Cross-Cohorts Changes in Physical and Cognitive Functions of Oldest-Old Aged 80-105 in China

Yi Zeng, Qiushi Feng, Qihua Tan, Kaare Christensen and James Vaupel

Abstract

This study compares three groups of cohorts born 10 years apart and aged 80-89, 90-99 and 100-105 at interview, using data from Chinese Longitudinal Healthy Longevity Surveys. We found that death rates at oldest-old ages among later cohorts was substantially reduced, compared to cohorts born 10-years earlier; later cohorts had significantly reduced impairment in activities of daily living(ADL) than did cohort born 10 years earlier. We discovered that cognitive functional scores and objective physical performance test scores(stand-up from a chair, pick-up a book from floor, turning around 360^0) were all significantly worse in later cohorts, compared to the cohorts born 10 years earlier. Three factors may explain our findings: (1) the mixed effects of the two opposing processes of the success-of-success and failure-of-success when human life span is prolonging; (2) Differences in disability measurements of ADL and objective physical performance tests; (3) Cross-cohorts differences in educational levels and childhood conditions.

Introduction

It is well known that the population of China, which is about one-fifth of the world's total, is aging rapidly (Banister and Bloom 2010). The amount and proportion of Chinese oldest-old aged 80+, who most likely need daily life assistance, are expected to increase much more dramatically than the young-old aged 65-79 in the next a few decade. Under the medium or low mortality assumptions, the total number of elderly aged 65+ in China is estimated to increase dramatically from 111 million in 2010 (8.2% of the total population) to 337 to 400 million in 2050 (23.9% to 26.9% of the total population); the number of oldest-old aged 80+ were about 19.3

million in 2010, but it will climb extraordinarily to about 107 to 150 million in 2050, respectively (Zeng and George, 2010). The average annual rate of increase of the oldest-old from 2000 to 2050 is about 4.4~5.1 percent in China, more than twice that of the U.S., EU and other industrialized countries (U.N. 2011). The main reason why the number of Chinese oldest-old will climb so quickly especially after the year 2030 is that China's “baby boomers,” who were born in the 1950s and 1960s, will fall into the category of the "oldest-old" after 2030, as well as the substantial decrease or mortality among senior adults especially among the oldest-old.

The dramatically rapid growth of the oldest old population is a main challenge for the public health system as health generally declines in the late life, especially at oldest-old ages. However, there are active debates on two contrasting scenarios of health changes in the aging population. An view is that the advancement of medical technology, improvement of health lifestyle, and socioeconomic development will postpone the onset of disability and chronicle diseases among the elderly so that morbidity will be “compressed” in late age (Fries, 1980). On the contrary, some other scholars believe that the same improvements in social/medical services could lower the mortality by saving more frail elderly with health problems, thus worsen the health and disability status of the whole elderly population (Gruenberg 1977). The former view indicates a “success of success”, i.e. better health with prolonged life in the old age, whereas the latter view represents a “failure of success”, i.e. poorer health with prolonged life in the old age (Waidmann, Bound, and Schoenbaum 1995). Some scholars argued that the two mechanisms of “success of success” and “failure of success” may coexist and interplay in reality (Manton 1982; Robine and Michel 2004). Such speculations have been extensively reflected in the mixed findings on the disability trends of the elderly populations. The current literature in this field consistently showed that different societies, even with similar level of socioeconomic development, may have opposite disability trends in the same period, whereas within one society, the disability trend could also be reversed at the different observation periods (Robin & Michel 2004; Christensen et al. 2009; Fuller-Thomson et al. 2009).

Research on the cohort and temporal changes of health in physical and cognitive functions among elderly are very limited in China, particularly for the oldest-old group. Du and Wu (2006) reported that the Activities of Daily Living (ADL) disability prevalence increased from 1994 to 2004 among the Chinese 60+ elderly. In contrast, Gu and Zeng (2006) reported a declining trend of the ADL disability among Chinese elders aged 65+ from 1992 to 2002. Feng and colleagues (2013) used the local survey data from Shanghai, the largest city with the highest socioeconomic level and largest proportion of older adults in China, and found that the ADL and IALD (Instrumental Activities of Daily Living) disability both declined in the elderly aged 65+ from 2002 to 2008. Martin, Feng, Schoeni and Zeng (2013) recently conducted a cross-sectional analysis on weighted average prevalence rates of self-reported ADL, IADL and self-reported whether being able to walk for one kilometers, to carry 5 kilograms and to crouch anfd stand 3 times among Chinese oldest-old aged 80-105 (without distinguishing sub age groups), using the data from multiple waves of the Chinese Longitudinal Healthy Longevity Survey (CLHLS) conducted in 1998, 2000, 2002, 2005, and 2008. The results showed that the weighted average scores of ADL and IADL limitations among Chinese oldest-old aged 80+ declined by 2.2 and 2.4 percent per year from 1998 to 2008. Males did not experience improvement in self-reported ability to carry out physical functions over the period 1998-2008, but females did. Martin et al. (2013) provided some strong evidence to support the “success of success” hypothesis, but did not test the hypothesis of “failure of success” and the mixed effects of both “success of success” and “failure of success”. As far as we know, all of the previously published/reported studies on Chinese elderly disability trends focused on period changes in self-reported disabilities, but did not investigate the cross-cohorts changes in objective physical performance, cognitive function, self-reported disabilities and mortality. Our present study intends to fill in this research gap.

In their most recent Lancet paper, Christensen et al. (2013) compared the Danish cohort born in 1905 and assessed at age 93 years to later cohort born in 1915 and assessed at age 95 years. They found that the 1915 cohort had significantly better activities of daily living and cognitive function scores than did the 1905 cohort, but the cohorts did not differ consistently in the physical performance tests. Compared to the earlier cohort, for example, Danish later cohort had significantly worse capacity of standing-up from a chair and significantly higher per cent of not being able to walk for 3 meters for both sexes combined and for females, but marginally significant or not significant for males.

The Danish study has provided support to the theoretical speculation of the mixed effects of both “success of success” and “failure of success” among nonagenarians in a developed country. However, do such mixed effects also exit in developing countries such as China? If it does exist, to what extend it is similar or different from what occurred in the developed countries such as Denmark? Our present study intends to address these interesting and important questions by testing the following hypotheses, which were based on relevant literature review including the Danish study and our own previous research.

