PAA paper for Session 130:
Methodological Issues in Health and Mortality

Mortality Measurement at Advanced Ages

Leonid A. Gavrilov, Natalia Gavrilova,
Center on Aging, NORC and the University of Chicago

Abstract

Estimation of hazard rates at extremely old ages poses serious challenges to researchers. Using data from the Social Security Administration Death Master File allowed us to estimate hazard rates for a set of very large single-year extinct birth cohorts (1884-1889) on a monthly basis. We found that mortality at advanced ages follows the Gompertz law up to the ages 102-105 years without a significant deceleration. Additional measures of quality control demonstrated that population subgroups with better age reporting do not show mortality deceleration (deviation of mortality from the Gompertz law). Earlier reports of mortality deceleration at ages below 100 appear to be artifacts of mixing together several birth cohorts with different mortality levels and using cross-sectional instead of cohort data.

Introduction

Accurate estimates of mortality at advanced ages are essential to improving forecasts of mortality and the population size of the oldest old age group. It is now considered as an established fact that mortality at advanced ages has a tendency to deviate from the Gompertz law, so that the logistic model often is used to fit human mortality (Horiuchi, Wilmoth, 1998). The estimates of mortality force at extreme ages are difficult because of small numbers of survivors to these ages in most countries. Data for extremely long-lived individuals are scarce and subjected to age exaggeration. Traditional demographic estimates of mortality based on period data encounter well known denominator problem. More accurate estimates are obtained using the method of extinct generations (Vincent, 1951). In order to obtain good quality estimates of mortality at advanced ages researches are forced to pool data for several calendar periods. Single-year life tables for many countries have very small numbers of survivors to age 100 that makes estimates of mortality at advanced ages unreliable. The aggregation of deaths for several calendar periods however creates a heterogeneous mixture of cases from different birth cohorts. Mortality deceleration observed in these data might be an artifact of data heterogeneity. In addition to that, many assumptions about distribution of deaths in the age/time interval used in mortality estimation are not valid for extreme old ages when mortality is particularly high and grows rapidly.

Thus the estimation of hazard rates at extremely old ages poses several serious challenges to researchers:

(1) The observed mortality deceleration may be at least partially an artifact of mixing different birth cohorts with different mortality (heterogeneity effect);

(2) Standard assumptions of hazard rate estimates may be invalid when risk of death is extremely high at old ages;

(3) Ages of very old people may be exaggerated.

One way of obtaining estimates of mortality at extreme ages is to pool together international records of persons surviving to extreme ages with subsequent efforts of strict age validation (Robine and Vaupel 2001; Robine, Cournil et al. 2005). This approach helps to resolve the third problem mentioned above but does not allow researchers to resolve the first two problems because of inevitable data heterogeneity when data for people belonging to different birth cohorts and countries are pooled together. In this paper we propose an alternative approach, which allows us to resolve partially the first two problems. This approach is based on using data from the Social Security Administration Death Master File (DMF), which allows compiling data for large single-year birth cohorts. Some birth cohorts covered by DMF could be studied by the method of extinct generations. Availability of month of birth and month of death information provides a unique opportunity to obtain hazard rate estimates for every month of age. Possible ways of resolving the third problem of hazard rate estimation are also elaborated.

Hazard rate estimation at advanced ages

A conventional way to obtain estimates of mortality at advanced ages is a construction of demographic life table with probability of death (q_x) as one of important life table functions. Although probability of death is a useful indicator for mortality studies, it may be not the most convenient one for studies of mortality at advanced ages. First, the values of q_x depend on the length of the age interval Δx for which it is calculated. This hampers both analyses and interpretation. For example, if one-day probability of death follows the Gompertz law of mortality, probability of death calculated for other age interval does not follow this law (see Gavrilov and Gavrilova, 1991 and Le Bras, 1976). Thus it turns out that the shape of age-dependence for q_x depends on the arbitrary choice of age interval. Also, by definition q_x is bounded by unity, which makes it difficult to study mortality at advanced ages.

More useful indicator for mortality studies at advanced age is instantaneous mortality rate or hazard rate, μ_x which is defined as follows:

$$\mu_x = -\frac{dN_x}{N_x dx}$$

where N_x is a number of living individuals at age x.

Hazard rate does not depend on the length of the age interval (it is measured at the instant of time x), has no upper boundary and has a dimension of rate ($time^{-1}$). It should also be noted that the famous law of mortality, the Gompertz law, was proposed for fitting the hazard rate rather than probability of death (Gompertz, 1825).

