In Y. Zeng, D.L. Poston, D.A. Vlosky, and D. Gu. (eds.). Healthy Longevity in China: Demographic, Socioeconomic, and Psychological Dimensions (Pp39-59). Dordrecht, The Netherlands: Springer.

Chapter 3

General Data Quality Assessment of the CLHLS[1]

DANAN GU

Center for the Study of Aging and Human Development, Medical School of Duke University

Abstract

This chapter provides a comprehensive review of data quality of the third wave of the Chinese Longitudinal Healthy Longevity Survey (CLHLS) in 2002 in terms of proxy use, nonresponse rate, sample attrition, and reliability and validity of major health measures. The results show that the data quality of the 2002 wave of the CLHLS is generally good. Some recommendations in use of the dataset are provided.

Keywords: Accuracy of imputation, bias, convergent validity, Cronbach's alpha coefficient, data assessment, discriminant validity, don't know answer, factual question, full proxy response, homogeneity, imputation, inconsistent responses, internal consistency, item nonresponse, item-total correlations, knowledgeable proxy, minimum reliability coefficient, missing completely at random, missing item, missing value, multiple imputation, multiple item scale, next-of-kin, nonresponse, nonresponse rate, objective question, proxy, proxy reporter, proxy response, proxy use, reliability, response pattern, response structure, sample attrition, significant other, sources of error, unit nonresponse, validity

1. INTRODUCTION

This chapter provides a comprehensive review of the quality of the data from the third wave of the Chinese Longitudinal Healthy Longevity Survey (CLHLS) in 2002 in terms of proxy use, nonresponse rate, sample attrition, and reliability and validity of major health measures. A data quality assessment for the first wave in 1998 may be found elsewhere (Zeng et al. 2001).[2] The third wave of the CLHLS gathered extensive questionnaire data through interviewing 16,020 elderly aged 65+. The survey was conducted in randomly selected counties and cities in 22 of China's 31 provinces.[3] The 2002 wave extended its age range of the sampled elderly to include the age range from 65 to 79 who were not included in the first two waves in 1998 and 2000. Among the 16,020 sample persons in 2002, 3,189 were centenarians, 3,747 were nonagenarians, 4,239 were octogenarians, and 4,845 were aged 65-79.

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. Extensive data were collected including family structure, living arrangements and proximity to children, self-rated health, self-evaluation on life satisfaction,

---

[1] The research reported in this chapter is supported by the National Institute on Aging grant (R01 AG023627-01) and the National Natural Science Foundation of China key project grant (No. 70533010).

[2] There is no systematical data assessment publication in English for the 2000 wave. According to the Chinese publication (Gu and Zeng 2004), the data quality of the CLHLS in 2000 is good.

[3] 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, and 866 counties and cities in the 1998, 2000, and 2002 surveys, respectively. The increase in numbers of survey units in 2002 was mainly due to adding an elderly comparison group aged 65-79 who were not interviewed in 1998 and 2000; and

chronic disease, medical care, social activities, diet, smoking and alcohol drinking, psychological characteristics, economic resources, caregivers and family support, nutrition and other health-related conditions in early 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 initially planned, instrumental activities of daily living (IADL) questions were added in the 2002 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 chapter follows this framework in its attempt to assess the data quality of the CLHLS in 2002. All data analyses to this end are conducted using STATA 8.0 and SPSS 12.0. Assessments of the accuracy of age reporting, mortality, and morbidity in the first three waves of the CLHLS are presented in Chapters 4, 5, and 6 of this volume.

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).

---

partly due to an administrative boundary change in the later wave; or some selected counties/cities that had no centenarians in an earlier wave but had centenarians in a later wave.

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 2002 wave. Consistent with the first two 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. Our analysis further shows 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).

--- Tables 1 and 2 about here------

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.[4] 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).[5]

The unit nonresponse rate among the Chinese oldest-old was very low, about 4 percent, in the first three 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

[4] Proxies for the prevalence of chronic diseases and primary cause of death for the decedent persons are not reliable as indicated in Chapter 6 of this volume.

