

A cohort is not representative of humanity

Ilya Kashnitsky

1. Doctoral researcher at Ageing and Longevity department, Netherlands Interdisciplinary Demographic Institute, 2511CV, 19 Lange Houtstraat, The Hague, The Netherlands & Faculty of Spatial Sciences, University of Groningen, 9701AD Landleven 1, Groningen, The Netherlands.
2. Junior Research Fellow at Institute for Social Policy, National Research University Higher School of Economics, 101000, 20 Myasnitskaya str., Moscow, Russia.

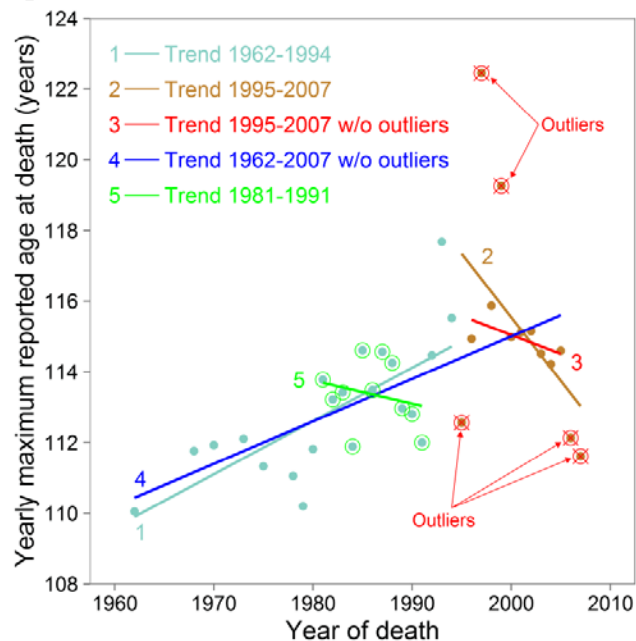
E-mail: kashnitsky@nidi.nl

In the freshly published research letter¹, Dong, Milholland, and Vijg (DMV) reported that they found strong evidence for a limit to human lifespan. Analyzing data from International Database on Longevity², they found that the yearly maximum reported age at death (MRAD, i.e. age at death of the world’s oldest person died in a specific year) stopped increasing from the mid-1990-s reaching a plateau at around 115 years. Even though the authors acknowledge that the data on “the supercentenarians <...> are still noisy and made of small samples”, they feel safe to conclude that “the results strongly suggest that the human lifespan has a natural limit”. I argue that the results and conclusions of the study are likely to be caused by just a data artifact, and that they are hardly generalizable for the humanity.

The authors chose to divide the study period at the year 1995, which is an arbitrary decision. Yet, this decision imposed a strong effect on the results and conclusions. The conclusions are based primarily on the basis of liner regression trends for the two sub-periods (figure 1, lines 1 and 2). The main conclusion is derived from the negative slope for the second sub-period (figure 1, line 1), which is largely explained by the high outliers (1997 and 1999) and low outliers in the first and the last two years (1995, 2006 and 2007). When those outliers are omitted, the slope for the second sub-period flattens greatly (regression coefficients for lines 2 and 3 are -0.36 and -0.11).

With the outliers omitted, there is only a tiny difference between the trend lines for the first sub-period and the whole period (figure 1, lines 1 and 4, regression coefficients are 0.15 and 0.12).

Figure 1: Reported age at death of supercentenarians.