The empirical estimates of hazard rates are often based on suggestion that age-specific mortality rate or death rate (number of deaths divided by exposure) is a good estimate of theoretical hazard rate. One of the first empirical estimates of hazard rate was proposed by George Sacher (Sacher, 1956; 1966):

$$\mu_x = \frac{1}{\Delta x}(ln\ l_{x-\frac{\Delta x}{2}} - ln\ l_{x+\frac{\Delta x}{2}}) = \frac{1}{2\Delta x}\ ln\frac{l_{x-\Delta x}}{l_{x+\Delta x}}$$

This estimate is unbiased for slow changes in hazard rate if $\Delta x \Delta \mu_x << 1$ (Sacher, 1966).

A simplified version of Sacher estimate (for small age intervals equal to unity) often is used in biological studies of mortality: $\mu_x = -\ln(1-q_x)$. This estimate is based on the assumption that hazard rate is constant over age interval.

At advanced ages when death rates are very high, the assumptions about small changes in hazard rate or a constant hazard rate within the age interval becomes questionable. Violation of these assumptions may lead to biased estimates of hazard rates calculated on annual basis. Fortunately narrowing of the age interval for hazard rate estimation from one-year to one-month might help to resolve this problem.

Social Security Administration Death Master File as a source of mortality data for advanced ages.

Social Security Administration Death Master File (DMF) is a publicly available data source that allows a search for individuals using various search criteria: birth date, death date, first and last names, social security number, place of last residence, etc. This resource covers individual deaths that occurred in the period 1937-2007 (see Faig, 2002 for more details). Many researchers suggest that the quality of SSA/Medicare data is superior to vital statistics records because of strict evidentiary requirements in application for Medicare while age reporting in death certificates is made by proxy informant (Kestenbaum, 1992; Kestenbaum, Ferguson, 2002; Rosenwaike et al., 1998; Rosenwaike, Stone, 2003).

Social Security Administration Death Master File (DMF) was used in the study of mortality kinetics after ages 85-90 years. The advantage of this data source is that some birth cohorts covered by DMF could be studied by the method of extinct generations (Vincent 1951; Kannisto 1988; Kannisto 1994). Availability of month of birth and month of death information provides a unique opportunity to obtain hazard rate estimates for every month of age, which is important given extremely high mortality after age 100 years (see Table 1).

The information from the DMF was collected for individuals who lived 90 years and over and died before 2007. DMF database is unique because it represents mortality experience for very large birth cohorts of the oldest-old persons. In this study mortality measurements were made for cohorts, which are more homogeneous in respect to the period of birth and historical life course experiences.

Table 1. Information available in the SSA Death Master File.

1. first, last names, SSN
2. date, month, year of birth
3. month, year of death
4. state of the SSN issuance
5. town, county, state, zip code of the last residence
6. death date verification code

The DMF collects deaths for persons who receive SSA benefits and currently covers over 90 percent of deaths occurring in the United States (Faig, 2002) and 93 percent to 96 percent of deaths of individuals aged 65 or older (Hill, Rosenwaike, 2001). Despite certain limitations, this data source allows researchers to obtain detailed estimates of mortality at advanced ages.

We already used this data resource for centenarians' age validation in the study of centenarian genealogies (Gavrilova, Gavrilov, 2007). This data resource is also useful in mortality estimates for several extinct or almost extinct birth cohorts in the United States.

Hazard rate estimates at advanced ages using data from the Social Security Death Master File

In this study we collected information from the DMF publicly available at Rootsweb.com. The total number of records collected over 9 million with more than 900,000 records for persons who lived 100 years and over. Several birth cohorts (those born before 1890) may be considered extinct or almost extinct, so it is possible to apply the method of extinct generations (Vincent, 1951) and estimate mortality kinetics at very advanced ages up to 115-120 years.

We obtained data for persons who died before 2008 and were born in 1875-1887. There were no persons born before 1882 and only a few persons born in 1883-1887 who died in 2006 or 2007. Thus, the 1875-1887 birth cohorts in this sample may be considered extinct or almost extinct. Assuming that the number of living persons belonging to these birth cohorts in 2007 is close to zero, it is possible to construct a cohort life table using the method of extinct generations, which was suggested and explained by Vincent (1951) and developed further by Kannisto (1994). In the first stage of our analyses we calculated an individual life span in completed months:

