The case for monitoring life-span inequality

Alyson A. van Raalte[1], Isaac Sasson[2], Pekka Martikainen[1,3,4]

[1]Max Planck Institute for Demographic Research, Rostock 18057, Germany.
[2]Department of Sociology and Anthropology and the Herczeg Institute on Aging, Tel Aviv University, Tel Aviv 6997801, Israel.
[3]Population Research Unit, Faculty of Social Sciences, University of Helsinki, Helsinki 00014, Finland.
[4]Department of Public Health Sciences, Stockholm University, SE-106 91 Stockholm, Sweden.

Correspondence to: vanraalte@demogr.mpg.de

Accepted version of manuscript.

The case for monitoring life-span inequality

Alyson A. van Raalte, Isaac Sasson, Pekka Martikainen

Inequality in length of life is the most fundamental of all inequalities; every other type of inequality is conditional upon being alive. As has long been recognized in studies of economic inequality, we can compare populations based on per capita gross national income, but there is a pressing need to further examine how income varies within populations via Gini coefficients and percentile-based metrics. Mortality inequalities should be approached in the same way. Human population health is generally monitored by average mortality levels, typically in terms of life expectancies, which belie substantial variation in length of life. Variation in ages at death, captured by a metric of life-span variation, should be used to supplement measures of average longevity when comparing or monitoring societies and population subgroups (1). Although life-span variation has historically been strongly inversely correlated with life expectancy (2, 3), we are beginning to see this relationship reversed, resulting in positive correlation in some countries or subnational populations. Often these changes reflect midlife mortality crises with roots in stratified education and wealth. We discuss these measures and trends and how they can have profound implications for how individuals might plan and live their lives, and for how societies might organize and manage health care, insurance, pensions, and other social policies and programs.

Life expectancy at birth (or simply life expectancy, as we refer to it in the rest of the text) is the most common metric of survival. It is the hypothetical average age at death given age-specific death rates in a given year. Life-span variation, the variability in ages at death around that average, can be measured by using an index of variation or inequality— for instance, the standard deviation, Gini coefficient, or interquartile range. To illustrate, consider age-at-death distributions of non-Hispanic black and non-Hispanic white men in the United States based on 2012–2016 death rates. The life expectancy from this distribution is 72 years for blacks and 77 years for whites [see supplementary materials (SM)]. But the timing of death was variable, skewed below the average in both groups, meaning that deaths were more spread out below the life expectancy than above it. Among blacks, the spread in survival was noticeably wider. Men in the 25th to 75th percentile (the interquartile range) died between 63 and 85 years in the black distribution, whereas those in the white distribution died between 69 and 88 years. Although life expectancy for blacks was only 6% lower than for whites, the age window over which these deaths occurred was 17% larger for blacks.

Some early efforts by the Organization for Economic Co-operation and Development (OECD) to monitor within-group variability included a one-off report that measured life-span variation conditional upon survival to age 10 (4). However, although we currently monitor life expectancy at birth in all countries of the world, which captures between-country differences in average mortality, no international organization regularly monitors and compares the within-country variation in age at death. Likewise, many countries evaluate health and social policies by their success in eliminating gaps in life expectancy between race, ethnic, or socioeconomic groups. But few countries monitor the variation in age at death within and across such groups, which ignores an important and substantial part of the inequalities in mortality.

TRENDS IN LIFE-SPAN VARIATION

In high-income countries, life expectancy has doubled over the past century and a half. Life expectancy at birth in 2014 was around 81 years for people living in a country belonging to the OECD (SM). Averting deaths at any age increases life expectancy. But for life-span variation to decrease when life expectancy is increasing, more deaths need to be averted at younger than older ages. This compresses the age-at-death distribution, making ages at death more similar. Whether a death is considered younger or older depends on a threshold age that varies in time and between groups. This age generally sits below the life expectancy, and is specific to the age pattern of mortality (2).

In the past, death rates declined faster at younger ages compared to older ages as populations transitioned through different epidemiologic environments (5). Gains to life expectancy and concomitant declines in life-span variation resulted from strong reductions in infectious disease, maternal and child mortality, injuries, and more recently, cancers (6). Reduced circulatory disease mortality, by contrast, accounted for the bulk of life expectancy gains in the final decades of the 20th century, but its impact on life-span variation trends was moderate. The averted deaths from declining circulatory disease rates occurred at ages above and below the threshold age, which led to either minor increases or decreases in life-span variation depending on the sex, period, and population examined.

As a result of these historical changes, the inverse correlation between life-span variation and life expectancy was so strong that it was reasonable to question how much more information could be gained by looking at life-span variation in addition to life expectancy. Yet, it is becoming increasingly apparent that there are considerable differences between populations with respect to life-span variation at the same levels of life expectancy. For example, a U.S. man retiring at age 67 can expect to live another 16.8 years—almost as long as his British counterpart (17.0 years) (SM). But the American is facing substantially greater variability about that mean with a standard deviation that is 13.8% higher.

Two emerging phenomena, related to age-specific mortality changes below and above the threshold age, may weaken the historical association between life expectancy and life-span variation. One is that some subpopulations are experiencing stalls in reducing mortality at younger ages or even experiencing midlife mortality increase (7). The other is that mortality decline is occurring at ever higher ages in many countries (8). Further gains in life expectancy in low-mortality countries may increasingly come from disproportionate reductions in old-age mortality. These phenomena acting alone or in concert could result in increases in life-span variation alongside increases in life expectancy; i.e., changing the historically observed correlation from negative to positive.

This is precisely what is being seen at the national level in the United States. From 1980 to 2014, life expectancy increased by about 10% for men and 5% for women (although both sexes experienced declines from 2014 to 2016 (fig. S1)). Life-span variation, however, fluctuated sharply over the period, with sustained increases observed in the late 1980s, early 2000s, and the 2010–2016 period (fig. S2). Not coincidentally, these years were marked by mortality episodes that primarily affected young adults—the HIV/AIDS and crack-cocaine epidemics of the 1980s, and more recently, the increase in accidental poisonings, particularly from opioids (SM). Had we been monitoring life-span variation as closely as life expectancy, the U.S. midlife mortality crisis would possibly have been uncovered earlier in an attempt to understand why life-span variation was increasing at the dawn of the 21st century.

Further evidence is observed among subpopulations in Finland, a country with an exceptionally long and good-quality time series of population register data, including socioeconomic status. Over the period 1971 to 2014, we see increasing life expectancy for women (see the figure, top) and men (fig. S3, top) aged 30 and above according to education, occupational class, and income quintile. The increases in life expectancy over time run mostly in parallel with one another, with some divergence because the lower socioeconomic groups experienced slower improvements. A policy implication is that the groups with a lower socioeconomic status are falling behind the upper classes, a lag that may be closed by investing in health and social policies to reduce mortality in the lower socioeconomic groups at all ages. However, trends in life-span variation (see the figure, bottom; fig. S3, bottom) reveal a more worrisome pattern. The less-advantaged groups were not only dying earlier than advantaged groups, on average, but they faced greater variation in the eventual time of death—a double burden of inequality—which has increased over time. The diverging trends in life-span variation mainly resulted from differences in the pace of early and midlife mortality decline, in particular slow improvements in early and midlife mortality in the lower socioeconomic groups. The pace of mortality decline at older ages was similar across occupational groups (9). A policy implication is that investing in efforts to drive down deaths in midlife—deaths that are typically considered premature and avoidable (e.g., deaths from accidents, violence, or substance abuse)—can reduce life-span variation.

[FIGURE 1 ABOUT HERE]

This is not just a Finnish phenomenon. Life-span variation is diverging between social groups wherever it has been examined. In the United States, differentials in life-span variation between those who completed college and those who only completed high school doubled over the 1990–2010 period, mainly because of substantial increases in variability among the high school educated. At the same time, life expectancy differentials by the same two groups widened from 2.2 to 5.7 years (10). In Denmark, life-span variation increased among the lowest-income quartile and decreased among the other three quartiles over the 1986–2014 period, although all income groups experienced increases in life expectancy (11). It is important to note that in most of these examples of increasing life-span variation, mortality declined over postretirement ages, even among the lowest socioeconomic groups. The increases in lifespan variation among lower socioeconomic groups thus occurred because mortality reduction over preretirement ages was modest or absent.

INDIVIDUAL AND POPULATION IMPACTS

Like life expectancy, life-span variation constitutes a useful summary measure of mortality regimes (i.e., age-specific death rates). We argue that it has two important implications: one at the individual (micro) level and the other at the population (macro) level. Life-span variation reflects both individual uncertainty in the timing of death and heterogeneity in underlying population health. At the micro level, life-span variation reflects individual discrepancies in the risk of death. In other words, it measures uncertainty in the timing of death. Economic models have shown that because individuals are inherently risk averse, they would forego additional years of expected life to reduce uncertainty in age at death (12). From this perspective, diverging life-span variation between socioeconomic groups means that an overlooked dimension of social inequality in health is increasing—those from more-advantaged groups can more effectively plan their life course, whereas less-advantaged groups face greater and increasing uncertainty about their survival.