H_0: Based on comparative analysis of cohorts of the oldest-old born in 1909-1918 vs 1919-1928, born in 1899-1908 vs 1909-1918, and born in 1893-1898 vs 1903-1908, the later cohorts might have significantly lower mortality rates and lower prevalence of disability in activities of daily living, compared to the earlier cohorts, due to effects of “success of success”, but the later cohorts might have significantly poorer objective physical performance tests scores and lower cognitive function scores, due to the effects of “failure of success”.

Methods

Data source

Data used in this article are from the 1998 and 2008 waves of the Chinese Longitudinal Healthy Longevity Survey (CLHLS). The CLHLS baseline survey was conducted in 1998, and the follow-up surveys were conducted in 2000, 2002, 2005, 2008 and 2011. The CLHLS has been conducted in randomly selected half of the counties and cities in 22 of the 31 provinces and municipalities of China (Liaoning, Jilin, Heilongjiang, Hebei, Beijing, Tianjin, Shanxi, Shaanxi, Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Guangxi, Sichuan, and Chongqing). The survey areas covered 985 million persons in the 1998 baseline year, which was 85 percent of the total population of China. The 1998 baseline and 2000 follow-up surveys included oldest-old interviewees aged 80 and above and did not include young elders aged 65-79. The CLHLS survey has been expanded since the 2002 wave to cover ages 65 and above through the addition of a new sub-sample of interviewees aged 65-79.

CLHLS tried to interview all centenarians who voluntarily agreed to participate in the study in the sampled counties and cities; for each centenarian interviewee, CLHLS tried to interview one nearby octogenarian, one nearby nonagenarian and one nearby young-old aged 65-79 of predefined age and sex. “Nearby” is loosely defined – it could be in the same village or in the same street, if available, or in the same town or in the same sampled county or city district. The predefined age and sex are randomly determined, based on the randomly assigned code numbers of the centenarians, to have comparable numbers of males and females at each of the age groups. At each follow-up wave, survivors were re-interviewed. Those elderly who were interviewed at an age younger than 100 but subsequently died before the next wave were replaced by new interviewees of the same sex and age as the deceased (or within the range of minus or plus two years old) in the follow-up waves conducted in 2000, 2002, 2005 and 2008. The CLHLS 2011 wave conducted follow-up interviews for those surviving participants, without replacement for deceased and lost-to-follow-up elders. Information about the date of death and health status of the elderly who were interviewed in the previous wave but died before the subsequent survey was collected by interviewing a close family member.

CLHLS is a national representative survey on determinants of healthy longevity, with over-sampling the oldest-old aged 80+ especially males, plus comparative samples of the young-old aged 65-79. The target was to interview more or less equal numbers of males and females at each age groups of 65-79, 80-89 and 90-99 and almost all centenarians in the sampled counties/cities. CLHLS does not follow the proportionally sampling framework in order to avoid too small sub-sample size for the oldest-old especially males. Consequently, appropriate weights based on the census and the CLHLS data need to be used to compute the averages of the age groups below age 100, but no weights are needed when computing the average of the centenarians. The method for computing the age-sex and rural-urban specific weights and the associated discussions are presented in the Appendix of Chapter Two of the book by Zeng et al. (2008) and available at the CLHLS Webpage.

The CLHLS questionnaire was initially translated from the instruments of the Danish longevity survey which provided data for the analysis reported in Christensen et al. (2013), with revisions in language and expressions adapted to the Chinese culture and socioeconomic context, while sustaining the substantive contents. Thus, the CLHLS data are adequately compatible to the corresponding data derived from Danish and other international surveys on longevity.

The systematic assessments on data quality concerning accuracy of age-reporting, reliability, validity, consistency of the main measures, and randomness of attrition show reasonably good quality in the CLHLS data sets (Goodkind, 2009; Gu and Dupre, 2008; Gu, 2008; Chen 2010; Shen 2010). For example, the reliability coefficients of the 10 categories of variables are reasonable (see Table 3 in Zeng and Vaupel et al. 2001). The ADL (activities of daily living) reliability coefficient is 0.88 in the 1998 baseline survey, as compared to 0.87 in the Duke Older American Resources and Services Program survey (Fillenbaum 1988) and 0.89 in the Canadian 1991-1992 elderly survey (Penning and Strain 1994). The factor analysis (e.g. Anita et al. 1992) demonstrates that in the CLHLS 1998 baseline survey interviewees' answers to questions of the same category but different aspects are generally consistent. The rates of logically inconsistent answers seem reasonably low; and the rates of "Don't know" and

“Missing” answers are also relatively small (see Tables B-1, B-2, B-3 in Appendix B in Zeng and Vaupel et al. 2001).

Study population of the cohorts in our comparative analysis

We compare the following three pairs of the cohorts of octogenarians, nonagenarians and centenarians, and the two cohorts in each pair of the comparison were born ten years apart, with the same age at the time of the CLHLS surveys which were conducted 10 years apart:

(1) Octogenarians: comparison between cohort born in 1909-1918 (assessed at ages 80-89 with mean age 83.1 in 1998 survey, n=3,235) and cohort born in 1919-1928 (assessed at ages 80-89 with mean age 83.0 in 2008 survey, n=4,053).

(2) Nonagenarians: comparison between cohort born in 1899-1908 (assessed at ages 90-99 with mean age 92.1 in 1998 survey, n=2,896) and cohort born in 1909-1918 (assessed at ages 90-99 with mean age 92.2 in 2008 survey, n=4,338).

(3) Centenarians: comparison between cohort born in 1893-1898 (assessed at ages 100-105 with mean age 101.1 in 1998 survey, n=2,197) and cohort born in 1903-1908 (assessed at ages 100-105 with mean age 101.7 in 2008 survey, n=2,809).

The data for our cross-cohorts comparative analysis were from 8,328 valid cases of the participants aged 80-105 in the CLHLS 1998 baseline survey and 11,200 valid cases of the participants aged 80-105 in the CLHLS 2008 follow-up survey (with replacements of deceased and lost-follow-up participants). The frequency distributions of the sample by the dependent variables and the covariates are listed in Table 1.

