DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 29, ARTICLE 19, PAGES 507-520
PUBLISHED 17 SEPTEMBER 2013
http://www.demographic-research.org/Volumes/Vol29/19/
DOI: 10.4054/DemRes.2013.29.19

Descriptive Finding

Minor gradient in mortality by education at the highest ages: An application of the Extinct-Cohort method

Roland Rau

Magdalena M. Muszyńska

Paul H.C. Eilers

Table of Contents

Demographic Research: Volume 29, Article 19
Descriptive Finding

Minor gradient in mortality by education at the highest ages: An application of the Extinct-Cohort method

Roland Rau [1]

Magdalena M. Muszyńska [2]

Paul H.C. Eilers [3]

Abstract

BACKGROUND
Socioeconomic mortality differentials are known to exist almost universally. Many studies show a trend towards convergence with increasing age. Information about the highest ages is very rare, though.

OBJECTIVE
We want to find out whether socioeconomic factors determine the chance of death in the United States among the oldest people.

METHODS
Based on official death count records, we employ the extinct cohort method to estimate the age-specific probability of dying by level of education.

RESULTS
We present evidence that socioeconomic differentials in mortality exist even at the highest ages (95+), although the gap is small.

[1] University of Rostock, Department of Sociology & Demography, Rostock, Germany and Max Planck Institute for Demographic Research, Rostock, Germany. E-mail: roland.rau@uni-rostock.de.
[2] Warsaw School of Economics, Warsaw, Poland, E-mail: mmuszyns@sgh.waw.pl.
[3] Erasmus Medical Center, Rotterdam, The Netherlands. E-mail: p.eilers@erasmusmc.nl.

1. Introduction

Few relationships in demography are as well established as the one between socioeconomic status (SES) and mortality: people with lower SES suffer from higher mortality (and vice versa). This finding is rather universal, irrespective how SES is measured (e.g., income, education, occupation) or for which countries the study is conducted (e.g., Kitagawa and Hauser 1973; Rogers, Hummer, and Nam 1995; Kunst 1997; Hummer, Rogers, and Eberstein 1998; Pappas et al. 1993; Mackenbach et al. 1999; Goldman 2001).[4] Many studies show that excess mortality of disadvantaged groups decreases with increasing age (e.g., Doblhammer, Rau, and Kytir 2005). This tendency towards convergence is at least partly caused by compositional changes as a result of mortality selection. Other explanations typically refer to policies, such as Medicare, that reduce inequalities or to the larger influence of biological factors at older ages (see, for instance, Hoffmann (2008) for an overview of causal explanations).

In this paper we analyze whether socioeconomic mortality differentials still exist at ages 95 and higher in the United States. Our goal is not to specify whether selection effects are stronger than policy or biological effects, or are potentially counteracted by an accumulation of detrimental effects acquired over the life-course (Ross and Wu 1996).

Despite the fact that the number of people at the highest ages is increasing rapidly (e.g. Rau et al., 2008), little is known about whether socioeconomic mortality differences still persist at those ages. Using data from the "National Long Term Care Survey" (NLTCS) of the United States, Manton, Stallard, and Corder (1997) found differences in remaining life expectancy at age 95 of less than one year between people with high and low education. More recently Zhu and Xie (2007) found that education — in conjunction with urbanity — plays a major role in Chinese mortality at ages 90+ and 100+ based on data from the Chinese Longitudinal Healthy Longevity Survey.

2. Data and method

For this study, we used the "Multiple Cause of Death Data" from the National Vital Statistics System of the National Center for Health Statistics of the United States. These data can be downloaded from the website of the National Bureau of Economic Research (National Center for Health Statistics 2013). The data are currently available for the years 1959–2010 and list individual deaths by a multitude of covariates such as sex, age at death, cause of death, year of death, or state of residence. Since 1989, these data also contain information on education, which we used as our indicator for socioeconomic sta-

[4] We would like to point out that the "universality" does not refer to calendar time, as recently shown by Bengtsson and van Poppel (2011).

tus. Besides this practical reason of data availability, education might be the best marker for SES as it can be expected to remain relatively stable throughout the adult life course, especially among the oldest-old. Additionally, from a causal perspective, it can be expected to predate other measures of SES (such as occupation or income), which might be difficult to measure at ages 90+ at all.