For the individual, this can have profound consequences. One's subjective assessments of his or her survival are instrumental when making decisions about lifecycle investment and consumption, including education, training, and retirement (12). The Survey of Health, Ageing and Retirement in Europe, as well as its U.S. progenitor the Health and Retirement Study, routinely ask their respondents about their future expectations. These questions probe how long responders expect to live and when they expect to die. Answers to these questions underlie answers to others, such as: Will they work for pay after age 70? How much money will they spend this year? Will they leave an inheritance or help their children financially? Although often framed around financial decisions, these survey questions are attempts to examine responders' expectations about the family sphere (e.g., time spent with family members) and work (e.g., time to retirement), and how these expectations may be influenced by anticipation of one's own mortality.

Individuals are of course mostly unaware of life-span variability statistics, but experience of survival chances of friends and relatives will influence perceptions of survival expectations. For example, because of larger life-span variability, U.S. blacks are more likely than U.S. whites to experience an early death of close relatives (13). Studies have shown that these subjective survival expectations predict actual mortality and mirror known socioeconomic disparities in death rates (1, 14). Furthermore, to the extent that mortality regimes (i.e., death rates) become inscribed in individuals' own mortality expectations—by which we mean that excess mortality experienced by certain social groups affects subjective survival assessment and future life course expectations for the members of that group—higher life-span variation also constitutes a form of disadvantage.

At the macro level, life-span variation is an indicator of heterogeneity in underlying population health. Understanding this heterogeneity is crucial for accurate forecasts in insurance and annuity markets, for public provision of medical care, and for creating equitable pension schemes when ages at death vary. Increasing life-span variation among disadvantaged groups implies that the individuals belonging to such groups are living increasingly diverse lives. This increasing heterogeneity mirrors diverging variability in other social realms, including increasing variation in participation in work and family life (15). Increasing life-span variation signals uneven age patterns of mortality decline, with faster declines at older rather than at younger ages, or even rising early and midlife mortality. Therefore, monitoring life-span variation may facilitate early detection of adverse mortality developments and warrant social interventions at younger ages.

TOWARD A COMPREHENSIVE MEASUREMENT OF MORTALITY

In our world in which reams of granular data related to health are routinely collected, easy-to-understand summary metrics are needed to set health targets, to evaluate policy outcomes, to uncover emerging threats, and to compare levels and trends in health and mortality across populations. Life expectancy is monitored almost everywhere. But despite growing awareness, life-span variation has not been systematically monitored by any country in the world. We suggest four reasons why this might be the case, and also demonstrate why this does not need to persist. First, given that we already monitor life expectancy, it was unclear to policy-makers whether there was an added benefit to monitoring life-span variation. The examples shown here demonstrate the independence of the two indices. Although life expectancy might still be the best metric for the speed of survival improvements in the population as a whole, life-span variation is its complement, needed to monitor the equality in survival improvement.

Second, as with studies of income inequality, it remains unclear which indicator for measuring life-span variation is the most appropriate. Although there is no gold standard, the high correlation between indices suggests that trends would be broadly similar no matter which index is chosen. Third, officials in statistical offices are not trained to measure life-span variation. Although this is true, all statistical offices produce life tables. Calculating an index of life-span variation is a relatively straightforward procedure from the life table. Fourth, metrics of variability are more difficult to understand than metrics of average levels. But this criticism could equally be leveled at indices of income differentials, yet the widespread adoption of the Gini coefficient by statistical offices suggests that the benefit of monitoring income variation outweighs the disadvantage of a more theoretically complex measure. In our view, this is also true for monitoring mortality.

Many countries and international organizations monitor premature death, but the cut-off age varies across jurisdictions. For instance, Statistics Canada defines premature deaths as those under age 75, the U.S. Centers for Disease Control and Prevention uses age 80 as the cut-off, while the New York City Department of Health uses age 65 (SM). Alternatively, a metric of life-span variation could be easy to understand, be used in various mortality regimes, and be sensitive to deaths that we would normally consider premature, but use mortality over the whole age range to capture this.

Life-span variation might be reduced by decreasing disparities in the social environment, medical care, and how individuals behave and interact within their social sphere. If an average level of health is deemed important enough to regularly monitor and report, we should also regularly summarize the spread around this average. A healthy population is one in which people live for a long time on average—and long lives are enjoyed by everyone.

REFERENCES AND NOTES

1. R. D. Edwards, S. Tuljapurkar, Popul. Dev. Rev. 31, 645 (2005).
2. J.W. Vaupel, Z. Zhang, A.A. van Raalte, BMJ Open, 1, e000128 (2011).
3. F. Colchero et al., Proc. Natl. Acad. Sci. U.S.A. 113, E7681 (2016).
4. OECD, Society at a Glance 2006: OECD Social Indicators. Organization for Economic Co-operation and Development (2007).
5. J.-M. Robine, Population 13, 173 (2001).
6. B. Seligman, G. Greenberg, S. Tuljapurkar, Proc. Natl. Acad. Sci. U.S.A. 113, 8420 (2016).
7. A. Case, A. Deaton, Proc. Natl. Acad. Sci. U.S.A. 112, 15078 (2015).
8. R. Rau, C. Bohk-Ewald, M.M. Muszyńska, J.W. Vaupel., Surface Plots of Rates of Mortality Improvement, in Visualizing Mortality Dynamics in the Lexis Diagram (Springer, 2018).
9. A.A. van Raalte, P. Martikainen, M. Myrskylä, Demography 51, 73 (2014).
10. I. Sasson, Demography 53, 269 (2016).
11. H. Brønnum-Hansen, BMJ Open 7, e014489 (2017).
12. R. D. Edwards, J. Popul. Econ. 26, 1485 (2013).
13 D. Umberson et al., Proc. Natl. Acad. Sci. U.S.A. 114, 915 (2017).
14. M. D. Hurd, K. McGarry, J. Hum. Resour. 30, S268 (1995).
15. D. C. Brown et al., Demography, 49, 819 (2012).

ACKNOWLEDGMENTS

We thank J. W. Vaupel, J. P. Mackenbach, and S. Klüsener for helpful comments and L. M. Junna for help with harmonizing the Finnish mortality data. This research was supported by a starting grant from the European Research Council awarded to A.A.vR. (grant no. 716323), grant no. I-2475-119.4/2017 from the German-Israeli Foundation for Scientific Research and Development (GIF) awarded to I.S., and funding to P.M. from the Academy of Finland (grant nos. 1294861 and 1308247).

FIGURES

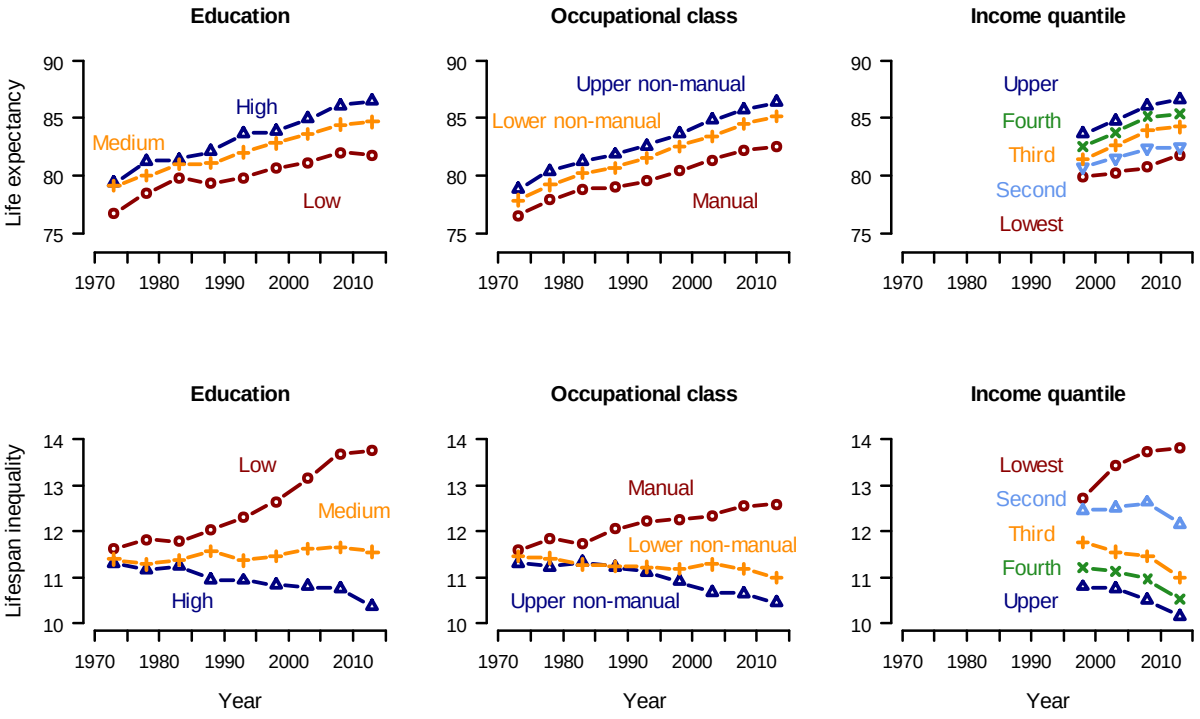

Figure 1: Trends in life expectancy and life-span variation for Finnish females, 1971–1975 to 2011–2014. Life expectancy is the average age at death, and lifespan inequality (or lifes-span variation) is the standard deviation, conditional upon survival to age 30, with age-specific death rates frozen at those observed in the given year. See supplementary materials for data and methods, including trends for males (which are qualitatively similar) and robustness checks using alternative measures of life-span variation.