[5] 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

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 65-79 (5.1 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.[6] 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.

---

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.

[6] 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). No difference is observed between the 2002 wave and the previous two waves. Table 4 summarizes the variables with incomplete answers of 2 percent or more.[7] Variables with the highest incomplete rate are “parents’ ages at death”. Although the incomplete rate declined in the 2002 wave compared to those in the 1998 and 2000 waves, it remains at more than 25 percent. Hence, extreme caution is recommended in dealing with such 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, which is consistent with some previous studies in Western nations (Francis and Busch 1975; Herzog and Rodgers 1981).

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 how the respondents who give answers differ from all those who are eligible to be interviewed

[7] 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

(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. 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.

--- Tables 3 and 4 about here------

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

---

listed in Table 4 since they were not asked in the 2000 and 2002 waves.

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.[8] 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 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 11,162 interviewees in 2000, 6,291 (56.3 percent) were still alive at the 2002 wave, 3,335 (29.9 percent) died before the interview was held in 2002, and 1,536 (13.8 percent) were lost.[9] The proportion of attrition was higher between the 2000 and 2002 waves than between the 1998 and 2000 waves (9.6 percent). The true reason for the higher sample attrition in the period of 2000-2002 is not known. The frequency distribution indicates that the urban sample had a higher attrition in the 1998-2000 period than in the 2000-2002 period. 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, physically and cognitively impaired, and with low social contacts are associated with higher attrition rates.

[8] Those who died but followed-up at the subsequent wave is not considered as a type of sample attrition in this study.
[9] Those who were lost to the follow-up also include some who actually died.

This is consistent with prior findings in the literature (e.g., Powell et al. 1990; Sugisawa et al. 1999). However, unlike some previous research that finds older age is associated with a higher attrition (e.g., Slymen et al. 1996), we find that the age pattern is not significant from Wave Two to Wave Three.[10]

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. Furthermore, 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. It is interesting to note that respondents using a proxy are less likely to be lost to follow-up in the following wave, possibly because they are less mobile, which makes them easier to locate.

-- Table 5 about here--

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-

[10] There is a significant age pattern of sample attrition between Wave One and Wave Two: younger oldest-old are more likely to be lost to follow-up.

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 the 2002 wave compared with levels in the previous waves, although the inconsistency given by interviewers is lower in the 2002 wave. Inconsistent responses seem to increase slightly with age. The difference between genders is trivial.[11]

-- Tables 6 and 7 about here---

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 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,

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

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. One unique feature of the CLHLS is that relatively comprehensive information on the extent of disability and suffering before dying was obtained by interviewing a close family member (next-of-kin).The 2002 wave gathered 3,340 questionnaires for the deceased aged 80 and over who died between the 2000 and the 2002 waves. The remaining parts of this chapter aim to provide a relatively detailed assessment on the reliability and validity for all major health domains in the CLHLS.

6.1 Internal Consistency of Multiple Items Scales

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,[12] implying that they might not be appropriate to use in scale generation.[13] It is also interesting to note that the reliability for some scales and variables among young elders is slightly lower than those for the oldest-old. The reasons for this are unclear. Smaller sample sizes among the younger elderly could be a possible cause. Further research is clearly warranted.

-------- Table 8 about here---

6.2 Homogeneity

The homogeneity of the ADL, IADL, and MMSE scales, and other potential scales is assessed by evaluating item-total correlations. Item-total correlations compute the correlation between an item and its own scale with the item of interest eliminated from the calculation of the score. Although some researchers argue that it is considered satisfactory if the item-total correlation reaches 0.40 or more for the purpose of comparison (Ware et al. 1980), other scholars suggest that the criterion for item-total correlations might be efficient if it exceeds 0.20, especially for categorical items that define the extremes of the scale range (Streiner and Norman 1995). Our results show that all item-total correlations (see Table 8) are over the minimum requirement suggested by Streiner and Norman (1995), and even over the 0.40 criterion, except those for personality if persons unable to answer questions are excluded. Note that for the ADL scale, if continence is deleted, both the reliability coefficients and the item–total correlations will be substantially improved, and the item-total correlation will pass 0.40 for young elders in the 2002

[12] 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.

