Trends in life expectancy and lifespan variation by level of education in Europe and the USA

Alyson A. van Raalte[1], Rosie Seaman[1,2], José Rubio Valverde[3], Isaac Sasson[4], Wilma Nusseldar[3], Johan P. Mackenbach[3] and the LIFEPATH consortium[5]

[1]Max Planck Institute for Demographic Research, Rostock Germany
[2]University of Stirling, Stirling Scotland
[3]Erasmus University, Rotterdam the Netherlands
[4]Tel Aviv University, Tel Aviv Israel
[5]https://www.lifepathproject.eu/. The full list of participating authors will be available at the time of EPC.

Short Abstract

In Finland, Denmark, Spain and the USA, lifespan variation has increased among the most disadvantaged socioeconomic groups while declining among the most advantaged groups. This has occurred despite increases in life expectancy for all groups, except in the USA. Currently it is unknown if such trends are unique to the aforementioned countries or are ubiquitous across Europe. In this study we examine trends in life expectancy and lifespan variation for around a dozen countries or regions in Europe and the USA, covering between two to five decades depending on the population. We find that widening differentials in lifespan variation is a fixture of almost all European countries and the USA, particularly when using an absolute measure of lifespan variation. Trends in lifespan variation are increasingly becoming divorced from trends in life expectancy. This furthers the argument that we should be monitoring both metrics for a full picture of population health.

Long Abstract

Introduction

Lifespan variation, i.e. the spread in ages at death across all individuals in the population, is receiving increasing attention as an important metric of population health (van Raalte, Sasson, and Martikainen 2018). For much of recent history it has been inversely correlated with life expectancy. In other words, populations have experienced increasing average ages at death with declining variability (Wilmoth and Horiuchi 1999, Vaupel, Zhang, and van Raalte 2011, Smits and Monden 2009, Colchero et al. 2016, Németh 2017, Permanyer and Scholl 2019).

There are exceptions however, notably across socioeconomic groups. In Finland (van Raalte, Martikainen, and Myrskyla 2014, van Raalte, Sasson, and Martikainen 2018), Denmark (Brønnum-Hansen 2017), Spain (Permanyer et al. 2018) and the USA (Sasson 2016), lifespan variation has increased among the most disadvantaged socioeconomic groups while declining among the most advantaged groups. The same pattern is evident in Scotland for quintiles of area-level deprivation (Seaman et al. 2019). These diverging patterns in lifespan variation have occurred despite increases in life expectancy for all groups, except in the USA. Such findings are especially worrisome because lifespan variation translates to uncertainty in the timing of death at the individual level, meaning that ages at death are becoming increasing less predictable for the lower social classes, but more predictable for the higher classes. At the population level it implies that the lower classes are becoming more heterogeneous while the upper classes are becoming more homogenous.

Currently it is unknown if such trends are unique to the aforementioned countries or are ubiquitous across a wider range of high longevity countries. In this study we examine trends in life expectancy and lifespan variation for more than a dozen countries or regions in Europe as well as the USA, covering a time period of between two and five decades depending on the country.

Data and Methods

Data

For European data, census-linked death and exposure counts by country or region, sex, time period, five-year age group (ages 30-34, … , 80-84, 85+) and level of education (Low=ISCED[1] 0-2, Medium=ISCED 3-4, High=ISCED 5-6) were aggregated and harmonized by the LIFEPATH consortium[2]. These data were collected in a longitudinal study design between national censuses, to maximize portability across

[1] International Standard Classification of Education 1997: http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm
[2] https://www.lifepathproject.eu/

different study objectives. Censuses were typically held 10 years apart. According to this design, subjects in each of the five-year age groups at the time of each census were followed for the intercensal period, which was split into two 5 year-periods. Deaths were coded according to the observed age at death, and person-years split and distributed into the appropriate 5-year age, 5-year period category [3]. As a result of this study design, in the second 5-year period of each intercensal interval, there were no exposures in the 30-34 year age group, and the 35-39 year olds were entirely contained within the upper Lexis triangle. To minimize bias, we present results from ages 40 plus, for which complete data is available in the usual age-period square formats.

The periods under analysis vary from population to population, with the longest time series beginning in the 1970s. We excluded country time-periods when data was not available for the 85+ age category, or when the missing data on education was unevenly distributed across age. For Poland, only unlinked data was available, opening up the possibility of numerator-denominator biases known to result from using unlinked data (Shkolnikov et al. 2007, Lerch et al. 2017). For now we have included this country in our study, but may drop it at a later stage of the analysis if it does not pass planned further consistency checks. For England & Wales, education was grouped differently with low=ISCED 0-3 and high=ISCED 4-6. No medium educational category was created. For most countries or regions we had data on the entire population, but for England & Wales we only had a 1% sample.

For the USA, we will use previously published harmonized data by education (Sasson 2016, Sasson and Hayward 2019), grouped into the same format as the European data, from the early 1990s to mid-2010s for all racial groups combined. Unfortunately this dataset is also unlinked, but alternative existing survey-based datasets do not provide a sufficient sample size to adequately estimate lifespan variation by level of education within reasonable confidence intervals (Sasson 2017). This data has not yet been included in our dataset, but will be by the time of the EPC.