SUPPLEMENTARY MATERIALS

The published supplementary materials are available here:

http://www.sciencemag.org/content/362/6418/1002/suppl/DC1

The accepted version of supplementary materials are as follows:

Materials and Methods

1. Calculating life expectancy and the modal age at death within the OECD

Averaging life expectancies among population subgroups can be done in one of two ways: either by aggregating the deaths and exposures of the subgroups, or by the arithmetic mean of life expectancies, weighted by the proportion of births. The two methods produce different results and, for the period mortality context examined here, the more appropriate method remains a 'philosophical question' depending in large part on how summary indices of period mortality are interpreted (*16*).

We used the aggregation method for two reasons: First, because we interpret period life expectancy as a snapshot of the overall mortality conditions faced by individuals of different ages at the period in question, and second, because we were also interested in examining the shape of the resultant death density. Death and exposure-to-risk counts for each country of the OECD were available from the Human Mortality Database (*17*), with the exception of Mexico, South Korea, and Turkey. We used the period counts by single year of age for both sexes combined from the year 2014, or latest year available (Chile (2008), Canada (2011), and Iceland, Greece and New Zealand (2013)). We then calculated a life table using standard methods (*18*), presented below (Table S1).

Age	mx	qx	ax	lx	Dx	Lx	Tx	ex
0	0.00431	0.00430	0.06	100000	430	99595	8089881	80.9
1	0.00031	0.00031	0.5	99570	31	99555	7990286	80.2
2	0.00019	0.00019	0.5	99540	19	99530	7890731	79.3
3	0.00015	0.00015	0.5	99521	15	99514	7791201	78.3
4	0.00012	0.00012	0.5	99506	12	99500	7691687	77.3
5	0.00011	0.00011	0.5	99494	11	99489	7592187	76.3
6	0.00010	0.00010	0.5	99484	10	99479	7492698	75.3
7	0.00009	0.00009	0.5	99473	9	99469	7393220	74.3
8	0.00009	0.00009	0.5	99464	9	99460	7293751	73.3
9	0.00009	0.00009	0.5	99455	9	99451	7194291	72.3
10	0.00008	0.00008	0.5	99446	8	99442	7094840	71.3
11	0.00010	0.00010	0.5	99438	10	99433	6995398	70.3
12	0.00010	0.00010	0.5	99428	10	99423	6895965	69.4
13	0.00013	0.00013	0.5	99418	13	99412	6796542	68.4
14	0.00015	0.00015	0.5	99406	15	99398	6697130	67.4
15	0.00019	0.00019	0.5	99390	19	99381	6597732	66.4
16	0.00025	0.00025	0.5	99371	25	99359	6498351	65.4
17	0.00031	0.00031	0.5	99346	31	99330	6398993	64.4
18	0.00041	0.00041	0.5	99315	41	99295	6299662	63.4
19	0.00047	0.00047	0.5	99274	47	99251	6200368	62.5
20	0.00051	0.00051	0.5	99227	50	99202	6101117	61.5
21	0.00054	0.00054	0.5	99177	54	99150	6001915	60.5
22	0.00057	0.00057	0.5	99123	57	99095	5902765	59.5
23	0.00059	0.00059	0.5	99067	58	99038	5803669	58.6
24	0.00059	0.00059	0.5	99009	58	98980	5704632	57.6
25	0.00061	0.00061	0.5	98950	60	98920	5605652	56.7
26	0.00063	0.00063	0.5	98890	62	98859	5506732	55.7
27	0.00063	0.00063	0.5	98828	62	98798	5407872	54.7
28	0.00065	0.00065	0.5	98767	65	98734	5309075	53.8
29	0.00068	0.00068	0.5	98702	67	98669	5210340	52.8
30	0.00070	0.00070	0.5	98635	69	98600	5111672	51.8
31	0.00074	0.00074	0.5	98566	73	98529	5013071	50.9
32	0.00076	0.00076	0.5	98493	75	98455	4914542	49.9
33	0.00080	0.00080	0.5	98418	79	98378	4816087	48.9

34	0.00086	0.00086	0.5	98339	84	98297	4717709	48.0
35	0.00089	0.00089	0.5	98255	87	98211	4619412	47.0
36	0.00093	0.00093	0.5	98168	91	98122	4521201	46.1
37	0.00098	0.00098	0.5	98076	96	98028	4423079	45.1
38	0.00106	0.00106	0.5	97981	103	97929	4325050	44.1
39	0.00112	0.00112	0.5	97877	109	97822	4227121	43.2
40	0.00121	0.00121	0.5	97768	118	97709	4129299	42.2
41	0.00135	0.00135	0.5	97649	132	97584	4031590	41.3
42	0.00144	0.00144	0.5	97518	141	97447	3934007	40.3
43	0.00158	0.00158	0.5	97377	154	97300	3836559	39.4
44	0.00174	0.00174	0.5	97223	169	97139	3739259	38.5
45	0.00189	0.00189	0.5	97054	183	96963	3642121	37.5
46	0.00209	0.00209	0.5	96871	203	96770	3545158	36.6
47	0.00228	0.00228	0.5	96669	220	96558	3448388	35.7
48	0.00255	0.00254	0.5	96448	245	96326	3351830	34.8
49	0.00282	0.00282	0.5	96203	271	96068	3255504	33.8
50	0.00314	0.00313	0.5	95932	300	95782	3159436	32.9
51	0.00345	0.00345	0.5	95632	330	95467	3063654	32.0
52	0.00380	0.00379	0.5	95302	361	95122	2968187	31.1
53	0.00418	0.00417	0.5	94941	396	94743	2873065	30.3
54	0.00459	0.00458	0.5	94545	433	94329	2778322	29.4
55	0.00501	0.00499	0.5	94112	470	93877	2683994	28.5
56	0.00549	0.00547	0.5	93642	512	93386	2590117	27.7
57	0.00596	0.00594	0.5	93130	553	92853	2496731	26.8
58	0.00643	0.00641	0.5	92576	593	92280	2403879	26.0
59	0.00694	0.00691	0.5	91983	636	91665	2311599	25.1
60	0.00755	0.00752	0.5	91347	687	91004	2219934	24.3
61	0.00817	0.00813	0.5	90661	737	90292	2128930	23.5
62	0.00877	0.00873	0.5	89923	785	89531	2038638	22.7
63	0.00931	0.00927	0.5	89138	826	88725	1949107	21.9
64	0.01000	0.00995	0.5	88312	879	87873	1860382	21.1
65	0.01077	0.01071	0.5	87434	937	86965	1772509	20.3
66	0.01165	0.01158	0.5	86497	1002	85996	1685544	19.5
67	0.01260	0.01252	0.5	85495	1071	84960	1599548	18.7
68	0.01365	0.01356	0.5	84424	1145	83852	1514588	17.9
69	0.01481	0.01471	0.5	83280	1225	82667	1430736	17.2
70	0.01631	0.01618	0.5	82055	1328	81391	1348068	16.4
71	0.01765	0.01750	0.5	80727	1413	80021	1266677	15.7
72	0.01923	0.01904	0.5	79314	1510	78559	1186656	15.0
73	0.02083	0.02062	0.5	77804	1604	77002	1108097	14.2
74	0.02300	0.02274	0.5	76200	1733	75334	1031095	13.5
75	0.02542	0.02510	0.5	74467	1869	73533	955761	12.8
76	0.02822	0.02782	0.5	72598	2020	71588	882229	12.2
77	0.03131	0.03083	0.5	70578	2176	69490	810641	11.5
78	0.03478	0.03418	0.5	68402	2338	67233	741151	10.8
79	0.03936	0.03860	0.5	66064	2550	64789	673919	10.2
80	0.04404	0.04309	0.5	63514	2737	62145	609130	9.6
81	0.04943	0.04823	0.5	60777	2932	59311	546985	9.0
82	0.05597	0.05445	0.5	57845	3149	56270	487674	8.4
83	0.06278	0.06087	0.5	54695	3329	53031	431404	7.9
84	0.07146	0.06899	0.5	51366	3544	49594	378373	7.4
85	0.07999	0.07691	0.5	47822	3678	45983	328779	6.9
86	0.09016	0.08627	0.5	44144	3808	42240	282796	6.4
87	0.10133	0.09644	0.5	40336	3890	38391	240556	6.0
88	0.11357	0.10747	0.5	36446	3917	34487	202165	5.5
89	0.12709	0.11949	0.5	32529	3887	30586	167678	5.2
90	0.14297	0.13344	0.5	28642	3822	26731	137092	4.8
91	0.16031	0.14841	0.5	24820	3684	22978	110361	4.4
92	0.17897	0.16427	0.5	21137	3472	19401	87383	4.1
93	0.19733	0.17961	0.5	17664	3173	16078	67982	3.8
94	0.22235	0.20010	0.5	14492	2900	13042	51904	3.6
95	0.23453	0.20991	0.5	11592	2433	10375	38862	3.4
96	0.26247	0.23202	0.5	9159	2125	8096	28487	3.1
97	0.28739	0.25128	0.5	7034	1767	6150	20391	2.9
98	0.31438	0.27167	0.5	5266	1431	4551	14241	2.7
99	0.34043	0.29091	0.5	3835	1116	3278	9690	2.5
100	0.37632	0.31673	0.5	2720	861	2289	6413	2.4
101	0.40590	0.33742	0.5	1858	627	1545	4124	2.2
102	0.44162	0.36174	0.5	1231	445	1009	2579	2.1
103	0.46306	0.37600	0.5	786	295	638	1570	2.0
104	0.49477	0.39664	0.5	490	195	393	932	1.9
105	0.51882	0.41195	0.5	296	122	235	539	1.8
106	0.52882	0.41824	0.5	174	73	138	304	1.7
107	0.58791	0.45435	0.5	101	46	78	167	1.6
108	0.61078	0.46789	0.5	55	26	42	88	1.6
109	0.60763	0.46604	0.5	29	14	23	46	1.6
110	0.66922	1.00000	1.49	16	16	23	23	1.5