[13] 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.

dataset, indicating the possibility of removing continence from ADL scales, as some recent studies have done (e.g., Jagger et al. 2001).

The percentage of respondents at the highest possible (ceiling) or lowest possible (floor) scores also should be noted. As for scales or variables related to functional limitations, a negative skew pattern is observed, indicating distributions with respondents scoring towards the positive end of the scales in the CLHLS. This is anticipated for a generally well elderly subpopulation.

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.[14]

---Table 9 about here---

8. CONCLUDING REMARKS

This chapter has examined the data quality of the 2002 wave of the CLHLS, mainly on proxy use, item incompleteness, sample attrition, and the dimensionalities of reliability and the validity

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

of health condition measurements. 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 chapter 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.[15]

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 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.

[15] 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

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

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.

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., 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

Jagger, C., A.J. Arthur, N.A. Spiers, and M. Clark (2001), Patterns of onset of disability in activities of daily living with age. *Journal of American Geriatrics Society* 49, pp. 404-409

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

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.

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 (2002), Version 11.5. www.SPSS.com

Sprangers, M.A.G., and N.K. Aaronson (1992), The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *Journal of Clinical Epidemiology* 45, pp. 743–760

Streiner, D.L., and G.R. Norman (1995), *Health measurement scales: a practical guide to their development and use*. *2nd edition.* Oxford: Oxford University Press

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

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

Ware, J.E., R.H. Brook, A. Davies-Avery, K. Williams, A.L. Stewart, W.H. Rogers, C.A. Donald, and S.A. Johnston (1980), *Model of health and methodology*. Santa Monica, CA: RAND Corporation

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

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

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

Note: the number of questions in the 2002 wave is slightly more than in the previous two waves, which might cause a relatively high % of use of proxy. Full proxy means all questions except those that must be answered by the sampled person are answered by the proxy.

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

	1998 Wave	2000 Wave	2002 Wave [a]	2002 Wave [b]
Spouse	5.55	5.18	5.02	32.00
Child or spouse of children	74.01	67.41	67.68	50.53
Grandchild or spouse of grandchild	12.37	16.56	15.92	8.00
Great grandchild or spouse of great grandchild	0.28	0.81	0.95	0.00
Sibling	0.24	0.18	0.20	0.84
Caregiver	2.76	4.34	5.27	2.11
Others	4.79	5.54	4.97	6.53

Note: (1) a, for age 80-105; b, for age 65-79. (2) 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, and 3.9% and 2.6% for ages 80-105 and 65-79 in the 2002 wave, respectively.

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

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.

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

	1998	2000	2002 [a]	2002 [b]
Eating style (D4 sets)	1.2-4.3	<2.0	<2.0	<1.0
Habit (i.e., smoking, drinking, exercise, physical laboring)	1.0-4.0	<2.0	<1.5	<1.5
Marriage history	>5.0[c]	>4.0[c]	>1.5[c]	>0.1[c]
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
Birth order and # of sibling	2.0-3.0	2.2	2.2	<0.5
Sibling information	>7.0[c]	>3.4[c]	>3.7[c]	>0.5[c]
Children's information	>3.0[c]	>2.7[c]	>2.2[c]	>1.0[c]
Blood pressure	3.0	3.2	0.9	<0.2
Height (Acromion-processus styloideus ulnae; right knee to the floor)	5.0	NA	0.0	0.0
Weight	8.0	1.7	0.0	0.0
Chronic diseases	7.0-10.0	5.0-8.0	3.6-7.5	2.4-5.0
Intergenerational transfers (upward)	NA	NA	6.5	7.5
Intergenerational transfers (downward)	NA	NA	6.3	7.0

Note: (1) a, age 80-105; b, age 65-79; c, 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.