Lifespan in months = (death year – birth year) x 12 + death month – birth month

Then it is possible to estimate the hazard rate at each month of age using standard methods of survival analysis. All calculations were done using the Stata statistical package, procedures 'stset' and 'sts' (Stata Corp, 2005). This software provides nonparametric estimates of major survival functions including the Nelson-Aalen estimates of hazard rate (force of mortality). Note that a hazard rate, in contrast to a probability of death, q(x), has a dimension of time frequency, because of the time interval in the denominator (reciprocal time, time^{-1}). Thus the values of hazard rates depend on the chosen units of time measurement (day^{-1}, month^{-1} or year^{-1}). In this study survival times were measured in months, so the estimates of hazard rates initially had a dimension of month^{-1}. For the purpose of comparability with other published studies, which typically use the year^{-1} time scale, we transformed the monthly hazard rates to the more conventional units of year^{-1}, by multiplying these estimates by a factor of 12 (one month in the denominator of hazard rate formula is equal to 1/12 year). Also note that a hazard rate, in contrast to a probability of death can be greater than 1, and therefore its logarithm can be greater than 0 (and we indeed observed this at extreme old ages in some rare cases as will be described later). We estimated hazard rates for four single-year birth cohorts—those born in 1884-1891.

Results of the hazard rate estimates for 1885 birth cohort is presented in Figure 1.

Figure 1 About Here

Measures to improve the data quality.

Recent study of age validation among supercentenarians (Rosenwaike, Stone, 2003) showed that age reporting among supercentenarians in SSA database is rather accurate with exception of persons born in the Southern states. In order to improve the quality of our dataset when estimating mortality rates, we excluded records for those persons who applied for social security number in the Southeast (AR, AL, GA, MS, LA, TN, FL, KY, SC, NC, VA, WV)and Southwest (AZ, NM, TX, OK) regions, Puerto Rico and Hawaii. This step of data cleaning however did not change significantly the overall trajectory of mortality at advanced ages, but decreased the number of too low mortality estimates and increased the number of higher mortality estimates after age 105 years (see Figure 2).

Figure 2 About Here

These figure demonstrates that for single-year birth cohort mortality agrees well with the Gompertz law up to very advanced ages. Previous studies of mortality at advanced ages used aggregated data combining several birth cohorts with different mortality and this aggregation apparently resulted in early mortality deceleration and subsequent leveling-off as it was demonstrated by heterogeneity model (Beard, 1971). Mortality deceleration and even decline of mortality often is observed for data with low quality. On the other hand, improvement of data quality results in straighter mortality trajectory in semi-log scale (Kestenbaum, Ferguson, 2002). In our study population of non-Southern states with presumably better data quality (because of more accurate age registration) demonstrates starter mortality dependence on age in semilog scale (verified using quadratic fit) (Figure 3).

Figure 3 About Here

In our study more recent 1889 birth cohort demonstrates straighter trajectory and lower statistical noise after age 105 than older 1884 one (see Figure 4). Thus, we may expect that cohorts born after 1891 would demonstrate even better fit by the Gompertz model than the older ones because of improved quality of age reporting. Testing this hypothesis now is hampered by the problem of data truncation for non-extinct birth cohorts.

Conclusion

We may conclude that populations with better quality of age reporting demonstrate mortality trajectories at advance ages more close to the Gompertz law than populations with lower quality of data. Earlier reports of mortality deceleration (deviation of mortality from the Gompertz law) at ages below 100 appear to be artifacts of mixing together several birth cohorts with different mortality levels and using cross-sectional instead of cohort data.

Previous studies showed that the period of mortality deceleration in mammals is very short (Sacher, 1966, Lindop, 1961) compared to lower organisms (Vaupel et al., 1998; Gavrilov, Gavrilova, 2006). It appears to be relatively short in humans too. This observation agrees well with the prediction of the reliability theory of aging that more complex living systems/organisms with many vital subsystems (like mammals) may experience very short or no period of mortality plateau at advance ages in contrast to more simple living organisms (Gavrilov, Gavrilova, 1991; 2001; 2006).

Acknowledgements

This study was supported by the National Institute on Aging (R01 AG028620 grant).

References

Allison P. (1995). Survival Analysis Using the SAS ® System: A Practical Guide. SAS Institute.

Austad, S. N. (2001). Concepts and theories of aging. In E. J. Masoro and S. N. Austad. *Handbook of the biology of aging*. San Diego, CA: Academic Press, 3-22.

Beard, R. E. (1959). Note on some mathematical mortality models. In: G. E. W. Wolstenholme & M.O'Connor (Eds.), *The lifespan of animals* (pp. 302-311). Boston: Little, Brown.

Beard, R. E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. In. W. Brass (Ed.), *Biological aspects of demography* (pp. 57-68), London: Taylor & Francis.

Costa, D.L. and Lahey, J. (2003). Becoming Oldest-Old: Evidence from Historical U.S. Data. *NBER Working Paper* No. W9933. http://ssrn.com/abstract=439614

Depoid F. (1973). La mortalite des grands viellards. *Population*, 28: 755-92.