Methods

For each sex, country or region, and period data was smoothed using a non-parametric penalized composite link model (pclm), designed for ungrouping binned data (Rizzi, Gampe, and Eilers 2015). The pclm treats death counts as poisson-distributed with the population exposure as offset, and maximizes a penalized likelihood by an iteratively reweighted least-squares algorithm. Optimal values of the smoothing parameter were found by minimizing the Bayesian Information Criterion. The analysis was done using the R-package 'ungroup' (Pascariu et al. 2018). The output of this model was smoothed death rates by single year of age from age 40 to a specified maximum final age, detailed below.

Indices of lifespan variation are more sensitive to changes at the tails of the age-at-death distribution than life expectancy (van Raalte and Caswell 2013). For this reason, the most critical modeling decision was how to close the life table, given a relatively young open-ended age 85+, which often fell below the modal age at death in our data. Death rates at the most advanced ages are generally assumed to

plateau due to differences in individual frailty (Vaupel, Manton, and Stallard 1979, Barbi et al. 2018), however this theory remains controversial (Gavrilov and Gavrilova 2019).

Within the pclm modeling framework, changing the shape of mortality at advanced ages is made by changing the maximum age parameter of the open-aged interval. Since this value is unknown, we experimented with 4 different maximum ages: 100, 110, 115 and 120. A maximum age of 100 would consistently produce rates that increased in a more or less exponential manner and 120 would impose a plateau, with intermediate ages producing intermediate choices.

The model was tested by binning data from the Human Mortality Database (HMD 2019) into the same age format of our LIFEPATHS data, and comparing the output derived from the pclm with the observed rates over ages 85-89, 90-94, 95-99, 100-104 from the HMD, as well as to the modeled HMD life expectancies conditional upon survival to age 40 (e_{40}), for the same countries and periods as were present in our dataset. Generally the pclm model closely matched the observed death rates from the HMD at higher ages when the maximum age chosen was 110, 115 or 120, and e_{40} values from life tables based on the pclm models were in almost all cases within 0.3 years of the HMD estimates, and usually within 0.1 years. Age 115 resulted in the closest match overall (so far based on eyeballing, by EPC a more systematic evaluation will be done). On the other hand, a maximum age of 100 resulted in death rates that were consistently higher than observed death rates over the 90+ age range, and e_{40} estimates that were sometimes as much as 1 year lower than the HMD estimates. However, we should bear in mind that the HMD estimated life expectancies themselves are based on modeled death rates at older ages (beyond age 95 or when male death counts first fall below 100), from a parametric Kannisto model which imposes a plateau (Wilmoth et al. 2017).

Thus for each country or region, sex, period, and educational level, four sets of modeled single year of age death rates were created by varying the maximum age as above. These death rates were used as inputs to life tables. From the life table age-at-death distribution, we calculated life expectancy conditional upon survival to age 40 (the mean of this distribution), as well as two indices of variability, the standard deviation and the Gini coefficient. The standard deviation measures absolute inequality in ages at death and is expressed in years of age, while the Gini coefficient measures inequality relative to mean ages at death, with higher inequality levels signaling greater variation in age at death.

In what follows, the results are based on a maximum age of 115 from our pclm model. Robustness checks with other maximum ages will be available by EPC. For now, we present and discuss male results only, with female results available in the appendix. In most cases the trends are qualitatively similar, though levels differ with men having lower life expectancy and higher lifespan inequality.

Results

Figure 1 shows trends in the male period average age at death conditional upon survival to age 40 (i.e. e_{40} +40). Apart from the medium and low educated in Lithuania over the 2000s, and all groups in

England & Wales during the 1980s, all educational subgroups experienced monotonically increasing life expectancy during the periods under study. Differences between educational groups were for the most part remarkably constant during the long periods under observation, although the Nordic countries and Lithuania experienced a noticeable increase in inequalities. Inequalities were consistently largest in Central and Eastern Europe and smallest in the southern European regions.

Figure 2 presents trends in the male period standard deviation in age at death conditional upon survival to age 40. In every country, except Lithuania and Estonia, differences in the standard deviation increased between educational groups over the time period observed. This was because the highest educated experienced reductions in the standard deviation in survival age, while the lowest educated experienced increases in the standard deviation. The gap between the highest and lowest educated grew most notably in the Nordic countries and southern European regions.

Figure 3 examines the same pattern as Figure 2, but using a metric of relative variation (the Gini coefficient). Differences between educational groups also widened according to the Gini, but by far less than when measured by the standard deviation. Overall trends were mixed for the Gini, with most low educated groups exhibiting stagnation or small declines in lifespan variation, and most high educated showing steeper declines.

Figures 4 and 5 show the association between life expectancy and lifespan variation within each country for the standard deviation (Fig 4) and the Gini coefficient (Fig 5). Here country differences become more apparent. In some countries, such as Finland, the trajectories of lifespan variation at different levels of life expectancy are clearly diverging, whereas in others, it depends on whether absolute or relative metrics of lifespan variation are examined. Low educated groups showed much higher lifespan variation levels at similar life expectancy levels according to the absolute measure (standard deviation), but differences were more modest according to the relative measure (Gini coefficient).