Table S1: The OECD life table for both sexes combined. Mexico, South Korea, and Turkey were excluded.

2. Male American age-at-death distributions by race

American data by race and Hispanic status were downloaded from the CDC Wonder website of the Centers for Disease and Control and Prevention (*19*), which contained death and population exposure-to-risk counts by single year of age up to open-age intervals of

100+ (death counts) and 85+ (exposure counts). We downloaded the data for the combined last 5 years available (2012-2016), to reduce random fluctuation. Since more than a third of white deaths and around a quarter of black deaths occurred above age 85, it was necessary to extend the last open-age interval to determine the interquartile range, and obtain a more reliable estimate of life expectancy (*20*). We estimated mortality by single year of age up to age 110 by fitting a Kannisto parametric model over ages 75 to 84 and extrapolating to ages 110. This model is commonly used for closing life tables with right-censored data, for instance by the Human Mortality Database team (*21*). Life tables were created with the observed mortality to age 84 and fitted mortality from ages 85 to 110+. The interquartile range was estimated by fitting a cubic spline interpolation to the survivorship column of the life table and taking the difference between the 25^{th} and 75^{th} percentile of survival age, to two decimals. The data and R-Code used to calculate life tables, life expectancy, and the interquartile range is provided to replicate these results. The data is in the accompanying .xlsx file, while the R-Code is embedded in this document, following the reference list.

To test the robustness of our treatment of the right-censored data, we also fitted a univariate penalized composite link model (pclm) (*22*), implemented with the R-package 'ungroup' (*23*). The pclm has been shown to outperform other non-parametric models in the treatment of right-censored mortality data (*24*). Life expectancy from the pclm model was 0.03 years higher for blacks and 0.33 years lower for whites. The interquartile range was 0.02 years higher for blacks and 0.35 years lower for whites.

3. Life expectancy and lifespan variation in Finland by sex and socioeconomic status

Data

Education: We used death and exposure counts for the entire population of Finland by sex, 5-year age interval (30-34, …, 100-104, 105+), calendar year (1971-2014), and education group obtained from the population register of Statistics Finland. Completed educational level was grouped according to the ISCED (International Standard Classification of Education). For the purposes of this paper we aggregated these levels into the following categories: *Low* (levels 1 to 2, or less than 10 years of education), *Medium* (levels 3-4, or 10-12 years of education), and *High* (levels 5 to 8, corresponding to any tertiary education).

Occupational Status: We used death and exposure counts for the entire population of Finland by sex, single year of age (ages 30-105+), calendar year (1971-2014), and occupational class group obtained from the population register of Statistics Finland. Occupations were grouped into *manual workers* (e.g. construction workers, bus drivers, cleaners), *lower non-manual workers* (e.g. shop attendants, nurses), *upper non-manual workers* (e.g. doctors and teachers) and *others*. The classification was retroactive for pensioners and the unemployed. Household workers were classified according to the occupation of the head of the household. The *others* group included farmers, students over 30 years of age, and self-employed workers that were difficult to classify elsewhere. This

group has undergone substantial compositional change over the 43 years of our study, which hampers any meaningful interpretation of trends. Thus, we did not present results for the *others* group. Over years 1971-2000, deaths and exposures were not disaggregated by occupational status for ages 100+. For these ages, we proportioned out the death and exposure counts into the different occupational status groups according to their proportions over ages 95-99.

Income: We used death and exposure counts for the entire population of Finland by sex, 5-year age interval (30-34,…,95-99, 100+), calendar year (1986-2014), and income decile (which we grouped into income quintiles). The income data was obtained from registers of the Finnish Tax Administration and the National Social Insurance Institution by Statistics Finland, and linked to death records from the same period via personal identification codes. Income was defined as household disposable income per consumption unit. Incomes consisted of the all wages, capital income and income transfers as well as paid taxes of all household members. The income measure was further adjusted to account for differences in household composition. This was done by dividing household income by the total number of consumption units in the household using the OECD equivalence scale (*25*). For each of 1995, 2000, 2005, and 2010, separate income quintiles were calculated for men and women.

Grouping data by year, ungrouping data by age

To reduce random fluctuation in the trends for life expectancy and lifespan variation, we grouped years into 5-year intervals (1971-75, 1976-1980, …, 2006-2010, 2011-2014). Note that the last group is a 4-year interval.

Data for income and education, as well as years 2011-2014 for occupational status, came grouped in 5-year age intervals. Data was ungrouped to obtain mortality estimates by single year of age using a penalized composite link model (pclm) (*22*). For the education data, we also had data by single year of age available for some years. As a robustness check, we estimated summary measures of life expectancy and lifespan variation from both the modeled mortality by single year of age and the raw mortality by single year of age over years 2001-2005 and 2006-2010 for all sex/education groups. Life expectancy differences were small and averaged 0.23 years and differences in standard deviations were also small and averaged 0.02 years between the modeled and raw mortality data.

Smoothing old age mortality

At oldest ages death rates can fluctuate due to stochastic processes. As is common in life table methodology (*21*), we smoothed raw death rates parametrically at oldest ages using a Kannisto function. The parametric curve was fit to ages 75-104, and we used the resultant smoothed data above age 85 for all life table calculations. We did not fit Kannisto models to income, education or the 2011-2014 occupational class mortality data, since these data had already been modeled using the pclm (see above section).

Robustness checks using the Gini coefficient

Figure 2 was replicated using the Gini coefficient of survival (*26*). This index was chosen because it measures relative variation (i.e. differences in age at death relative to life expectancy) as opposed to the standard deviation which is an absolute measure of variation. Additionally, the indices differ in their sensitivity to mortality change at different ages. Differences in standard deviations tend to be weighted toward mortality differences at younger ages than the Gini (*27*). The Gini can vary between 0 and 1, with higher values indicating greater variation in age at death. The interpretation of the Gini is a little less straightforward than the standard deviation of survival—it is the average difference in age at death between any two random individuals in the population divided by the life expectancy at birth. While some differences in patterns can be seen in comparison to the standard deviation figures (Fig. 2, fig. S3), the main message that differentials in lifespan variation by socioeconomic group are increasing more quickly than differentials in life expectancy still holds.

4. Subjective survival expectations and aggregate mortality

Several studies have shown that subjective survival assessments, reflecting how long individuals expect to live, correlate with socioeconomic status (SES) and predict actual mortality. Specifically, individuals with higher socioeconomic status report higher survival probabilities than their lower-SES counterparts (in answer to the question "What are the chances that you will live to be age [75/80/…/110/120] or more?"), both in the United States (*14*) and across Europe (*28*). These subjective assessments were predictive of actual mortality, both at the individual (*29*) and aggregate level (*14*). In other words, evidence suggests that people indeed assess how long they will live, even if only vaguely, and that on average their assessments mirror objective social disparities in mortality.

A number of studies have further suggested that subjective survival assessments play a role in lifecycle decisions, including consumption and saving, health behaviors, investment in education, and retirement (*12, 14, 30, 31*). For example, in the Netherlands, subjective survival expectations factors into retirement decisions among older employees (*32*), and in the United States employees with lower survival expectations tend to retire and claim their social security benefits earlier (*33*). Subjective survival assessments are important not only for decision-making, but also for one's wellbeing. Among Israeli older adults, subjective assessments of 'nearness to death' predicted psychological distress, net of health status and socio-demographic factors (*34*).