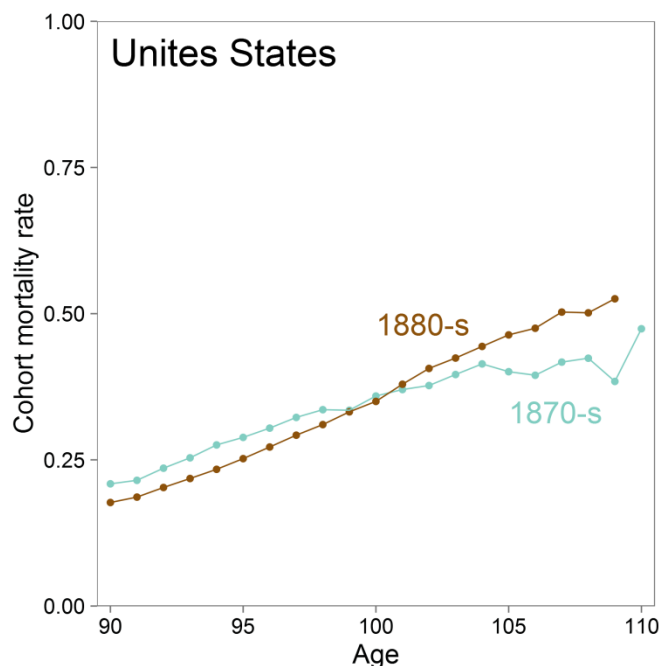


NOTE. The yearly maximum reported age at death (MRAD). The lines represent the functions of linear regressions. All data were collected from the IDL database (all 15 countries included, 1968–2007, $n = 668$). Unlike DMV, I use all the death records from IDL available on 9 October 2016, not just the data for France, Japan, UK, and US. Unlike DMV, I take MRAD values as age-at-death in *days* divided by 365.25 to convert into years, not the *rounded to years* values. The above mentioned data decisions of DMV are not explained and justified in the paper, and, in my view, are not optimal. The outliers are identified with Cook’s distance in two steps. Cook’s distances are 0.57 and 0.10 for the observations 1997 and 1999 in the model 2. After the removal of the high outliers, Cook’s distance for the observations 1995, 2006, and 2007 are 1.50, 0.19, and 0.46, correspondingly.

Seeing how volatile the data are, it seems too hurriedly to drive humanity-wide conclusions based on the presented type of analysis. Imagine, we would now have these data just until 1991. The similarly arbitrary division of the study period at the year 1981 would have shown that the growth in MRAD had stopped and even reverted (figure 1, line 5). We would have then concluded that there was strong evidence for “the limit of human lifespan” at around 113.5 years. Yet, the following one and a half decade would have proven us to be misinterpreting the development of MRAD.

However, the haste of conclusions based on highly vulnerable linear regression estimates is not the only caveat of the presented paper. I believe, it is essentially incorrect to draw conclusions about such an ecumenical concept as human lifespan limit based on so sporadic and erratic data. Most likely, the slowdown in MRAD in 2000-s is a cohort effect. Namely, the difference behind the top-survivors’ data analyzed here is the *increase* in old-age cohort age-specific mortality rates that took place in the United States in the cohorts born in 1880-s as compared to the cohorts born in 1870-s (figure 2).

Figure 2: Cohort mortality rates in the United states.



NOTE. Data are cohort age-specific mortality rates (CASMR) from Human Mortality Database ³. The lines represent the average CASMR over two groups of birth cohorts, ten 1-year birth cohorts each: born in 1870-s and 1880-s. See the similar graphs for France, Great Britain, Japan, and Sweden in Extended Data Figure 2.

The US had exceptionally low old-age mortality in the cohorts of 1870-s. That is why the majority of MRAD supercentenarians in 1980-s and 1990-s were from the US (see extended data table 1). For some reasons (which are beyond the scope of the present letter), maximal longevity of the US 1880-s births cohorts tuned out worse than that of the 1870-s birth cohorts (with the average difference in mortality rates of 0.05 at the ages 100-110; p-value < 0.0001). Higher mortality in succeeding cohorts as compared to the preceding cohorts is a relatively rare case; as we may see from the Extended Data Figure 2, similar rise in old-age mortality did not happen in France, Great Britain, Japan, or Sweden. So, quite a local effect of the increase in the US cohort mortality was uplifted by DMV to report the presence of a natural limit for human longevity.

Summing up, the presented evidence does not clearly “suggest that the maximum lifespan of humans is fixed and subject to natural constraints”. I believe, we better put faith in the results of demographers (e.g. Vaupel⁴ or Vallin & Meslé⁵), who draw much more optimistic projections based on the population-wide analyses. The dynamics of human mortality show that increased percentages of human populations reach more and more mature ages; eventually the 115-year “limit” will fall, and, in some years, we will likely speculate over the 120-year limit. And so on.