--Table 1 about here---

Assessment procedure

The CLHLS 1998 and 2008 surveys used almost exactly the same ascertainment and assessment protocols. A proxy responder who was a close family member of the interviewee was encouraged to participate in the interview for the subjective questions if the cohort member was unable to answer because of mental or physical handicap. No proxy was used for the objective questions, such as cognitive function and physical performance assessments, self-reported health and life satisfaction. In most cases, the survey took place in the participants' homes and was done by one of the well-trained interviewers who were selected from the local networks of aging committees, centers for disease prevention and control (CDC), or university students. The assessment consisted of an interview, physical and cognitive tests. The data variables analyzed in this article are described below.

Mean annual death rates

As described in the "Data source" section above, information on date of death were collected for the interviewees who were interviewed in 1998 or 2008, but died in the inter-wave period 1998-2000 or 2008-2011. We estimated the age-sex-specific weighted mean annual death rate for each of the cohorts by dividing the weighted total number of deaths occurred to the cohort members in the follow-up period by the weighted total number of person-years lived by all of the cohort members (including those survived and died).

Activities of Daily Living (ADL) Disability

The ADL functional statuses of six daily activities of eating, dressing, transferring, using the toilet, bathing, and continence are used to measure the elders' status of independence in daily living. If the participant did not need or need assistance in perform each of the six daily activities, he or she was given a ADL disability score of 0 or 1. The lowest and highest ADL disability score is 0 and 6. ADL is a good measurement of functional capacity, a reasonable proxy of health status, and one of the key elements in efforts to measure quality of life (e.g., Katz et al. 1983; Spitzer 1987; Wiener et al. 1990; Gillen et al 1996; Muldoon et al. 1998; Christensen et al. 2013). ADL is closely related to care-giving needs (e.g., Branch et al 1988; Fredman et al. 1992;

Slivinske et al 1998) and it has important implications for public policy concerning the health unitization of older adults (Wolinsky et al. 1996). In this study, we follow the ADL capacity group classification widely adopted in the other studies (e.g. Christensen et al. 2013): if none or one of the six ADL activities is impaired, the oldest-old is classified as “normal”; if two activities are impaired, the oldest-old is classified as “moderately disabled”; “severely disabled” refers to those elders who have three or more activities impaired.

Cognitive function measured by Mini-Mental State Examination (MMSE)

The MMSE, a global assessment test of cognitive function (Folstein, Folstein and McHugh 1975; Christensen, et al 2013), was adapted to the Chinese cultural context and was carefully tested in the pilot survey (Zeng and Vaupel 2002). The testing protocol includes 24 items regarding orientation, registration, attention, calculation, recall and language, with a total score ranging from 0 to 30. Following the practice widely adopted in the other studies (e.g. Christensen et al. 2013), we use the MMSE cutoffs to define cognitive function as: severe impairment (0-17), mild impairment (18-22), normal (23-27) and maximum (28-30). Note that a zero score was given to those items to which the interviewee was not able to answer or perform the test, purely due to his or her mental or physical impairment (rather than not willing to answer or perform the test), and no proxy was allowed in performing the MMSE tests.

Physical performance tests

Self-reported subjective measures of disability in activities of daily living were criticized for their potential to be affected by differences in availability of associated facilities and perceptions of the participants; and the objective performance-based tests are highly recommended as complementary measures in examining the physical functions (Daltroy et al. 1995; Elam et al 1991; Melzer et al 2004; Reuben et al 2004). In the Chinese elderly population, the objective performance-based tests have been recently valued as important complementary measures for routinely-used ADL, which help clarify the intrinsic psychological impairment of the elderly and environmental barriers of their daily activities (Feng et al. 2010; Purser et al 2012).

Three objective physical performance tests were administrated in the CLHLS surveys. The first task asked the respondent to stand from a chair. This test have three levels of outcomes, i.e. “can without using arms” (coded as 1), “can using arms” (coded as 0.5), and “cannot” (coded as 0). The second task is to pick up a book from the floor, and respondents are “can while standing” (coded as 1), “can while sitting” (coded as 0.5), and “cannot” (coded as 0). The last one is test if the respondent could turn 360 degrees (yes vs. no, coded as 1 or 0).

Statistical analyses

Similar to what have been done in Christensen et al. (2013), we conducted the comparisons of the weighted means of the cohorts' physical and cognitive functions and other characteristics for men and women separately and for both sexes combined. We did the standard statistical χ^2 tests or t tests with the assumption of either equal or unequal variance, without controlling for the other covariates. We also conducted multivariate regression to explore the changes in physical and cognitive functions between the oldest-old cohorts born 10 years apart, adjusted for the covariates of age, rural/urban residence, marital status and education.

Results

Tables 2, 3 and 4 present the detailed results of cross-cohorts changes in physical and cognitive functions and follow-up death rates of oldest-old aged 80-105, through nine sets of comparisons of cohorts born 10 years apart. Table 5 presents the summary results of the cross-cohorts changes. In general, our interesting findings can be summarized into a few key points as follows.

(1) The age-specific follow-up death rates among Chinese oldest-old aged 80-105 were reduced in the later cohorts, compared to the earlier cohorts born 10 years earlier. All of the nine sets of the comparisons among different cohorts by sex showed follow-up mortality reduction in the range of annual decrease rate of -0.2% to -1.3%. Adjusted for the age, gender, education, rural/urban residence and marital status, the cross-cohort differences in mortality rates were statistically significant with $p<0.01$ in one comparison, significant with $p<0.05$ in one comparison, marginally significant with $p<0.1$ in four comparisons, and three other comparisons were not statistically significant (see Tables 2, 3, 4 and 5).

(2) The activities of daily living function disability of the Chinese oldest-old were significantly reduced in the later cohorts with different ages, compared to the earlier cohorts born 10 years earlier. All of the nine sets of the comparisons of the cohorts by sex showed substantial reduction in ADL disability in the range of annual decrease rate of -0.8% to -2.8%. Adjusted for the age, gender, education, rural/urban residence and marital status, the cross-cohort differences in ADL disability scores were statistically significant with $p<0.001$ in four comparisons, significant with $p<0.05$ in four comparison, and another comparison was not statistically significant (see Tables 2, 3, 4 and 5).

