# Mortality at old ages: Evaluation of data quality

*D.A. Jdanov*[1,2] *and D. Jasilionis*[3]

[1]*Max Planck Institute for Demographic Research, Rostock, Germany; Jdanov@demogr.mpg.de*
[2]*New Economic School, Moscow, Russia; djdanov@nes.ru*
[3]*Max Planck Institute for Demographic Research, Rostock, Germany; Jasilionis@demogr.mpg.de*

The old age population in developed countries has increased remarkably. Yet internationally comparable high quality data on oldest-old mortality remain relatively scarce. Our study addresses several data quality issues relevant to population and death statistics above the age 80. Following previous studies by V.Kannisto, we apply the same set of measures and introduce new methods. Using data from the Human Mortality Database, we update prior findings by extending the analyses made so far to thirty five countries and by adding data on longer historical periods. We propose a systematic classification of country- and period-specific data, thus simultaneously accounting for each indicator of data quality. The results show that the data quality varies across countries and time periods. We found that data for some countries such as Chile, Canada, Germany, and the USA, mortality indicators for the most advanced ages should be treated with caution even for the most recent decades.
Keywords: old age mortality, quality of statistics, population estimates

## 1. Introduction

The old age population in developed countries has increased very rapidly throughout the second half of the 20th century. Improvements in survival are pushing new limits: today more than half of all males and two thirds of all females born in Western countries may reach their 80th birthday. The proportion of centenarians increased about ten times over the last thirty years, and more and more people celebrate their 100th birthday (Robine & Vaupel, 2001). Despite these remarkable developments, internationally comparable high quality demographic data on old-age populations remain insufficient.

The most detailed and systematic assessment of international old-age mortality data was published by Väinö Kannisto (Kannisto, 1994; 1999). Introducing a set of data quality criteria, Kannisto (1994) divides thirty developed countries into four groups according to data quality: good, acceptable, conditionally acceptable, and weak quality data. Thatcher, Kannisto & Vaupel (1998) defined the good quality data as follows: no evidence of age heaping; no obvious age overstatement; plausible sex ratios; internal consistency in death rates, age pattern and sex ratio over time; comparability of country-specific data to international trends. In addition, the authors stress compulsory birth registration (at least with a history of more than 100 years) as an important factor in data reliability for highest ages.

Irregular old-age mortality patterns due to age reporting problems have been identified for the former USSR countries, the USA, Canada, Spain, Portugal, and some other countries (Anderson & Silver, 1997; Bourbeau & Lebel, 2000; Vaupel, Wang, Andreev & Yashin, 1997; Coale & Kisker, 1986; Elo & Preston, 1994; Preston, Elo & Stewart, 1999; Vaupel, Rau, Camarda & von Kistowski, 2006). Perhaps the most in-depth country-case studies were conducted in the USA, disclosing a complexity of problems that may influence old-age mortality statistics. Relying on individual-level data, Preston, Elo & Stewart (1999) have shown that age misreporting occurs in both directions – age overstatement and age understatement. The authors conclude that irrespective of the direction age misreporting takes, it leads to a substantial underestimation of mortality levels at ages over 80 (Preston, Elo & Stewart, 1999). To compensate for weak data quality, additional data sources such as Medicare Social Security Administration records have been widely applied to derive USA mortality estimates for oldest ages (Kestenbaum, 1992). A more favorable situation has been observed among countries with a long tradition of fully functioning population registries, such as Sweden, Denmark, and the Netherlands. But even here, age overstatement and age heaping in deaths and the population stopped only by the beginning of the 20th century (Skytthe & Jeune, 1995; Skytthe, Hauge & Jeune, 1999; Lundström, 1995; Jeune & Vaupel, 1999).

Studies on the validation of centenarian and super-centenarian (aged 110 and older) data require separate attention. Jeune & Vaupel (1999) suggest that the majority of countries have reported data on centenarians and super-centenarians that are questionable (particularly for cohorts born before reliable birth registrations began). The age validation of centenarians remains problematic for many countries since they do not have fully functioning population registration systems – so far, these are available only in a few countries. In other countries such as France and Italy, alternative data sources have been widely used (municipal election registers or birth registration systems) (Jeune & Vaupel, 1999). The most difficult situation is for countries with a relatively short history of official birth registration (for example, the USA and China). In these cases, Kannisto's (1994, 1999) criteria and other indirect indicators of the accuracy of age reporting have been used (Wang, Zeng, Jeune & Vaupel, 1999).

In our article, we study old-age mortality patterns in the 37 countries that are currently included in the Human Mortality Database (HMD). We identify dubious or irregular mortality patterns, suggesting that there are potential problems of data quality. We update previously published findings by extending the analyses conducted so far to 37 countries and by adding data for longer historical periods. Applying a cluster analysis, we propose a systematic classification of country- and period-specific data, thus simultaneously accounting for the whole set of data quality criteria. In addition, we examine the reliability of official population estimates by comparing the data series provided by statistical offices with those obtained from the HMD.

## 2. Data and methods

***Description of the database***

For our analyses we used data from the HMD. This unique database provides a set of mortality indicators as well as original population statistics. The main advantage of the HMD is that country-specific old-age mortality estimates were calculated by applying uniform methods. Consequently, artificial cross-country differences resulting from a variety of methods applied for the estimation of mortality surfaces have been avoided.

At least some of the potential old-age population data quality problems can be solved by using the method of extinct generations (Vincent, 1951). As the quality of population statistics has been continuously improving (in most cases), retrospective recalculations of population counts according to the most recent data allow us to adjust age-specific population counts for previous time periods. The recalculations are based on death counts only, but negligible international migration at these ages is not taken into account. This approach has been widely used to validate official population figures (Rosenwaike, 1979; Kannisto, 1988; Thatcher, 1992, 2001; Hill et al., 2000). Special adjustments to the method were introduced for "almost-extinct" cohorts (with a relatively small proportion of survivors), e.g., cohorts over age 90 or 95 (Thatcher, 1999).

The methodology applied to the HMD for the re-estimation of population counts above age 80 is a combination of methods built around Paul Vincent's basic method of the retrospective revision of population estimates, using data on deaths counts (extinct cohort and almost extinct cohort method), with further modifications to it by Väinö Kannisto, Roger Thatcher, and John Wilmoth (Wilmoth et al., 2007).

***Methods***

We apply several indicators that identify irregular patterns in death and population counts above age 80. Our system of data quality checks is essentially based on the previously published studies by Kannisto (1994, 1999). The outcomes of these indirect data quality tests have been validated by studies that used individual-level population register data or alternative data sources, such as social security records, municipal election registers, or birth registration system data (Kestenbaum, 1992; Jeune & Vaupel, 1999). Since the death registration is considered almost 100% complete in the HMD countries, the data quality tests do not include criteria of the coverage of death registration.