REFERENCES

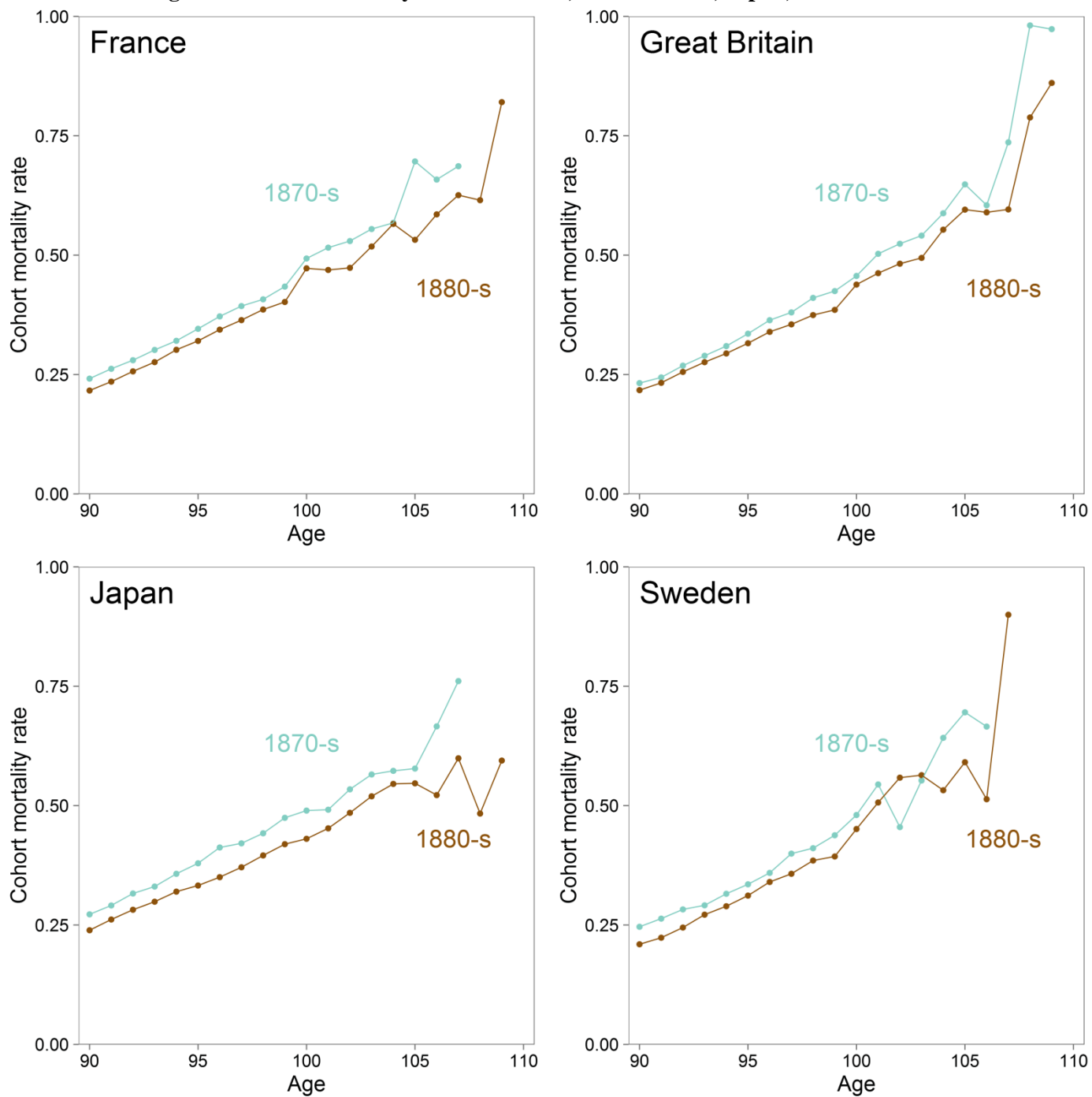
1. Dong, X., Milholland, B. & Vijg, J. Evidence for a limit to human lifespan. *Nature* (2016). doi:10.1038/nature19793
2. Maier, H. et al. Supercentenarians (Springer, 2010). doi:10.1007/978-3-642-11520-2_2
3. *The Human Mortality Database* (<http://www.mortality.org>, 2016).
4. Vaupel, J. W. Biodemography of human ageing. *Nature* **464**, 536–542 (2010).
5. Vallin, J. & Meslé, F. The Segmented Trend Line of Highest Life Expectancies. *Population and Development Review* **35**, 159–187 (2009).