But how are subjective survival assessments constructed? Aside from one's own health status, they are shaped and constantly updated by the experience of death in one's immediate social circle, particularly the loss of a parent or a spouse (*19, 35*). Because kinship and other social ties run along ethnic and class divisions (*36*), members of different social groups are effectively exposed to different everyday experiences of mortality among family and friends. As a result, we would expect that the greater the exposure to highly varying

ages at death within an individual's social group, the more uncertainty an individual would place on surviving to older ages, although this has never been empirically tested to the best of our knowledge.

5. Increasing lifespan variation among national populations

Two major contemporary instances of increasing lifespan variation stand out at the population level—Central and Eastern Europe (CEE) over much of the past 50 years and the United States in the wake of the current opioid epidemic.

Since the 1960s, age patterns of mortality decline in CEE have more often than not diverged, not only in intensity but often even in direction, for instance mortality increased over midlife and declined at young and old ages. A detailed study is forthcoming on the relationship between life expectancy and lifespan variation in CEE since 1960 (*37*). Over this time, lifespan variation and life expectancy moved independently, showing that additional information could be gained from a summary measure of variability. During the protracted life expectancy stagnation of the 1960s and 1970s, 55 % of yearly changes in the two metrics occurred in the same direction (i.e. what would be expected from a positive rather than a negative correlation). This changed to around 21 % during the sharp 1988-1994 mortality crisis following the break-up of the former Soviet Union, when mortality increased over virtually all age groups. Over the 1960-2015 period, yearly proportional change in lifespan variation was larger than for life expectancy, because mortality fluctuation was highest over early adulthood, and lifespan variation is more sensitive to mortality change over these ages than life expectancy is (*27*).

As for the United States, the scale of the current ongoing midlife mortality crisis in the wake of the opioid epidemic is now well known, but was not part of the national discourse until recently (*7*, *38*, *39*). In part this is because of our reliance on life expectancy as a summary measure of overall population health, which as the fig. S1 below shows, increased near-linearly from 1980 to 2010. Lifespan variation, on the other hand, experienced two major increases—the first during the mid-1980s (mostly affecting men) and the second starting around the late 1990s, albeit with a decline during the late 2000s. While we have not fully investigated the ages and causes behind these lifespan variation trends, fig. S2 below shows that both periods of lifespan variation increase coincided with unfavorable mortality developments for early adulthood. The first increase in the 1980s was plausibly related to increased mortality from crack cocaine and the HIV/AIDS epidemic (with the decline in 1996 coinciding with the introduction of antiretroviral treatments (*40*) and waning homicide mortality attributable to crack cocaine (*41*), while the second increase coincides with increasing mortality from accidental poisonings, particularly from drug overdoses (*7, 42, 43)*. The brief period of lifespan variation decline in the late 2000s coincides with a period of rapid decline in infant mortality, following a 10-15 year stagnation in infant mortality rates.

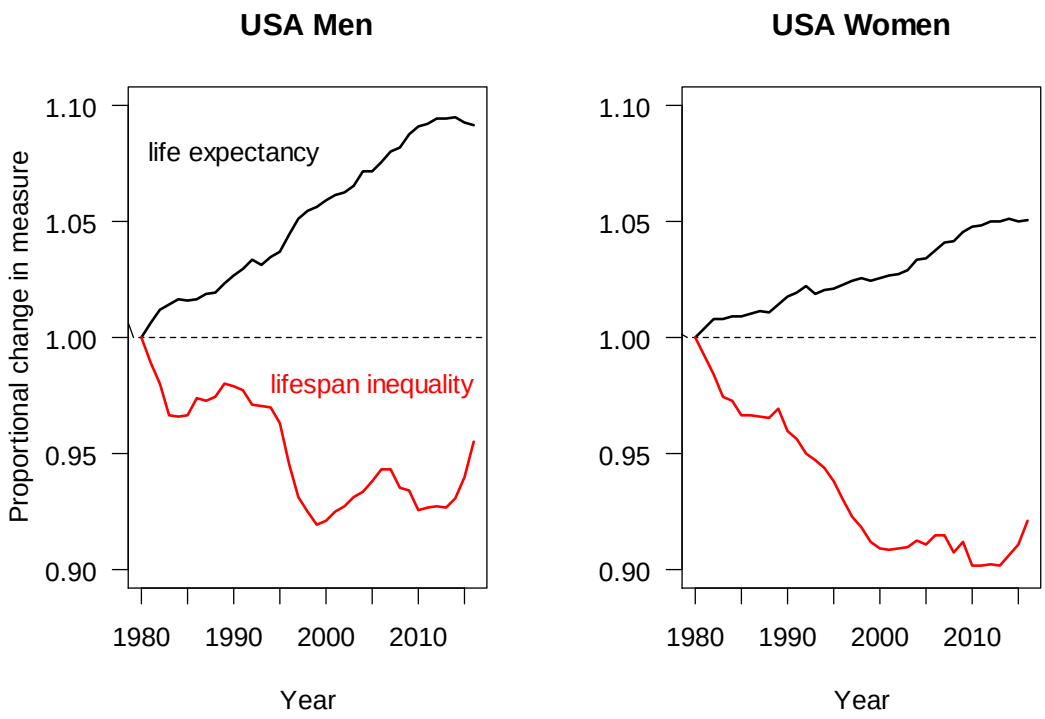

Figure S1: The proportional change in life expectancy and lifespan variation measured by standard deviation since 1980. Data are period single-year by single-age life table data from the Human Mortality Database (*17*). Lifespan variation measured by the Interquartile range and the life disparity index (*2*) showed similar temporal patterns (not shown).

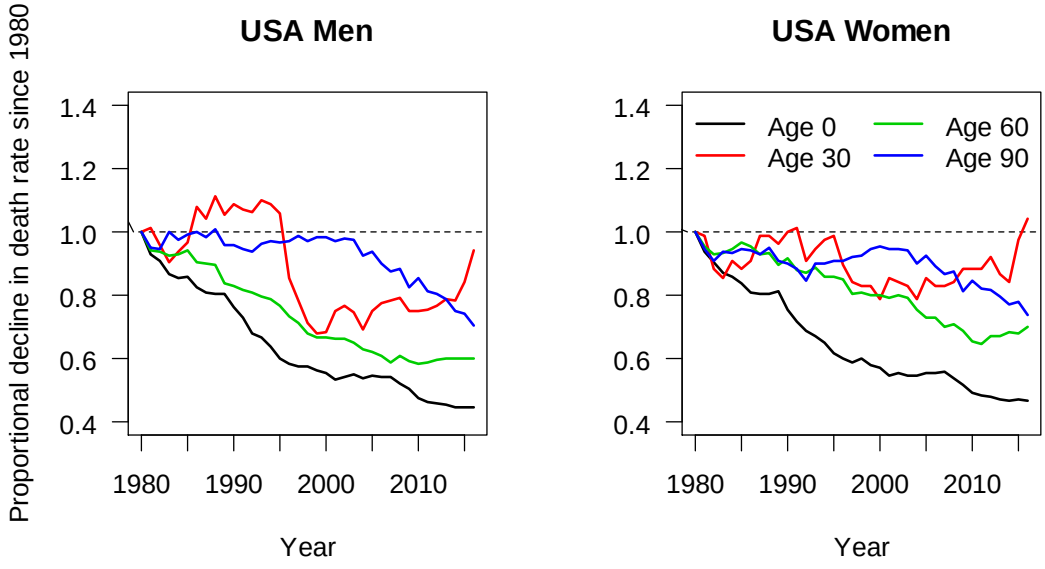

Figure S2: The proportional decline in death rates since 1980 for selected ages. Data are period single-year by single-age life table death rates from the Human Mortality Database (17).

In terms of the implications for policy, we would argue that had we been monitoring lifespan variation as closely as life expectancy, the US midlife mortality crisis would plausibly have been uncovered earlier in an attempt to understand why lifespan variation was increasing at the dawn of the 21st Century. Earlier guidelines to curb the over-prescription of opioids and boosted funding to social and public health interventions over working ages might have prevented or tempered the observed midlife mortality increases.

More generally, both the CEE example and the US example highlight the independence and importance of lifespan variation as a summary measure of population health. It is rare to observe increases in death rates at some ages alongside rapid declines at other ages. These examples highlight the importance of an equitable balance of social and health policies and spending across ages.

6. Previous studies of lifespan variation

The earliest known work examining variability in the distribution of ages at death come from Lexis (*44*) in the 19th Century. Lexis theorized that if populations were purged of extrinsic mortality (i.e. age-independent mortality from predominately infectious diseases), ages at death would be normally distributed. This line of thinking was later taken up by Dublin and Lotka in the 1940s (*45*), Comfort in the 1950s (*46*), and most famously by Fries in 1980, in his compression of mortality hypothesis (*47*). Fries foresaw a future where premature mortality would be reduced to negligible levels, and older age mortality would become normally distributed around a life expectancy of 85 years with a standard deviation of 4 years. Empirically, Fries's predictions have not borne out. Old age survival improvements far outpaced those envisioned by Fries with no looming limit to lifespan in sight (*48*, *49*). Meanwhile, although lifespan variation has continued to decline in most high income countries of the world, the levels remain far higher than what Fries envisioned (*2*, *50, 51*).