We begin our analyses with describing *age overstatement* in death and population statistics among the HMD countries. The probability of age overstatement becomes more pronounced with increasing age. This leads to implausible age-specific distributions of deaths and population at old ages (Kannisto, 1999). Typically, such distortions are identified by comparing the observed age distributions to those obtained by using indirect methods, mathematical models or population registration system based estimates that are known as accurate (Kannisto, 1999; Perls et al., 1999; Coale & Kisker, 1986; Coale & Li, 1991; Kestenbaum, 1992; Preston, Elo, Stewart, 1999; Bennett & Garson, 1983). The validation studies reveal very significant distortions in centenarian statistics. For example, Perls et al. (1999) show that only 46 persons out of an initial list of 289 people qualify to be counted as centenarians in a local area of eight suburbs around Boston, USA.

The first criterion of age overstatement examines whether the proportion of the population at the extreme old ages of the whole old age population is plausible. The proportion is calculated as the ratio of the total life table person-years lived at age 100 to the total life table person-years lived at age 80 (T100/T80). The final data quality index is obtained by dividing country and period-specific ratios (T100/T80) to the corresponding ratios for Sweden (the "golden standard"). Significantly higher scores (e.g., the ratios 1.5 or more times exceed those observed in Sweden) signal about potential age overstatement in the population data.

We are aware that in some cases higher ratios simply point to the fact that mortality above age 80 becomes lower than in Sweden (Cheung & Robine, 2007). This is applicable only to some successful countries (such as Japan and France) during the last two decades showing a greater progress in reducing mortality than in the reference country, Sweden. Therefore, we validate the T100/T80 criterion by performing an additional test for mortality crossovers. The mortality crossovers, occurring as the consequence of data inaccuracies at old age, refer to the cases when mortality rates at old ages are surprisingly low despite high mortality at young and adult ages (Coale & Kisker, 1986). Following Coale & Kisker (1986), we examine graphically the feasibility of relationships between survival probabilities between ages 10 and 55 and values of life expectancy at age 80. We consider that age overstatement is present only in the cases when: 1) the ratio T100/T80 is significantly (at least 1.5 times) higher than in Sweden; 2) the country shows implausibly high values of e(80) (e.g., close or even higher than in Sweden) despite high (if compared to the majority of other countries) mortality in adult ages.

The second criterion of age overstatement deals with deaths at the most advanced ages and refers to the ratio between observed deaths at ages 105+ to observed deaths at ages 100+ and the ratio between observed deaths at ages 110+ to observed deaths at ages 105+ (Kannisto, 1994, 1999). The ratios with considerably higher values than those for golden standard (Sweden) are considered as evidence of age overstatement.

Age heaping (digit preference) in death and population counts is assessed by applying Whipple's Index of age accuracy. Usually, the index is calculated as the ratio of the sum of population counts at ages ending with 5 and 0 to the total of sum population counts within a given age interval. Following Wang, Zeng, Jeune, & Vaupel (1999), we employ a modified version of this measure to the assessment of death data quality at oldest-old ages (Equation (1)). Again, we assume that significant deviations of this indicator from the "golden standard" (Sweden) suggest that there are possible age reporting problems.

$$WI = \frac{D_{95} + D_{100} + D_{105}}{\sum_{i=93}^{i=107} D_i} * 100 * 5, \qquad (1)$$

where $D_i$ is number of deaths at age *i*.

We further investigate the quality of age reporting by applying several age heaping measures proposed by Kannisto (1994, 1999) and Vaupel, Wang, Andreev, & Yashin (1997).

Slightly more complex indicators are used for assessing the digit preference for age reporting (age heaping) for ages 90 and 100 (Kannisto, 1999):

$$AHI_i = \frac{D_i}{\exp\left(\frac{1}{5}\sum_{y=i-2}^{i+2} \ln(D_y)\right)}, \qquad (2)$$

where $D_i$ is the number of deaths at age *i*.

In many cases, the literature has not clearly described precise criteria for distinguishing between "good" and "bad" data quality. Such criteria have been available for defining data quality groups according to age heaping and age-sex accuracy indicators. For the remaining criteria, we applied conventional procedures of cluster analysis. This procedure allows to avoid arbitrary definitions of quantitative borders between the data quality categories. Therefore, after dividing country-specific data series into 10-year periods, each country is assigned to one of the four data quality clusters[1] according to every data quality criterion described above. The first cluster refers to the best quality data and the fourth cluster describes the worst quality data. The maximum average values of indicator within the first three clusters across the whole period have been considered as limits $B_i$ defining the groups of data quality:

[1] We used a classical hierarchical clustering with single linkage (nearest neighbor) algorithm and restricted number (four) of clusters (see Everitt et all (2001)). All calculations have been performed using Matlab v.7.01, function *clusterdata* (see for detailed description http://www.mathworks.com/).

$$B_i = \max_{1900 \le y \le 2000} \left( \frac{1}{N_i^y} \sum_{j=1}^{N_i^y} v_j^y \right), \quad i = 1,2,3 \qquad (3)$$

where $N_i^y$ refers to a number of countries assigned to cluster $i$ in 10-year period $y$, $v_j^y$ is a value of indicator for respective country and time period, and $j$ denotes an ordering number of a country (according to an alphabetical order). To simplify the classification procedure we used rounded $B_i$. These values are used for defining limits of the data quality groups. Following the prior studies (Kannisto, 1994, 1999), we assume that the increasing deviation from the best group (exceeding $B_1$) signals about the deterioration of the data quality. Our further analyses assume that data quality indicators exceeding the maximum average values of the third cluster ($B_3$) point to serious problems in data quality.

## 3. Results

This section presents a set of age overstatement and age heaping indicators allowing for the identification of potential data quality problems for each HMD country. Each measure points to very specific aspects of data quality problems, thus a relatively complex approach is needed to draw final conclusions on the reliability of population and death figures for a particular country. Note that in some cases, the impact of certain data deficiencies may have a negligible effect on aggregate mortality measures (e.g., age heaping at age 100 has rather negligible impact life expectancy at age 80). Thus, several indicators simultaneously should be taken into account.

Below, we provide outcomes from our data quality validation study, with a special focus on countries that have problematic data on old age. The country-periods are divided into the following groups: good quality, acceptable quality, conditionally acceptable quality, and weak quality. We introduce a "golden standard" to some cases in order to reveal unusual patterns of age reporting indicators. Following numerous prior studies, we consider Sweden as having the best data quality in the world (Wang, Jeune & Vaupel, 1999). Accordingly, we calculate the relative ratios between country-specific and Swedish indicators. We assume that countries with higher data quality follows the pattern of data quality measures shown by the "golden standard" data.

**Age overstatement.** Our first measure of age overstatement is the ratio between T100 (life table person-years lived above age 100) to T80 (life table person-years lived above age 80) (T100/T80). We investigate the ratio of T100 to T80 in the country under consideration to the corresponding ratio of T100 to T80 in Sweden.