Figure 6 presents the association between life expectancy and lifespan variation measured by standard deviation for different educational groups. The negative association between life expectancy and lifespan variation is strong for the high educated, but virtually non-existent for the low educated. This can also be seen by following the points longitudinally---for the high educated there is a clear movement from the top left to the bottom right of the figure for each country (indicated the line joining a group of points). For the low educated, the trend lines indicate a positive association, i.e. increasing lifespan variation with increasing average ages at death. Figure 7 presents the same association but with lifespan variation measured by the Gini coefficient. A negative correlation between the two metrics is clearly evident for all educational groups, but the correlation is stronger for the high educated, who have also experienced sharper declines in the Gini between average ages at death of 70 and 75. Also worth noting, in both Fig 6 and Fig 7, differences between countries in lifespan variation were substantially larger for the low educated than for the high educated.

Discussion

Summary

We found diverging trends in lifespan variation in all countries examined, even in instances when life expectancy was increasing in a near-parallel fashion for all educational subgroups. This provides the most comprehensive evidence to date that the highest educated are becoming increasingly homogenous in their survival profiles, allowing them to more effectively plan their working life, their retirement years, and ultimately their own passing. The opposite is occurring for the low educated. Although the average age at death is continuing to increase in most countries, often at the same pace as the highest educated, the variation is increasing when measured by the standard deviation, and either slowly declining or stagnating when measured by the Gini coeffient.

Limitations

Our estimates are model based and given the abridged nature of the data, the choice of model is critical. We chose to model death rates non-parametrically, which follows the data points closely, thus assuming minimal data errors. An alternative would have been a parametric smoothing approach, for instance from a Gamma-Gompertz model, which has also been shown to ungroup abridged life tables reasonably well (Missov, Nemeth, and Danko 2016, Németh and Missov 2018). However the close correspondence between the 'b' parameter of the Gompertz model and lifespan variation (Tuljapurkar and Edwards 2011, Vaupel 1986, Wrycza 2014) cautioned us against this approach, which would likely impose a higher inverse correlation between life expectancy and lifespan variation than the reality.

Caution must always be taken in interpreting long term trends in life expectancy by level of education, given the substantial expansion of education that has taken place in the past 50 years (Hendi 2015, Dowd and Hamoudi 2014). For instance, among Finnish females in 2010-2014, 9 percent of females aged 85+ were high educated compared to 53 percent aged 40-44 in our data. Using a life table approach implicitly assumes that these differences do not matter. It is noteworthy, however, that although the highest educated groups have experienced the largest proportional changes in their group membership, they are also the group that is consistently becoming more homogenous in their age-at-death profiles, while the lowest educated are becoming increasingly heterogeneous.

Future work between now and EPC

Between now and the EPC, we will run more diagnostic checks on the modeled death rates. The lead author has access to these same Finnish educational data by single year of age up to age 95+. While our assumptions seem to hold well when tested with national data from the HMD, we will additionally test whether the model works equally well for all educational subgroups and time periods in Finland and investigate alternatives. Moreover, we will compare our model-based estimates of life expectancy and

lifespan variation at the national level (i.e. all educational groups combined) with the corresponding values for each period from the HMD. Second, since lifespan variation is known to be sensitive to the starting age (Robine 2001, Engelman, Canudas-Romo, and Agree 2010, Seaman et al. 2019), we will also run all analysis from ages 35 onward, either making corrections to the data points with the missing lower Lexis triangle described in the data section, or omitting the effected data points positioned at the latter half of each census. Third, we will add the USA to our results. Fourth, we will run additional analyses to gain a better understanding of the role played by the changing educational composition on all of the trends presented, as well as the different trends found using the standard deviation and Gini coefficients. Finally, for all of these countries, we have data by cause of death, which we have grouped into 4 main causes: circulatory disease, cancer, external mortality and other. We will run decompositions to better understand the differential impact of these causes of death on trends in life expectancy and lifespan variation.

Conclusions

We find that widening differentials in lifespan variation is indeed a fixture of almost European countries, and that trends in lifespan variation are increasingly becoming divorced from trends in life expectancy. This furthers the argument that we should be monitoring both metrics for a full picture of population health.

Acknowledgements

AvR and RS are funded by a starting grant from the European Research Council to AvR (Grant # 716323). Data harmonization was conducted under the LIFEPATHS and DEMETRIQ projects, funded by the European Commission (Grant # 633666 and FP7-CP-FP 278511).

References

Barbi, Elisabetta, Francesco Lagona, Marco Marsili, James W Vaupel, and Kenneth W Wachter. 2018. "The plateau of human mortality: Demography of longevity pioneers." *Science* 360 (6396):1459-1461.

Brønnum-Hansen, Henrik. 2017. "Socially disparate trends in lifespan variation: a trend study on income and mortality based on nationwide Danish register data." *BMJ Open* 7 (5):e014489.

Colchero, Fernando, Roland Rau, Owen R Jones, Julia A Barthold, Dalia A Conde, Adam Lenart, Laszlo Nemeth, Alexander Scheuerlein, Jonas Schoeley, and Catalina Torres. 2016. "The emergence of longevous populations." *Proceedings of the National Academy of Sciences* 113 (48):E7681-E7690.