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

Variables	Lost to follow-up in 2000	Lost to follow-up in 2002
Females (males)	1.21*	1.20*
Age 90-99 (age80-89)	0.68***	1.03
Age 100-105 (age80-89)	0.51***	0.90
Rural (urban)	0.46***	0.70***
Minority ethnicity (Han)	0.46***	0.62**
1+ schooling (no schooling)	1.49***	1.23*
Currently married (not married)	1.04	1.07
Living alone (others)	1.19	1.17#
High proximity with children (low)	0.82*	0.80**
Bad ADL (good ADL)	1.14	1.32***
Bad MMSE (good MMSE)	1.09	1.20**
Self reported bad health (good health)	0.97	1.00
Proxy (no proxy)	0.89	0.89#
Missing group 2 (missing group1)	1.11	1.22**
Missing group 3 (missing group 1)	1.13	1.33***
2000 newly interviewed (1998 interviewed)	---	1.31***
N	8,805	10,844
-2LL	5298.1***	8506.6***

Note: (1) age 80-105 only. (2) Three missing groups are classified based on the missing rate of each respondent. Group1, <2%; Group 2, 2-5%; Group 3, >5%.

Table 6. Inconsistent responses for selected items in the CLHLS

	1998		2000		2002 [a]		2002 [b]	
Inconsistent items	#	%	#	%	#	%	#	%
1. ADL fully dependent but can pick-up a book while standing	112	1.27	110	1.00	108	0.99	0	0.00
2. ADL fully independent but can't stand up from a chair	50	0.57	83	0.76	230	2.11	96	1.98
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
4. Reported bedsores but does housework or field work everyday	6	0.07	26	0.24	6	0.05	1	0.02
5. Had a proxy for answering some questions but interviewer didn't mark*	891	10.12	544	4.96	248	2.27	86	1.78

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?".

Table 8. Reliability coefficients and validity for selected measures in the 2002 wave

Scales and measures*	Cronbach's alpha coefficient	N	Range of alphas if individual item deleted	Range of item-total correlations	% at floor+	% at ceiling+
Age 80-105						
Functioning of upper extremities (3)	0.833	10912	0.753-0.794	0.668-0.710	5.3	80.0
Functioning of body mobility (2)	0.762	10905	NA	0.617	7.8	44.8
Negative personality related variables (3)	0.891	10953	0.815-0.892	0.753-0.822	--	--
Negative personality related variables (3) #	0.662	9157	0.463-0.696	0.391-0.550	0.0	2.2
Positive personality related variables (4)	0.918	10953	0.880-0.917	0.750-0.857	--	--
Positive personality related variables (4) #	0.453	8838	0.344-0.416	0.237-0.307	0.5	10.0
ADL (6)	0.867	10905	0.815-0.876	0.508-0.839	0.6	59.5
ADL for deceased persons between 2000 to 2002 (6)	0.939	3188	0.916-0.930	0.624-0.902	15.9	22.9
IADL (8)	0.937	10951	0.924-0.932	0.727-0.832	18.8	14.3
Mini-Mental State Examination (MMSE) (22)	0.984	10945	0.983-0.984	0.771-0.901	10.9	19.7
Mini-Mental State Examination (MMSE) (22)#	0.888	6971	0.877-0.887	0.286-0.659	0.1	30.6
Age 65-79						
Functioning of upper extremities (3)	0.892	4843	0.839-0.858	0.775-0.798	1.0	92.5
Functioning of body mobility (2)	0.585	4843	NA	0.418	0.6	86.4
Negative personality related variables (3)	0.655	4845	0.482-0.690	0.392-0.526	--	--
Negative personality related variables (3)#	0.625	4758	0.438-0.678	0.351-0.500	0.2	15.2
Positive personality related variables (4)	0.630	4845	0.530-0.596	0.414-0.490	--	--
Positive personality related variables (4)#	0.456	4684	0.323-0.428	0.224-0.326	0.0	3.2
ADL (6)	0.817	4843	0.745-0.854	0.355-0.826	0.1	93.1
IADL (8)	0.862	4843	0.835-0.903	0.469-0.730	1.2	66.9
Mini-Mental State Examination (MMSE) (22)	0.952	4844	0.948-0.951	0.574-0.777	0.6	54.6
Mini-Mental State Examination (MMSE) (22)#	0.788	4461	0.764-0.789	0.120-0.568	0.0	59.3

