Data Quality of Oldest-Old Population in Taiwan: 2003 Census for Ages 89 and Above

Jack C. Yue
Department of Statistics
National Chenghi University
Taipei, Taiwan, R.O.C. 11632

Presented at The Living to 100 and Beyond Symposium
Sponsored by the Society of Actuaries

Orland, Fla.

January 12-14, 2005

Copyright 2005 by the Society of Actuaries.
All rights reserved by the Society of Actuaries. Permission is granted to make brief excerpts for a published review. Permission is also granted to make limited numbers of copies of items in this monograph for personal, internal, classroom or other instructional use, on condition that the foregoing copyright notice is used so as to give reasonable notice of the Society's copyright. This consent for free limited copying without prior consent of the Society does not extend to making copies for general distribution, for advertising or promotional purposes, for inclusion in new collective works or for resale.

Summary

The mortality data of the elderly, usually in insufficient sizes and with questionable data quality, often are not easy to verify if they satisfy certain mortality assumptions (such as the Gompertz law). In particular, it is difficult to deal with the data quality of the elderly. In fact, although the census data are one of the major sources for constructing life tables, the mortality rates of the elderly are seldom calculated solely from the census data. One of the reasons for not using the census data alone is the low response rate of the elderly. For example, in the 2000 population census (mail-back questionnaire) of the Taiwan area, more than 93 percent of the elderly between ages 90 and 99 responded, but about 80 percent responded for ages 100 and over.

In order to increase the data quality of the elderly and the accuracy of their mortality rates, the Taiwan government conducted a census (face-to-face interviews) for people ages 89 and over, a population around 50,000 persons. Due to the large size of population for face-to-face interviews, the census period is six months, between August 2003 and February 2004. The new census data will be used to modify the census records and the 1999-2001 complete life tables in the Taiwan area. In this paper, we will introduce the details of the census for the elderly and discuss the difficulty in collecting the data. We will also compare the differences of the elderly data between these two census (2000 and 2003) records and check if the mortality rates of the elderly in Taiwan satisfy frequently used mortality assumptions.

1. Introduction

The proportion of the elderly in Taiwan area has increased dramatically in recent years. The proportion of people ages 65 and over first reached the 7 percent mark in 1993 and has continued to grow since then with an increment of over 0.2 percent annually. It surpassed 9 percent in 2003, and the time needed to double the percentage from 10 percent to 20 percent is expected to be 20 years, even faster than the 22 years in Japan (Council for Economic Planning and Development, Executive Yuan, R.O.C.). The population aging is one of the major concerns of Taiwan government policy-making.

The prolonging life expectancy and declining mortality rates in the Taiwan area is one of two major causes in population aging (the other is declining fertility rates). For example, the life expectancy of the male in the Taiwan area was 53.38 in 1951 and then increased to 74.00 in 2003, which is equivalent to 0.4 year annually. Although the life expectancy has the largest increase between 1950 and 1970, the increment is still larger than 0.2 year annually between 1991 and 2001. (This number is calculated via the 1989-1991 and the 1999-2001 complete life tables in the Taiwan area.) The decline of mortality rates of the elderly is the major reason of prolonging life expectancy in recent years, and the population of ages 65 and beyond has contributed more than 60 percent of the increase in life expectancy since 1980. (See Yue, 2001, for details.) As a result, the study of the elderly, including their mortality rates and living arrangements, becomes one of the popular research topics since the turn of the 21st century.

However, like many countries in the world, the study of the elderly faces two major difficulties: the size of the elderly population and the data quality. The population size of Taiwan (both sexes) is around 23 million in 2004, but the size of the population ages 85 and over (i.e., the oldest-old) is much smaller, and is even smaller in the past. For example, the male and female populations for ages 85 and over in 1980 were 7,716 and 15,932, and even in 2000 were 46,746 and 59,362. On average, there are fewer than 3,000 persons for each age. In other words, the fluctuations of estimated mortality rates are large, and it is difficult to proceed with the study of elderly mortality rates based on this size of data. There are several methods to deal with the problem of insufficient data, such as accumulating several years (such as 10 years) of data (Kannisto, 1996). However, combining years of data might pool data of different attributes and shall be handled with care (Yue, 2002).