(3) The cognitive function measured by MMSE test score of the Chinese oldest-old were significantly worse in the later cohorts with different ages, compared to the earlier cohorts born 10 years earlier. All of the nine sets of the comparisons among different cohorts by sex showed significantly reduced MMSE score of cognitive function, in the range of annual declining rate of -0.7% to -2.2%. Adjusted for the age, gender, education, rural/urban residence and marital status, the cross-cohort differences in cognitive functional scores were statistically significant with $p<0.001$ in all of the nine comparisons (see Tables 2, 3, 4 and 5).

(4) The objective physical performance test scores of standing-up from a chair, picking-up a book from the floor and turning-around 360^0 of the Chinese oldest-old were all significantly worse in the later cohorts with different ages, compared to the earlier cohorts born 10 years

earlier. Adjusted for the age, gender, education, rural/urban residence and marital status, the cross-cohort differences in the objective physical performance test scores were statistically significant with p<0.001 in 25 comparisons, and significant with p<0.01 in the other two comparisons (see Tables 2, 3, 4 and 5).

--Tables 2, 3, 4 and 5 about here—

Discussions

Christensen et al. (2013) compared one pair of Danish cohorts born 10 years apart and aged 92-95 at the time of the surveys. Our present study, however, compares three pairs of Chinese cohorts born 10 years apart and aged 80-89, 90-99 and 100-105, respectively, at the time of the surveys. Our study in general reconfirmed some of the major findings of Christensen et al. (2013), while some interesting differences were also observed.

Both our Chinese study and the Danish study found that death rates at oldest-old ages among the later cohorts of the oldest-old was substantially reduced, compared to the cohorts born 10 years earlier; and the later cohorts of the oldest-old had significantly reduced impairment in activities of daily living than did the earlier cohort born 10 years earlier.

On the other hand, our study discovered that the objective physical performance test scores (stand-up from a chair, pick-up a book from the floor, and turning around 360^0) of the Chinese oldest-old were all significantly worse in the later cohorts with different ages, compared to the corresponding cohorts born 10 years earlier (p<0.001 in 25 comparisons and p<0.01 in the other two comparisons). As shown in Table 3 of Christensen et al. (2013), the Danish later cohort had substantially worse capacity of standing-up from a chair, compared to the earlier cohort: statistically highly significant for both sexes combined (p=0.002) and for females (p=0.015), and marginally significant for males (p=0.077). Compared to the earlier cohort, the Danish later cohort had significantly higher percent of not being able to walk among females (p=0.006) and for both sexes combined (p=0.026), while no significant difference for males.

Why did both our present study and the Danish study (Christensen et al. 2013) indicated the seemingly (at a first glance) contradictory findings: the likelihood of survival and self-reported ADL disability of both Chinese and Danish oldest-old were significantly improved in later cohorts than the earlier cohorts born 10 years ago, but the opposite was true for the objective physical performance tests? We believe that two underlining factors may help to understand this phenomenon. The first is the mixed effects of the two opposing processes of the success-of-success and failure-of-success. On one hand, a later cohort might benefit from progress resulted from more effective disease prevention and treatment, healthier lifestyles, and improved standards of living. Such progressive process known as the “success of success” helps the members of the later cohort to reduce mortality rates and reach older ages with improved health and functional capacity in daily living. On the other hand, as compared to the earlier cohort, the later one includes more surviving members at oldest-old ages who had been saved (or otherwise who already died if no substantial morality decline), but the additional survivors might be in relatively poor health. This process is known as the “failure of success” effect, namely, saving lives might reduce the overall physical functional capacity measured by score of objective physical performance tests (Christensen 2013: 5).

The second underline factor is associated with different types of disability measurements. The self-reported ADL disability depends not only on health status but also on facilities to assist the activities of daily living, such as transferring, using the toilet, and bathing. But the objective physical performance tests do not depend on facilities. The fact that the objective physical performance was reversed in comparison with trends of ADL among the Chinese cohorts of the oldest-old may suggest that the improvement of ADL could be partly due to the rapid changes in the living environment and facilities occurred in the past a couple of decades. For example, the proportion of disability in bathing, using the toilet and transferring, which also depend on facilities of bath rooms, toilets and moving chairs etc., declined by 26.1%, 21.3% and 19.1%, respectively, in the later Chinese cohorts compared to the earlier cohorts. In contrast, the proportion of disability in eating and dressing, which do not (or less) depend on the facilities, declined by 6.2% and 12.7%, respectively, in the later Chinese cohorts compared to the earlier cohorts. Thus, self-reported ADL disability scores may not be considered as an accurate indicator of health status, while it can be used as a good measurement of assistance needs in elders' daily living activities. The objective physical performance tests have "added predictive value" beyond the self-reported measures of disability, and they might be useful for evaluating the true health status changes and making health care and interventions decisions.

It is interesting to note that the weighted average MMSE score of cognitive function among the Chinese oldest-old were significantly worse in the later cohorts with different ages, compared to the earlier cohorts born 10 years earlier, but the Danish 1915 cohort scored significantly better on the MMSE than did the 1905 cohort. How to explain such disparity of changes in cognitive functions between the Chinese and Danish cohorts of the oldest-old? We believe that two main factors may shed light on understanding such disparity. First, the cross-cohorts educational difference and their childhood conditions may provide partial explanation. As shown in Table 1, for the two sexes combined, the average education levels of the three latter cohorts born in 1903-1908, 1909-1918 or 1919-1928 were significantly lower than that of the three corresponding cohorts born 10 years earlier, adjusted for age, gender, rural/urban residence and marital status, with a statistical level of $p = 0.002$; $p=0.026$; $p < 0.001$, respectively. The retrospectively self-reported weighted average proportion of frequently going to bed hungry as a child among the later Chinese cohorts were 30.5% higher than that in the earlier cohorts (data derived from the CLHLS but not shown in Table 1). Such cross-cohorts differentials in educational attainments and childhood conditions were due to more internal wars during the periods when the later Chinese cohorts were in childhood, compared to the earlier cohorts. This implied that the poorer childhood conditions experienced by the later Chinese cohorts might contribute to their lower cognitive function score, according to the existing literature (Zhang et al. 2008; Zeng et al. 2007; Shen and Zeng 2013). However, the Danish

1915 cohort's average education level was significantly better than the 1905 cohort (p=0.006), which implied that the Danish later cohort might enjoy better childhood conditions than the earlier cohort. This may explain the disparities of cohort changes in cognitive function among the oldest-old between the Chinese and Danish cohorts.