For some countries with higher ratios of T100 to T80 than in Sweden, age overstatement cannot be confirmed according to the additional test for the presence of mortality crossovers. The following countries-periods have been assigned to the good data quality group despite higher ratios of T100 to T80: Australia (1971-2000 for males and 1981-2000 for females), France (1991-2000 for males), Iceland (1981-2000 for males), Italy (1991-2000 for males), Japan (1981-2000 for males and 1991-2000 for females), Luxemburg (1971-1980 for males), and Spain (1971-2000 for males). It can be seen that many of these cases represent countries showing the most rapid mortality improvements during the last two-three decades (e.g., countries of southern Europe and Japan). Results for Luxemburg and Iceland are probably attributable to the large fluctuations due to small numbers.

Following this criterion, the weak quality group includes data for Canada (1921-1950 for males, 1921-1940 for females), Chile (1991-2000 for males), Spain (1921-1940 for females), Portugal (1941-1950 for females), and Lithuania (1971-1980 for males). The worst indicators are observed for Lithuanian males for 1971-1980 (6.5) and Spanish females for 1931-1940 (5.8).

Data for almost all countries included in the HMD have improved in quality over time. For example, a remarkable progress has been observed for Lithuanian males, who show a decrease in the age overstatement indicator from 6.5 in 1971-1980 to 1.8 in 1991-2000. For USA (both sexes), Canada (both sexes), and Spanish males, however, the relative difference from the Swedish standard have remained above the level of 2.0 (conditionally acceptable quality) throughout the whole period covered (1900-2000) (with the exception of lower rates for Spain in 1951-1970). The data for the 1990s shows that almost all countries are assigned to the good or acceptable data quality group, with the exception of Canada (conditionally acceptable quality for both sexes), Chile (weak quality for males, conditionally acceptable quality for females), Spanish males (conditionally acceptable quality for males), and the USA (conditionally acceptable quality for both sexes).

The consequence of age overstatement is underestimation of the overall level of mortality at old ages. Following Preston, Elo & Stewart (1999), we performed several simulations attempting to estimate how age overstatement may affect estimates of life expectancy at age 80. We have found that the decrease in the age overstatement index (T100/T80 ratio for Canada to the corresponding ratio for Sweden) from 2.0 to 1.5 times leads to the drop in life expectancy at age 80 (0.05-0.2 years or by 0.5-2%). A more notable effect (decrease of 0.5 years or by 25%) has been observed for life expectancy at age 90.

The next two indicators of age overstatement in deaths are <u>the ratio of deaths at ages 105+ to deaths at ages 100+ and the ratio of deaths at ages 110+ to deaths at ages 105+</u>.

The ratios of deaths at ages 105+ to deaths at ages 100+ are very high for Chile (1981-1990 for both sexes), New Zealand (1951-1960 for both sexes), New Zealand Non-Maori (1951-1960 for males), and Portugal (1941-1950 for both sexes), suggesting that significant age overstatement in deaths has occurred. Exceptionally unfavorable indicators are identified for Chilean males and females for 1981-1990 (48.1 for males and 25.3 for females). At the same time, the corresponding ratios for Sweden have never reached the level of 6.0 throughout the whole period. In addition, the following country-period data are classified as only conditionally acceptable: Chile (1991-2000 for males), Lithuania (1981-1990 for males, 1971-1990 for females), New Zealand (1961-1970 for both sexes), New Zealand Non-Maori (1921-1930 for males), Portugal (1951-1960 for both sexes), Spain (1931-1940, 1951-1960, and 1981-1990 for males, 1911-1920, 1931-1950, and 1981-1990 for females), the USA (1961-1980 for both sexes). With the exception of Canada, the classification of countries generally agrees with the groups identified by the first indicator of age overstatement (T100/T80).

An interpretation of the indicator of age overstatement for super-centenarians (the death ratio at ages 110+ to ages 105+) is not so straightforward due to the small numbers. In addition, it was not possible to calculate this indicator for several countries-periods due to the fact that denominators (deaths at ages 105+) were equal to zero. However, we found that a similar set of countries as in the previous analyses consistently show substantially higher death ratios than does Sweden. For example, weak quality or conditionally acceptable quality groups include Chile (1981-1990 for males), New Zealand (1951-1960 and 1971-1980 for males, 1951-1970 for females), New Zealand Non-Maori (1951-1960 and 1971-1980 for males), Portugal (1961-1970 for males, 1941-1960 for females), and the USA (1961-1980 for males, 1961-1970 for females). The most strikingly high ratios are found for New Zealand Non-Maori males for 1971-1980 (100.0), Finnish females for 1961-1970 (33.3), and New Zealand males for 1971-1980 (33.3). These results (especially for Finland) should be treated with caution, however, as they are probably due to random fluctuations in small death numbers.

**Age heaping.** Digit preference or age heaping in death and census records is another important source of errors in population and death statistics. Following formulae proposed by Kannisto (1999), we estimate the age heaping index (defined by the equation 3) for deaths at ages 90 and 100. Age heaping at these ages is verified by comparing death numbers at these ages to the expected number of deaths at adjacent ages.

Our results indicate serious age heaping at age 80 in Ireland (1951-1980 for both sexes), Portugal (1941-1950 for males, 1941-1960 for females), and Spain (1911-1960 for both sexes). The most significant age heaping indicators are identified for Spanish males and females in 1911-1920. Canada (1931-1940 for males and 1931-1950 for females) and New Zealand Non Maori (1921-1930 and 1941-1950 for males and 1921-1950 for females) show moderate age heaping indicators.

As for age heaping at age 90, Portugal (1941-1950 for both sexes) and Spain (1911-1940 for males and 1911-1950 for females) again are assigned to the weak data quality group. Canadian data (1921-1930 for females), Finnish data (1901-1910 for both sexes), Icelandic data (1961-1970 for females), Irish data (1951-1970 for both sexes), and Portuguese data (1951-1960 for females), Spanish data (1941-1960 for males, 1951-1960 for females) show moderate age heaping levels.

Concerning age heaping at age 100, the majority of countries show weak data quality at least for some decades preceding the 1980s. It was not possible to calculate this indicator for several countries-periods due to the absence of deaths at the most advanced ages. The data for the 1990s show serious age heaping at age 100 for Estonian males, Finish males, Iceland, Latvian males, Scottish males, and Slovenian females. The following countries are assigned to the conditionally acceptable quality group, covering the same period: Chile (females), Germany (males), Italy (males), Latvia (females), Lithuania (females), Netherlands (males), New Zealand (females), Non Maori New Zealand (females), Norway (females), Poland (both sexes), Portugal (males), Slovakia (females), Slovenia (males), and Switzerland (males) These results (especially for smaller countries, such as Iceland or Estonia) should be treated with caution due to the very small numbers of deaths at these very advanced ages. However, our results on larger countries, such as France or Japan, confirm the importance of age heaping at age 100 in the past. Overall, our study shows that age reporting tends to improve with time in almost all countries included in the HMD. We do not find evidence of age heaping at ages 80 and 90 for the 1990s. However, the quality of age reporting at age 100 remains problematic for a relatively large number of countries.