REPRODUCIBILITY

The code to reproduce this analysis is published at github: <https://git.io/vSMYx>

The data is free to download upon registration at <http://www.supercentenarians.org/>

Extended Data Figure 2: Cohort mortality rates in France, Great Britain, Japan, and Sweden.



NOTE. Data are cohort age-specific mortality rates (CASMR) from Human Mortality Database (<http://www.mortality.org>). The lines represent the average CASMR over two groups of birth cohorts, ten 1-year birth cohorts each: born in 1870-s and 1880-s.

Extended Data Table 1. Yearly maximal reported age at death of supercentenarians.

Year of death	Age at death, years	Sex	Country of birth	Country of death	Date of birth	Date of death
1962	110.0	M	CAN	CAN	1852-02-26	1962-03-16
1968	111.8	M	GBR	GBR	1856-06-15	1968-03-21
1970	111.9	F	GBR	GBR	1858-02-06	1970-01-11
1973	112.1	F	GBR	GBR	1861-07-10	1973-08-18
1975	111.3	F	GBR	GBR	1864-08-25	1975-12-24
1978	111.1	F	GBR	GBR	1867-01-23	1978-02-12
1979	110.2	F	GBR	GBR	1868-11-30	1979-02-12
1980	111.8	F	GBR	GBR	1868-12-26	1980-10-20
1981	113.8	F	USA	USA	1867-07-01	1981-07-01
1982	113.2	F	USA	USA	1869-07-01	1982-07-01
1983	113.4	F	USA	USA	1870-07-01	1983-07-01
1984	111.9	F	USA	USA	1872-07-01	1984-07-01
1985	114.6	M	USA	USA	1870-07-01	1985-07-01
1986	113.5	F	USA	USA	1873-07-01	1986-07-01
1987	114.6	F	GBR	GBR	1873-06-02	1987-12-27
1988	114.3	F	USA	USA	1873-07-01	1988-07-01
1989	113.0	F	USA	USA	1876-07-01	1989-07-01
1990	112.8	M	GBR	GBR	1877-08-19	1990-06-10
1991	112.0	M	USA	USA	1879-07-01	1991-07-01
1992	114.5	F	USA	USA	1877-07-01	1992-07-01
1993	117.7	F	USA	USA	1875-07-01	1993-07-01
1994	115.5	F	USA	USA	1878-07-01	1994-07-01
1995	112.6	F	USA	USA	1883-07-01	1995-07-01
1996	114.9	F	USA	USA	1881-07-01	1996-07-01
1997	122.4	F	FRA	FRA	1875-02-21	1997-08-04
1998	115.9	F	USA	USA	1882-07-01	1998-07-01
1999	119.3	F	USA	USA	1880-07-01	1999-07-01
2000	115.0	F	GBR	GBR	1885-11-08	2000-11-02
2001	115.1	F	FRA	FRA	1886-04-25	2001-06-06
2002	115.2	F	USA	USA	1887-07-01	2002-07-01
2003	114.5	F	JPN	JPN	1889-03-23	2003-09-28
2004	114.2	M	ESP	ESP	1889-12-15	2004-03-05
2005	114.6	F	JPN	JPN	1890-08-30	2005-04-05
2006	112.1	F	GBR	GBR	1894-01-12	2006-03-01
2007	111.6	F	ESP	ESP	1895-08-27	2007-04-08

NOTE. The table provides the information on yearly maximum reported age at death (MRAD) extracted from the whole the International Database on Longevity (<http://www.supercentenarians.org>, all 15 countries included, 1968–2007, $n = 668$) obtained on 9 October 2016. Color background delimitates 1870-s birth cohorts from 1880-s birth cohorts.