The data quality is a bigger concern for the countries with larger populations or the countries with rapid population-aging (such as Taiwan). For example, in the 2000 population census (mail-back questionnaire) of the Taiwan area, more than 93 percent of the elderly between ages 90 and 99 responded, but about 80 percent responded for ages 100 and over. As a result, the data quality of the elderly mortality becomes less reliable as the age increases. In order to check if the low response rate of the elderly in the 2000 population census has a significant influence on the estimation of the elderly mortality, the Ministry of the Interior in Taiwan conducted a census for people ages 89 and over in 2003. In this study, we will introduce the details of this elderly census (which will be called the "2003 elderly census" for the rest of this study) and calculate mortality rates based on the 2003 elderly census data. The background of the 2003 elderly census is given in section 2, followed by the details of conducting the census in section 3. The mortality rates based on the 2003 elderly census, with the comparison to those from the 2000 census, is in section 4.

2. Population Data in Taiwan

According to Brown (1997), the majority of demographic data come from censuses, survey and vital statistics registration systems. In Taiwan, a population registration system that continuously traces records of each individual is also used, in addition to the regular vital statistics used in many countries. This population registration system, also known as household registration system, is the main source of the official statistics in the Taiwan area. It is mandatory for the people in Taiwan to report all the changes in a household, including migrations, births and deaths. In addition to the mandatory reports, the records in household registration are also updated by the population censuses (at the national level, every 10 years) and the household registration corrections (usually at the county level, usually every five to 10 years).

In the past, household registration, censuses and registration corrections together constantly provided fairly reliable data quality in the Taiwan area. However, like the low response rates of population censuses in the United States, the population data in Taiwan from all three data sources show similar problems since 1990. The data quality is especially questionable at the group of the elderly due to the following reasons.

First, although it is mandatory that the residents in Taiwan report household changes, people usually don't file for changes unless the changes are beneficiary to the household. For example, the Taiwan government is running a national health-care system, and children under three years old don't need to pay for health care. Therefore, unless the infants are born dead, almost all births are registered. Keeping the elderly data updated has become even more difficult since the people ages 65 and over are entitled to receive a monthly living expense (citizen's annuity) starting in 2001. There are several cases in which the records of the elderly show a person alive, after being dead for years.

There are problems in the population censuses and registration corrections as well. The population censuses usually evaluate the usual residents in a household, instead of the registered population in the household registration system. The results from the population censuses often are different from those from the household registration. In the past, because the population censuses are most updated, the results from the censuses were used to modify the records in household registration system and to construct complete life tables. Unfortunately, the budget of the 2000 population census was largely reduced, due to the 921 earthquake (Richter scale 7.3) that killed more than 2,000 persons in 1999. As a result, the post-enumeration of the 2000 population census was cancelled, and the census results were never verified. Many researchers believe that the non-response rates of the elderly are larger than those from the official released. (For example, as mentioned earlier, 6.63 percent of the elderly between ages 90 and 99 did not respond, and 19.26 percent did not respond for ages 100 and over.)

Similar problems occur in registration corrections. The corrections are supposed to be face-to-face interviews, and usually the head of the household fills in the questionnaires for the elderly (due to their high illiteracy and health problems). Since the people in Taiwan usually trust each other and do not question the accuracy of the data, the interviewers seldom check the information filled by the household head, unless there are obvious mistakes. Also, similar to the population censuses, the registration corrections have been experiencing low response rates for the recent years.

Because these three factors combined cause researchers and policymakers to be skeptical of the quality of the elderly data, the Ministry of the Interior issued a census for the people ages 89 and over in 2003, in order to update the records in the household

registration system. In the following section, we shall introduce the details of this census, including the census plan and the difficulty in conducting the 2003 elderly census.

3. The 2003 Elderly Census

The Ministry of the Interior is in charge of the household registration, the registration corrections and the 1990 population census, but not the 2003 population census (which was conducted by the Directorate General of Budget Accounting and Statistics). Because the data sources are not under the same department, it is likely that there exist data discrepancies. Therefore, the primary goal of the 2003 elderly census is to collect the information of the elderly who are missing or dead, in order to correct the records in the household registration system. Two other goals of the 2003 elderly census are to study the current lifestyles and the needs of the elderly. The questions related to the lifestyles in the questionnaire of the elderly census include the health, daily living and choices of the elder care. Most of the questions related to needs are related to social welfare. Since the interest of this paper is on the elderly mortality, we will focus on updating the records of the elderly for the rest of this paper.

