0.21	0.21	1.52	4.55
2005 Wave				
Without proxy	82.74	54.75	33.84	16.29
Mix	17.22	45.22	65.60	82.31
Full proxy*	0.04	0.03	0.56	1.40

Note: (1) the number of questions in the 2005 wave is slightly more than in the previous three waves. Full proxy means all questions except those that must be answered by the sampled person are answered by the proxy. (2) *, excluding those questions must be answered by the sampled persons.

Table 2. Distribution of proxy subjects in the CLHLS (%)

	1998 Wave Age 80+	2000 Wave Age 80+	2002 Wave Age 80+	2002 Wave Age 65-79	2005 Wave Age 80+	2005 Wave Age 65-79
Spouse	5.55	5.18	5.02	32.00	4.69	28.86
Child or spouse of children	74.01	67.41	67.68	50.53	71.94	53.20
Grandchild or spouse of grandchild	12.37	16.56	15.92	8.00	13.66	7.33
Great grandchild or spouse of great grandchild	0.28	0.81	0.95	0.00	0.58	0.00
Sibling	0.24	0.18	0.20	0.84	0.31	1.25
Caregiver	2.76	4.34	5.27	2.11	3.73	1.56
Others	4.79	5.54	4.97	6.53	5.10	7.80

Note: (1) Based on two questions addressed to the interviewer "did anyone help the interviewee to answer any question?" and "who helped the interviewee to answer questions?" There is inconsistency between these two questions and actual proxy use in the questionnaire due to a misunderstanding of the questions by the interviewer (Zeng et al. 2001:112). The inconsistency rates are 13.2% and 6.2% for the 1998 and 2000 waves, 3.9% and 2.6% for ages 80-105 and 65-79 in the 2002 wave, respectively. This inconsistency rates in the 2005 wave are around 5.3% for elders aged 65-79 and 7.5% for the oldest-old aged 80 and older.

The CLHLS did not obtain data comparing the agreement of responses between the proxy and the respondent. However, the small amount of full proxy responses indicates the results may not be a big problem even if the bias between the respondent and proxy exists. Researchers may also add an indicator variable for proxy use (i.e., whether the proxy is used or not for the sampled person) to adjust for such a bias if they think proxy answers could be problematic, as some other previous studies have done (e.g., Jenkins and Fultz 2005).

3. NONRESPONSE RATE AND INCOMPLETE DATA

Nonresponse is an important indicator of data quality because it can bias survey estimates (Jay et al. 1993). Numerous studies indicate that nonresponse is greater for older adults than for younger adults (Herzog and Rodgers 1988), and nonresponse is a serious problem among older age groups and may be particularly high among those ages 85 and older (Herzog and Rodgers 1992). There are two types of nonresponse, namely, unit nonresponse and item nonresponse (Mohadjer, Bell, and Waksberg 1994).[3]

The unit nonresponse rate among the Chinese oldest-old was very low, about 4 percent, in the first four waves. This is because the Chinese oldest-old, in general, like to and have the time to talk to outside people, as they are at home without a job or other responsibilities. Many of the respondents and their family members may also feel honored to be interviewed about healthy longevity, as they may be proud of being a member of a long-lived group. Many of the seriously disabled oldest-old agreed to participate through proxy assistance provided by a close family member. Unit nonresponse rates tended to increase slightly among younger interviewees aged

[3] The line between item and unit nonresponse is sometimes not clear. For example, if a completed questionnaire requires 90 percent of all possible items to be answered, it is possible that a number of partial interviews would be treated as unit nonresponses. On the other hand, if the required level of item responses is 80 percent for a completed questionnaire, the number of partial interviews treated as unit nonresponses would decrease and the unit response rate would increase (OMB 2001). In this study, a respondent who answered 60% or more of all possible items is coded as a valid unit response.

65-79 (5-6 percent) because some of them apparently did not want to devote their time to the interview. One Japanese study of the elderly also finds a higher nonresponse rate in the lower age categories (Sugisawa et al. 1999). The amount of unit nonresponse error is difficult to measure, and thus efforts are often directed to minimize its occurrence (OMB 2001).

More recently, Lindner, Murphy and Briers (2001) recommend that steps should be taken to account for possible nonresponse error when a unit response rate is less than 85 percent. Although the CLHLS has a unit response rate higher than 85 percent, attention should be paid to item nonresponses, because a low unit nonresponse rate does not guarantee a low item nonresponse rate. Most data failures are due to a failure to obtain or record all-item information. A large amount of incomplete data for a particular item may indicate a problem with the translation of the item. Incomplete data might also indicate that respondents do not understand how to complete that part of the questionnaire. Data incompleteness can be classified into "Don't Know" (DK) and "Missing" categories, when the respondent refuses to answer or for other reasons.[4] DK usually occurs on questions related to historical information when the sampled person suffers recall problems, or when the proxy does not know about actual facts of the sampled person. Francis and Busch (1975) find that the oldest-old tend to give DK answers, and Herzog and Rodgers (1981) find that the oldest-old give DK answers more frequently on questions related to attitudes, feelings and expectations.