Note: (1) *, some are not designed as scales in the questionnaire, but they are related variables to measure similar functioning. Our purpose here is to examine their reliability to see the possibility of generating scales later on. (2) +, percentage of subjects with worst and best possible scores, respectively. #, 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". (3) The figure in the parentheses indicates the number of items in corresponding scales or groups. (4) NA, not applicable; --, not calculated since those close-end questions contain the "unable to answer" answer, which didn't provide any information regarding personality. (5) Four newly added questions in the cognitive function section in the 2002 questionnaire are not included in generating the MMSE scale. Four newly added questions are mainly designed for the elderly aged 65 to 79 by the CLHLS research team. We suggest users drop these four variables in creating MMSE scores since they are not included in the MMSE scale proposed by Folstein et al. (1975). (6) Minor difference could be found from those in Zeng et al. (2001) due to different age ranges used in this report.

Table 9. Convergent and discriminant validity for selected measures in the 2002 wave

Scales and measures*	FU	FB	NP	PP	ADL	MMSE	IADL
Age 80-105							
Functioning of upper extremities (3)	**0.68-0.72**	0.21-0.27	0.07-0.12	0.05-0.08	0.13-0.26	0.03-0.18	0.16-0.22
Functioning of body mobility (2)	0.21-0.27	**0.54**	0.12-0.16	0.09-0.13	0.19-0.40	0.08-0.26	0.37-0.48
Negative personality related variables (3) #	0.07-0.12	0.12-0.16	**0.31-0.57**	0.11-0.33	0.04-0.11	0.02-0.13	0.07-0.18
Positive personality related variables (4) #	0.05-0.08	0.09-0.13	0.11-0.33	**0.16-0.32**	0.02-0.12	0.00-0.15	0.03-0.18
ADL (6)	0.13-0.26	0.19-0.40	0.04-0.11	0.02-0.12	**0.23-0.63**	0.08-0.20	0.16-0.52
Mini-Mental State Examination (MMSE) (22)#	0.03-0.18	0.08-0.26	0.02-0.13	0.00-0.15	0.08-0.20	**0.09-0.78**	0.09-0.30
IADL (8)	0.16-0.22	0.37-0.48	0.07-0.18	0.03-0.18	0.16-0.52	0.09-0.30	**0.48-0.69**
Age 65-79							
Functioning of upper extremities (3)	**0.77-0.80**	0.12-0.14	0.02-0.06	0.00-0.07	0.02-0.19	0.00-0.09	0.13-0.17
Functioning of body mobility (2)	0.12-0.14	**0.38**	0.03-0.10	0.01-0.09	0.07-0.29	0.01-0.13	0.25-0.40
Negative personality related variables (3)#	0.02-0.06	0.03-0.10	**0.30-0.58**	0.11-0.33	0.00-0.10	0.00-0.10	0.02-0.17
Positive personality related variables (4)#	0.00-0.07	0.01-0.09	0.11-0.33	**0.13-0.26**	0.00-0.08	0.00-0.09	0.03-0.15
ADL (6)	0.02-0.19	0.07-0.29	0.00-0.10	0.00-0.08	**0.14-0.60**	0.00-0.18	0.10-0.44
Mini-Mental State Examination (MMSE) (22)#	0.00-0.09	0.01-0.13	0.00-0.10	0.00-0.09	0.00-0.18	**0.01-0.60**	0.01-0.18
IADL (8)	0.13-0.17	0.25-0.40	0.02-0.17	0.03-0.15	0.10-0.44	0.01-0.18	**0.30-0.68**

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.