Dowd, JB, and A Hamoudi. 2014. "Is life expectancy really falling for groups of low socio-economic status? Lagged selection bias and artefactual trends in mortality." *International journal of epidemiology* 43 (4):983.

Engelman, Michal, Vladimir Canudas-Romo, and Emily M. Agree. 2010. "The Implications of Increased Survivorship for Mortality Variation in Aging Populations." *Population and Development Review* 36 (3):511-539.

Gavrilov, Leonid A, and Natalia S Gavrilova. 2019. "Late-life mortality is underestimated because of data errors." *PLoS biology* 17 (2):e3000148.

Hendi, Arun S. 2015. "Trends in US life expectancy gradients: the role of changing educational composition." *International Journal of Epidemiology* 44 (3).

HMD. 2019. Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. Data downloaded on May 3, 2019.

Lerch, Mathias, Adrian Spoerri, Domantas Jasilionis, and Francisco Viciana Fernandèz. 2017. "On the plausibility of socioeconomic mortality estimates derived from linked data: a demographic approach." *Population health metrics* 15 (1):26.

Missov, Trifon I, Laszlo Nemeth, and Maciej J Danko. 2016. "How much can we trust life tables? Sensitivity of mortality measures to right-censoring treatment." *Palgrave Communications* 15049 (doi:10.1057/palcomms.2015.49).

Németh, László. 2017. "Life expectancy versus lifespan inequality: A smudge or a clear relationship?" *PloS one* 12 (9):e0185702.

Németh, László, and Trifon I. Missov. 2018. "Adequate life-expectancy reconstruction for adult human mortality data." *PLOS ONE* 13 (6):e0198485. doi: 10.1371/journal.pone.0198485.

Pascariu, Marius D, Maciej J Dańko, Jonas Schöley, and Silvia Rizzi. 2018. "ungroup: An R package for efficient estimation of smooth distributions from coarsely binned data." *The Journal of Open Source Software* 3 (29):937. doi: 10.21105/joss.00937

Permanyer, Iñaki, and Nathalie Scholl. 2019. "Global trends in lifespan inequality: 1950-2015." *PloS one* 14 (5):e0215742.

Permanyer, Iñaki, Jeroen Spijker, Amand Blanes, and Elisenda Renteria. 2018. "Longevity and Lifespan Variation by Educational Attainment in Spain: 1960–2015." *Demography*. doi: 10.1007/s13524-018-0718-z.

Rizzi, Silvia, Jutta Gampe, and Paul HC Eilers. 2015. "Efficient estimation of smooth distributions from coarsely grouped data." *American journal of epidemiology* 182 (2):138-147.

Robine, Jean-Marie. 2001. "Redefining the Stages of the Epidemiological Transition by a Study of the Dispersion of Life Spans: The Case of France." *Population: An English Selection* 13 (1):173-193.

Sasson, Isaac. 2016. "Trends in Life Expectancy and Lifespan Variation by Educational Attainment: United States, 1990–2010." *Demography* 53 (2):269-293.

Sasson, Isaac. 2017. "Reply to Trends in Education-Specific Life Expectancy, Data Quality, and Shifting Education Distributions: A Note on Recent Research." *Demography* 54 (3):1215-1219. doi: 10.1007/s13524-017-0583-1.

Sasson, Isaac, and Mark D Hayward. 2019. "Association Between Educational Attainment and Causes of Death Among White and Black US Adults, 2010-2017." *Jama* 322 (8):756-763.

Seaman, Rosie, Tim Riffe, Alastair H Leyland, Frank Popham, and Alyson van Raalte. 2019. "The increasing lifespan variation gradient by area-level deprivation: A decomposition analysis of Scotland 1981–2011." *Social Science & Medicine* 230:147-157.

Shkolnikov, Vladimir M., Domantas Jasilionis, Evgeny M. Andreev, Dmitri A. Jdanov, Vladislava Stankuniene, and Dalia Ambrozaitiene. 2007. "Linked versus unlinked estimates of mortality and length of life by education and marital status: Evidence from the first record linkage study in Lithuania." *Social Science & Medicine* 64 (7):1392-1406.

Smits, Jeroen, and Christiaan Monden. 2009. "Length of life inequality around the globe." *Social Science & Medicine* 68 (6):1114-1123.