Table 3. Average percentage of item incompleteness of each respondent in the CLHLS (%)

Age	Males			Females		
	DK	Missing	Total	DK	Missing	Total
1998 Wave *						
80-89	4.64	0.48	5.12	5.45	0.63	6.07
90-99	4.58	0.65	5.22	7.23	0.75	7.98
100-105	6.03	0.75	6.78	8.53	1.03	9.56
2000 Wave						
80-89	2.26	1.51	3.78	2.91	1.97	4.87
90-99	2.61	2.06	4.67	3.54	2.82	6.37
100-105	3.22	2.50	5.72	4.42	3.52	7.94
2002 Wave						
65-79	2.09	0.95	3.04	2.32	1.33	3.65
80-89	2.74	1.58	4.32	4.01	2.23	6.24
90-99	3.80	2.01	5.81	4.84	2.69	7.52
100-105	4.60	1.95	6.55	5.95	2.74	8.69
2005 Wave						
65-79	1.32	5.06	6.39	1.42	5.31	6.74
80-89	2.05	5.21	7.26	2.52	5.59	8.12
90-99	2.95	5.01	7.96	3.80	5.28	9.08
100-105	4.48	4.54	9.02	5.21	4.61	9.82

Note: (1) Percentage of incomplete items (including don't know and refusal to answer) of each respondent is calculated based on the number of items that could be answered and the number of items answered by each respondent. (2) Numerator in DK does not include 'unable to answer' questions, which should be answered by the interviewee only. (3) *, the results for 1998 are different from Zeng et al (2001) since the results of Zeng et al. (2001) did not include 'don't know' in chronic diseases and did not include non-reported information about siblings or children.

[4] Unlike most other studies, data incompleteness due to DK and missing in this paper is separately discussed from sample attrition.

Table 3 shows that the average proportion of incompleteness of an item rated for each respondent in the CLHLS is less than 10 percent, much lower than some previous studies have reported (Wallace, Kohout, and Colsher 1992:132). Difference in total proportion of incompleteness between the 2005 wave and previous waves is small. However, the missing rate in the 2005 wave is slightly larger than earlier waves. Table 4 summarizes the variables with incomplete answers of 2 percent or more.[5] Variables with the highest incomplete rate are "parents' ages at death". As compared to the previous waves, the incomplete rates for children and siblings seem increased in the 2005 wave. Further, the incomplete rates for parental ages at death are still high, especially among the oldest-old, suggesting that extreme caution should be given to these variables. If item nonresponses are missing completely at random, the estimates will not be biased (Allison 2002). The estimates might be biased, however, if item nonresponses are not completely at random. In such a case, tests should be made to detect any correlates. Prior studies have suggested that factors that might pertain to item nonresponse include age, sex, education, geographic region, and urban/rural residence (Jay et al. 1993). Our multivariable logistic results reveal that factors such as ethnicity, marital status, urban/rural residence, cognitive functioning, and self-reported health are all correlated to aggregated item nonresponses in each of the first three waves in the CLHLS. Those who are older, female, urban residents, of a minority ethnicity, not currently married, and in bad health are more like to have incomplete items (not shown), which is consistent with some previous studies in Western nations (Francis and Busch 1975; Herzog and Rodgers 1981).

Table 4. Distribution of variables with more than 2% incomplete answers in the CLHLS (%)

	1998 wave Age 80+	2000 wave Age 80+	2002 Age 80+	2002 Age 65-79	2005 Age 80+	2005 Age 65-79
Habit (i.e., smoking, drinking, exercise, physical laboring)	1.0-4.0	<2.0	<1.5	<1.5	2.7	2.2
Marriage history	>5.0[c]	>4.0[c]	>1.5[c]	>0.1[c]	2.4	0.7
Parents' age at death, and respondent's age at parents' death	30.0-40.0	27.0-35.0	25.0-30.0	7.8-10.5	21.1	8.6
Birth order and # of sibling	2.0-3.0	2.2	2.2	<0.5	4.9	3.0
Sibling information	>7.0[a]	>3.4[a]	>3.7[a]	>0.5[a]	>4.7[a]	>1.7[a]
Children's information	>3.0[a]	>2.7[a]	>2.2[a]	>1.0[a]	>8.9[a]	>6.7[a]
Blood pressure	3.0	3.2	0.9	<0.2	1.3	1.1
Height (Acromion-processus styloideus ulnae; right knee to the floor)	5.0	NA	0.0	0.0	1	0.9
Weight	8.0	1.7	0.0	0.0	0.5	0.2
Chronic diseases	7.0-10.0	5.0-8.0	3.6-7.5	2.4-5.0	2.1	1.7

Note: (1) a, no upper boundary was provided here since the number of items that could be answered by each respondent is different, and the aggregated incomplete proportion is high for some items although the absolute number is not large due to the very small number of eligible respondents. (2) NA, not applicable since there was no such question in 2000.

Could item nonresponses that are conditional on a set of covariates introduce a bias in the estimation? Some studies argue that the effect of item nonresponses on outcomes does not depend on the difference between who gives the answers and who does not; rather, it depends on

[5] In the 1998 survey, there are 22% of respondents who did not know the name of the county in which they were born. There are 16% of respondents' with missing lung flow data in the 1998 wave. These two variables are not listed in Table 4 since they were not dropped from the questionnaire since the 2000 wave.

how the respondents who give answers differ from all those who are eligible to be interviewed (Norris and Goudy 1986; Kempen and van Sonderen 2002). In other words, if the response structure or pattern for those who answered the question is the same as the response structure or pattern for all sampled persons if they all could provide answers, then the estimates based on only those without nonresponse would be the same as the estimates based on the whole sample if all persons could answer the question.

Given that the CLHLS survey has encountered some level of nonresponse both in unit and item nonresponses, two general approaches could be applied to compensate for these nonresponses, namely, weight adjustment and imputation. Kalton and Kasprzyk (1986) note that weight adjustments are primarily used to compensate for unit nonresponses while imputation procedures are more likely to be used to compensate for missing items. In the CLHLS, a weight matching the post hoc distribution of age-sex-urban/rural residence in the sample with the distribution of the total population in the sampled 22 provinces is employed to reflect the unique sample design and compensate for unit nonresponses. This post hoc weight takes both the special design of the CLHLS and unit nonresponses of three basic demographic variables (i.e., age, sex, and urban/rural residence) into consideration (see Zeng et al. 2001; Zeng 2008 for detail). However, this weight has no relationship with other factors since their frequency distributions for the population are difficult to obtain and the weighting adjustment has the disadvantage of taking too many factors into consideration (Lepkowski, Kalton, and Kasprzyk 1989). Researchers can create other weighting schemes if they have a reliable distribution for the total population in those 22 sampled provinces.