Doblhammer, G. (1999). Longevity and month of birth: evidence from Austria and Denmark. *Demographic Research* [Online] 1, 1-22. Available: http://www.demographic-research.org/Volumes/Vol1/3/default.htm

Doblhammer, G. (2003). The late life legacy of very early life. Rostock, *MPIDR Working Paper WP-2003-030*.

Doblhammer G, Vaupel JW (2001). Lifespan depends on month of birth. *Proc. Natl. Acad. USA* 98: 2934-2939.

Economos, A.C. (1979). A non-gompertzian paradigm for mortality kinetics of metazoan animals and failure kinetics of manufactured products. *AGE*, 2, 74-76.

Economos, A.C. (1980). Kinetics of metazoan mortality. *J. Social Biol. Struct.*, 3, 317-329.

Economos, A.C. (1983). Rate of aging, rate of dying and the mechanism of mortality. *Arch. Gerontol. and Geriatrics,* 1, 3-27.

Economos, A.C. (1985). Rate of aging, rate of dying and non-Gompertzian mortality - encore… *Gerontology*, 31, 106-111.

Faig K. (2002). Reported deaths of centenarians and near-centenarians in the U.S. Social Security Administration's Death Master File. In: Proceedings of the Society of Actuaries "Living to 100 and Beyond International Symposium", Orlando, FL.

Fukui, H. H., Ackert, L., & Curtsinger, J. W. (1996). Deceleration of age-specific mortality rates in chromosomal homozygotes and heterozygotes of Drosophila melanogaster. *Experimental Gerontology*, 31, 517-531.

Fukui, H. H., Xiu, L., & Curtsinger, J. W. (1993). Slowing of age-specific mortality rates in Drosophila melanogaster. *Experimental Gerontology*, 28, 585-599.

Gavrilov, L.A. (1980). Study of life span genetics using the kinetic analysis. Ph.D. Thesis, Moscow, Russia: Moscow State University.

Gavrilov, L. A. (1984). Does a limit of the life span really exist? *Biofizika*, 29, 908-911.

Gavrilov, L.A., Gavrilova, N.S. (1991). *The Biology of Life Span: A Quantitative Approach*, Harwood Academic Publisher, New York.

Gavrilov, L.A. & Gavrilova, N.S., 1999. Season of birth and human longevity. *Journal of Anti-Aging Medicine* 2: 365-366.

Gavrilov, L.A. & Gavrilova, N.S. (2001). The reliability theory of aging and longevity. *J. Theor. Biol.* 213: 527-545.

Gavrilov LA, Gavrilova NS. (2002). Early-Life Seasonal Programming of Adult Lifespan: Evidence from the 19th century birth cohorts. *Annual Meeting of the Social Science History Association,* St. Louis, 24-27 October 2002.

Available at: http://www.ssha.org/abstract2002/abs348.html

Gavrilov L.A. & Gavrilova N.S. (2003a). The quest for a general theory of aging and longevity. *Science's SAGE KE (Science of Aging Knowledge Environment)* for 16 July 2003; Vol. 2003, No. 28, 1-10.

Gavrilov, L.A., Gavrilova, N.S. (2003b). Early-life factors modulating lifespan. In: Rattan, S.I.S. (Ed.). *Modulating Aging and Longevity*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 27-50.

Gavrilov LA, Gavrilova NS. (2004). Early-Life Programming of Aging and Longevity: The Idea of High Initial Damage Load (the HIDL Hypothesis). *Annals of the New York Academy of Sciences*, 1019: 496-501.

Gavrilov LA, Gavrilova NS. Reliability Theory of Aging and Longevity. In: Masoro E.J. & Austad S.N.. (eds.): *Handbook of the Biology of Aging*, Sixth Edition. Academic Press. San Diego, CA, USA, 2006, 3-42.

Gavrilova, N. S., Gavrilov, L. A., (2005). Search for predictors of exceptional human longevity: Using computerized genealogies and Internet resources for human longevity studies. In: Proceedings of the Society of Actuaries “Living to 100 and Beyond International Symposium”, Orlando, FL.

Gavrilova, N. S., Gavrilov, L. A., Evdokushkina, G. N., Semyonova, V. G. (2003). Early-life predictors of human longevity: Analysis of the 19th Century birth cohorts. *Annales de Demographie Historique*, 2, 177-198.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality and on a new mode of determining life contingencies. Philos.Trans.Roy.Soc.London A, **115:** 513-585.

Greenwood, M. & Irwin, J. O. (1939). The biostatistics of senility. *Hum. Biol.,* 11, 1-23.