More recently, attention has shifted to looking at lifespan variation as a marker of heterogeneity in population health at the population level and uncertainty in the timing of death at the individual level. Le Grand (*52*) and Wilmoth and Horiuchi (*53*) were among the first to argue that lifespan variation was a useful summary measure of mortality, that could move independently of life expectancy. Shkolnikov et al. (*26*) showed that lower educational groups in Russia experienced higher lifespan variation than higher educational groups. This finding was replicated for numerous countries of Europe by van Raalte et al. (*54*), and for the United States by Edwards and Tuljapurkar (*1*) and by Brown et al. (*15*). Trends in lifespan variation by a dimension of individual (*9-11*, *26*, *55*) or area-level (*56*) socioeconomic status have revealed widening between-group disparities in lifespan variation—a finding that we also replicate in the current analysis for three dimensions of Finnish socioeconomic status. In all of the aforementioned cases, the gap in lifespan variation widened at a proportionally

higher rate than the gap in life expectancy. The independence of life expectancy and lifespan variation is especially salient for age-at-death distributions conditional upon surviving childhood , and is sensible for examining trends in adult variability over time periods when infant and child mortality are rapidly declining (*1*, *57*). However, the choice of starting age of a conditional age-at-death distribution can have a large impact on the direction of trends in variability, and cannot be compared to unconditional lifespan variation (*5*, *50*, *58*).

Note that most of the work to date on lifespan variability has been conducted with period death rates (i.e. on ages at death that would prevail if death rates were fixed to the levels observed in a given year). Trends in cohort lifespan variation might be qualitatively different. Engelman et al. (*50*) argue that as death rates have declined over childhood, frailer individuals who normally would not have survived environments with high disease loads are surviving to older ages. As a result, we might expect a more heterogeneous old age population, and increased variation in older age mortality across birth cohorts.

7. Definitions of premature mortality

It has long been recognized that premature mortality is not a fixed concept, but differs depending on the overall level of mortality. As deaths over increasingly older age groups become rarer with improvements in survival, we are more inclined to consider them premature. However even within similarly developed populations, we found substantial differences in which ages were being defined as contributing to premature mortality. New York City monitors . Statistics Canada defined premature deaths as those of individuals who were younger than age 75 (*59*). A recent report by the CDC used a cut-off of age 80 (*60*), arguing that premature mortality should roughly reflect deaths below the life expectancy. Meanwhile, New York City continues to use a cut-off of age 65 (*61*). The arbitrariness of choosing a fixed cut-off age hampers both continuous time series of trends, as well as international and subnational comparisons of premature mortality.

A further advantage of lifespan variation as a summary metric is that it can be decomposed into younger and older age mortality components separated by a moving threshold age (*2*, *27*, *62*). Thus the component of lifespan variation from younger-aged mortality is not bound by a fixed upper age limit, but rather by a moving threshold age that increases gradually with improvements in survival. Lifespan variation as a whole remains sensitive to premature mortality. This younger-aged mortality component to lifespan variation is a comparable metric to other indicators of premature mortality.

8. The standard deviation above age 67 in the USA and in England & Wales

Please see the accompanying Excel spreadsheet for all calculations. Data were downloaded from the Human Mortality Database (*17*).

9. Additional figures

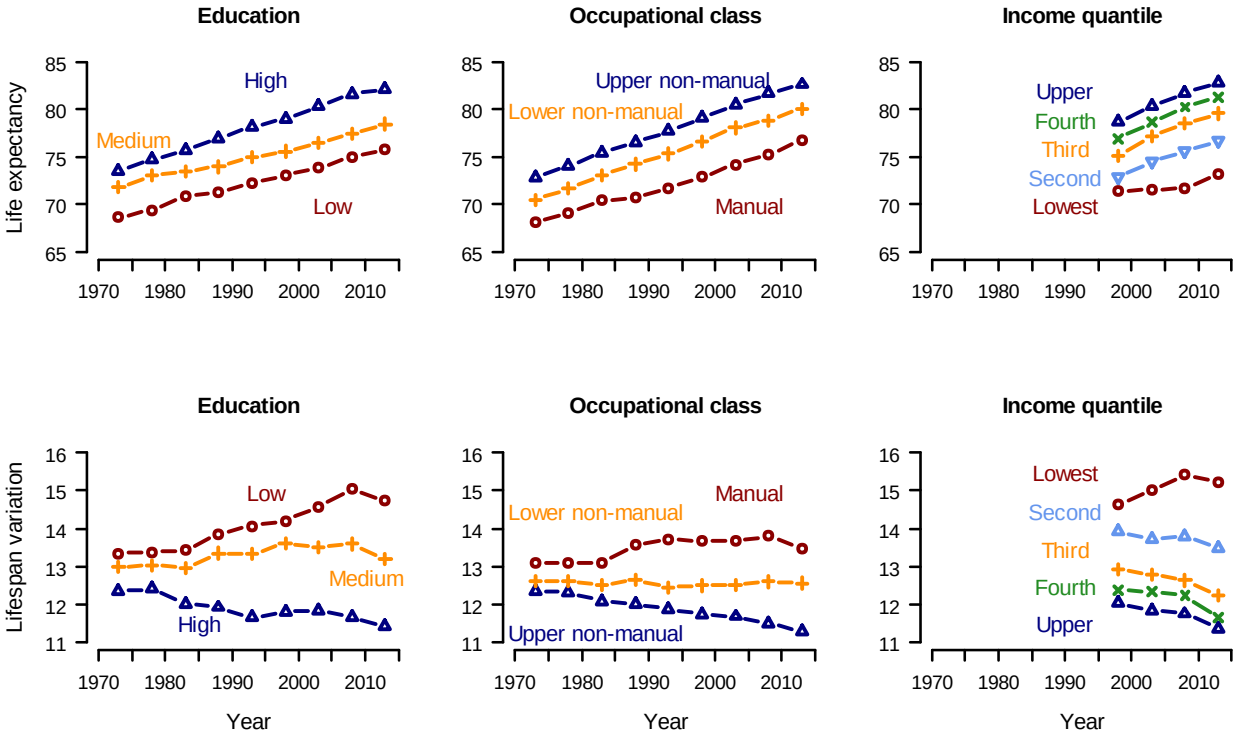

Figure S3: Equivalent of Figure 2 for Finnish men. Life expectancy is the average age at death conditional upon survival to age 30, with age-specific death rates frozen at those observed in the given year. Lifespan variation is the standard deviation in survival above age 30 in years, with age-specific death rates frozen at those observed in the given year. More details on the data and methods are provided in section 3.Note: the y-scales are different to those in Fig. 2 reflecting lower life expectancy and greater lifespan variation among men.