For compensating item nonresponses, Landerman, Land, and Pieper (1997) suggest using the mean if the incomplete rate of a particular variable is less than 2 percent; however, they argue that it is better to use regression or maximum likelihood methods to estimate nonresponse values when the incomplete rate is 2 to 5 percent, and to use multiple imputation to get estimates for nonresponse values when the incomplete rate exceeds 5 percent. With regression, maximum likelihood, or multiple imputation, biases in estimation can be lessened. Other strategies for dealing with this problem such as trimming bounds have also been suggested (Lee 2002). Other studies suggest treating the missing value as a special category if one is unable to ensure the accuracy of imputation (e.g., Hayward and Gorman 2004; Zimmer, Martin, and Chang 2002). The released CLHLS dataset does not provide imputed values for those variables with item nonresponses. If users of the CLHLS dataset want to impute the variable, they should follow the recommended approaches of Allison (2002), who provides a simple and very good theoretical background for how to handle item nonresponses. Most statistical packages such as SPSS, SAS, and STATA are capable of handling imputations or multiple imputations for item nonresponses. Our testing analyses show that the difference across different imputation approaches is not substantial, especially when the item nonresponse rate is less than 5 percent.

4. SAMPLE ATTRITION

In longitudinal surveys, sample attrition (or data attrition, i.e., respondents lost in a follow-up survey) occurs when previous respondents migrated, refused to participate in the survey, became hospitalized, moved, or the address of a previous respondent was not sufficiently detailed.[6] Sample attrition is one of the most serious problems associated with longitudinal survey data. Similar to item nonresponse, sample attrition may distort the treatment/control

[6] Those who died but followed-up at the subsequent wave is not considered as a type of sample attrition in this study.

comparison, depending on the type of attrition that takes place. If attrition is completely random with respect to all factors relevant to the outcome being measured, it leads to less precise estimates of program impacts (due to the reduction of the sample size), but does not lead to biased estimates (Mossel and Brown 1984). However, biased estimates might occur if sample attrition is correlated with some particular attributes, which may result in a lack of generalizability.

Out of a total of 16, 064 interviewees in 2002, 8,167 (50.8 percent) were still alive at the 2005 wave, 5,878 (36.6 percent) died before the interview was held in 2005, and 2,019 (12.6 percent) were lost.[7] This proportion of attrition was slightly lower than that between the 2000 and 2002 waves but higher than that between the 1998 and 2000 waves (9.6 percent). The true reason for the higher sample attrition in the period of 2002-2005 is not known. We suspect this had something to do with more frequent resettlement of urban residents due to municipal construction and/or more frequent changes in re-delimiting the administrative boundary of counties and/or districts in the period, which would tend to cause more difficulty in locating previously sampled persons. Other possible reasons include unfavorable weather, transportation difficulties, and so forth. Compared with data attrition in surveys conducted in Western countries, the CLHLS has a similar proportion of data attrition. For instance, the proportion lost to follow-up in the two-year interval in the second, third and fourth waves of the Longitudinal Study of Aging in the USA was 7.6 percent, 12.1 percent, and 16.0 percent, respectively (Mihelic and Crimmins 1997). The proportion of respondents lost to a 2-year follow-up was 17.8 percent in a survey of Mexican elderly (Vellas et al. 1998).

Table 5 indicates that significant associations between sample attrition and variables in the model are observed except in self-reported health. Respondents who are female, living in urban areas, physically impaired, and with low social contacts are associated with higher attrition rates. This is consistent with prior findings in the literature (e.g., Powell et al. 1990; Sugisawa et al. 1999). Attrition patterns between the period of 2000-2002 and the period of 2002-2005 are more or less same among the oldest-old except for gender. We also found a minor difference in the pattern in association with attrition in the period of 2002-2005 between young elders and the oldest-old. It is noteworthy that centenarians are less likely to have a higher rate of attrition in the CLHLS, which differs from previous research showing that older age is associated with a higher attrition (e.g., Slymen et al. 1996). This is because the CLHLS stressed the importance of the high follow-up rate for centenarians, and thus interviewers might did their best to find the sampled centenarians.

[7] Those who were lost to the follow-up also include some who actually died.

Table 5. Odds ratios of lost to follow-up by selected variables in the CLHLS

Variables	Lost to follow-up in 2000 Age 80+	Lost to follow-up in 2002 Age 80+	Lost to follow-up in 2005 Age 80+	Lost to follow-up in 2005 Age 65-79
Females (males)	1.21*	1.20*	1.02	1.23*
Age 90-99 (age80-89)	0.68***	1.03	1.09	---
Age 100-105 (age80-89)	0.51***	0.90	0.82*	---
Rural (urban)	0.46***	0.70***	0.47***	0.42***
Minority ethnicity (Han)	0.46***	0.62**	0.39***	0.39***
1+ schooling (no schooling)	1.49***	1.23*	1.14*	0.97
Currently married (not married)	1.04	1.07	1.02	0.83#
Living alone (others)	1.19	1.17#	1.19*	0.90
High proximity with children (low)	0.82*	0.80**	0.72***	0.50***
Poor ADL (good ADL)	1.14	1.32***	1.36***	1.31#
Poor MMSE (good MMSE)	1.09	1.20**	1.00	0.90
Self reported poor health (good health)	0.97	1.00	1.01	1.00
Proxy (no proxy)	0.89	0.89#	0.91	0.95
Missing group 2 (missing group1)	1.11	1.22**	1.22**	1.36**
Missing group 3 (missing group 1)	1.13	1.33***	1.01	1.18
newly interviewed at current wave (except for the 1998 wave)	---	1.31***	1.33***	1.48
N	8,805	10,844	11,175	4,840
-2LL	5298.1***	8506.6***	7,592.6***	3,53.3***

Note: Three missing groups are classified based on the missing rate of each respondent. Group1, <2%; Group 2, 2-5%; Group 3, >5%.