Hill M.E., Rosenwaike I. (2001). The Social Security Administration's Death Master File: the completeness of death reporting at older ages. *Soc Secur Bull*. 64: 45-51.

Horiuchi S, Wilmoth JR. (1998). Deceleration in the age pattern of mortality at older ages. *Demography*, 35: 391-412.

Kestenbaum B. (1992). A description of the extreme aged population based on improved Medicare enrollment data. *Demography*, 29: 565-80.

Kestenbaum B., Ferguson B.R. (2002). Mortality of the extreme aged in the United States in the 1990s, based on improved Medicare data. In: Proceedings of the Society of Actuaries “Living to 100 and Beyond International Symposium”, Orlando, FL.

Laake K., Sverre JM (1996). Winter excess mortality: a comparison between Norway and England plus Wales. Age and Ageing, 25: 343-348.

Le Bras, H. (1976). Lois de mortalité et age limité. *Population*, 31, 655-692.

Lindop, P.J. (1961). Growth rate, lifespan and causes of death in SAS/4 mice. *Gerontologia*, 5: 193-208.

Olshansky, S. J. (1998). On the biodemography of aging: a review essay. *Population and Development Review*, 24, 381-393.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63, 12-57.

Robine J-M, Vaupel J. W. (2001). Supercentenarians: Slower ageing individuals or senile elderly? Experimental Gerontology, 36: 915-930.

Rosenwaike I., Logue B. (1983). Accuracy of death certificate ages for the extreme aged. *Demography*, 20: 569-85.

Rosenwaike I., Hill M., Preston S., Elo I. (1998). Linking death certificates to early census records: the African American Matched Records Sample*. Historical Methods*, 31: 65-74.

Rosenwaike I., Stone L.F. (2003). Verification of the ages of supercentenarians in the United States: Results of a matching study. *Demography,* 40: 727-739.

Sacher, G. A. (1956). On the statistical nature of mortality, with especial reference to chronic radiation mortality. *Radiology*, 67: 250-257.

Sacher, G.A. (1966). The Gompertz transformation in the study of the injury-mortality relationship: Application to late radiation effects and ageing. In P. J. Lindop and G. A. Sacher (eds.) *Radiation and ageing* (pp. 411-441). Taylor and Francis, London.

Shrestha L.B., Preston S. H. (1995). Consistency of census and vital registration data on older Americans: 1970-1990. *Survey Methodology*, 21: 167-177.

Tatar, M., Carey, J. R., & Vaupel, J. W. (1993). Long-term cost of reproduction with and without accelerated senescence in *Callosobruchus maculatus*: Analysis of age-specific mortality. Evolution, 47, 1302-1312.

Thatcher A.R. (1999). The long-term pattern of adult mortality and the highest attained age. *J. R. Statist. Soc.* A, 162, Part 1, 5-43.

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

Vincent P. (1951). La mortalite des viellards. *Population*, 6, 181-204.

Wilkinson P., Pattenden S., Armstrong B., Fletcher A., Kovats R.S., Mangtani P., McMichael A.J. (2004). Vulnerability to winter mortality in elderly people in Britain: population based study. BMJ, 329: 647.

Wilmoth, J. R. (1997). In search of limits. In K. W. Wachter & C. E. Finch (Eds.), Between Zeus and the salmon. The biodemography of longevity (pp. 38-64). Washington, DC: National Academy Press.