The 2003 elderly census was conducted from August 11, 2003, to February 29, 2004. The standard date of the census is June 30, 2003. The population of the people ages 89 and over in Taiwan was 50,169 in 2003, according to the records of the household registration. The 2003 elderly census was conducted according to the 50,169-person list from the household registration, and all people on the list were interviewed, no matter if they were in nursing homes or institutes. Comparing to the 2000 population census, tracing about 22 million persons in about two weeks, there was more time and fewer people to trace in the 2003 elderly census. There were 236 interviewers involved in the 2003 elderly census, while there were about 90,000 interviewers in the 2000 population census.

There are three main differences in conducting the 2003 elderly census and the 2000 population census. First, the goal of the 2003 elderly census is to locate all the people ages 89 and over on the list of the household registration, while the goal of the 2000 population census is to count the usual residents in a household. Second, both censuses are face-to-face interview, but the 2003 elderly census particularly emphasizes

that the interviewers need to see the elderly in person. There are several ways for identifying the elderly in a face-to-face interviewer. The methods include identification cards, health-care cards, household registration records and personal proof from, for example, relatives, household registration officials and health-care providers. The third difference is the data verification after the enumeration. There was no post-enumeration survey in the 2000 population census, but there was a two-stage verification in the 2003 elderly census. At each stage, 10 percent of the questionnaires were randomly selected and were verified by telephone or a face-to-face interview.

TABLE 1
Response and Non-Response Rates in the 2003 Elderly Census

	Age 89	**Ages 90~94**	**Ages 95~99**	**Ages 100+**	**Total**
Total Population	12,597	23,898	11,190	2,484	50,169
Response	92.17%	91.39%	88.38%	66.39%	89.67%
Non-Response	7.83%	8.61%	11.62%	33.61%	10.33%

Note that the address of each elderly is the only information on which the interviewers can rely. In addition, all available government information is used to acquire the most updated address and to reduce the non-response rates. Other than the records in the household registration, the information used includes the national health care, the worker pension and the citizen annuity. Even with all the efforts, there are still 5,169 persons (about 10.33 percent) who could not be traced (details in Table 1). The non-response rate increases as the age increases. Note that the non-response rates in the 2003 elderly census are even larger than those in the 2000 population census. For example, the non-response rate for ages 100 and beyond is 19.26 percent in 2000 and almost doubles (33.61 percent) in 2003. With six months of census time and a great deal of effort for the 2003 elderly census, it is very difficult to believe that the non-response is even higher. This is one of the reasons why the researchers in Taiwan question the data quality in the 2000 population census.

Among the non-response people, the majority (4,302 persons, accounting for 83.23 percent of the non-response) are "no such person for the recorded address." The second-largest group of the non-response is "unknown address" (656 persons, accounting for 12.69 percent). This indicates that the factor of address accounts for the majority of the non-response (95.92 percent), which means that there are problems in

the current household registration system. The household registration system needs some work in order to provide good quality of data.

Because the non-response rates in the 2003 elderly census are higher than those in the 2000 population census, this provides us stronger motivation for checking the mortality rates in the 1999-2001 Taiwan Complete Life Table. However, the overall 10.33 percent of non-response rate is not ignorable. Thus, in order to simplify the discussion, we assume that the mortality rates of the non-response are similar to those who responded. The estimates of the mortality rates will be given in the following section.

4. Mortality Rates of the Elderly

In this section, the results from the 2003 elderly census will be used to modify the 1999-2001 Taiwan Complete Life Table and to check the mortality assumptions, such as the Gompertz law. We shall discuss modifying the life table first.

Note that the standard date of the 2003 elderly census is June 30, and the census time is between August 11, 2003, and February 29, 2004. There are 2,290 persons who died in this census period, 953 male and 1,337 female. The exposures of the elderly will be based on those who were alive between the standard date and the time of census. On average, the time between the standard date and the time of census is approximately 4.5 months.

Because the time between the standard date and the time of census is short, the exposure of each age group will be small as well. Smaller exposures indicate larger variances, and the mortality estimates from the elderly census is likely to be highly fluctuated. Therefore, the age groups with smaller exposures cannot be used to estimate the mortality rates. Because the exposures of both sexes at each age are fewer than 50 (unit: person-year) for people ages 100 and over, we will focus on the mortality rates for the people between ages 89 and 99.