Urban respondents in the CLHLS are more likely to be lost to follow-up, partly because of changes that were made in administrative zones in urban areas as indicated earlier, and partly because urban respondents have a higher mobility than their rural counterparts. Respondents of minority ethnicities are less frequently lost in follow-up surveys compared with Han respondents. In the 2002 wave, respondents who have missing items of 5 percent or more have 13-33 percent more chance of being lost to follow-up compared with those respondents with missing values of less than 2 percent, after controlling for sociodemographic attributes and health conditions at previous waves. However, it was not the case in the 2005 wave. But respondents with missing items of 2-5 percent are 22 percent more likely to being lost to follow-up than those with missing value of less than 2 percent.

As reported by Norris and Goudy (1986) and Kempen and van Sonderen (2002), the effects of sample attrition on outcomes depends on how re-interviewed respondents differ from all those who are eligible to do so. Furthermore, the strong linkage between sample attrition and its associates does not necessarily mean that the coefficients of predictors for outcomes of interest must be affected by sample attrition. Kempen and van Sonderen (2002) demonstrate that attrition might not always be a serious problem when associations between variables are the focus of a study, particularly when the proportion of dropouts is not too large, although a cross-sectional descriptive analysis at a later wave may be more affected by attrition. Therefore, it is unlikely that there will be significant problems in estimations in the CLHLS, with its relatively low sample attrition. All compensation approaches mentioned above for nonresponse items are fully applicable to deal with sample attrition wherever necessary.

5. LOGICAL ERROR (INCONSISTENT RESPONSE)

Logical errors might occur across all questions due to inconsistent answers provided by interviewees, the carelessness of interviewers, and mistyping or miscoding of data entries. Table 6 and Table 7 show that the inconsistency of responses given by interviewees or proxies is slightly higher in both the 2002 and 2005 waves compared with levels in the previous waves, and the inconsistency given by interviewers rebounded in 2005 after a decline in the 2002 wave. Inconsistent responses seem to increase slightly with age. The difference between genders is trivial.[8]

Table 6. Inconsistent responses for selected items in the CLHLS

Inconsistent items	1998 wave Age 80+		2000 wave Age 80+		2002 wave Age 80+		2002 wave Age 65-79		2005 wave Age 80+		2005 wave Age 65-79	
	#	%	#	%	#	%	#	%	#	%	#	%
1. ADL fully dependent but can pick-up a book while standing	112	1.27	110	1.00	108	0.99	0	0.00	14	0.14	1	0.02
2. ADL fully independent but can't stand up from a chair	50	0.57	83	0.76	230	2.11	96	1.98	258	2.49	94	1.86
3. Can't stand up from a chair but does housework or fieldwork everyday	6	0.07	4	0.04	117	1.07	66	1.36	12	0.12	8	0.16
4. Had a proxy for answering some questions but interviewer didn't mark*	891	10.12	544	4.96	248	2.27	86	1.78	772	7.45	266	5.27

Note: (1) a, Age 80-105; b, Age 65-79. (2) *, This might be caused by interviewer's misunderstanding the question "Did anyone help the interviewee to answer any question?". They might have mistakenly understood it as referring only to those questions that must be answered by interviewee.

Table 7. Distribution of Inconsistent responses in the CLHLS

Ages	Males 1998	Males 2000	Males 2002	Females 1998	Females 2000	Females 2002
65-79			4.72			4.94
			(2.75)			(2.95)
80-89	12.53	7.70	6.25	14.70	8.24	6.35
	(2.41)	(2.35)	(3.90)	(2.93)	(2.49)	(3.60)
90-99	14.16	8.27	6.94	13.59	7.94	7.81
	(3.93)	(2.92)	(4.42)	(3.44)	(3.78)	(5.59)
100+105	13.97	11.75	9.92	13.34	8.54	8.23
	(5.46)	(6.39)	(7.24)	(4.71)	(4.61)	(6.65)

Note: (1) The figures in parentheses do not include interviewer's misunderstanding the question "Did anyone help the interviewee to answer any question?".

6. RELIABILITY OF MAJOR HEALTH MEASUREMENTS

In the CLHLS, health has been conceptualized in a multidimensional manner, with a general emphasis on physical and mental domains. The use of existing standardized instruments

---

[8] A similar pattern was also observed in the National Long-Term Care Survey in the U.S. (Wallance et al. 1992:133).

has the benefit of prior experience and information on measurement properties (Wallace and Herzog 1995), which is increasingly advocated as key outcome measures in health surveys (McHorney et al. 1994). Several translated Chinese versions of activities of daily living (ADL), instrumental activities of daily living (IADL), and the Mini-Mental State Examination (MMSE) have been developed and have been shown to be reliable and valid (e.g., Chou 2003; Zhang 1993; Zhang, Zhu, and Chen 1998). However, these scales mainly focus on young adults or young elders. Their appropriateness for the oldest-old has not been determined. The CLHLS provides an opportunity to examine their reliabilities and validities among this rapidly growing sub-population who need the most help but about whom we know very little.

Internal-consistency reliability for selected measurements was estimated using Cronbach's alpha coefficient (Cronbach 1951). A minimum reliability coefficient of 0.70 has been recommended for group-level analyses, while reliability coefficients of 0.90 or greater have been suggested for individual-level analyses (Nunnally 1994; Stewart, Hays, and Ware 1992).