The second factor may be related to the fact that the relative speed of changes in improvements of living conditions and medical services in China who experienced exceptionally rapid socioeconomic development in the past a couple of decades was substantially faster than that in Denmark. Consequently, compared to the Danish case, relatively more frail Chinese oldest-old were "saved" from dying but their health including cognitive function were poor.

While we are satisfied with the interesting, unique and factual findings with some speculative explanations reported in this article, we must be aware of that our study has important limitations and more further research are needed. For example, additional in-depth studies are warranted to develop a deeper understanding of the mechanisms and causalities of how and why the mortality risk and ADL disability significantly declined due to "success of success", while the cognitive functional score and objective physical performance test scores were significantly reduced due the effects of "failure of success". Further studies need to extend the analysis to cover all elderly age groups (i.e., from age 65 to 110) to fully understand the process of healthy aging in a cohort life course perspective. We also hope that other studies on healthy longevity involving similar samples of different cohorts of the oldest-old aged 80+ born 10 or more or less years apart may replicate and validate our findings.

References

Banister J., David E. Bloom, and Larry Rosenberg (2010). Population Aging and Economic Growth in China. Harvard University Program on the Global Demography of Aging Working Paper No. 53, http://www.hsph.harvard.edu/pgda/working.htm; accessed July 20, 2013.

Daltroy LH, Philips CB, Eaton HM et al. Objective measuring physical ability in elderly persons: The physical capacity evaluation. American Journal of Public Health 1995; 85:558-560.

Elam JT, Craney MJ, Beaver T et al. Comparison of subjective ratings of function with observed functional ability of frail older persons. American Journal of Public Health 1991; 81:1127-1130.

Feng, Qiushi, Helen M Hoenig, Danan Gu, Yi Zeng, and Jama L Purser. 2010. "Impact of New Disability Subtypes on Three-year Mortality in Chinese Older Adults." Journal of the American Geriatrics Society 58(10):1952-1958.

Melzer D., Lan TY, Tom BDM. Variation in thresholds for reporting mobility disability between national population subgroups and studies. J Gerontol A Biol Sci Med Sci 2004; 59A(12): 1295-1303.

Purser, Jama L, Qiushi Feng (corresponding author), Zeng Yi, and Helen M Hoenig. 2012. "A New Classification of Function and Disability in China: Subtypes Based on Performance-based and Self-reported Measures." Journal of Aging and Health 24(5):779-798.

Reuben DB, Seeman TE, Keeler E et al. Refining the categorization of physical functional status: The added value of combining self-reported and performance-based measures. Journal of Gerontology: Biological Science and Medical Science. 2004; 59(10):1056-1061.

U.N. (United Nations) (2011), World Population Prospects: The 2010 Revision. New York: United Nations.

Zeng, Y. and L. K, George (2010), Population ageing and old-age insurance in china. In D. Dannefer & C. Phillipson (Eds.), The SAGE handbook of social gerontology (pp.420-430).London: SAGE Publications Ltd.

Table 1. Demographic characteristics of cohorts born in 1909-1918 vs 1919-1928, born in 1899-1908 vs 1909-1918, and born in 1893-1898 vs 1903-1908

	Years of birth		p-value		Years of birth		p-value		Years of birth		p-value	
	1909-1918 (n=3235)	1919-1928 (n=4053)	Unadjusted	Adjusted¶	1899-1908 (n=2896)	1909-1918 (n=4338)	Unadjusted	Adjusted¶	1893-1898 (n=2197)	1903-1908 (n=2809)	Unadjusted	Adjusted¶
Age in years, mean(sd)	83.07 (2.59)	82.98 (2.57)	0.1516*		92.11 (2.13)	92.24 (2.19)	0.0109*		101.15(1.34)	101.72(1.55)	<0.0001‡	
Women, n(%)	1995 (61.7%)	2362 (58.3%)	0.0033£	0.024	2102 (72.6%)	3144(72.5%)	0.9246£	0.545	1652(75.2%)	2254(80.2%)	<0.0001£	0.004
Rural residence (urban)												
Both sex, n (%)	2135(66.0%)	2186(53.9%)	<0.0001£	<0.001	1770(61.1%)	2314(53.3%)	<0.0001£	<0.001	1342(61.1%)	1466(52.2%)	<0.0001£	<0.001
Women, n (%)	1058(66.4%)	1108(54.8%)	<0.0001£	<0.001	1007(60.9%)	1377(54.5%)	<0.0001£	0.001	1019(57.9%)	1237(56.0%)	0.2189£	0.121
Men, n (%)	1072(65.3%)	1072(52.8%)	<0.0001£	<0.001	767(61.7%)	911 (50.3%)	<0.0001£	<0.001	310(70.6%)	220 (36.7%)	<0.0001£	<0.001
Married												
Both sex, n (%)	902(27.9%)	1424 (35.1%)	<0.0001£	<0.001	280 (9.7%)	497 (11.5%)	0.0156£	0.025	74 (3.4%)	90 (3.2%)	0.7840£	0.621
Women, n (%)	225 (14.1%)	451 (22.3%)	<0.0001£	<0.001	47 (2.8%)	139 (5.5%)	<0.0001£	<0.001	5 (0.3%)	25 (1.2%)	0.0028£	0.001
Men, n (%)	822 (50.3%)	1077 (53.1%)	0.0920£	0.262	345 (27.7%)	490 (27.1%)	0.6884£	0.537	55 (12.6%)	70 (11.6%)	0.6371£	0.187
Education												
Both sex												
Not educated, n (%) §	2006 (62.2%)	2525 (62.3%)			2129 (73.8%)	3233(74.8%)			1763(81.1%)	2389(85.4%)		
Primary school, n (%) §	900 (27.9%)	1160 (28.6%)			593 (20.6%)	856(19.8%)			335 (15.4%)	319(11.4%)		
Above primary school, n (%) §	320 (9.9%)	362 (8.9%)	0.4225†	<0.001	164 (5.7%)	231 (5.4%)	0.6755†	0.026	76 (3.5%)	90 (3.2%)	0.0007†	0.002
Missing, n (%)	11 (0.3%)	6 (0.2%)	0.3903£		17 (0.6%)	16 (0.4%)	0.2279£		25 (1.2%)	11 (0.4%)	0.0011£	
Women												
Not educated, n (%) §	1286 (81.0%)	1269 (80.7%)			1439 (87.4%)	2180(86.6%)			1594(91.8%)	2045(92.8%)		
Primary school, n (%) §	227 (14.3%)	315 (15.6%)			174 (10.6%)	283 (11.2%)			118 (6.8%)	127 (5.8%)		
Above primary school, n (%) §	75 (4.8%)	76 (3.8%)	0.3301†	0.085	34 (2.0%)	55 (2.2%)	0.8279†	0.701	25 (1.4%)	32 (1.4%)	0.5171†	0.348
Missing, n (%)	7 (0.5%)	4 (0.2%)	0.1195£		12 (0.7%)	10 (0.4%)	0.1867£		24 (1.4%)	6 (0.3%)	0.0001£	
Men												
Not educated, n (%) §	525 (32.0%)	748 (36.9%)			469 (37.8%)	792 (43.9%)			214 (49.1%)	328 (55.2%)		
Primary school, n (%) §	815 (49.8%)	951 (47.0%)			582 (46.9%)	764 (42.4%)			180 (41.2%)	204 (34.4%)		
Above primary school, n (%) §	198 (18.2%)	327 (16.1%)	0.0228†	<.001	189 (15.3%)	246 (13.7%)	0.0166†	<.001	43 (9.7%)	62 (10.5%)	0.1029†	0.013
Missing, n (%)	2 (0.1%)	2 (0.1%)	1.0000£		4 (0.3%)	6 (0.3%)	1.0000£		2 (0.5%)	5 (0.9%)	0.4547£	