Tuljapurkar, Shripad, and Ryan D Edwards. 2011. "Variance in death and its implications for modeling and forecasting mortality." *Demographic Research* 24 (21):497-526.
van Raalte, A. A., P. Martikainen, and M. Myrskyla. 2014. "Lifespan variation by occupational class: compression or stagnation over time?" *Demography* 51 (1):73-95. doi: 10.1007/s13524-013-0253-x.
van Raalte, Alyson A., and Hal Caswell. 2013. "Perturbation analysis of indices of lifespan variability." *Demography* 50 (5):1615-1640. doi: http://dx.doi.org/10.1007/s13524-013-0223-3.
van Raalte, Alyson A., Isaac Sasson, and Pekka Martikainen. 2018. "The case for monitoring life-span inequality." *Science* 362 (6418):1002-1004. doi: 10.1126/science.aau5811.
Vaupel, J. W. 1986. "How Change in Age-Specific Mortality Affects Life Expectancy." *Population Studies* 40 (1):147-157.
Vaupel, James W, Kenneth G Manton, and Eric Stallard. 1979. "The impact of heterogeneity in individual frailty on the dynamics of mortality." *Demography* 16 (3):439-454.
Vaupel, James W., Zhen Zhang, and Alyson A. van Raalte. 2011. "Life expectancy and disparity: an international comparison of life table data." *BMJ Open* 1 (1):1-6.
Wilmoth, John R., Kirill Andreev, D. Jdanov, Dana Glei, and Timothy Riffe. 2017. "Methods Protocol for the Human Mortality Database (version 6)."
Wilmoth, John R., and Shiro Horiuchi. 1999. "Rectangularization Revisited: Variability of Age at Death within Human Populations." *Demography* 36 (4):475-495.
Wrycza, Tomasz. 2014. "Entropy of the Gompertz-Makeham mortality model." *Demographic Research* S8 (49):1397-1404.

Results

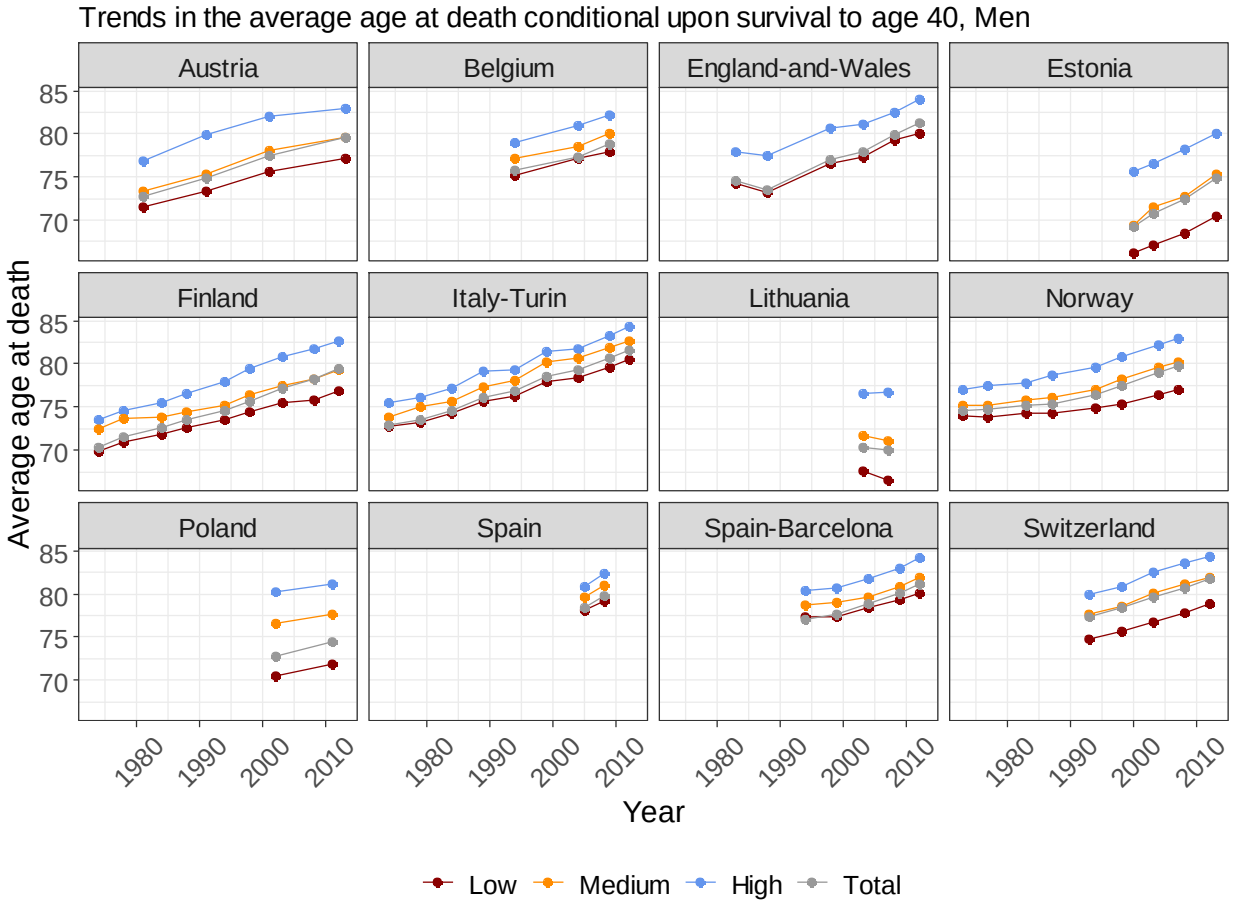

Figure 1: Male trends in the average age at death conditional upon survival to age 40.

Notes: The data points generally depict a 5-year range in years, with the point centered on the middle year.