To estimate mortality from such right-truncated data, we employed the so-called Extinct-Cohort method pioneered by Vincent (1951) and Depoid (1973): Starting at the highest age the last surviving member of a birth cohort has reached, this method estimates the age-specific probability of dying, $q(x)$, backwards using the number of people who died at a given age as the numerator and the cumulative number of people who died at that age or older as the denominator. We expect the requirement of a stationary level of education to estimate mortality accurately to be met. Other requirements (Dinkel 1997) for this method are only partially fulfilled: Whereas death registration is virtually complete and international migration is negligible, deaths are not registered by age and year of birth. This problem is outlined in Figure 1: The area on a Lexis surface covered by deaths of a given year of birth and a given age at death is illustrated with a green trapezoid. Unfortunately, the data only provide information on deaths by year of death and age at death. We have exploited the additionally provided information on month of birth to approximate birth cohorts: The left red column represents deaths in January of the given year at the selected age. Assuming a uniform distribution of deaths within a month, most deaths in that column can be attributed to the older cohort as indicated by the part of that column above the intersecting 45-degree line. Only a small proportion belongs to the younger cohort below the diagonal. We attributed $11/12 + 0.5/12 = 11.5/12 \approx 95.83\%$ of deaths in January to the older cohort and $0.5/12 \approx 4.17\%$ to the younger cohort. In February, we partitioned death $10/12 + 0.5/12 = 10.5/12 = 87.5\%$ to the older cohort and 12.5% to the younger cohort, etc..

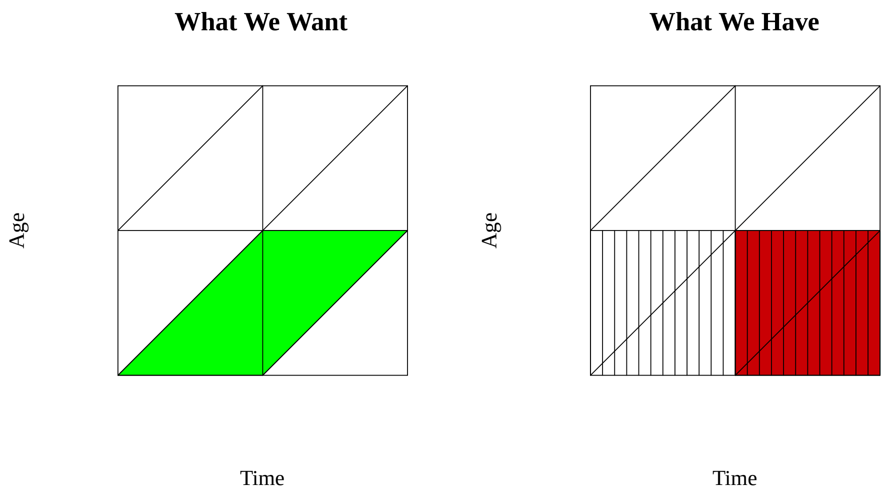
For our mortality estimations, we selected birth cohorts 1895 and 1896 (see Figure A-1 in the appendix).[5] Having information available from 1989 through 2010, we were able to reconstruct mortality for birth cohort 1895 for ages 94 through 115 and for birth cohort 1895 for ages 93 through 114, as indicated by the green and blue trapezoids, respectively. For comparative purposes we set an upper limit of 114 for the highest age any individual can reach. It can safely be assumed that both cohorts are virtually extinct when they reach that age: The Kannisto-Thatcher Database on Old Age Mortality ("KTDB")[6] estimates for the US in their most recent decennial life-table (for 2000–2007) that only 0.0118% of women alive at age 95, the starting age for our mortality analysis, survive un-

[5] We also estimated mortality for younger birth cohorts with earlier "cut-off" ages. The results, which can be obtained from the corresponding author, are remarkably similar.

[6] This database is accessible via the webpage of the Max Planck Institute for Demographic Research at http://www.demogr.mpg.de.

til age 114; for female centenarians, the probability to live another 14 years was 0.056%. According to the corresponding life table for men, there are no male survivors left at age 114.