Results are all weighted outcomes.
*Test of equal mean, with an assumption of equal variance (test of equal variance is not rejected)
‡Test of equal mean, without an assumption of equal variance (test of equal variance is rejected)
†Test of equal proportions (χ^2 test)
£ Test of equal proportions (Z test)
§ Missing data are excluded from the totals for percentage calculations
¶ Test of logistic or ordered logistic model adjusted by gender, age, education, rural/urban residence, and marital status

Table 2. Differentials in mortality and of cognitive and physical functions between cohorts born in 1909-1918 (aged 80-89 with mean age 83.1 in 1998 survey, n=3235) and born in 1919-1928 (aged 80-89 with mean age 83.0 in 2008 survey, n=4053)

	Two sexes combined				**Men**				**Women**			
	Years of birth		p-value		Years of birth		p-value		Years of birth		p-value	
	1909-1918	1919-1928	Unadjusted	Adjusted¶	1909-1918	1919-1928	Unadjusted	Adjusted¶	1909-1918	1919-1928	Unadjusted	Adjusted¶
Annual death rate	10.3%	9.6%	0.3550£	0.060	12.5%	10.9%	0.1646£	0.067	9.0%	8.7%	0.7698£	0.264
MMSE score score												
Mean (range=0-30)	24.82 (5.37)	22.87 (7.27)	<0.0001‡		25.86 (5.09)	24.18 (6.32)	<0.0001‡		24.17 (5.44)	21.93 (7.56)	<0.0001‡	
Missing, n (%)	26 (0.8%)	9 (0.2%)	0.0004£		13 (0.8%)	4 (0.2%)	0.0081£		12 (0.8%)	5 (0.3%)	0.0380£	
Grouped results, n (%) §			<0.0001†	<0.001			<0.0001†	<0.001			<0.0001†	<0.001
0-17	285 (8.9%)	658 (16.3%)			97 (6.0%)	206 (10.2%)			169 (10.7%)	416 (20.7%)		
18-22	437 (13.6%)	744 (18.4%)			148 (9.1%)	308 (15.2%)			260 (16.4%)	418 (20.7%)		
23-27	1328(41.4%)	1439 (35.6%)			642 (39.5%)	791 (39.0%)			674 (42.6%)	668 (33.2%)		
28-30	1157(36.1%)	1200 (29.7%)			737 (45.4%)	722 (35.6%)			479 (30.3%)	513 (25.5%)		
ADL disability score												
Mean (range=0-6)	0.36 (1.06)	0.28 (1.01)	0.0006‡		0.32 (1.01)	0.24 (0.93)	0.0219‡		0.39 (1.08)	0.30 (1.06)	0.0163*	
Missing, n (%)	18 (0.6%)	1 (0.0%)	<0.0001£		10 (0.6%)	1 (0.0%)	0.0005£		8 (0.5%)	0 (0.0%)	0.0000£	
Grouped results, n (%) §			0.0015†	0.005			0.2077†	0.071			0.0055†	0.029
0-1	2974(92.4%)	3823 (94.4%)			1525(93.5%)	1927(95.0%)			1456(91.8%)	1899(93.9%)		
2	80 (2.5%)	51 (1.3%)			33 (2.0%)	28 (1.4%)			44 (2.8%)	23 (1.2%)		
≥3	164 (5.1%)	178 (4.4%)			73 (4.5%)	73 (3.6%)			86 (5.5%)	100 (5.0%)		
Physical performance score												
Stand-up from chair												
Mean (range= 0-1)	0.92 (0.21)	0.86 (0.28)	<0.0001‡		0.93 (0.20)	0.87 (0.26)	<0.0001‡		0.92 (0.22)	0.84 (0.29)	<0.0001‡	
Missing, n (%)	15 (0.5%)	1 (0.0%)	<0.0001£		9 (0.6%)	0 (0.0%)	0.0005£		6 (0.4%)	1 (0.1%)	0.0628£	
Pick-up a book from floor												
Mean (range= 0-1)	0.90 (0.25)	0.85 (0.29)	<0.0001‡		0.91 (0.23)	0.87 (0.27)	<0.0001‡		0.89 (0.26)	0.84 (0.29)	<0.0001‡	
Missing, n (%)	18 (0.6%)	0 (0.0%)	<0.0001£		14 (0.8%)	0 (0.0%)	0.0001£		6 (0.4%)	0 (0.0%)	0.0044£	
Turn-around 360^0												
Mean (range= 0-1)	0.91 (0.28)	0.81 (0.39)	<0.0001‡		0.92 (0.27)	0.84 (0.37)	<0.0001‡		0.91 (0.29)	0.79 (0.41)	<0.0001‡	
Missing, n (%)	11 (0.3%)	1 (0.0%)	0.0005£		10 (0.6%)	0 (0.0%)	0.0005£		2 (0.1%)	1 (0.0%)	0.1549£	

Results are all weighted outcomes.
*Test of equal mean, with an assumption of equal variance (test of equal variance is not rejected)
‡Test of equal mean, without an assumption of equal variance (test of equal variance is rejected)
†Test of equal proportions (χ^2 test)
£ Test of equal proportions (Z test)
§ Missing data are excluded from the totals for percentage calculations
¶ Multivariate model test of the difference between the cohorts, adjusted for by age, education, rural/urban residence, and marital status; for the two-sex combined, the gender was also adjusted for. The tests for annual death rates are based on parametric survival model with Weibull distribution; all other tests are based logistic regression models.