Figures

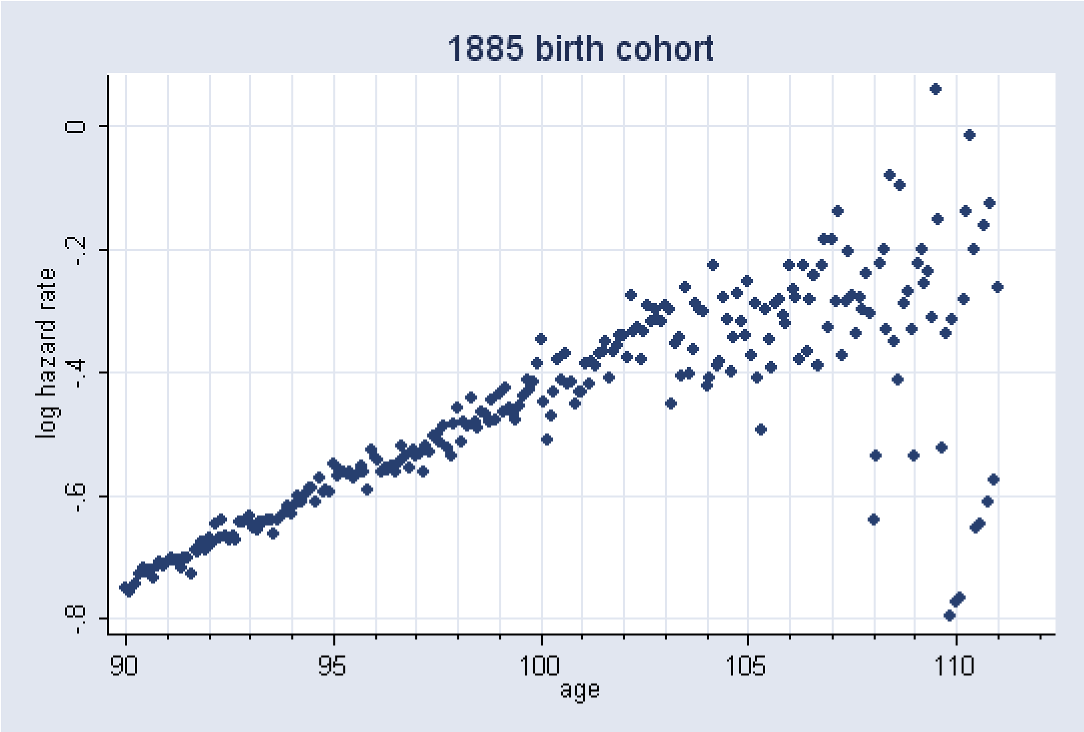

Figure 1.

Hazard rate (per year) for 1885 birth cohort.

Data from the Social Security Administration Death Master File.