Figure 1: Left panel: The Extinct-Cohort-Method requires data by age at death and year of birth as outlined by the green trapezoid on a Lexis surface. Right panel: The data are listed by age at death and year of death as outlined by the red square on a Lexis surface. The additional information on month of death allows us to approximate deaths by birth cohort



We are confident that our reconstruction of birth cohorts and the resulting estimates of mortality are appropriate. Table A-1 in the appendix compares our raw estimates of the probability of dying at ages 95 through 110 with those of the KTDB. In most cases the two estimates resemble each other very closely, with errors typically smaller than one percent.

Because there are relatively few cases when broken further down by level of education, the probabilities of dying have been smoothed to control for binomial noise. Instead of employing a parametric model — which might impose a shape not warranted by the data — we have chosen the nonparametric P-spline approach to smooth data introduced by Eilers and Marx (1996a). This method has been widely adopted due to its desirable properties in comparison to other smoothing methods, such as the speed of fitting, the relatively straightforward implementation, being able to work with non-Gaussian data, conservation of moments, and the lack of boundary effects to name just a few (see Eilers and Marx (1996b) for a more in-depth discussion of the advantages of smoothing with P-splines).

3. Results

Table 1 depicts the number of people alive at age 95 by birth cohort, sex, and level of education. One seeming disadvantage of our partitioning method comes immediately to light: Because of the outlined procedure to assign deaths to two different birth cohorts, we obtain "fractional persons". This does not affect the validity of our estimates, though, as we show in Table A-1. Indeed, our reconstructed birth cohorts and those estimated by the Kannisto-Thatcher-Database are very similar. For instance, the number of men from birth cohort 1895 who are alive at age 95 is 17,779.79 according to our reconstruction and 17,826 according to the KTDB. The latter database uses a different approach than us to attribute death counts to birth cohorts, which could explain those minor differences.[7] Since we have additional information on month of death, we believe that our approach yields more accurate results than that pursued by the KTDB, which only has death counts available by age at death and year of death.

Probabilities of dying have been estimated and smoothed for persons with eight years of elementary school and four years or more of college or university education. The first group can be labeled as persons who have finished elementary school; the second group represents people with at least a bachelor's degree. These two categories are depicted in bold face in Table 1. The smoothed probabilities of dying are plotted in Figure 2 as solid lines, jointly with the observed probabilities of dying, denoted as "+" symbols. The dashed lines represent 95% confidence intervals.

The upper panel (A) illustrates the results for birth cohort 1895, whereas the lower panel (B) depicts the corresponding results for birth cohort 1896. The estimates for women and men in both birth cohorts are in accordance with the literature: People with lower education face a higher risk of death than do individuals with higher education — even at those advanced ages. Those differences are minor, though. The 95% confidence intervals illustrate that statistically significant differences exist only for women at ages 95 to approximately 100. The mortality gaps among men are not statistically significant at all at the chosen level. As shown in Table 2, the estimated probabilities of dying translate to a

[7] See http://www.demogr.mpg.de/databases/ktdb/xservices/method.htm for details on the methodology used by the KTDB.

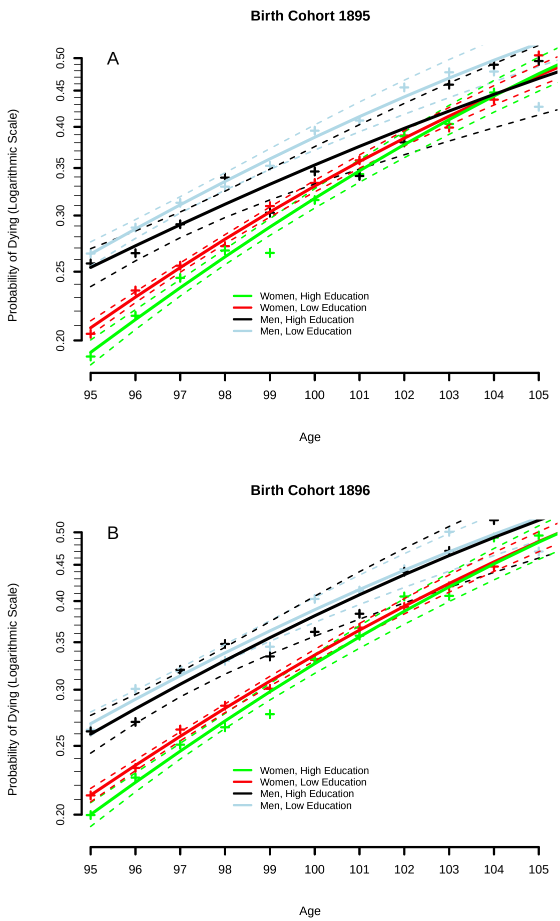
remaining life expectancy at age 95 of 4.20 years for women with low education and 4.39 years for those with high education for birth cohort 1895. The difference of 0.19 years corresponds to sightly more than two months. The results for men and for birth cohort 1896 are even smaller (Men 1895: 0.18 years; Women 1896: 0.14 years; Men 1896: 0.08 years). With the exception of men born in 1895, remaining life expectancy at age 100 differs only by about two to three weeks.