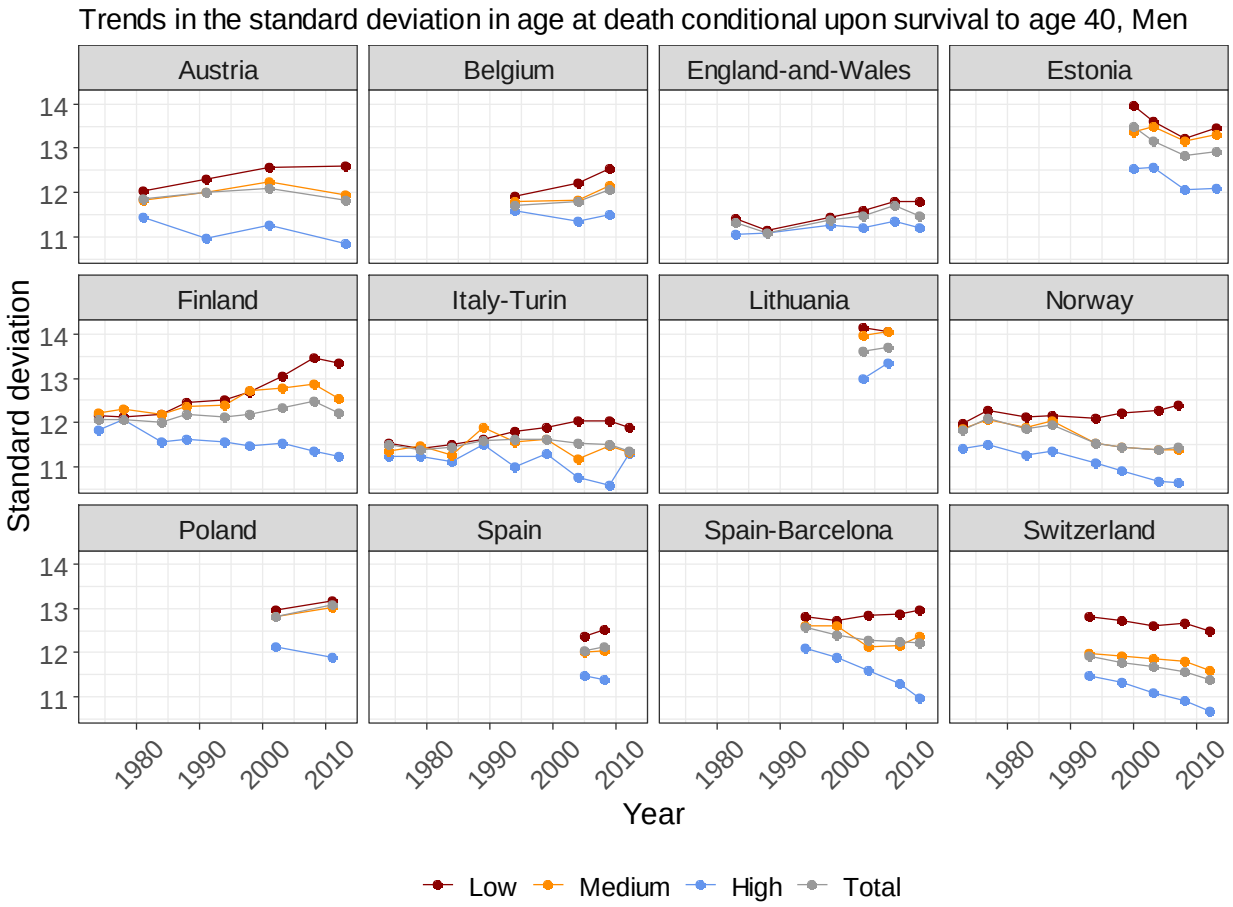

Figure 2: Male trends in the standard deviation in age at death conditional upon survival to age 40.

Notes: The data points generally depict a 5-year range in years, with the point centered on the middle year.