Whipple's Index for centenarians is an additional measure of accuracy in age reporting in death counts at the most advanced ages (see methods section for more details). According to Whipple's Index, the weak and conditionally acceptable quality groups include Canada (1921-1930 for males), Finland (1921-1940 for males, 1911-1920 for females), East Germany (1961-1970 for males), Iceland (1971-1980 for females), Luxemburg (1971-1980 for males), Portugal (1941-1950 for both sexes), and Spain (1911-1930 for males, 1911-1930 and 1941-1950 for females). The most significant deviations from the Swedish "golden standard" have been found for Spain (16.5 for males in 1911-1920) and Portugal (13.7 for males in 1941-1950).

***Classification of country- and period-specific statistics***

In this section, we introduce a more systematic classification of country- and period-specific data simultaneously accounting for the selected data quality criteria used in our study. The age overstatement criteria include T100/T80 criterion (validated by the test for mortality crossovers) and D105+/D100+ criterion. Age heaping criteria include the Whipple index and age heaping indexes (for ages 80 and 90 only). Several criteria (ratio of deaths at ages 110+ to deaths at ages 105+, age heaping indicator for age 100) discussed in the previous sections have not been considered for the final classification due to a larger degree of uncertainty (a consequence of small numbers at the most advanced ages).

A number of points according to the selected data quality criteria have been assigned to each country-period data. The following rule has been applied: the good quality group refers to zero points, the acceptable quality group refers to one point, the conditionally acceptable quality group corresponds to two points, and the weak quality group refers to three points. Finally, the country-period data are appointed by a maximum of points collected according the whole set of criteria. Due to the fact that the data for both sexes are pooled together, the countries are classified by choosing either a male or female indicator showing maximum points. The data-quality groups and their labels were determined on the basis of clustering results, numerical rounding, and previous studies that used those indicators.

It can be seen that for most of the HMD countries, the reliability of old age mortality statistics has improved. Ireland, Lithuania, New Zealand, New Zealand Non Maori, and Portugal, for example, have shown weak data quality in the past but now they are assigned to the acceptable quality group according to more recent indicators. The data for some other countries, however, have not improved over time: Canada and the USA, for example, remain in the conditionally acceptable group throughout the second half of the 20th century. The Chilean data is the most problematic, as are the historical series for some other countries, such as Canada, Ireland, Finland, Lithuania, New Zealand (Non-Maori), Norway, Portugal, Spain, and the USA. As expected, the data for the countries with a long history of fully functioning national population registration systems (Sweden and Denmark) or municipal population registration systems (the Netherlands, Belgium) show the best quality.

Overall, the country-period data can be summarized by applying the following schema:

- **Best data quality group**. It shows the highest quality throughout the whole period. The group includes Belgium, France, the Netherlands, and Sweden. The Danish, Finnish (since 1951-1960), Italian, Japan (since1971-1980), Swiss, Polish, and Western German data are also assigned to this category as they show best data quality throughout the whole period covered with exceptions of one or two periods (before 1991-2000) with acceptable data quality.

- **Acceptable data quality group**. It consist of Australia, Austria, the Czech Republic, England & Wales, Estonia, East Germany (except 1961-1970), Hungary, Iceland, Ireland (from 1981-1990) Japan (before 1970), Latvia (from 1971-1980), Luxemburg (from 1981-1990), New Zealand (from 1961-1970), New Zealand Non-Maori (from 1961-1970), Norway (from 1961-1970), Portugal (from 1961-1970), Scotland, Slovakia (from 1971-1980), Spain (from 1961-1970), and Slovenia.
- **Conditionally acceptable group**. It includes countries consistently showing moderate data quality problems (with a possible short-term improvement or weakening in data quality): Canada (from 1951-1960), Finland (before 1951-1960), Latvia (before 1971), Lithuania (from 1981-1990), Luxembourg (before 1981-1990), New Zealand Non Maori (before 1961-1970), Norway (before 1961-1970), and the USA.
- **Weak data quality group.** It incorporates countries and periods showing very serious data quality problems: Canada (before 1951-1960), Chile, Ireland (before 1981-1990), Lithuania (1971-1980), New Zealand (1951-1960), Portugal (before 1961-1970), and Spain (before 1961-1970).

### *A comparison between the HMD and official statistics*

Our results, presented in the previous sections, are based on the HMD data series. As described in the section on data and methods, a set of methods allowing improvements to be made to the quality of the mortality estimates has been applied within the framework of the HMD methods protocol. Therefore, country-specific population data from the HMD may differ from the official population figures provided by national statistical offices. It has been shown that the official population estimates often have various deficiencies, whereas the corresponding (adjusted) HMD data demonstrate more plausible patterns. Jdanov, Scholz & Shkolnikov (2005), for example, found evidence of population overestimation for West German males (1980s-1990s). The relative difference between the official and recalculated population estimates (using the extinct and almost extinct cohort method) manifested at ages 90+ in the beginning of the 1970s. This gap increased further between 1971 and 1987. The difference between the official and re-estimated population can be used as an indicator of potential data quality problems in official population statistics.

The relative weighted difference is calculated as

$$\delta(x,t) = \sum_{x=80,85,90,95+} w_x \frac{P_5^{KTD}(x,t) - P_5^{official}(x,t)}{P_5^{KTD}(x,t)},$$

where $P_5^{KTD}$ and $P_5^{official}$ refer to a population in a 5-year age group in the HMD and official estimates, respectively, $w_x$ denotes the population weights in the Swedish population averaged for 1950-2000.

Our analyses showed that for majority of countries the difference between the official statistics and the HMD estimates tend to diminish to the level of zero. However, at least some small differences are observed for most of the countries at some points in the past. In addition, part of the decrease in the gap between the two series of the population estimates in the 1990s can be explained by the fact that the official population estimates (for the most recent years) have been used as input data for the backward re-estimation by extinct and almost extinct cohort methods.

Deviations between the population estimates are significantly associated with sex, early calendar periods, and high ages. The differences are more notable among males than among females. This can be explained by the stronger impact of small numbers affecting the magnitude of the absolute difference between the official and HMDB data series among males. For the period around WW1, the quality of old age statistics for males is considerably lower than for other periods. The difference between official data and extinct and almost extinct cohort estimates dramatically increases with age, but this gradient becomes less pronounced in time. In general, all factors explaining the gap between the two series tend to loose their importance. The only exception is the period 1941-1950.

The countries with fully functioning population registration systems (the Nordic countries, the Netherlands, and Belgium) show almost no difference between the official estimates and the HMD population estimates for the last few decades. The introduction of the population registration systems seems to enhance the accuracy of population estimates and also to narrow differences between the official estimates and the estimates based on the HMD methodology. Finland is an illustrative example. Compared to the HMD data, this country shows a significant undercount of its old age population in the 1960s. However, the discrepancy continuously decreased over the 1970s and become very small afterwards.

Among the countries with good quality data, the most notable difference between the official and the HMD series has been found for Slovakia. The gap tends to diminish significantly for the population census years and this suggests that the quality of the post-censal population estimates tend to deteriorate with each subsequent year after the census. As for the countries assigned to the acceptable quality group, the most pronounced disagreement between the two sources has been identified for East Germany. Here, the gap is much more pronounced than in the Nordic countries but it is smaller than in Slovakia or New Zealand.