Table 1: Number of people alive at age 95 by birth cohort, sex, and level of education

Level of Education	**Cohort 1895**			**Cohort 1896**		
	Men	**Women**	**Total**	**Men**	**Women**	**Total**
No Formal Education	287.21	867.29	1,154.50	269.83	925.29	1,195.12
Elementary School						
1–7 Years	3,130.00	8,078.54	11,208.54	3,113.54	8,158.17	11,271.71
8 Years	**3,987.87**	**14,607.62**	**18,595.50**	**4,163.67**	**15,106.54**	**19,270.21**
High School						
1–3 Years	1,210.54	4,429.46	5,640.00	1,260.83	4,807.42	6,068.25
4 Years	3,375.17	15,345.92	18,721.08	3,666.29	16,707.04	20,373.33
College and University						
1–3 Years	1,157.12	5,913.12	7,070.25	1,162.75	6,278.58	7,441.33
4 Years or more	**1,676.54**	**4,781.71**	**6,458.25**	**1,729.50**	**5,100.12**	**6,829.62**
Not Stated	2,955.33	10,197.17	13,152.71	2,493.83	9,017.00	11,510.83
$\sum$	17,779.79	64,221.04	82,000.83	17,860.25	66,100.16	83,960.41

We also estimated mortality for individuals who attended high school for four years (“medium”). In addition to life expectancy estimates in Table 2, the estimated probabilities of dying are given in Table A-2 in the appendix and follow the general trends for younger ages found in the literature: The risk of death of individuals with four years of high school education is typically lower than the risk of people with less education, but higher than the risk of individuals with more years spent in formal education.

Figure 2: Estimated observed ('+') and smoothed (incl. 95% confidence intervals) probabilities of dying at ages 95 and higher for women and men with low and high education born in 1895 (upper panel) and in 1896 (lower panel)



Source: Authors' estimates based on data from the National Center for Health Statistics (2013).

4. Summary and discussion

We believe that our contribution is unique for two reasons: First, we are not aware of any previous studies of socioeconomic mortality differences at these advanced ages, based on a (reconstructed) cohort follow-up of the total surviving population. Second, we could not find any references in the literature which employed the extinct-cohort method to analyze mortality by education.

Table 2: Remaining life expectancy at ages 95 and 100 for three levels of education for women and men born in 1895 and 1896

	Cohort 1895				Cohort 1896			
	Women		Men		Women		Men	
	Age 95	Age 100	Age 95	Age 100	Age 95	Age 100	Age 95	Age 100
Low	4.20	3.03	3.59	2.68	4.14	2.98	3.56	2.68
Medium	4.30	3.05	3.74	2.82	4.22	3.00	3.61	2.70
High	4.39	3.09	3.77	2.91	4.28	3.03	3.64	2.71
High ⇔ Low (in years)	0.19	0.06	0.18	0.23	0.14	0.05	0.08	0.04
High ⇔ Low (in days)	69.11	23.67	65.70	84.03	52.17	17.93	29.01	13.39