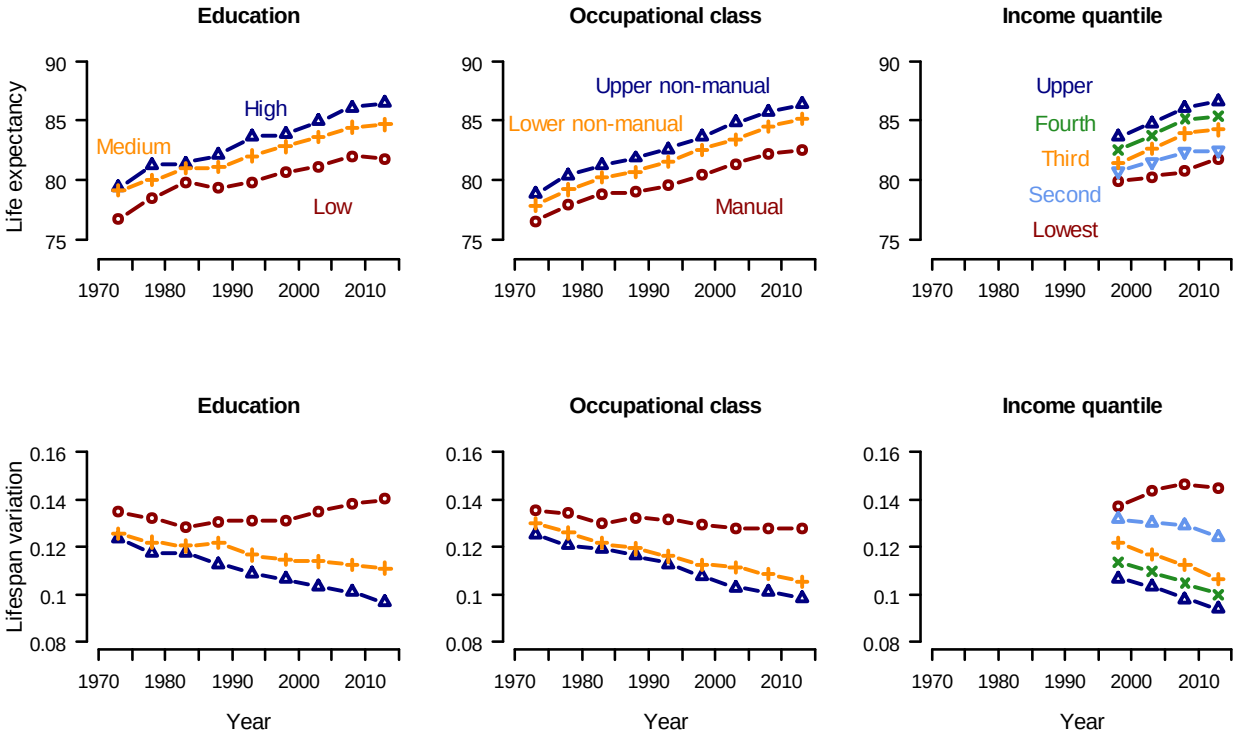

Figure S4: Equivalent figure to Figure 1 but measuring lifespan variation using the Gini coefficient of survival (*26*) instead of the standard deviation of survival. Life expectancy is the average age at death conditional upon survival to age 30, with age-specific death rates frozen at those observed in the given year. Lifespan variation is the standard deviation in survival above age 30 in years, with age-specific death rates frozen at those observed in the given year. More details on the data and methods are provided in section 3.

References and Notes:

1. R. D. Edwards, S. Tuljapurkar, *Population and Development Review* 31, 645-674 (2005).
2. J.W. Vaupel, Z. Zhang, A.A. van Raalte, *BMJ Open*, 1, 1-6 (2011).
3. F. Colchero *et al*., *PNAS*, 113, 7681-7690 (2016).
4. OECD, *Society at a Glance 2006: OECD Social Indicators*. Organization for Economic Co-operation and Development (2007).
5. J.-M. Robine, *Population*, **13**, 173 (2001).
6. B. Seligman, G. Greenberg, S. Tuljapurkar, *PNAS*, **113**, 8420 (2016).
7. A. Case, A. Deaton, *PNAS*, **112**, 15078 (2015).
8. R. Rau, C. Bohk-Ewald, M.M. Muszyńska, J.W. Vaupel., *Surface Plots of Rates of Mortality Improvement*, in *Visualizing Mortality Dynamics in the Lexis Diagram*. Springer: Cham., 43 (2018).
9. A.A. van Raalte, P. Martikainen, M. Myrskylä, *Demography* **51,** 73 (2014).
10. I. Sasson, *Demography*, **53**, 269 (2016).
11. H. Brønnum-Hansen, *BMJ Open*, **7,** e014489 (2017).
12. R. D. Edwards, *J. of Population Economics*, **26,** 1485 (2013).
13. D. Umberson, et al, *PNAS*, **114**, 915 (2017).
14. M. D. Hurd, K. McGarry, *The Journal of Human Resources*, **30,** S268 (1995).
15. D.C. Brown, et al, *Demography*, **49,** 819 (2012).
16. E. Andreev, W. Lutz, S. Scherbov, *IIASA Working Paper, WP-89-035*, 1-15, (1989).
17. *Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on June 12, 2018.*
18. S.H. Preston, P. Heuveline, M. Guillot, *Demography: measuring and modeling population processes*. Wiley-Blackwell (2001).
19. Centers for Disease Control and Prevention, National Center for Health Statistics. Multiple Cause of Death. Data downloaded on October 29, 2018.
20. T.I. Missov, L. Nemeth, M.J. Danko, *Palgrave Communications*, 15049 (2016).
21. Wilmoth, J.R., et al., *Methods Protocol for the Human Mortality Database (version 6).* (2017). http://www.mortality.org/Public/Docs/MethodsProtocol.pdf
22. S. Rizzi, J. Gampe, P.H. Eilers, *American journal of Epidemiology*, 182, 138-147 (2015).
23. M.D. Pascariu, S. Rizzi, M. Danko, *ungroup: R Package for Estimating Smooth Distributions from Coarsely Binned Data v.0.8.4* Available on Github, https://github.com/mpascariu/ungroup, 2018.
24. S. Rizzi, M. Thinggaard, G. Engholm, N. Christensen, T. B. Johannesen, J. W. Vaupel, , R. Lindahl-Jacobsen, *BMC medical research methodology*, 16, 59 (2016).

25. OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth. Available from: http://www.oecd.org/statistics/OECD-ICW-Framework-Chapter8.pdf. (2013).

26. V.M. Shkolnikov, E.M. Andreev, A.Z. Begun, *Demographic Research*, 8, 305-358 (2003).

27. A.A. van Raalte, H. Caswell. *Demography*, 50, 1615-1640 (2013).

28. A. Delavande, S. Rohwedder. *Demography*, 48, 1377–1400 (2011).

29. M. Hurd, K. McGarry. *The Economic Journal*, 112, 966–85 (2002).

30. M. Hurd, D. McFadden, L. Gan. In *Inquiries in the Economics of Aging*, edited by David A. Wise, 259–309. Chicago: University of Chicago Press. (1998).

31. B. Bosworth, G. Burtless, K. Zhang. *Economic Studies at Brookings*, 87 (2016).

32. H. van Solinge, K. Henkens. *The European Journal of Public Health*, 20, 47–51 (2010).

33. M. Hurd, J. P. Smith, J. M. Zissimopoulos. *Journal of Applied Econometrics*, 19, 761–75 (2004).

34. A. Shrira, E. Bodner, Y. Palgi. *Aging & Mental Health*, 18, 1066–70 (2014).

35. Y. Palgi, A. Shrira, G. Kavé, T. Rubinstein, D. Shmotkin. *The Journals of Gerontology: Series B*, gby048 (2018).

36. M. McPherson, L. Smith-Lovin, J.M. Cook. *Annual Review of Sociology*, 27, 415–44 (2001).

37. J. M. Aburto, A.A. van Raalte. *Demography* (in press).

38. A. Case, A. Deaton. Mortality and Morbidity in the 21(st) Century. *Brookings Papers on Economic Activity*, 397–476 (2017).

39. J. Y. Ho. *Health Affairs*, 32, 459–67 (2013).

40. Antiretroviral Therapy Cohort Collaboration. *The Lancet*, 372, 293–299 (2008).

41. R. Fryer, P. Heaton, S. Levitt, K. Murphy. Measuring the Impact of Crack Cocaine. w11318. Cambridge, MA: National Bureau of Economic Research. (2005).

42. R.K. Masters, A. M. Tilstra, D. H. Simon. *Biodemography and Social Biology*, 63, 31–37 (2017).

43. M.S. Shiels, P. Chernyavskiy, W.F. Anderson, A.F. Best, E.A. Haozous, P. Hartge, P.S. Rosenberg, D. Thomas, N.D. Freedman, A. Berrington de Gonzalez. *The Lancet*, 389, 1043–54 (2017).

44. W. Lexis. *Annales de démographie internationale, Paris*, 447-462 (1878).

45. L. I. Dublin, A. J. Lotka, *Length of life*. (Ronald Press, New York, 1949).

46. A. Comfort, *The Biology of Senescence*. (London: Routledge & Kegan Paul, 1956).

47. J. F. Fries, *New England Journal of Medicine*, 303, 130-135 (1980).

48. J. W. Vaupel, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 352, 1799-1804 (1997).

49. R. Rau, E. Soroko, D. Jasilionis, J. W. Vaupel, *Population and Development Review*, 34, 747-768 (2008).

50. M. Engelman, V. Canudas-Romo, E. M. Agree, *Population and Development Review*, 36, 511-539 (2010).

51. N. Ouellette, R. Bourbeau, *Demographic Research*, 25, 595-628 (2011).

52. J. Le Grand, *European Economic Review*, 31, 182-191 (1987).

53. J. R. Wilmoth, S. Horiuchi, Rectangularization Revisited: Variability of Age at Death within Human Populations. *Demography*, 36, 475-495 (1999).

54. A. A. van Raalte *et al.*, *International Journal of Epidemiology*, 40, 1703-1714 (2011).

55. I. Permanyer, J. Spijker, A. Blanes, E. Renteria, *Demography*, DOI: 10.1007/s13524-018-0718-z (2018).

56. R. Seaman, T. Riffe, H. Caswell *Poster presented at the 2018 Annual Meeting of the Population Association of America, Denver CO*. (2018).

57. J. Smits, C. Monden, *Social Science & Medicine*, 68, 1114-1123 (2009).

58. G. Myers, K. Manton, *Gerontologist*, 24, 346-353 (1984).

59. Statistics Canada. Table 13-10-0744-01 Premature and potentially avoidable mortality, Canada, provinces and territories.