The most pronounced difference between the HMD and official population estimates are in the third group of countries, assigned to the conditionally acceptable quality group. There are very large discrepancies between the two sources for pre-war Spain (even for the census years).

## 4. Conclusion

Our work was stimulated by prior substantial research done by Väinö Kannisto. Using his system of data quality indicators together with other internationally recognized criteria, we performed an analysis of data quality for all countries included in the most recent version of the HMD. We extended prior analyses by including 37 countries and by lengthening the time period back to include time since the beginning of the 20$^{th}$ century. Finally, applying cluster analysis, we introduced a more objective classification of countries and periods simultaneously accounting for a whole set of selected data quality indicators.

We assume that dubious or irregular mortality patterns identified in this study suggest (although indirectly) about data quality problems. The outcomes of the studies using indirect measures of age overstatement or age heaping have been proved by the findings based on individual or population registration system data (such as studies on age-validation of centenarians) (Perls et al., 1999; Kestenbaum, 1992).

Our results suggest that the data on oldest-old mortality in industrialized countries is quite reliable. However, the data quality varies substantially across countries and time periods. Although the data quality has improved in the majority of the HMD countries, some country-period data should be used with caution.

Several countries, e.g. Canada, Chile, Spain, and the USA, systematically demonstrate worse results up to the most recent periods. Researchers should be aware of very serious inaccuracies in the historical series of old-age mortality statistics for Canada, Ireland, Lithuania, New Zealand Non-Maori, New Zealand, Portugal, and Spain. Very pronounced data quality problems in the aforementioned countries lead to misleading old-age mortality estimates. Age overstatement results in underestimation of the overall mortality level at old age, while age heaping only distorts mortality estimates at certain ages

We suggest that data quality problems come both from inaccuracies in the death count data and errors in estimating populations within the inter-censal periods. After having compared the HMD population estimates derived by applying a set of extinct- and almost-cohort methods, we conclude that the biggest discrepancies between the two sources are found for countries with conditionally acceptable or weak quality data. The smallest differences and the most reliable official data have been found for the countries with fully functioning population registration systems.

## References


Anderson, B.A. & Silver, B.D. (1997). Issues of Data Quality in Assessing Mortality Trends and Levels in the New Independent States. In: J.L. Bobadilla, Ch.A. Costello & F. Mitchell (Eds.). *Premature Death in the New Independent States* (pp. 120-154). Washington: National Academy Press.

Andreev K., Jdanov, D., Soroko, E., & Shkolnikov, V. (2003). *Methodology of the Kannisto-Thatcher Database on old age mortality.* Retrieved 13 June 2006 (http://www.demogr.mpg.de/databases/HMDb/xservices/method.htm).

Bennett, N.G., Garson, L.K. (1983). The centenarian question and old-age mortality in the Soviet Union, 1959-1970. Demography, Vol.20, No. 4, pp.587-606.

Bourbeau, R. & Lebel, A. (2000). Mortality Statistics for the Oldest-Old: An Evaluation of Canadian Data. *Demographic Research*, Vol. 2, pp.1-36.

Cheung, S.L.K., Robine, J.-M. (2007). Increase in common longevity and the compression of mortality: The case of Japan. Population Studies, Vol.61, No. 1., pp.85-97.

Coale, A.J. & Kisker, E.E. (1986). Mortality crossovers: reality or bad data? *Population Studies*, Vol. 40, pp. 389-401.

Coale, A.J., Li, Sh. (1991). The effect of age misreporting in China on the calculation of mortality rates at very high ages. *Demography*, Vol. 28, No. 2, pp. 293-301.

Elo, I. & Preston, S. (1994). Estimating African-American mortality from inaccurate data. *Demography*, Vol.31, pp. 427-458.

Everitt, B., S. Landau, M. Leese "Cluster Analysiss", 4th edition (2001), 248 pages.Publisher**:** A Hodder Arnold Publication

Hill, M.E., Preston, S.H. & Rosenwaike, I. (2000). Age reporting among white Americans aged 85+: results of a record linkage study. *Demography*, Vol. 37 (2), pp. 175-186.

Jdanov, D.A., Scholz, R.D. & Shkolnikov, V.M. (2005). Official population statistics and the Human Mortality Database estimates of populations aged 80+ in Germany and nine other European countries. *Demographic Research,* Vol. 13 (14), pp. 335-362.

Jeune, B. & Vaupel, J. (1999). Species of Evidence of Exceptional Longevity. In: B. Jeune and J. Vaupel (Eds). *Monographs on Population Aging, Vol. 6.,* (pp. 11 - 22). Odense: Odense University Press.

Kannisto, V. (1988). On the survival of centenarians and the span of life. *Population Studies*, Vol. 42, pp. 389-406.

Kannisto, V. (1994). *Development of oldest-old mortality, 1950-1990. Evidence from 28 developed countries. Monographs on Population Aging. Vol. 1* (1-108 p.). Odense: Odense University Press.

Kannisto, V. (1999). Assessing the Information on Age at Death of Old Persons in National Vital Statistics. *Validation of Exceptional Longevity. Monographs on Population Aging. Vol. 6.* (pp. 240-249). Odense: Odense University Press.