Table 8 shows that all Cronbach's alpha coefficients for the ADL scale (consisting of bathing, dressing, toileting, indoor transferring, continence, and eating) and MMSE in the 2002 wave are above the 0.70 criterion suggested for group comparisons, indicating good internal consistency. It is worth noting that the reliability of ADL before dying is higher than that for survivors, although questions of ADL before dying were all answered by next-of-kin. The IADL items are a combination of different sources derived from major surveys for elders around the world. The reliability for eight IADL items in the 2002 wave is also high, indicating the possibility of creating a scale. On the other hand, the data reported in Table 8 indicate that the reliability coefficients for negative and positive personality variables are lower than 0.70 if we exclude those who are too sick to be able to answer questions,[9] implying that they might not be appropriate to use in scale generation.[10]

Table 8. Reliability coefficients and validity for selected measures in the 2005 wave

Scales and measures*	Cronbach's alpha coefficient	Scales and measures*	Cronbach's alpha coefficient
Age 80-105		Age 65-79	
Functioning of upper extremities (3)	0.91	Functioning of upper extremities (3)	0.84
Functioning of body mobility (2)	0.74	Functioning of body mobility (2)	0.76
Negative personality related variables (3) #	0.64	Negative personality related variables (3) #	0.65
Positive personality related variables (4) #	0.50	Positive personality related variables (4) #	0.50
Negative personality related variables (3)	0.69	Negative personality related variables (3)	0.88
Positive personality related variables (4)	0.69	Positive personality related variables (4)	0.91
ADL (6)	0.87	ADL (6)	0.89
IADL (8)	0.92	IADL (8)	0.94
Mini-Mental State Examination (MMSE) (22)	0.97	Mini-Mental State Examination (MMSE) (22)	0.99

Note: (1) #, excluding those persons who answered 'unable to answer' these questions. If persons are too sick to answer such questions that should be answered ONLY by interviewees, the answers for such questions are 'unable to answer".

[9] Questions related to personality and cognitive functioning must be answered by the interviewee themselves; no proxy is allowed in this regard. If the interviewee is too sick to answer a question, the interviewer marks 'unable to answer' for that question.

[10] In designing the personality scale, we did not follow existing scales because most scales are developed in Western countries, and might not be appropriate for use in China. We, therefore, selected some major items from various scales that we believed were appropriate for use with the Chinese elderly. Therefore, it is better to analyze these variables individually, which is confirmed by the results in Table 8.

7. VALIDITY OF MAJOR HEALTH MEASUREMENTS

The validity of a measure in the health field has often been evaluated by its content, construct, and criterion validity (Gandek and Ware 1998). Content validity examines the extent to which a measure or questionnaire represents the universe of concepts or domains; that is, whether the measure offers an adequate sample of the content of a construct (Steward et al. 1992). Construct validity is a process in which validity is evaluated in terms of the extent to which a measure correlates with variables in a manner consistent with theory (Stewart et al. 1992). Convergent and discriminant validity are at the foundation of construct validation. Convergent validity is supported when different methods of measuring the same construct provide similar results, whereas discriminant validity is supported when a measure of one underlying construct can be differentiated from another construct. In brief, high and consistent correlations were assumed between an item and its own scale, and significantly lower associations between that item and all other scales. If a scale is valid, items on which the scale is based should be related to each other (convergent validity) and not related to measures of different concepts (discriminant validity).

For establishing convergent and discriminant validity of the measures, relationships of selected scales and measures have been examined, and the results are presented in Table 9. ADL measures daily functioning in terms of eating, dressing, moving, toileting, continence, and bathing. IADL also measures daily functioning but with respect to more difficult tasks. If they are valid, they are expected to have a higher correlation between them and a higher correlation with functional capacity of extremities and body mobility than correlations with personality measures. On the other hand, if the personality measures are valid, positive and negative personality should have a higher correlation between each other than the correlations between them and other measures. Table 9 also presents the ranges of all possible correlation coefficients within scales (measurements) and across scales (measurements). It is apparent that all correlations between items within the same dimension or similar dimensions are much higher than correlations between items from different dimensions. Moreover, the correlations between the cognitive performance measures and the IADL index of cognitive functioning are positive but small, reflecting similar findings reported in the literature (Morris 1983; Wallace and Herzog 1995). It is clear that the results presented in Table 9 support a good convergent and discriminant validity for these measurements in the 2002 wave.

Another approach for testing construct validity of measures is factor analysis, which measures whether the same dimensional variables load on the same factor (Stewart et al. 1992). Our results support the good validity of these measures in the 2002 wave.[11]

[11] The results of factor analyses are not shown in this report due to limited space, but they are available upon request.

Table 9. Convergent and discriminant validity for selected measures in the 2005 wave