Table 3. Differentials in mortality and cognitive and physical functions between cohorts born in 1899-1908 (aged 90-99 with mean age 92.1 in 1998 survey, n=2896) and born in 1909-1918 (aged 90-99 with mean age 92.2 in 2008 survey, n=4338)

	Two sexes combined				**Men**				**Women**			
	Years of birth		p-value		Years of birth		p-value		Years of birth		p-value	
	1899-1908	1909-1918	Unadjusted	Adjusted¶	1899-1908	1909-1918	Unadjusted	Adjusted¶	1899-1908	1909-1918	Unadjusted	Adjusted¶
Annual death rate	24.1%	23.4%	0.5226£	0.065	27.1%	25.6%	0.3867£	0.118	23.0%	22.6%	0.7793£	0.186
MMSE score												
Mean (range=0-30)	20.62 (7.93)	17.41 (9.62)	<0.0001‡		22.95 (7.18)	19.81 (9.26)	<0.0001‡		19.73 (8.02)	16.50 (9.59)	<0.0001‡	
Missing, n (%)	39 (1.3%)	20 (0.5%)	0.0002£		16 (1.3%)	4 (0.2%)	0.0002£		23 (1.4%)	14 (0.6%)	0.0079£	
Grouped results, n (%)			<0.0001†	<0.001			<0.0001†	<0.001			<0.0001	<0.001
0-17	789 (27.6%)	1778 (41.2%)			206 (16.7%)	538 (29.8%)			516 (31.7%)	1145(45.5%)		
18-22	577 (20.2%)	936 (21.7%)			173 (14.1%)	366 (20.3%)			367 (22.6%)	559 (22.2%)		
23-27	952 (33.4%)	1004 (23.2%)			519 (42.2%)	530 (29.4%)			487 (30.0%)	526 (20.9%)		
28-30	537 (18.8%)	603 (14.0%)			332 (27.0%)	371 (20.6%)			255 (15.7%)	288 (11.4%)		
ADL disability score												
Mean (range=0-6)	0.94 (1.62)	0.74 (1.55)	<0.0001‡		0.74 (1.49)	0.59 (1.42)	0.0219‡		1.02 (1.66)	0.80 (1.59)	0.0163*	
Missing, n (%)	10 (0.3%)	0 (0.0%)	0.0003£		5 (0.4%)	0 (0.0%)	<0.0001£		5 (0.3%)	0 (0.0%)	0.0059£	
Grouped results, n (%)			0.0023†	<0.001			0.0238†	0.008			0.0236	0.002
0-1	2290 (79.4%)	3618 (83.4%)			1031 (83.3%)	1578(87.2%)			1283(77.9%)	2073(82.0%)		
2	152 (5.3%)	181 (4.2%)			59 (4.8%)	57 (3.2%)			90 (5.5%)	115 (4.6%)		
≥3	443 (15.4%)	538 (12.4%)			148 (11.9%)	175 (9.7%)			275 (16.7%)	340 (13.5%)		
Physical performance score												
Stand-up from chair												
Mean (range= 0-1)	0.80 (0.31)	0.72 (0.34)	<0.0001‡		0.84 (0.28)	0.77 (0.32)	<0.0001‡		0.78 (0.32)	0.71 (0.34)	<0.0001‡	
Missing, n (%)	24 (0.8%)	10 (0.2%)	0.0002£		10 (0.8%)	7 (0.4%)	0.1467£		14 (0.8%)	4 (0.2%)	0.0040£	
Pick-up a book from floor												
Mean (range= 0-1)	0.77 (0.33)	0.67 (0.37)	<0.0001‡		0.83 (0.30)	0.74 (0.35)	<0.0001‡		0.75 (0.34)	0.65 (0.38)	<0.0001‡	
Missing, n (%)	28 (1.0%)	3 (0.1%)	<0.0001£		13 (1.1%)	2 (0.1%)	0.0001£		16 (1.0%)	1 (0.1%)	<0.0001£	
Turn-around 360^0												
Mean (range= 0-1)	0.78 (0.41)	0.59 (0.49)	<0.0001‡		0.83 (0.38)	0.65 (0.48)	<0.0001‡		0.76 (0.43)	0.57 (0.50)	<0.0001‡	
Missing, n (%)	21 (0.7%)	2 (0.1%)	<0.0001£		7 (0.6%)	1 (0.0%)	0.0010£		13 (0.8%)	2 (0.1%)	0.0003£	

Results are all weighted outcomes.
*Test of equal mean, with an assumption of equal variance (test of equal variance is not rejected)
‡Test of equal mean, without an assumption of equal variance (test of equal variance is rejected)
†Test of equal proportions (χ^2 test)
£ Test of equal proportions (Z test)
§ Missing data are excluded from the totals for percentage calculations
¶ Multivariate model test of the difference between the cohorts, adjusted for by age, education, rural/urban residence, and marital status; for the two-sex combined, the gender was also adjusted for. The tests for annual death rates are based on parametric survival model with Weibull distribution; all other tests are based logistic regression models.