Our analysis showed that socioeconomic differences in mortality still exist at the highest ages in the United States. The differences are rather small and only significant for women at ages 95–100. Depending on sex and birth cohort, those with high education can expect to live between one to two months longer than their peers with low education. In three of the four cases presented, the probabilities of dying tend to converge. An instinctive explanation for the narrowing gap might claim that biological factors become more important than social factors at ages 95 and higher. Or one might attribute it to successful public policies aimed at providing equal opportunities for people, regardless of their social background. Both interpretations could be correct. However, "[c]hange in a population average can be accounted for in three alternative ways", as pointed out by Vaupel and Canudas-Romo (2002, p. 2), who labeled these Level-0, Level-1, and Level-2 explanations. The interpretations above are Level-1 explanations: a direct change in the phenomenon of interest. Level-0 explanations refer to problematic data. Although an impact of erroneous data can not be ruled out completely, our extensive methodological section — including the comparison of mortality estimates with the Kannisto-Thatcher-Database — demonstrates that such an error is rather negligible. The more difficult challenge is to differentiate between a "Level-1" and a "Level-2" explanation. The latter is the outcome of a compositional change of the population: Perhaps the observed convergence is simply the outcome of selective mortality because populations are heterogeneous and frailer individuals tend to die younger on average (e.g. Vaupel et al. 1998)? We are not able to answer this question. Nevertheless, we hope our novel results will serve as an incentive for mathematical demographers to develop a model that can disentangle compositional effects from "real" effects in the case of mortality convergence or crossovers.

5. Disclaimer

All analyses, interpretations, and conclusions were reached by the authors. The National Center for Health Statistics (NCHS) is responsible only for the initial data.

References

Bengtsson, T. and van Poppel, F. (2011). Socioeconomic inequalities from past to present: An introduction. *Explorations in Economic History* 48(3): 343–356. doi:10.1016/j.eeh.2011.05.004.

Depoid, F. (1973). La Mortalité des Grands Vieillards. *Population* 28: 755–792. doi:10.2307/1531256.

Dinkel, R.H. (1997). Sterblichkeitsmessung in den obersten Altersstufen. *Zeitschrift für Bevölkerungswissenschaft* 22: 493–509.

Doblhammer, G., Rau, R., and Kytir, J. (2005). Trends in educational and occupational differentials in all-cause mortality in Austria between 1981/82 and 1991/92. *Wiener Klinische Wochenschrift* 117(13–15): 468–479. doi:10.1007/s00508-005-0393-0.

Eilers, P.H.C. and Marx, B.D. (1996a). Flexible Smoothing with B-splines and Penalties. *Statistical Science* 11(2): 89–102. doi:10.1214/ss/1038425655.

Eilers, P.H.C. and Marx, B.D. (1996b). Flexible Smoothing with B-splines and Penalties: Rejoinder. *Statistical Science* 11: 115–121.

Goldman, N. (2001). Mortality Differentials: Selection and Causation. In: Smelser, N.J. and Baltes, P.B. (eds.). *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam, NL: Elsevier: 10068–10070. doi:10.1016/B0-08-043076-7/02120-3.

Hoffmann, R. (2008). *Socioeconomic differences in old age mortality*, vol. 25 of *The Springer Series on Demographic Methods and Population Analysis*. Springer.

Hummer, R.A., Rogers, R.G., and Eberstein, I.W. (1998). Sociodemographic differentials in adult mortality: A review of analytic approaches. *Population and Development Review* 24(3): 553–578. doi:10.2307/2808154.

Kitagawa, E.M. and Hauser, P.M. (1973). *Differential mortality in the United States: A study in socioeconomic epidemiology*. Cambridge, MA: Harvard University Press.

Kunst, A. (1997). Cross-national comparisons of socio-economic differences in mortality. [PhD Thesis]. Rotterdam, NL: Department of Public Health, Erasmus University Rotterdam.

Mackenbach, J.P., Kunst, A.E., Groenhof, F., Borgan, J.K., Costa, G., Faggiano, F., Józan, P., Leinsalu, M., Martikainen, P., Rychtarikova, J., and Valkonen, T. (1999). Socioeconomic inequalities in mortality among women and among men: An international study. *American Journal of Public Health* 89(12): 1800–1806. doi:10.2105/AJPH.89.12.1800.

Manton, K.G., Stallard, E., and Corder, E. (1997). Education-specific estimates of life expectancy and age-specific disability in the U.S. elderly population 1982 to 1991. *Journal of Aging and Health* 9(4): 419–450. doi:10.1111/j.1728-4457.2008.00249.x.