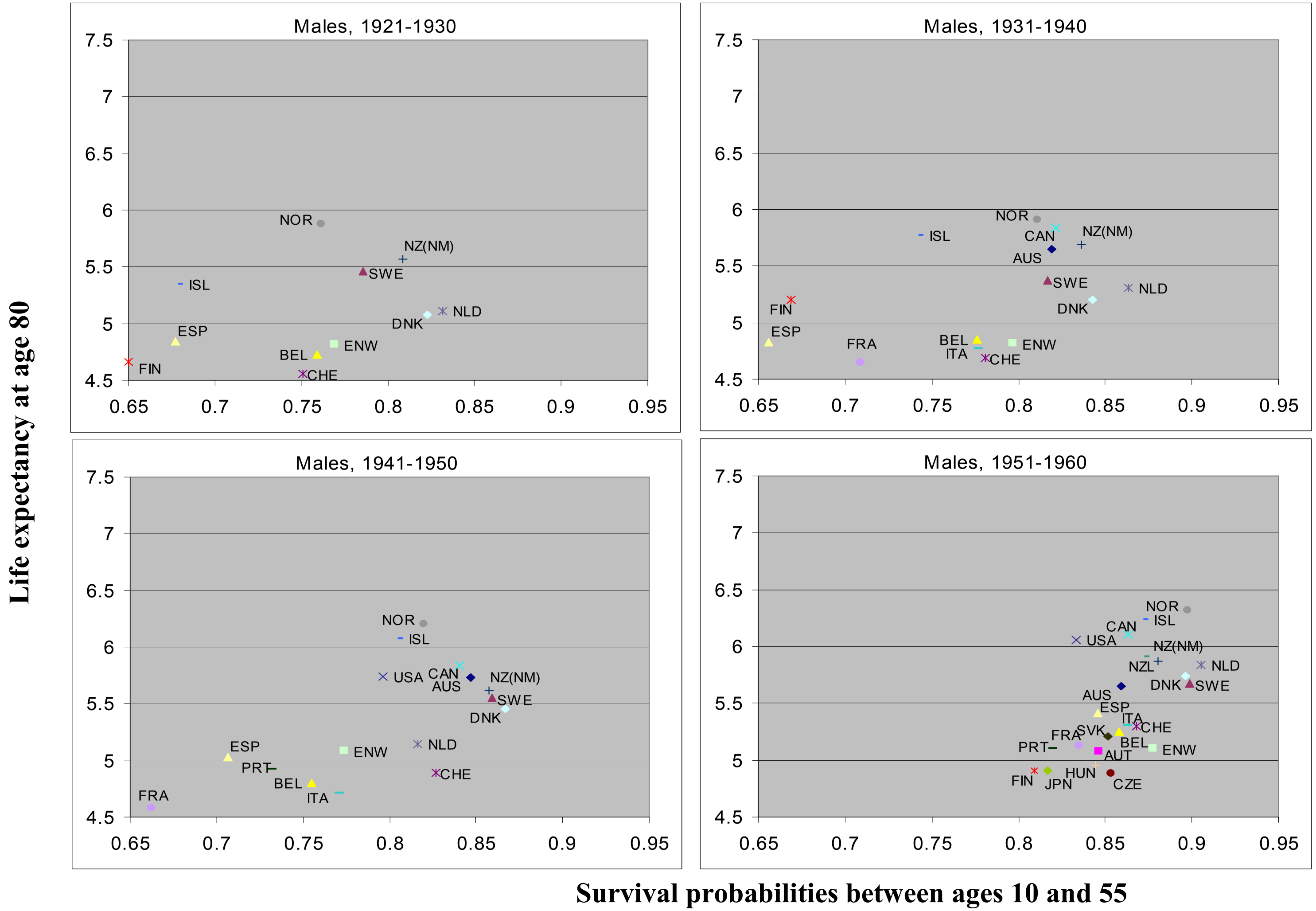
Kestenbaum, B. (1992). A description of extreme aged population based on improved Medicare enrollment data. *Demography*, Vol.29 (4), pp. 565-580.

Lundström, H. (1995) Record Longevity in Swedish Cohorts Born since 1770. . In: B. Jeune and J. Vaupel (Eds). *Monographs on Population Aging*. *Vol.2* (pp. 67 - 74). Odense: Odense University Press.

Perls, T.T., Bochen, K., Freeman, M., Alpert, L., Silver, M.H. (1999). Age Validation in the New England Centenarian Study. *Validation of Exceptional Longevity.*

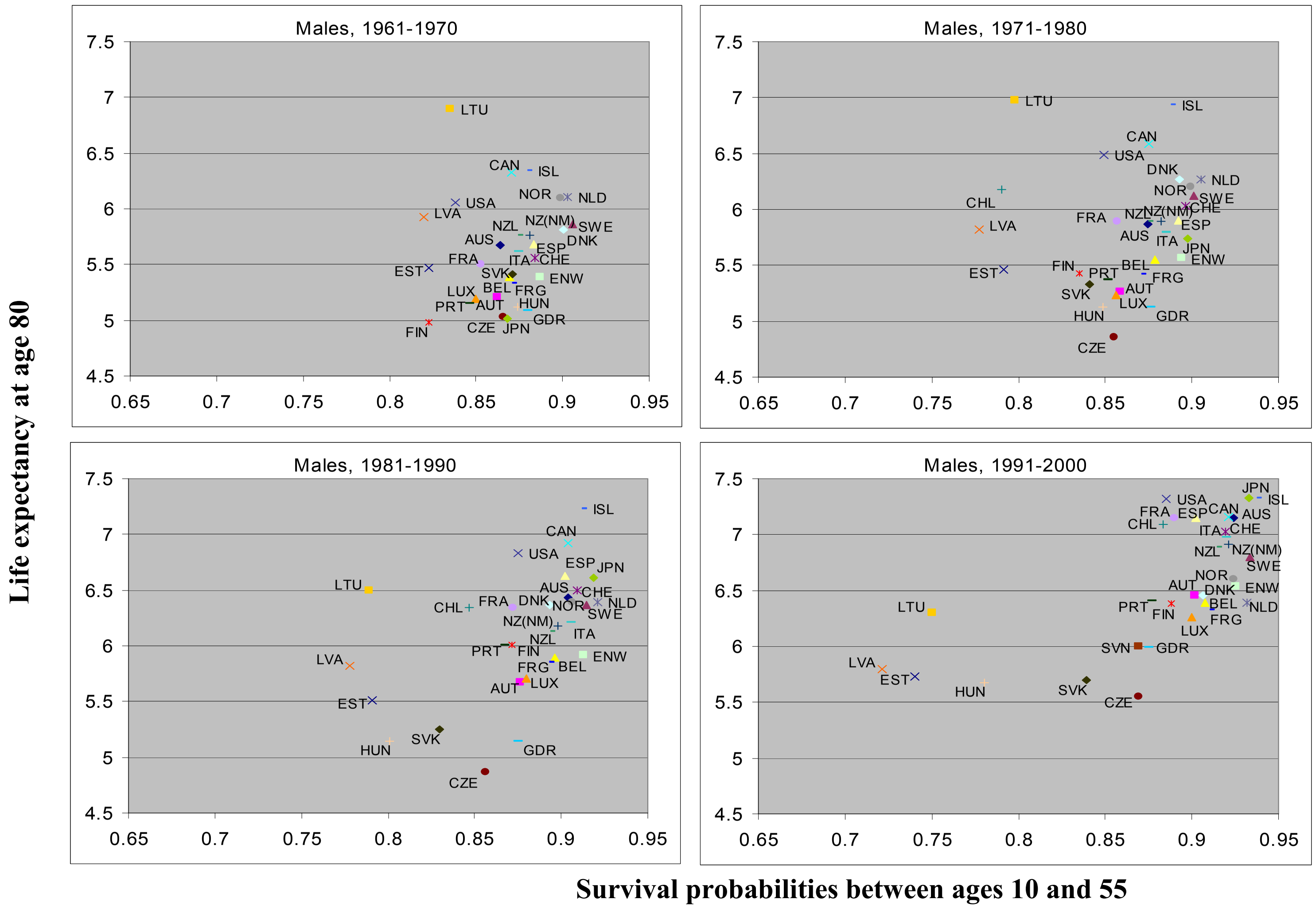
*Monographs on Population Aging. Vol. 6.* (pp. 195-214). Odense: Odense University Press.
Preston, S.H., Elo, I.T. & Stewart, Q. (1999). Effects of age misreporting on mortality estimates at older ages. *Population Studies*, Vol. 53, No. 2, pp. 165-177.
Robine, J.-M. & J. W. Vaupel (2001). Supercentenarians: slower ageing individuals or senile elderly? *Experimental Gerontology*, Vol. 36, pp. 915-930.
Rosenwaike, I. (1979). A new evaluation of United States census data on the extreme aged. *Demography*, Vol. 16 (2), pp. 279-288.
Skytthe, A., Hauge, G. & Jeune, B. (1999). Age Validation of Reported Centenarians before 1900 in Denmark. In: B. Jeune and J. Vaupel (Eds). *Monographs on Population Aging, Vol.6* (pp. 79-96). Odense: Odense University Press.
Skytthe, A. & Jeune, B. (1995). Danish centenarians after 1800. In: B. Jeune and J. Vaupel (Eds). *Monographs on Population Aging. Vol. 2* (pp. 55-66). Odense: Odense University Press.
Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society Series A*. Vol. 162, pp. 5-43.
Thatcher, A. R., Kannisto, V., Vaupel, J. W. (1998). The force of mortality at ages 80 to 120. *Monographs on Population Aging. Vol. 5*. Odense: Odense University Press.
Thatcher, A.R. (1992). Trends in numbers and mortality at high ages in England and Wales. *Population Studies*, Vol. 46 (3), pp. 411-426.
Thatcher, A.R. (2001). The demography of centenarians in England and Wales. *Population: An English Selection*, Vol. 13, No. 1, pp. 139-156.
Thatcher, A.R., Kannisto, V. & Andreev, K. (2002). The survivor ratio method for estimating numbers at high ages. *Demographic Research,* Vol. 6 (1), pp. 1-16.
Vallin, J. (1973). *La Mortalité par Génération en France. depuis 1899*. Paris: Presses Universitaires de France.
Vaupel, J.W., Wang, Z., Andreev, K.F. & Yashin, A.I. (1997). Population data at a glance: shaded contour maps of demographic surfaces over age and time. *Monographs on Population Aging. Vol.4* (1-98 p.). Odense: Odense University Press.
Vaupel, J.W. Rau, R., Camarda, C.G., von Kistowski, K. (2006): Can Heterogeneity of Populations Explain Differences in Mortality? *Center for Retirement Research at Boston College Working Paper WP #2006-10* (44 p.). Chestnut Hill, MA (USA): Center for Retirement Research, Boston College.
Vincent, P. (1951). La Mortalité des vieillards. *Population,* Vol. 6, No. 2, pp. 181-204.
Wang, Z., Zeng, Y., Jeune, B., & Vaupel, J. (1999). Age Validation of Han Chinese Centenarians. In: B. Jeune & J. Vaupel (Eds). *Monographs on Population Aging. Vol. 6* (pp. 195-214). Odense: Odense University Press.
Wilmoth, J.R., Andreev, K., Jdanov, D. & Glei, D. (2005). Methods protocol for the Human Mortality Database. *The Human Mortality Database*, 81 p. (retrieved from http://www.mortality.org/Public/Docs/MethodsProtocol.pdf on 07.06.2006).