Scales and measures*	FU	FB	NP	PP	ADL	IADL	MMSE
Age 80-105							
Functioning of upper extremities (3)	**0.79-0.80**	0.24-0.34	0.03-0.06	0.02-0.06	0.19-0.34	0.22-0.30	0.00-0.18
Functioning of body mobility (2)	0.24-0.34	**0.51**	0.10-0.15	0.06-0.13	0.27-0.48	0.35-0.55	0.01-0.47
Negative personality related variables (3) #	0.03-0.06	0.10-0.15	**0.29-0.51**	0.09-0.31	0.05-0.12	0.05-0.23	0.00-0.11
Positive personality related variables (4) #	0.02-0.06	0.06-0.13	0.09-0.31	**0.16-0.32**	0.04-0.08	0.06-0.12	0.00-0.13
ADL (6)	0.19-0.34	0.27-0.48	0.05-0.12	0.04-0.08	**0.35-0.89**	0.18-0.71	0.00-0.31
IADL (8)	0.22-0.30	0.35-0.55	0.05-0.23	0.06-0.12	0.18-0.71	**0.47-0.80**	0.01-0.29
Mini-Mental State Examination (MMSE) (22)#	0.00-0.18	0.01-0.47	0.00-0.11	0.00-0.13	0.00-0.31	0.01-0.29	**0.01-0.79**
Age 65-79							
Functioning of upper extremities (3)	**0.54-0.81**	0.35-0.43	0.08-0.13	0.06-0.09	0.19-0.33	0.22-0.33	0.01-0.18
Functioning of body mobility (2)	0.35-0.43	**0.62**	0.11-0.17	0.09-0.16	0.33-0.56	0.44-0.58	0.01-0.54
Negative personality related variables (3)#	0.08-0.13	0.11-0.17	**0.29-0.52**	0.11-0.30	0.05-0.14	0.06-0.20	0.00-0.12
Positive personality related variables (4)#	0.06-0.09	0.09-0.16	0.11-0.30	**0.15-0.32**	0.04-0.11	0.02-0.20	0.00-0.17
ADL (6)	0.19-0.33	0.33-0.56	0.05-0.14	0.04-0.11	**0.38-0.86**	0.20-0.63	0.00-0.29
IADL (8)	0.22-0.33	0.44-0.58	0.06-0.20	0.02-0.20	0.20-0.63	**0.54-0.81**	0.02-0.25
Mini-Mental State Examination (MMSE) (22)#	0.01-0.18	0.01-0.54	0.00-0.12	0.00-0.17	0.00-0.29	0.02-0.25	**0.00-0.74**

Note: (1) Correlation coefficients are Spearman Coefficients. (2) FU, functioning of upper extremities; FB, functioning of body mobility; NP, negative personality measures; PP, positive personality. (3) The main purpose of the table is to see the magnitude of correlation coefficients in term of an absolute number. Therefore, negative coefficients have been represented as an absolute number. (4) Other notations see the note in Table 8.

8. CONCLUDING REMARKS

This report has examined the data quality of the 2005 wave of the CLHLS, mainly on proxy use, item incompleteness, sample attrition, and the dimensionalities of reliability and the validity of health condition measurements.[12] Based on the results, we are generally pleased with the quality of the health indicators in the CLHLS. Analyses of health measures showed high reliability and validity on items that we were able to evaluate and exceeded widely used criteria. Therefore, we are confident that they are measuring meaningful underlying concepts and are doing so accurately, thereby permitting comparisons between groups. The results reported in this report also suggest that the Chinese translated version of the Katz ADL Index and the Chinese version of the MMSE are both reliable and valid for the oldest-old. Our results indicate no translation complications, which ensures data quality.[13]

We recommend that attention be given to the following issues and items. First, it is inappropriate and not recommended to generate a scale for personality measurements since their reliability is below the required cut off point. Second, higher proxy use is related to older age, lower education, rural residence, lower cognitive functioning, and higher disability. Therefore, it would be better to add an indicator variable (i.e., the presence or absence of a proxy) in the analysis when the aim of the proposed research is to examine the effects of these factors. Third, we find that item incompleteness and sample attrition are linked to age, gender, urban/rural

[12] For age validity for the CLHLS, please refer to Zeng and Gu (2008).

[13] The high quality of CLHLS data is also due to the data quality control program used in the CLHLS. Before data entry, a three-stage check is employed: a local site check, provincial check, and final check at the Mainland Information Company in Beijing. Questionnaires are returned to participating sites for correction if local and provincial supervisors or supervisors at the Mainland Information Company find them with missing items or errors. Data entry is conducted at Peking University. In data entry, specific logic, range, and consistency checks between related items are added to the data entry program using EPI 6.0 software. Data double-entry is conducted at Peking University under professional supervision to minimize entry errors. A questionnaire is returned to the participating site for correction if a logic error is detected in the questionnaire not due to a coding or entry error.

residence, ethnicity, and health conditions. Although it is unlikely that these limitations will significantly affect results, sufficient attention must be paid to them in verifying and reporting the outcomes. In sum, the evidence above has led us to believe that the data quality of the 2002 wave of the CLHLS is generally good.

REFERENCES:

Allison, P. (2002), *Missing data.* Thousand Oaks, California: Sage Publications

Basset, S.S., and J. Magaziner (1988), *The use of proxy responses on mental health measures for aged, community-dwelling women.* Paper presented at the 41dt annual scientific meeting of the Gerontological Society of America, San Francisco

Blazer, D., D.C. Hughes, and L.K. George (1987), The epidemiology of depression in an elderly community population. *Gerontologist* 27, pp. 281-287

Chou, K.L. (2003), Correlates of everyday competence in Chinese older adults. *Aging Mental Health* 7(4), pp.308-315

Coroni-Huntley, J., D.B. Brock, A.M. Ostfeld, J.O. Taylor, and R.B. Wallace (1986), *Established population or epidemiological studies of the elderly: resource data book.* Washington, D.C.: NIH Publication NO. 86-2443

Cronbach, L.J. (1951), Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334

DeHaan, R., N. Aaronson, M. Limburg, K. Langton-Hewer, and H. van Crevil (1993), Measuring quality of life in stroke. *Stroke 2*4, pp. 320-326

Folstein, M.F., S.E. Folsein, and P.R. McHugh (1975), "Mini-Mental State": a practical method for grading the cognitive state of pattern for clinician. *Journal of Psychological Research* 12, pp.189-198

Francis, J.D., and L. Busch (1975), What we know about "I don't know". *Public Opinion Quarterly* 39, pp. 207-18

Gandek, B., and J.E. Jr. Ware (1998), Methods for validating and norming translations of health status questionnaire: The IQOLA project approach. *Journal of Clinical Epidemiology* 51(11), pp. 953-959

Groves, R. M. (1987), Research on survey data quality. *Public Opinion Quarterly* 51, pp.S156-S172

Gu, D. (2008). General data quality assessment of the CLHLS. In Y. Zeng, D. L. Poston, D.A. Vlosky, and D. Gu (eds.). *Healthy longevity in China: Demographic, socioeconomic, and psychological dimensions*, pp. 39-59. Dordrecht, The Netherlands: Springer Publisher.