Table 4. Differentials in mortality and cognitive and physical functions between cohorts born in 1893-1898 (aged 100-105 with mean age 101.1 in 1998 survey, n=2197) and born in 1903-1908 (aged 100-105 with mean age 101.7 in 2008 survey, n=2809)

	Two sexes combined				**Men**				**Women**			
	Years of birth		p-value		Years of birth		p-value		Years of birth		p-value	
	1893-1898	1903-1908	Unadjusted	Adjusted¶	1893-1898	1903-1908	Unadjusted	Adjusted¶	1893-1898	1903-1908	Unadjusted	Adjusted¶
Annual death rate	40.7%	38.0%	0.0723£	0.003	45.7%	41.2%	0.1833£	0.056	39.1%	37.4%	0.3100£	0.016
MMSE score												
Mean (range=0-30)	14.63 (9.44)	11.63 (10.12)	<0.0001‡		17.92 (9.19)	14.95(10.45)	<0.0001‡		13.54 (9.27)	10.82 (9.87)	<0.0001‡	
Missing, n (%)	44 (2.0%)	55 (2.0%)	1.0000£		9 (2.0%)	18 (3.1%)	0.2742£		35 (2.0%)	37 (1.7%)	0.4932£	
Grouped results, n (%)			<0.0001†	<0.001			0.0022†	<0.001			<0.0001†	<0.001
0-17	1192 (55.3%)	1837 (66.5%)			172 (40.0%)	300 (51.2%)			1040(60.3%)	1528(70.2%)		
18-22	420 (19.5%)	378 (13.7%)			89 (20.6%)	87 (14.9%)			330 (19.1%)	291 (13.4%)		
23-27	386 (17.9%)	380 (13.8%)			107 (24.8%)	137 (23.5%)			270 (15.6%)	248 (11.4%)		
28-30	158 (7.3%)	168 (6.1%)			63 (14.6%)	60 (10.3%)			85 (4.9%)	109 (5.0%)		
ADL disability score												
Mean (range=0-6)	2.01 (2.09)	1.58 (2.00)	<0.0001‡		1.57 (1.91)	1.45 (1.97)	0.3043*		2.15 (2.12)	1.61 (2.00)	<0.0001‡	
Missing, n (%)	10 (0.5%)	0 (0.0%)	0.0002£		1 (0.3%)	0 (0.0%)	0.1749£		9 (0.5%)	0 (0.0%)	0.0014£	
Grouped results, n (%)			<0.0001†	<0.001			0.0204†	0.002			<0.0001†	<0.001
0-1	1186 (54.2%)	1820 (64.8%)			274 (62.6%)	424 (70.7%)			899 (51.4%)	1399(63.3%)		
2	219 (10.0%)	229 (8.2%)			43 (9.7%)	38 (6.4%)			176 (10.1%)	190 (8.6%)		
≥3	784 (35.8%)	759 (27.0%)			121 (27.6%)	137 (22.9%)			674 (38.5%)	620 (28.1%)		
Physical performance score												
Stand-up from chair												
Mean (range= 0-1)	0.62 (0.37)	0.57 (0.37)	<0.0001*		0.70 (0.36)	0.63 (0.37)	0.0017*		0.59 (0.37)	0.56 (0.37)	0.0021*	
Missing, n (%)	36 (1.7%)	12 (0.4%)	<0.0001£		8 (1.9%)	0 (0.0%)	0.0007£		27 (1.6%)	12 (0.5%)	0.0007£	
Pick-up a book from floor												
Mean (range= 0-1)	0.56 (0.39)	0.49 (0.40)	<0.0001*		0.66 (0.38)	0.57 (0.41)	0.0002‡		0.52 (0.39)	0.47 (0.39)	<0.0001*	
Missing, n (%)	51 (2.3%)	7 (0.3%)	<0.0001£		12 (2.7%)	0 (0.0%)	0.0001£		39 (2.2%)	7 (0.3%)	<0.0001£	
Turn-around 360^{0}												
Mean (range= 0-1)	0.52 (0.50)	0.37 (0.48)	<0.0001*		0.67 (0.47)	0.45 (0.50)	<0.0001*		0.47 (0.50)	0.35 (0.48)	<0.0001*	
Missing, n (%)	12 (0.6%)	2 (0.1%)	0.0019£		4 (1.0%)	0 (0.0%)	0.0141£		8 (0.4%)	1 (0.1%)	0.0595£	

Results are all weighted outcomes.
*Test of equal mean, with an assumption of equal variance (test of equal variance is not rejected)
‡Test of equal mean, without an assumption of equal variance (test of equal variance is rejected)
†Test of equal proportions (χ^2 test)
£ Test of equal proportions (Z test)
§ Missing data are excluded from the totals for percentage calculations
¶ Multivariate model test of the difference between the cohorts, adjusted for by age, education, rural/urban residence, and marital status; for the two-sex combined, the gender was also adjusted for. The tests for annual death rates are based on parametric survival model with Weibull distribution; all other tests are based logistic regression models.

Table 5. Annual rates of changes in mortality rates, physical and cognitive functions between the oldest-old cohorts born 10 years apart, who were interviewed with same age in CLHLS 1998 baseline and 2008 wave

	2 cohorts born 10 years apart and aged 80-89 at interviews conducted 10 years apart			2 cohorts born 10 years apart and aged 90-99 at interviews conducted 10 years apart			2 cohorts born 10 years apart and aged 100-105 at interviews conducted 10 years apart		
	2-sexes	men	women	2-sexes	men	women	2-sexes	men	women
Evidences may support "success of success"									
Follow-up death rate	-0.7%#	-1.3%#	-0.3%	-0.3%#	-0.6%	-0.2%	-0.7%**	-1.0%#	-0.4%*
Mean ADL disability score	-2.4%***	-2.8%*	-2.5%*	-2.3%***	-2.2%*	-2.3%*	-2.3%***	-0.8%	-2.8%***
Evidences may support "failure of success"									
Cognition (MMSE) score	-0.8%***	-0.7%***	-0.9%***	-1.6%***	-1.4%***	-1.7%***	-2.2%***	-1.7%***	-2.2%***
Physical performance score									
Stand-up from chair	-0.7%***	-0.6%***	-0.9%***	-1.0%***	-0.8%***	-0.9%***	-0.8%***	-1.0%**	-0.5%**
Pick-up book from floor	-0.6%***	-0.4%***	-0.6%***	-1.3%***	-1.1%***	-1.4%***	-1.3%***	-1.4%***	-1.0%***
Turn-around 360^0	-1.1%***	-0.9%***	-1.4%***	-2.7%***	-2.4%***	-2.8%***	-3.3%***	-3.8%***	-2.8%***

Note: the stars and # indicate statistical significance levels of the difference between the mean scores or death rate of the two cohorts born 10 years apart, who were interviewed with same age in CLHLS 1998 baseline and 2008 wave: # p<0.1; * p<0.05; ** p<0.01; *** p<0.001.