To get some brief idea of modifying the life table, we use the mortality rate for ages 90 to 99 as a demonstration to see if the results from the 2003 elderly census match those from the 2000 population census (or, the 1999-2001 Taiwan Complete Life Table). From Table 2, the mortality rates estimated from the two sources are very close to each other, with mortality rates from the elderly census being smaller. Thus, it is natural to

expect that the mortality rates estimated from the 2003 elderly census are close to those from the 2000 population census.

TABLE 2
The Estimated Mortality Rates of Ages 90~99

	1999-2001 Complete Life Table	**2003 Elderly Census**
Male	18.24%	18.15%
Female	19.18%	18.02%

Figure 1. Comparisons of Estimated Mortality Rates

Comparisons of the estimated mortality rates from the two censuses can be seen in Figure 1 as well. As expected, the estimated mortality rates have larger variances for the 2003 elderly census, due to smaller sample sizes. The female mortality rates are obviously smaller from the elderly census, while the male mortality rates have little differences. Therefore, the life table values using the 2003 elderly census data shall be fairly close to those constructed using the 2000 population census data. As a

demonstration of the comparison, the life expectancies using the two censuses' data are listed in Table 3. As expected, the difference of the male is smaller.

Because the differences in life table values are small, the Taiwan government decided not to change the life table values constructed using the 2000 population census data. Note that the graduation methods considered in smoothing the mortality rates of the elderly include Whittaker and WLS (Weighted Least Squares) of the Gompertz law (Yue, 2002). Both methods use the exposures (or sample size) as the weights and are generally more stable than the moving-weighted-average-type methods used by the Taiwan government before. (See London, 1985, or Yue, 1997, for a detailed discussion of graduation methods.)

TABLE 3
Comparison of the Life Expectancies From the Two Censuses

	1999-2001 Complete Life Table	**2003 Elderly Census**
Male	73.79	73.82
Female	79.63	79.71

As for verifying the Gomperzt law, we shall first look at the exposures of each age from the elderly census. The age group "90 years old" has the largest exposures, with exposure values in the 1,300 person-year. With this size of exposures, it is likely to have inconclusive results unless there is very strong evidence. The following developments of verifying the Gompertz law are summarized from Yue (2002), and readers shall refer to the original paper for the details.

The Gompertz law assumes that the force of mortality satisfies $\mu_x = BC^x$, where x is the age, and $B > 0$ and $C > 1$ are model parameters. The probability that an individual now aged x would survive to age $x+1$ (denoted by p_x) is

$$p_x = e^{-\int_x^{x+1} \mu_t dt} = e^{-\int_x^{x+1} BC^t dt}$$

$$= e^{-BC^x (C-1)/\log C},$$

or equivalently, $\log p_x = -BC^x(C-1)/\log C$ and $\log p_{x+1}/\log p_x = C$. We will use the ratio of $\log p_x$ to verify the Gompertz law.

The verification is via the bootstrap simulation. Figure 2 shows the (1,000 replications) bootstrap confidence intervals of the parameter C for male and female. If the confidence intervals of each age have common intersection, we do not reject the Gompertz law; otherwise, we reject the Gompertz law. In other words, the Gompertz law is rejected for the male but is not rejected for the female. The possible value of C for the female is between 1.02 and 1.12. The results of male and female bootstrap simulation from the 2003 elderly census are very similar to those from the household registration data (Yue, 2002), and we use the 2003 data in Figure 3 as a demonstration.