60. N. B. Johnson, L. D. Hayes, K. Brown, E. C. Hoo, K. A. Ethier, CDC National Health Report: leading causes of morbidity and mortality and associated behavioral risk and protective factors—United States, 2005–2013. (2014).

61. W. Li, *et al*. Summary of Vital Statistics, 2015. New York, NY: New York city department of health and mental hygiene, Bureau of vital statistics, 2017.

62. D. O. S. Gillespie, M. V. Trotter, S. Tuljapurkar, *Demography*, 51, 1003-1017 (2014).

R Code to calculate U.S. life tables by race from data downloaded from CDC Wonder

Output is life expectancy, IQR, and figures plotting the age-at-death distributions. We thank Ugofilippo Basellini for the visualization idea.

```
library(tidyverse)
library(readxl)
library(scales)

setwd("/your_path/your_folder/")

# reading in the data and creating a data frame "CDC"
# with columns Age, Deaths, Population, Race

nh_blacks <- read_excel("DataANDResults.xlsx",
                    sheet = 1, col_names = TRUE,
                    col_types = c("text","text","numeric","numeric","numeric","text"),
                    na = c("Not Applicable","NS"), skip = 0)[,2:5]

nh_whites <- read_excel("DataANDResults.xlsx",
                    sheet = 2, col_names = TRUE,
                    col_types = c("text","text","numeric","numeric","numeric","text"),
                    na = c("Not Applicable","NS"), skip = 0)[,2:5]

BNH <- nh_blacks[,2:4] %>%
      rename(Age="Single-Year Ages Code") %>%
      mutate(Race="BNH")

WNH <- nh_whites[,2:4] %>%
      rename(Age="Single-Year Ages Code") %>%
      mutate(Race="WNH")

CDC <- rbind(BNH,WNH)
CDC$Age <- as.integer(CDC$Age)



# Smoothing death rates above age 85, where population counts
# by single year of age are not available from the CDC Wonder database.
# We also assume that deaths with age not stated (70 nh_whites, 20 nh_blacks)
# are all above age 85. Making an alternative assumption of redistributing these
# deaths over ages proportional to death counts makes little difference to end result

# Smoothing is done by fitting a Kannisto model over ages 75 to 84
# and extrapolating from ages 85 to 110

KannistoLL <- function(theta,x,Dx,Nx){
  a <- theta[1]
  b <- theta[2]
  mux <- a*exp(b*x)/(1+a*(exp(b*x)))
  LL <- -sum(Dx*log(mux)-(Nx*mux))
  return(LL)
```

```
}
Blpar <- optim(par=c(0.00001,0.1),fn=KannistoLL,
            x=75:84,Dx=filter(CDC,Age>=75,Age<85,Race=='BNH')$Deaths,
            Nx=filter(CDC,Age>=75,Age<85,Race=='BNH')$Population)
Bla <- Blpar$par[1]
Blb <- Blpar$par[2]

Whpar <- optim(par=c(0.00001,0.1),fn=KannistoLL,
            x=75:84,Dx=filter(CDC,Age>=75,Age<85,Race=='WNH')$Deaths,
            Nx=filter(CDC,Age>=75,Age<85,Race=='WNH')$Population)
Wha <- Whpar$par[1]
Whb <- Whpar$par[2]


oldage <- 85:110
boldmx <- woldmx <- rep(0,length(oldage))

for(i in 1:length(oldage)){
  boldmx[i] <- Bla*exp(Blb*oldage[i])/(1+Bla*(exp(Blb*oldage[i])-1))
  woldmx[i] <- Wha*exp(Whb*oldage[i])/(1+Wha*(exp(Whb*oldage[i])-1))
}

# binding the raw death rates (ages 0:84) and the smoothed death rates (ages 85:110)

wmx <-
c(filter(CDC,Age<85,Race=="WNH")$Deaths/filter(CDC,Age<85,Race=="WNH")$Populati
on,
        woldmx)
bmx <-
c(filter(CDC,Age<85,Race=="BNH")$Deaths/filter(CDC,Age<85,Race=="BNH")$Populatio
n,
        boldmx)


# life table function

lifetable <- function(mx,Age,sex){
  # a function for a single year of age life table
  # need as input year, mx (death rates), Age (i.e. 0:110), and the sex ("m" or "f")

  n <- length(Age)
  qx <- dx <- lx <- Lx <- Tx <- ex <- rep(0,n)


  # determine the length of the interval lived by those who died (ax)
  ax <- rep(0.5,n)
  ax[1] <- ifelse(Age[1]==0,
               ifelse(sex=="f",
                    ifelse(mx[1] < 0.107,0.053+2.8*mx[1],0.35),
                    ifelse(mx[1] < 0.107,0.045+2.684*mx[1],0.33)),0.5)
  ax[n] <- 1/mx[n]
```

```
 # mx to qx conversion

 for(i in 1:length(Age)){
  qx[i] <- mx[i]/(1+(1-ax[i])*mx[i])
 }

 # number alive at age x
 lx[1] <- lx[1] <- 100000

 for(i in 1:(length(Age)-1)){
  lx[i+1] <- lx[i]*(1-qx[i])
 }

 # death density
 dx <- lx*qx

 # person-years lived in age interval [x, x+1)
 for(i in 1:length(Age)){
  Lx[i] <- lx[i]-(1-ax[i])*dx[i]
 }
 Lx[length(Age)] <- lx[length(Age)]*ax[length(Age)]

 # person-years remaining for individuals aged x
 for(i in 1:(length(Age)-1)){
  Tx[i] <- sum(Lx[i:length(Age)])
 }

 # calculate the remaining life expectancy
 ex <- Tx/lx

 # put all values in a dataframe with appropriate decimals
 lifetable <- data.frame(Age=Age,mx=round(mx,5),qx=round(qx,5),
                         ax=round(ax,2),lx=round(lx,0),dx=round(dx,0),
Lx=round(Lx,0),Tx=round(Tx,0), ex=round(ex,2))
 return(lifetable)
}

wLT <- lifetable(mx=wmx,Age=0:110,sex="m")
bLT <- lifetable(mx=bmx,Age=0:110,sex="m")

# creating a function to use a spline interpolation of the survivorship curve
# this is only needed to estimate the IQR in ages at death to decimal point accuracy

intFUN <- function(x, y, n, n.grid=10000){
 fit <- spline(x=x, y=y, n=n.grid)
 xi <- fit$x
 yi <- fit$y
 xn <- xi[which.min(abs(yi-n))]
 return(xn)
}
```

```
Age=0:110

wage25 <- intFUN(x=Age, y=wLT$lx, n=75000)
wage75 <- intFUN(x=Age, y=wLT$lx, n=25000)
wiqr <- wage75-wage25

bage25 <- intFUN(x=Age, y=bLT$lx, n=75000)
bage75 <- intFUN(x=Age, y=bLT$lx, n=25000)
biqr <- bage75-bage25
wdx <- wLT$dx/1000
bdx <- bLT$dx/1000


#---------plot

x11(width=7,height=4)
par(mar=c(4.5,4.5,4,1))

Age <- 0:110
plot(Age,wLT$dx/1000,xlab="Age",ylab="Percent of Deaths",
    t="l",col="blue",lwd=4,cex.lab=1.5,axes="F",ylim=c(0,4))
lines(Age,bLT$dx/1000,col="red",lwd=4)
axis(side=1, at=c(seq(0,100,20),110), labels=c(seq(0,100,20),""),cex.axis=1.3,lwd=2, las=1)
axis(side=2, at=seq(0,4,2),labels=seq(0,4,2),cex.axis=1.3,lwd=2, las=2)
abline(v=wLT$ex[1],col="blue",lwd=4,lty=6)
abline(v=bLT$ex[1],col="red",lwd=4,lty=6)

xpol = c(69,69:88,88)
ypol = c(0,wdx[70:89],0)
polygon(xpol,ypol,col=alpha("blue",0.2), border=NA )
polygon(c(63,63:85,85), c(0,bdx[64:86],0),col=alpha("red",0.2), border=NA )

text(wLT$ex[1]+18,3.9, "LE = 76.6",cex=1.3,col="blue")
text(bLT$ex[1]-18,3.9, "LE = 72.2",cex=1.3,col="red")
text(30,1,"IQR = 22.4 years",col="red",cex=1.3)
text(47,2.5,"IQR = 19.2 years",col="blue",cex=1.3)

arrows(x1=(wLT$ex[1]+6),x0=(wLT$ex[1]+2),code=1,y0=3.9,col="blue",lwd=3,length=0.05
)
arrows(x1=(bLT$ex[1]-6),x0=(bLT$ex[1]-2),code=1,y0=3.9,col="red",lwd=3,length=0.05)
arrows(x1=85,x0=66,code=2,y0=2.5,y1=3,col="blue",lwd=3,length=0.05)
arrows(x1=67,x0=48,code=2,y0=1,y1=1,col="red",lwd=3,length=0.05)

legend("topleft",bty='n',legend=c("blacks","whites"),lwd=3,col=c("red","blue"),cex=1.5)
```