National Center for Health Statistics (2013). Mortality data — vital statistics NCHS's multiple cause of death data, 1989–2010. http://www.nber.org/data/multicause.html.

Pappas, G., Queen, S., Hadden, W., and Fisher, G. (1993). The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *The New England Journal of Medicine* 329(2): 103–109. doi:10.1056/NEJM199307083290207.

Rau, R., Jasilionis, D., Soroko, E.L., and Vaupel, J.W. (2008). Continued Reductions in Mortality at Advanced Ages. *Population & Development Review* 34(4): 747–768. doi:10.1111/j.1728-4457.2008.00249.x.

Rogers, R.G., Hummer, R.A., and Nam, C.B. (1995). *Living and dying in the USA. Behavioral, health and social differentials of adult mortality*. San Diege, CA: Academic Press.

Ross, C.E. and Wu, C.L. (1996). Education, age, and the cumulative advantage in health. *Journal of Health and Social Behavior* 37(1): 104–120. doi:10.2307/2137234.

Vaupel, J.W. and Canudas-Romo, V. (2002). Decomposing demographic change into direct vs. compositional components. *Demographic Research* 7(1): 1–14. doi:10.4054/DemRes.2002.7.1.

Vaupel, J.W., Carey, J.R., Christensen, K., Johnson, T.E., Yashin, A.I., Holm, N.V., Iachine, I.A., Kannisto, V., Khazaeli, A.A., Liedo, P., Longo, V.D., Zeng, Y., Manton, K.G., and Curtsinger, J.W. (1998). Biodemographic trajectories of longevity. *Science* 280(5365): 855–860. doi:10.1126/science.280.5365.855.

Vincent, P. (1951). La Mortalité des Vieillards. *Population* 6(2): 181–204. doi:10.2307/1524149.

Zhu, H. and Xie, Y. (2007). Socioeconomic differentials in mortality among the oldest old in China. *Research on Aging* 29(2): 125–143. doi:10.1177/0164027506296758.