Gu, D., and Y. Zeng (2004), Data assessment of the CLHLS 1998, 2000, and 2002 waves. In: Y. Zeng, Y. Liu, C. Zhang and Z. Xiao (eds.): *Analyses of the determinants of healthy longevity*. Beijing, China: Peking University Press, pp. 3-22

Hayward, M. D. and B.K. Gorman (2004), The long arm of childhood: the influence of early-life social conditions on men's mortality. *Demography* 41(1), pp.87-107

Herzog, A.R., and W.L .Rodgers (1981), Age and satisfaction – data from several large surveys. *Research on Aging* 3 (2), pp. 142-165

Herzog, A.R., and W.L. Rodgers (1988), Age and response rates to interview sample survey. *Journal of Gerontology: Social Sciences* 43, pp. S200-S205

Herzog, A.R., and W.L. Rodgers (1992), The use of survey method in research on older Americans. In: R.B. Wallace and R.F. Woolson (eds.): *The epidemiologic study of the elderly*, pp.60-90. New York: Oxford University Press

Horiuchi, S., and J.R. Wilmoth (1998), Deceleration in the age pattern of mortality at older ages. *Demography* 35(4), pp. 391-412

Jay, G.M., J. Liang, X. Liu, and H. Sugisawa (1993), Patterns of nonresponse in a national survey of elderly Japanese. *Journal of Gerontology: Social Sciences* 48, pp. S143-S152

Jenkins, K.R., and N. H. Fultz. (2005). Functional impairment as a risk factor for urinary incontinence among older Americans. *Neurourology and Urodynamics* 24, pp.51-55

Kalton, G. and D. Kasprzyk (1986), The treatment of missing survey data. *Survey Methodology* 12, pp.1-16

Kannisto, V. (1994), *Development of oldest-old mortality, 1950-1990: evidence from 28 developed countries*. Odense, Denmark: Odense University Press.

Katz S., A.B. Ford, R.W. Moskowitz, B.A. Jackson, and M.W. Jaffe (1963), Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. *Journal of the American Medical Association* 185(12), pp. 914-919

Kempen, G.I.J.M., and E. van Sonderen (2002), Psychological attributes and changes in disability among low-functioning older persons: does attrition affect the outcomes? *Journal of Clinical Epidemiology* 55, pp. 224-229

Klinkenberg, M., J.H. Smit, J.H. Deeg, D.L. Willems, B.D. Onwuteaka-Philipsen, G. van der Wal (2003), Proxy reporting in after-death interviews: the use of proxy respondents in retrospective assessment of chronic diseases and symptom burden in the terminal phase of life. *Palliative Medicine* 17, pp. 191-201

Kolonel, L.N., T. Hirohata, and A.M.Y. Nomura (1977), Adequacy of survey data collected from substitute respondents. *American Journal of Epidemiology* 106, pp. 476-484

Lee, D.S. (2002), *Trimming for bounds on treatment effects with missing outcomes*. Center for Labor Economics, University of California, Berkeley. Working paper no. 51

Lepkowski, J., G. Kalton, and D. Kasprzyk (1989), Weighting adjustments for partial nonresponse in the 1984 SIPP panel. *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, pp. 296–301

Lindner, J. R., T.H. Murphy, and G.E. Briers (2001), Handling nonresponse in social science research. *Journal of Agricultural Education* 42(4), pp. 43-53

Landerman, L.R., K.C. Land, and C.F. Pieper (1997), An empirical evaluation of the predictive mean matching method for imputing missing values. *Sociological Methods and Research* 26(1), pp. 3-33.

Magaziner J., E.M.Simonsick, T.M.Kashner, and J.R. Hebel (1988), Patient proxy response comparability on measures of patient health and functional status. *Journal of Clinical Epidemiology* 41, pp. 1065-1074.

McHorney, C.A., J.E. Ware, J.F.R. Lu, and C.D. Sherbourne (1994), The MOS 36 item Short Form Health Survey (SF-36) III: tests of data quality, scaling assumptions and reliability across diverse patient groups. *Medical Care* 32, pp. 40–62

Mihelic, A.H., and E.M. Crimmins (1997), Loss to follow-up in a sample of Americans 70 years of age and older: the LSOA 1984-1990. *The Journals of Gerontology* 52B, pp. S37-S48

Mohadjer, L., B. Bell, and J. Waksberg (1994), *Accounting for item nonresponse bias—National Health and Nutrition Examination Survey III*. National Center for Health Statistics, Hyattsville, Maryland

Morris, P.E. (1983), The validity of subjective reports on memory. In: J.E. Harris and P.E. Morris (eds.): *Everyday memory: actions and absent-mindedness*. London: Academic Press, pp. 153-172

Mossel, P.A., and R.S. Brown (1984), *The effect of sample attrition on estimates of channeling's impact for an early sample.* Executive summary, U.S. DHHS. Washington, D.C.

Norris, F.H., and W.J. Goudy (1986), Characteristics of older nonrespondents over five waves of a panel study: comments. *Journal of Gerontology* 41, pp. 806-807

Nunnally, J.C. (1994). *Psychological Theory.* 3rd ed. New York: McGrawhill.