**Figure 1. Survival probabilities between ages 10 and 55 and corresponding values of life expectancy at age 80**



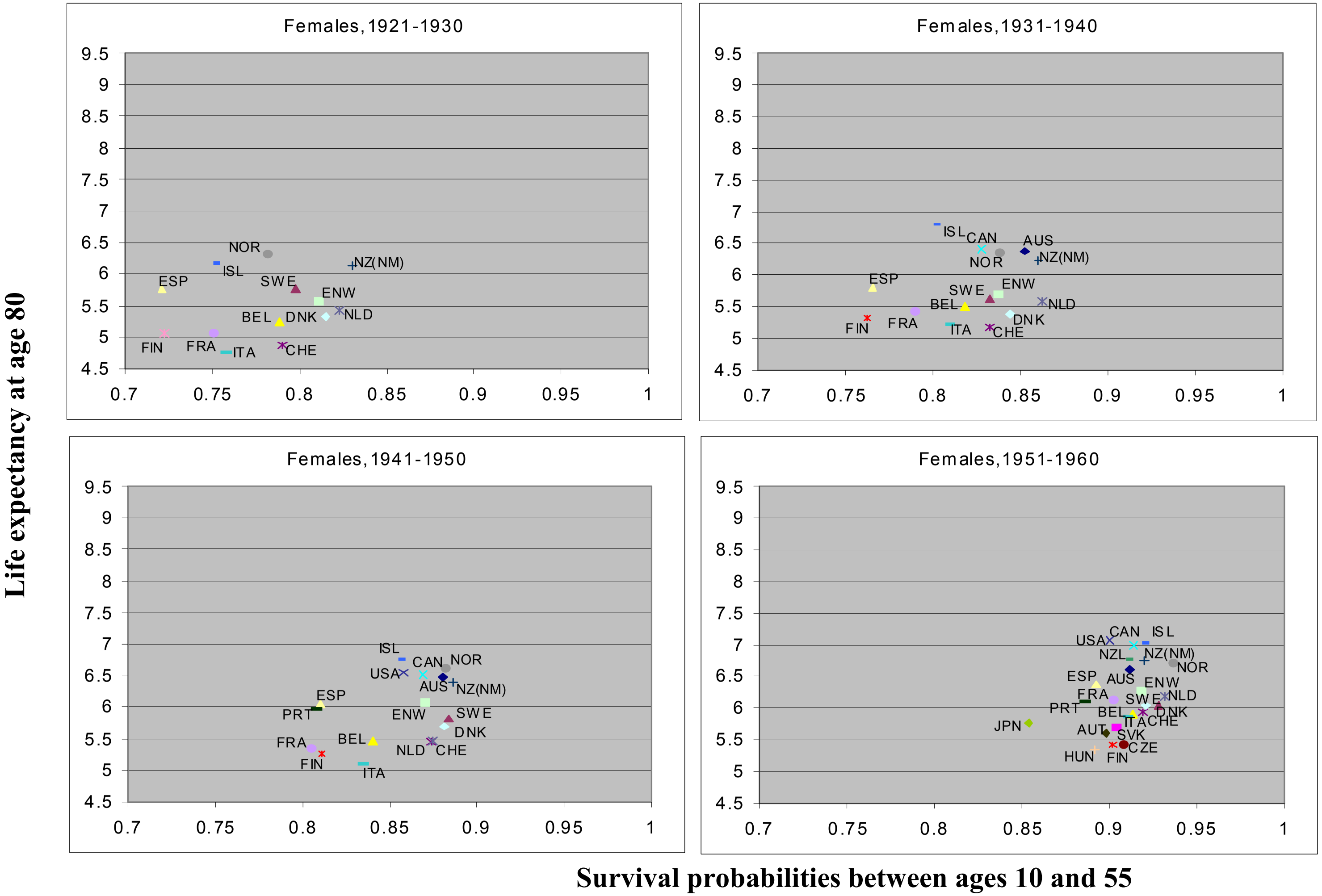

Note: Country-specific labels refer to the standard UN numeric code (ISO 3166-1 numeric-3) (http://www.nationsonline.org/oneworld/countrycodes.htm).