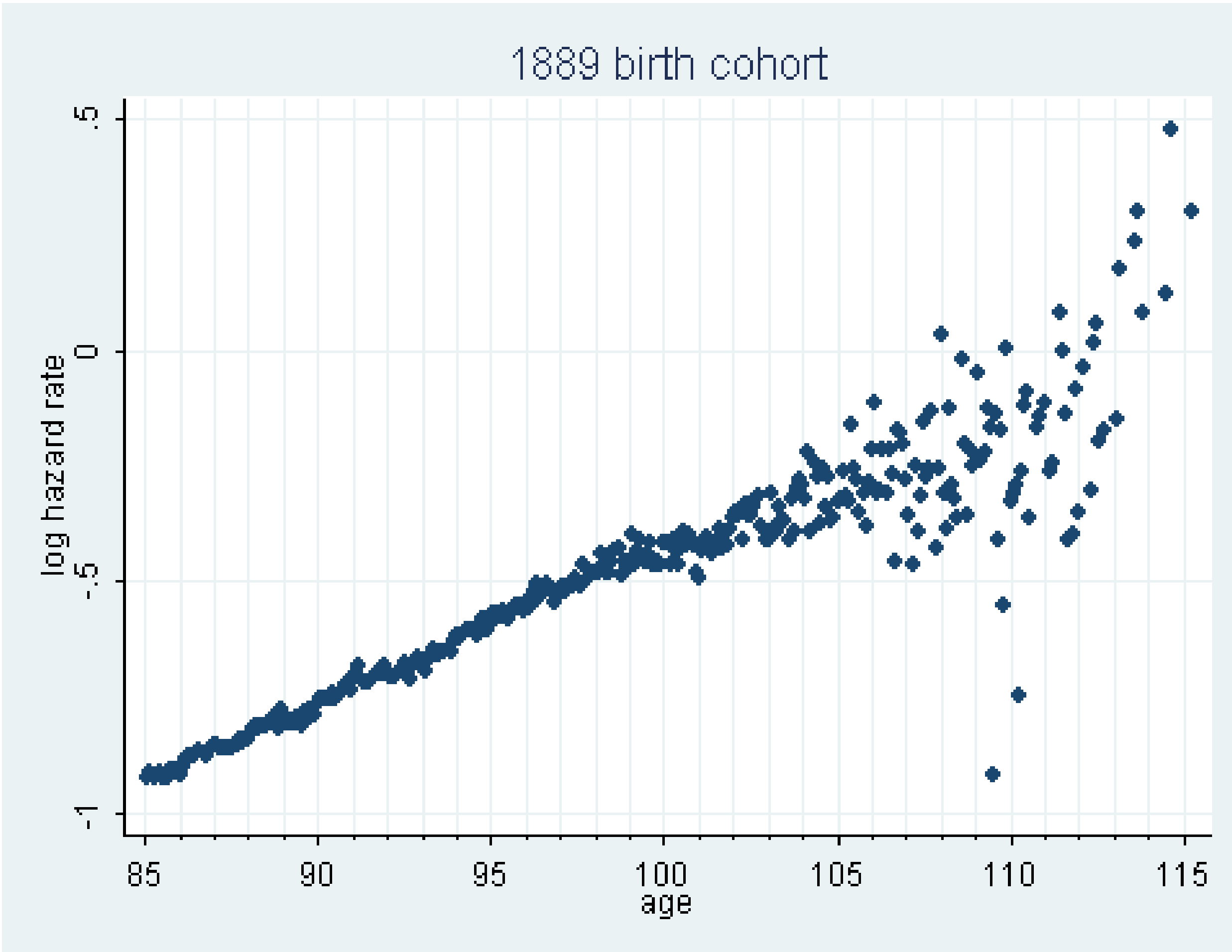

Figure 3.

Hazard rate (per year) for 1889 birth cohort. All data are included.

Data from the Social Security Administration Death Master File.

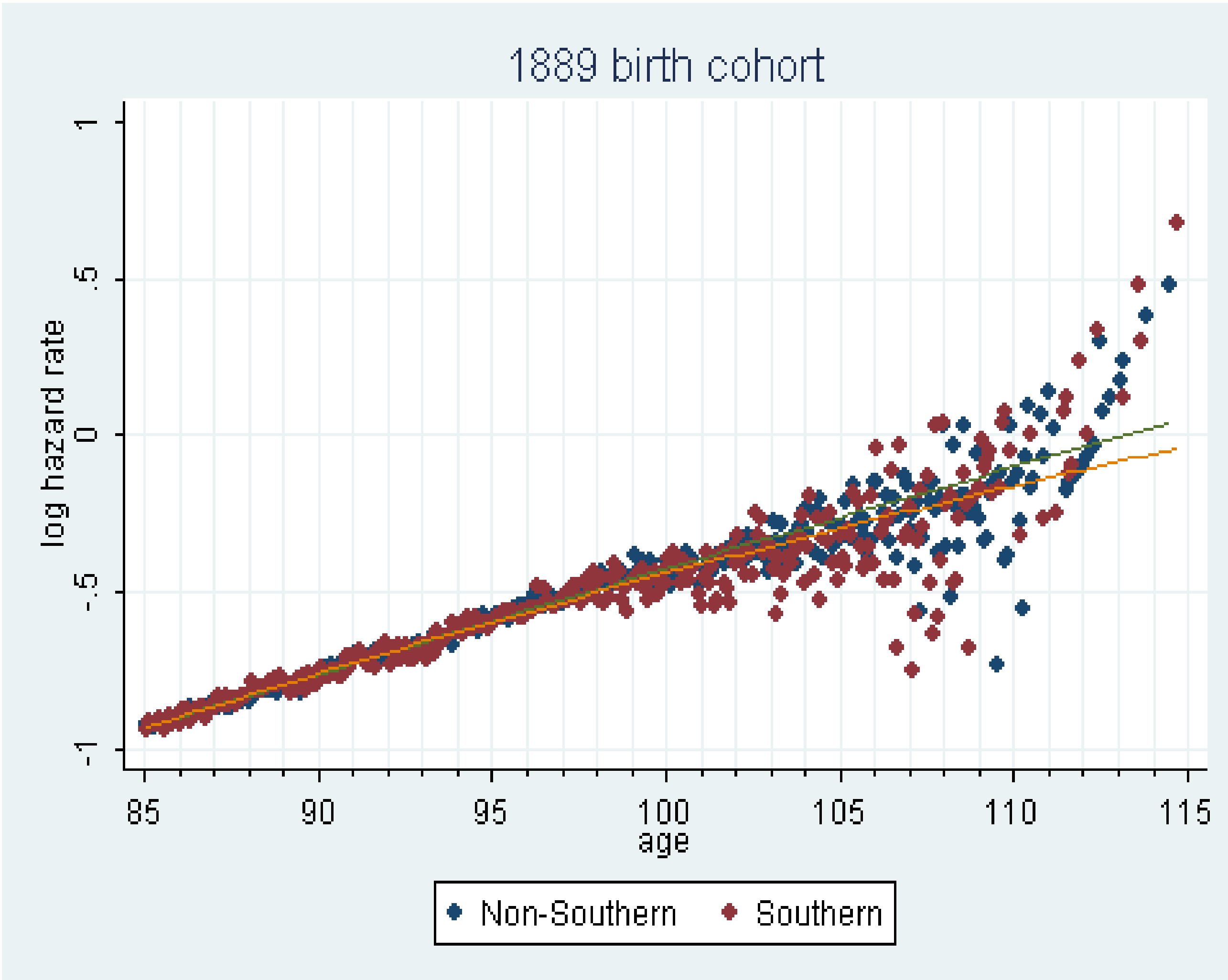

Figure 4. Mortality data with presumably different quality.