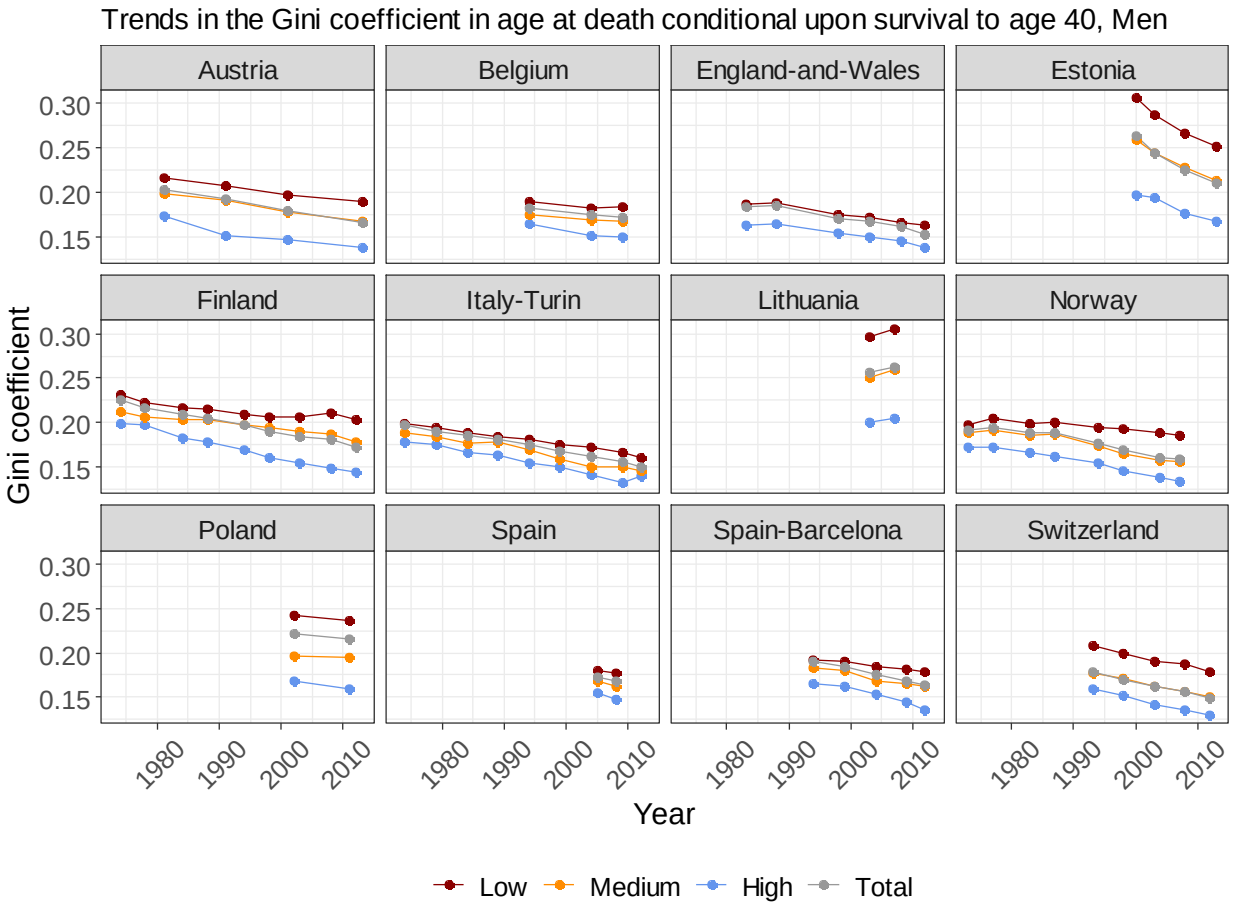

Figure 3: Male trends in the average age at death conditional upon survival to age 40.

Notes: The data points generally depict a 5-year range in years, with the point centered on the middle year.