**Figure 1 (continued). Survival probabilities between ages 10 and 55 and corresponding values of life expectancy at age 80**



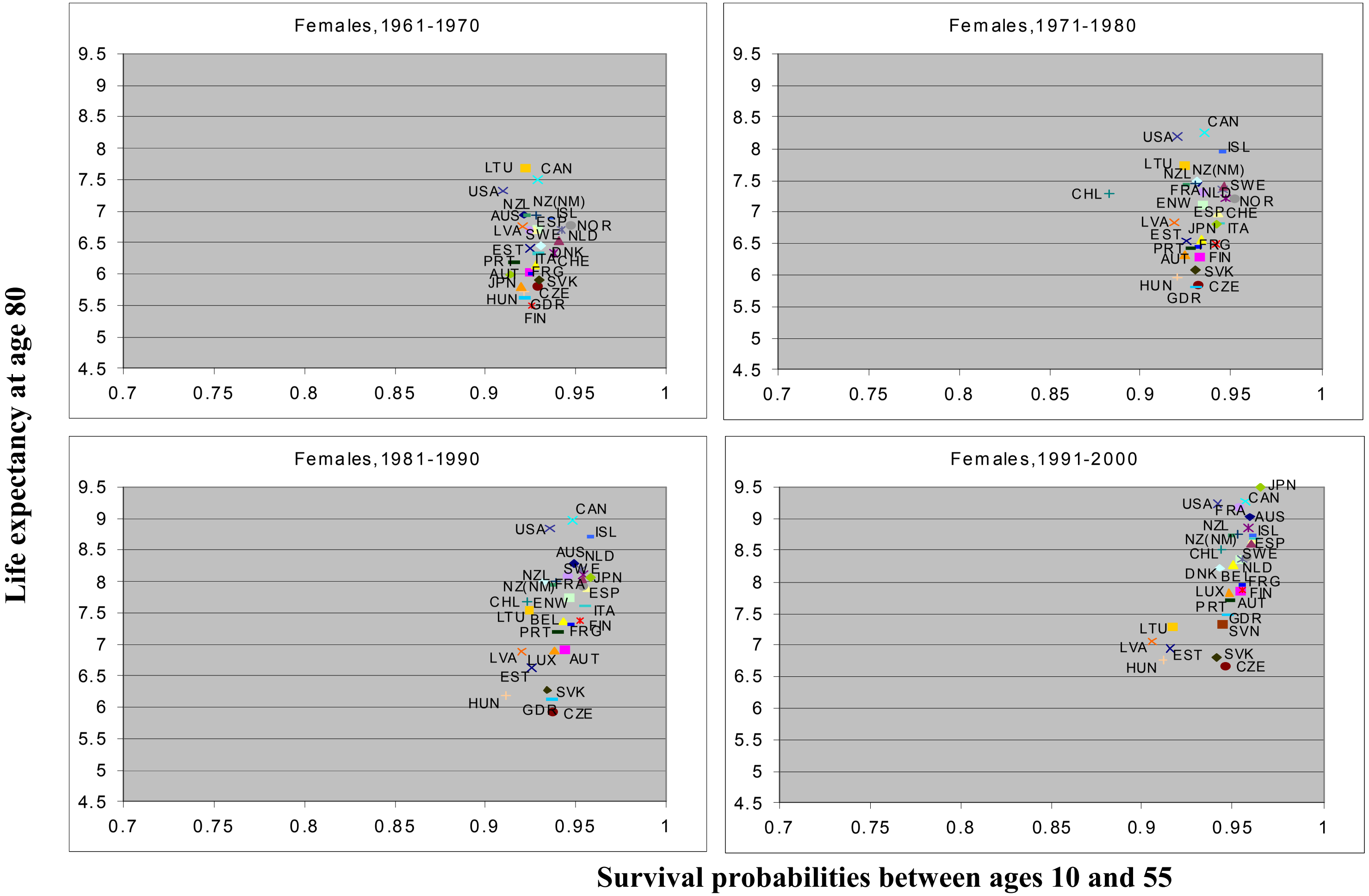

Note: Country-specific labels refer to the standard UN numeric code (ISO 3166-1 numeric-3) (http://www.nationsonline.org/oneworld/countrycodes.htm).

**Figure 1 (continued). Survival probabilities between ages 10 and 55 and corresponding values of life expectancy at age 80**




Note: Country-specific labels refer to the standard UN numeric code (ISO 3166-1 numeric-3) (http://www.nationsonline.org/oneworld/countrycodes.htm).

**Figure 1 (continued). Survival probabilities between ages 10 and 55 and corresponding values of life expectancy at age 80**




Note: Country-specific labels refer to the standard UN numeric code (ISO 3166-1 numeric-3) (http://www.nationsonline.org/oneworld/countrycodes.htm).

**Table 1. Outcomes of the OLS regression connecting the relative differences between the HMD and official estimates to the sex, and age group of the data by 10 year periods since 1900 (coefficient and standard error; statistically insignificant values with p>0.1 are in italic; points mean that data are not available).**

| | **1900-10** | **1911-20** | **1921-30** | **1931-40** | **1941-50** |
|---|---|---|---|---|---|
| Males | 0 | 0 | 0 | 0 | 0 |
| Females | -3.93 (2.04) | -5.57 (2.11) | *-1.06 (1.41)* | -4.13 (1.74) | -33.54 (9.69) |
| age 80-84 | 0 | 0 | 0 | 0 | 0 |
| age 85-89 | 5.17 (2.76) | *3.67 (2.72)* | 3.28 (1.84) | 4.73 (2.27) | *3.75 (12.75)* |
| age 90-94 | 12.79 (2.76) | 10.69 (2.99) | 4.33 (2.01) | 9.73 (2.44) | *18.14 (13.36)* |
| age 95+ | 46.81 (3.2) | 48.39 (3.43) | 32.98 (2.54) | 33.04 (3.21) | 134.44 (15.64) |
| Constant | -3.63 (2.66) | -4.95 (2.84) | -7.2 (2.05) | -4.64 (2.6) | *-17.79 (16.37)* |

| | **1951-60** | **1961-70** | **1971-80** | **1981-90** | **1991-2000** |
|---|---|---|---|---|---|
| Males | 0 | 0 | 0 | 0 | 0 |
| Females | *2.44 (1.98)* | *-0.1 (1.06)* | -1.62 (0.31) | -1.42 (0.41) | -1.63 (0.42) |
| age 80-84 | 0 | 0 | 0 | 0 | 0 |
| age 85-89 | *1.57 (2.60)* | *0.55 (1.38)* | *0.54 (0.41)* | *0.68 (0.56)* | *0.54 (0.57)* |
| age 90-94 | 4.89 (2.76) | 4.03 (1.50) | 3.00 (0.44) | 2.19 (0.57) | 2.05 (0.58) |
| age 95+ | 35.17 (3.12) | 22.94 (1.68) | 11.63 (0.49) | 12.51 (0.62) | 10.1 (0.63) |
| Constant | -7.9 (4.14) | -5.02 (2.43) | -2.31 (0.76) | -2.83 (1.03) | -2.13 (1.09) |