Office of Management and Budget (OMB). (2001). *Measuring and reporting sources of errors in survey*. Statistical Policy working paper No. 31. Office of Management and Budget, Washington D.C.

Perls, T.T., K. Bochen, M. Freeman, L. Alpert, and M.H. Sliver (1999). Validity of reported age and centenarian prevalence in New England. *Age and Ageing 28*, pp.193-197.

Pierre, U., S. Wood-Dauphinee, N. Korner-Bitensky, D .Gayton, and J. Hanley (1998), Proxy use of the Canadian SF-36 in rating health status of the disabled elderly. *Journal of Clinical Epidemiology* 51, pp. 983-990

Powell, D.A., E. Furchtgott, M. Henderson, L. Prescott, A. Mitchell, P. Hartis, J.D. Valentine, and W. L. Milligan (1990), Some determinants of attrition in prospective studies on aging. *Experimental Aging Research* 16, pp. 17-24

Rodgers, W.L. (1988), *Epidemiological Survey of Older Adults: Response rate, data quality and the use of proxies*. Paper presented at the 41st annual scientific meeting of the Gerontological Society of American. San Francisco

Rodgers, W.L., and A.R. Herzog (1992), Collecting data about the oldest: problems and procedures. In: R.M. Suzman, D.P. Willis, K.G. Manton (eds.): *The oldest old.* New York: Oxford University, pp.135-156

Rothman, M.L., S.C. Hedrick, K.A. Bulcroft, D.H. Hickam, and L.Z. Rubenstein (1991), The validity of proxy-generated scores as measures of patient health status. *Medical Care 29*, pp.115-124

Slymen, D.J., J.A. Drew, J.P. Elder, and S.J. Williams (1996), Determinants of non-compliance and attrition in the elderly. *International Journal of Epidemiology* 25, pp. 411-419

SPSS (2007), SPSS Version 15.01 www.SPSS.com

Stata Corp. (2007). Stata version 10.0. www.stata.com

Stewart, A.L., R.D. Hays, and J.E. Ware(1992), Methods of constructing health measure. In: A. L. Stewart and J.E. Ware (eds.): *Measuring function and well-being-The medical outcome study approach.* Duke University Press, NC, pp. 67-85

Sugisawa, H., H. Kishino, Y. Sugihara, H. Okabayashi, and H. Shibata (1999), Comparison of characteristics between respondents and nonrespondents in a national survey of Japanese elderly using six year follow-up study. *Nippon Koshu Eisei Zasshi* 46(7), pp. 551-562

Tang, S.T., and R. McCorkle (2002), Use of family proxies in quality of life research for cancer patients at the end of life: a literature review. *Cancer Investigation* 20(7-8), pp. 1086-1104

Thatcher, A.R., V. Kannisto, and J.W. Vaupel (1998), *The force of mortality at ages 80 to 120.* Odense: Odense University Press. Online at http://www.demogr.mpg.de/Papers/Books/Monograph5/ForMort.htm.

Vaupel, J.W., J.R. Carey, K. Christensen, T.E. Johnson, A.I. Yashin, N.V. Holm, I.A. Iachine, V. Kannisto, A.A. Khazaeli, P. Liedo, V.D. Longo, Y. Zeng, K.G. Manton, J.W. Curtsinger (1998), Biodemographic trajectories of longevity. *Science* 280(5365), pp. 855 – 860

Vellas, B.J., S.S. Wayne, P.J. Garry, and R.N. Baumgartner (1998), A two-year longitudinal of falls in 482 community-dwelling elderly adults. *The Journals of Gerontology* 53A(4), pp.M264-M274

Wallace, R.B., and A.R. Herzog (1995), Overview of the health measures in the health and retirement study. *The Journal of Human Resources* 30 (supplement), pp. S84-S107

Wallace, R.B., F.J. Kohout, and P.L. Colsher (1992), Observations on interview survey of the oldest old. In: R. M. Suzman, D.P. Willis, and K.G. Manton (eds.): *The oldest old*. New York: Oxford University, pp123-134

Zeng, Y., and D. Gu. (2008). Reliability of age reporting among the Chinese oldest-old in the CLHLS datasets. In Y. Zeng, D. L. Poston, D.A. Vlosky, and D. Gu (eds.). *Healthy longevity in China: Demographic, socioeconomic, and psychological dimensions*, pp. 61-79. Dordrecht, The Netherlands: Springer Publisher.

Zeng, Y., and J.W. Vaupel (2003), Oldest-old mortality in China. *Demographic Research* 8, pp.215-244

Zeng, Y., J.W. Vaupel, Z. Xiao, C. Zhang, and Y. Liu (2001), The healthy longevity survey and the active life expectancy of the oldest old in China. *Population* 13(1), pp. 95-116

Zhang, M.Y. (1993). *Manual of evaluation of scales in psychiatrics*. Changsha, China: Hunan Science and Technology Press

Zhang, M., Z. Zhu, and P. Chen (1998), Community investigation of the activities of daily living (ADL) and medical conditions of the elderly in Shanghai. *Chinese Journal of Medical Science* 78(2), pp. 124-127

Zhang, W., D. Gu, Y. Zeng, and Y. Liu (2006). *General data quality assessments for the 2005 CLHLS*. The CLHLS technical document 2006 No.1. The Center for Healthy Aging and Family Studies, Peking University, Beijing. July, 2006. (in Chinese)

Zimmer, Z., L. Martin, and M.C. Chang (2002), Changes in functional limitation and survival among older Taiwanese, 1933, 1996, and 1999. *Population Studies* 3, pp. 265-276

Zsembik, B.A. (1994), Ethnic and sociodemographic correlates of the use of proxy respondents: The National Survey of Hispanic Elderly People, 1988. *Research on Aging* 16 (4), pp. 401-414