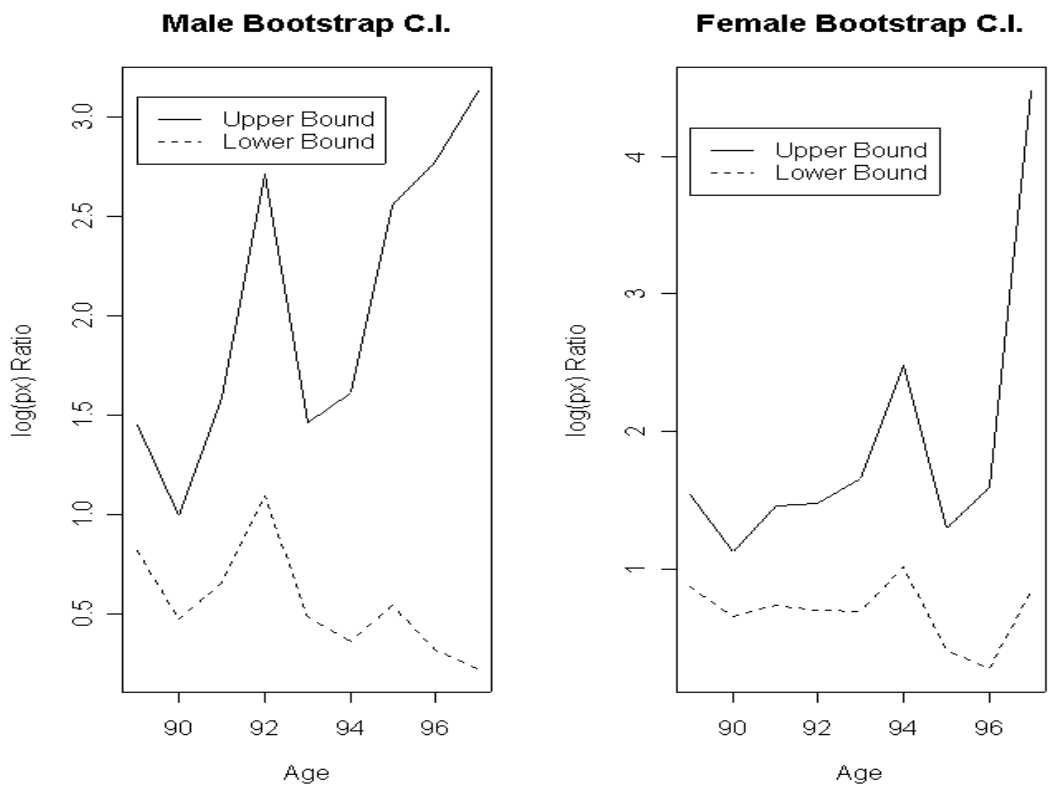

Figure 2. Bootstrap C.I. From the 2003 Elderly Census

However, we do not suggest rejecting (or not rejecting) the Gompertz law solely based on the 2003 elderly census data. From our experience, the rejection of the Gompertz law also depends on the data period used. For example, if we use only a

single year of mortality data in Taiwan, the Gompertz law is always rejected for the male and is often rejected for the female. But, after combining three years of mortality data (such as the 1999-2001 Complete Life Table), the Gompertz law is not rejected for both the male and female. Similar results appear in using the data of Japan and Sweden.

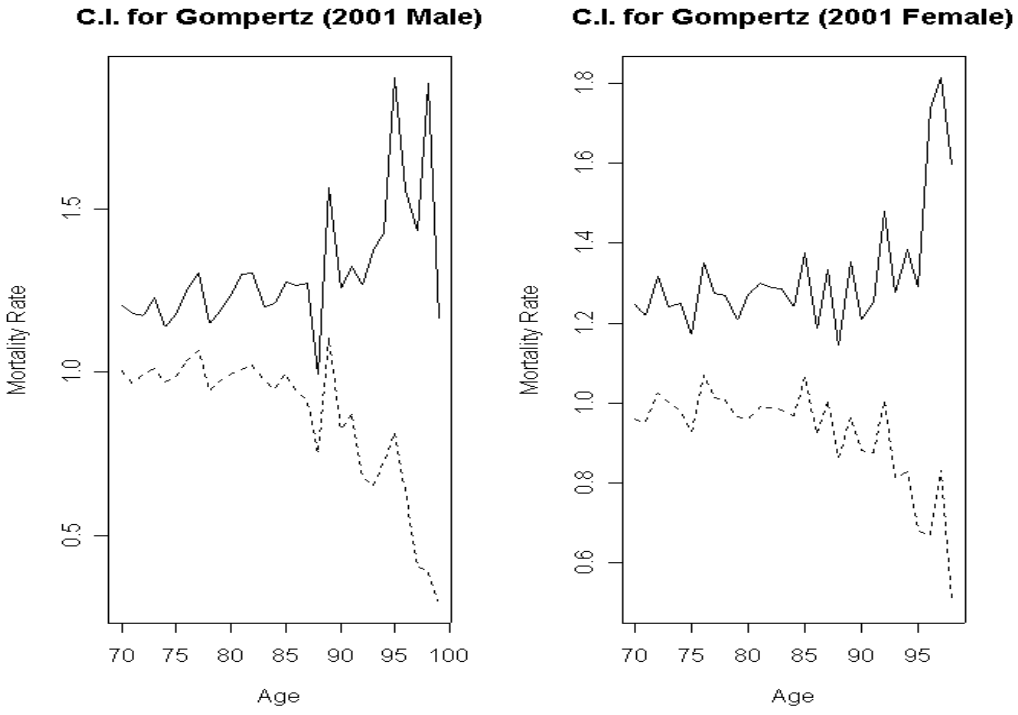

Figure 3. Bootstrap C.I. From the 2003 Taiwan Data

5. Discussion

Because the data quality of the 2000 population census in Taiwan has been questioned, the Taiwan government conducted a census for people ages 89 and over in 2003. The elderly census took six months to locate the elderly on the household registration list and another six months to verify and clean the data. With such a big effort, there are still problems in using the elderly census data directly.