Hazard rate (per year) for 1891 birth cohort. Comparison of less reliable (Southern states, Puerto Rico and Hawaii) and more reliable (non-Southern states) data. Solid lines show fit with quadratic function.

Data from the Social Security Administration Death Master File.

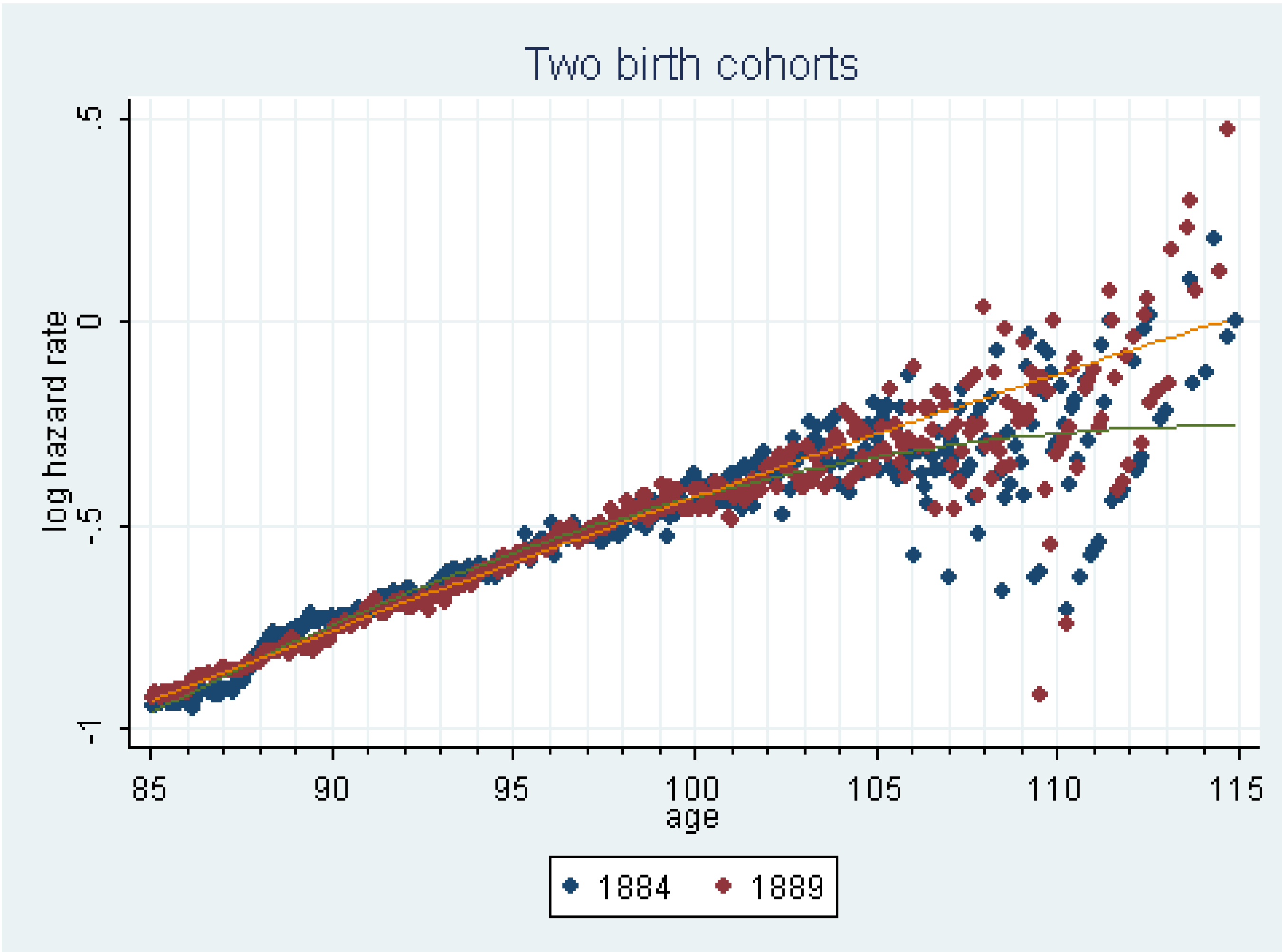

Figure 4. Mortality data with presumably different quality.

Hazard rate (per year) for 1891 birth cohort. Comparison of less reliable (Southern states, Puerto Rico and Hawaii) and more reliable (non-Southern states) data. Solid lines show fit with quadratic function.

Data from the Social Security Administration Death Master File.