Appendix

Figure A-1: Lexis surface of data availability for birth cohorts 1895 and 1896

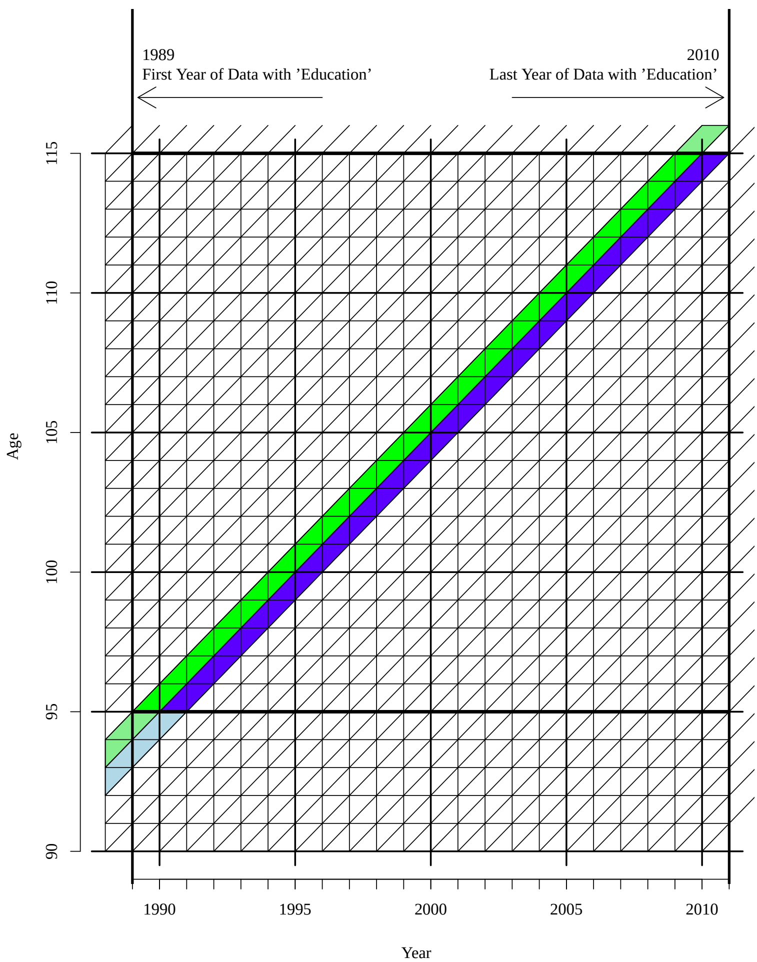


Table A-1: Comparison of our estimates ("Authors") of the age-specific probabilities of dying, $q(x)$, with estimates from the Kannisto-Thatcher-Database on old-age mortality

Age	Cohort							
	1895				1896			
	Women		Men		Women		Men	
	KTDB	Authors	KTDB	Authors	KTDB	Authors	KTDB	Authors
95	0.2145	0.2133	0.2585	0.2621	0.2132	0.2141	0.2693	0.2659
96	0.2322	0.2309	0.2889	0.2879	0.2320	0.2323	0.2868	0.2887
97	0.2514	0.2515	0.3083	0.3072	0.2599	0.2592	0.3133	0.3131
98	0.2755	0.2760	0.3326	0.3296	0.2757	0.2765	0.3253	0.3278
99	0.2936	0.2939	0.3363	0.3337	0.2921	0.2922	0.3377	0.3411
100	0.3282	0.3278	0.3615	0.3705	0.3260	0.3253	0.3789	0.3836
101	0.3431	0.3477	0.3834	0.3811	0.3532	0.3528	0.3932	0.3929
102	0.3717	0.3770	0.4162	0.4142	0.3764	0.3744	0.4189	0.4235
103	0.3712	0.3735	0.4471	0.4538	0.3859	0.3845	0.4479	0.4312
104	0.4250	0.4141	0.4344	0.4276	0.4101	0.4185	0.4543	0.4582
105	0.4437	0.4554	0.4348	0.4510	0.4514	0.4479	0.4555	0.4406
106	0.4639	0.4633	0.4957	0.4798	0.4417	0.4454	0.4808	0.4690
107	0.4764	0.4868	0.3559	0.3515	0.4870	0.4933	0.5185	0.5130
108	0.4549	0.4757	0.5263	0.5267	0.5297	0.5449	0.4231	0.3969
109	0.4803	0.5206	0.3889	0.4147	0.4505	0.4850	0.4667	0.4025
110	0.5303	0.6108	0.3636	0.4094	0.3770	0.4400	0.5000	0.3898

Table A-2: Age-specific probabilities of dying, $q(x)$, for three levels of education for women and men born in 1895 and 1896

Age	Cohort 1895						Cohort 1896					
	Women			Men			Women			Men		
	Low	Med.	High	Low	Med.	High	Low	Med.	High	Low	Med.	High
95	0.208	0.196	0.192	0.265	0.253	0.253	0.213	0.207	0.200	0.269	0.264	0.260
96	0.230	0.221	0.214	0.287	0.273	0.272	0.235	0.228	0.222	0.291	0.286	0.282
97	0.253	0.247	0.237	0.311	0.295	0.291	0.258	0.250	0.246	0.314	0.309	0.305
98	0.278	0.273	0.262	0.335	0.317	0.311	0.282	0.274	0.271	0.338	0.333	0.330
99	0.303	0.299	0.289	0.361	0.341	0.332	0.308	0.301	0.298	0.363	0.358	0.355
100	0.330	0.325	0.317	0.387	0.365	0.353	0.336	0.329	0.326	0.389	0.384	0.381
101	0.357	0.351	0.347	0.414	0.390	0.376	0.364	0.359	0.356	0.415	0.411	0.408
102	0.386	0.378	0.378	0.441	0.415	0.398	0.393	0.393	0.387	0.442	0.438	0.436
103	0.414	0.409	0.410	0.469	0.441	0.421	0.424	0.429	0.419	0.469	0.465	0.464
104	0.443	0.447	0.442	0.497	0.467	0.445	0.454	0.467	0.451	0.497	0.493	0.492
105	0.473	0.489	0.475	0.525	0.494	0.468	0.485	0.506	0.484	0.524	0.521	0.520