The elderly census rely too much on the registration list, which possibly contains unreliable data, and thus the non-response rate is higher than expected and can have non-ignorable influence on the estimates of the elderly mortality rates. At a glance, Taiwan is an island country and the elderly have less mobility. The elderly should not be difficult to locate. However, this is not the case. About 3 million people migrated to

Taiwan due to the political unrest in 1949, and a lot of these people still alive moved back (or travel back and forth) to mainland China since 1990, when the relationship between Taiwan and China became less tight. It is very difficult to locate these people who move or travel to China, at least for now, since China is the major thread to Taiwan. We expect that the difficulty of tracing the elderly would be similar for countries or areas with high mobility.

If we define the population of interest as those who do not travel a lot, the elderly census can provide us a fairly good support for modifying the population census. However, if the population of interest is all the citizens, then the large percentage of non-response would make a difference in the estimate of the mortality rates. The current household registration system alone cannot guarantee us to locate the elderly, and the Taiwan government needs to modify this system.

The time between the census time and the standard date is short, which means that the exposure is small and the estimate of the mortality has big variance. Still, the overall behavior of the mortality rates from the 2003 elderly census is similar to those from the 2000 population census. The life tables constructed from the two censuses have almost identical results.

On the other hand, the small exposures do have an impact on checking the Gompertz law. Although we reject the Gompertz law for the Taiwan male based on the bootstrap simulation, we suggest not overlooking the testing result. Because the exposures are small, the simulation testing would become less reliable. In fact, if we increase the simulation runs to 10,000, the Gompertz law is not rejected. This is different from the results in Yue (2002). Like combining three years of mortality data in constructing complete life tables, we do not recommend using the 2003 elderly census data solely to verify the Gompertz law, since the census period is too short and a lot of factors (such as cold winters or flues) would increase/decrease the number of deaths of the elderly (especially the oldest-old, who usually have poor health).

Acknowledgements

The author is grateful to the Ministry of the Interior for providing the data of the 2003 elderly census and the information of conducting this census.

References

Brown, R.L. (1997). *Introduction to the Mathematics of Demography*. 3rd edition, ACTEX Publications, Winsted, Connecticut.

Council for Economic Planning and Development, Executive Yuan, R.O.C., Population Projection of Taiwan Area in 2004 and 2051.

Kannisto, V. (1996). *The Advancing Frontier of Survival: Life Tables for Old Age*. Odense University Press.

London, D. (1985). *Graduation: The Revision of Estimates*. ACTEX Publications, Winsted, Connecticut.

Thatcher, A.R. (1999). "The Long-Term Pattern of Adult Mortality and the Highest Attained Age." *Journal of the Royal Statistical Society*, Series A, 162: 5-43.

Willets, R. (1999). "Mortality in the Next Millennium." Paper presented in *The Staple Inn Actuarial Society*, December 1999.

Wilmoth, J., Skytthe, A., Friou, D. and Jeune, B. (1996). "The Oldest Man Ever? A Case Study of Exceptional Longevity." *The Gerontologist*, 36(6): 783-788.

Yue, C. J. (1997). *Graduation: Applications of Statistical Methods in Insurance* (In Chinese). Yeh-Yeh Publications, Taipei.

Yue, C. J., Hu, D.Y. and Chang, C. (2001). "Projections for the Old-Age Mortality in Taiwan: A Method Based on Incomplete Data." *Journal of Population Studies*, Vol. 22, 1-18.

Yue, C. J. (2002). "Oldest-Old Mortality Rates and the Gompertz Law: A Theoretical and Empirical Study Based on Four Countries." *Journal of Population Studies*, Vol. 24, 33-57.

Yue, C. J. (2004). "1999-2001 Complete Life Tables in Taiwan Area." Project of the Ministry of the Interior.