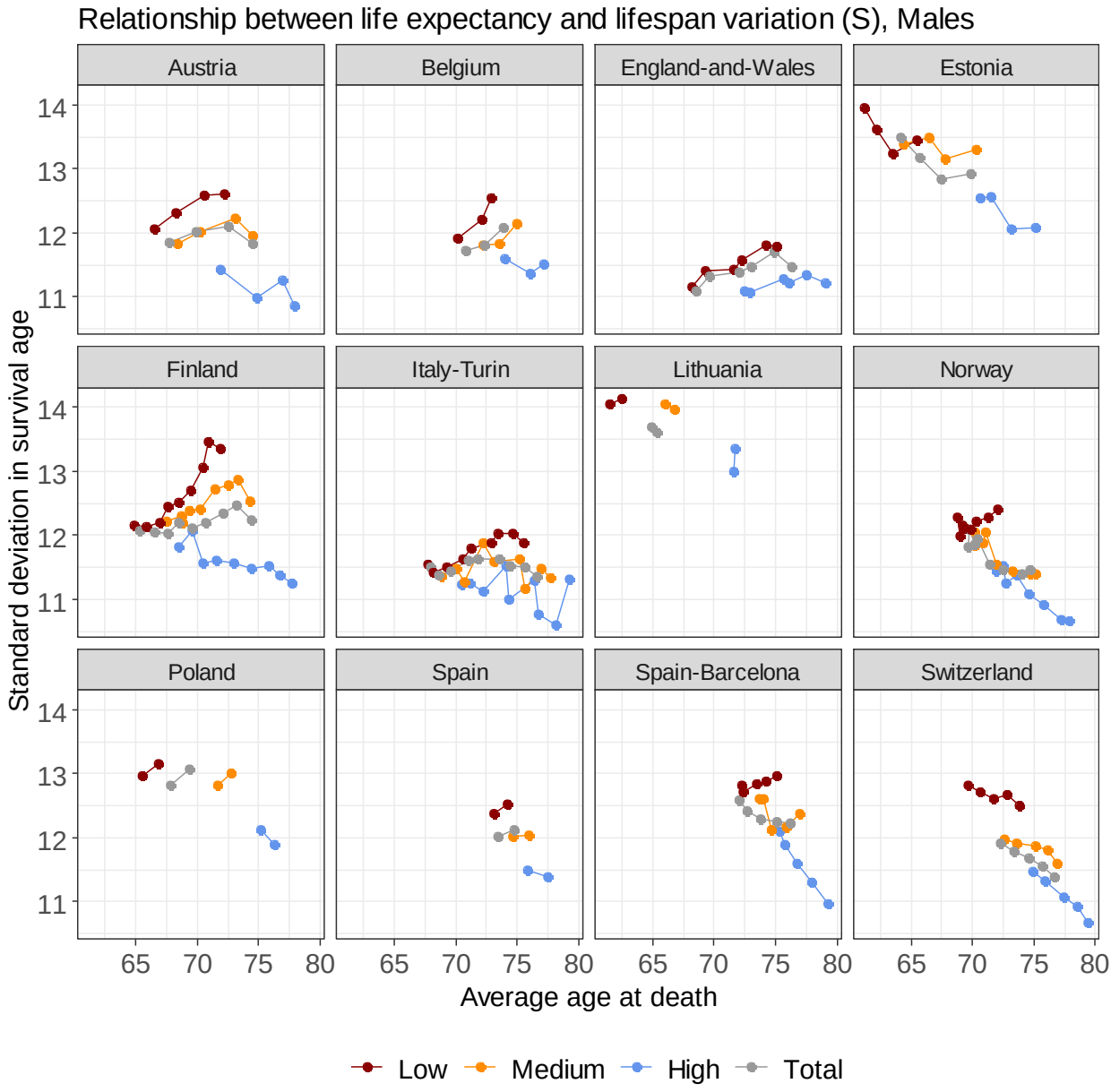

Figure 4: The association between the standard deviation and the average age at death, males

Notes: The data points generally depict a 5-year range in years, with the point centered on the middle year.

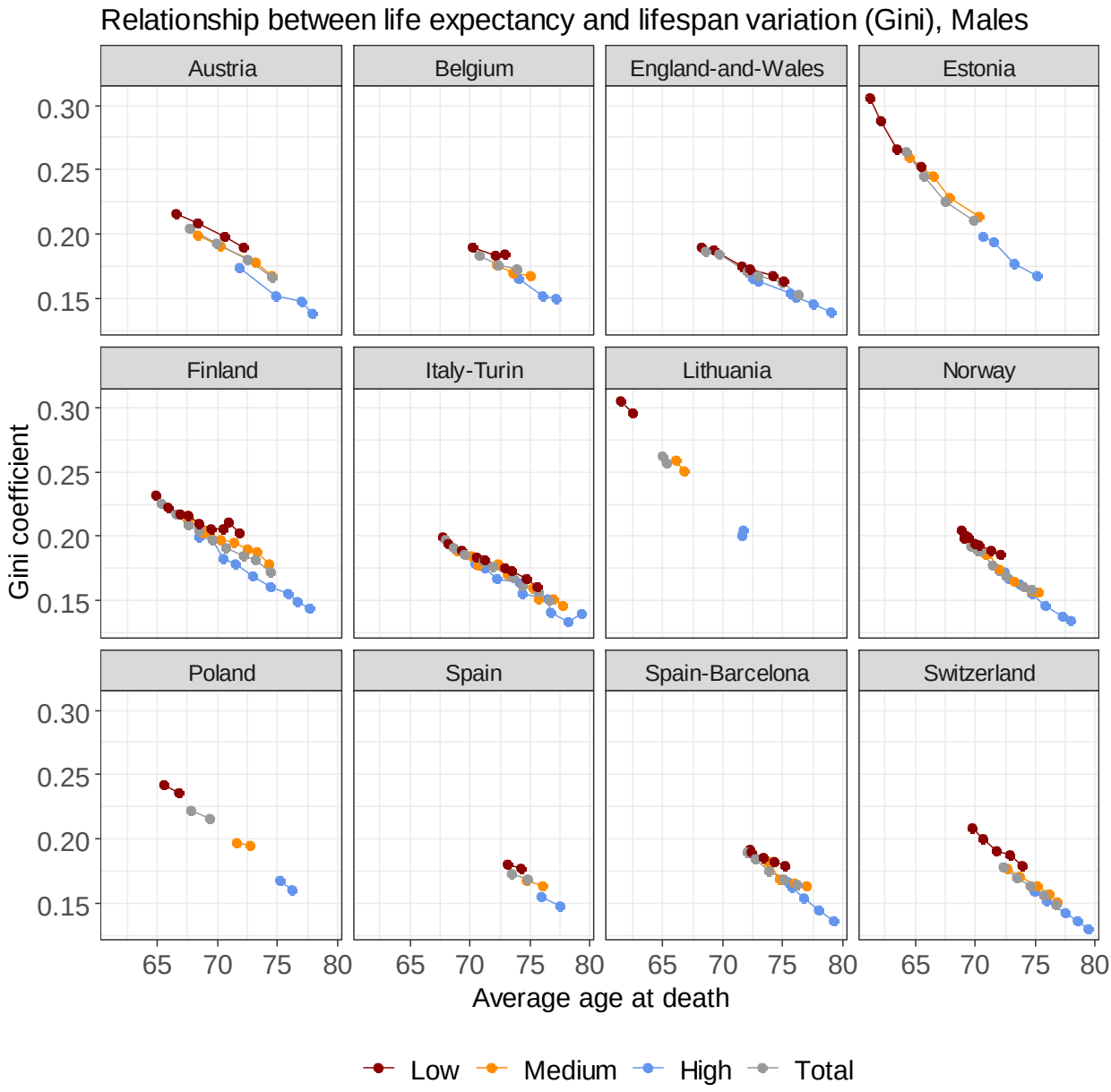

Figure 5: The association between the Gini coefficient and the average age at death, males

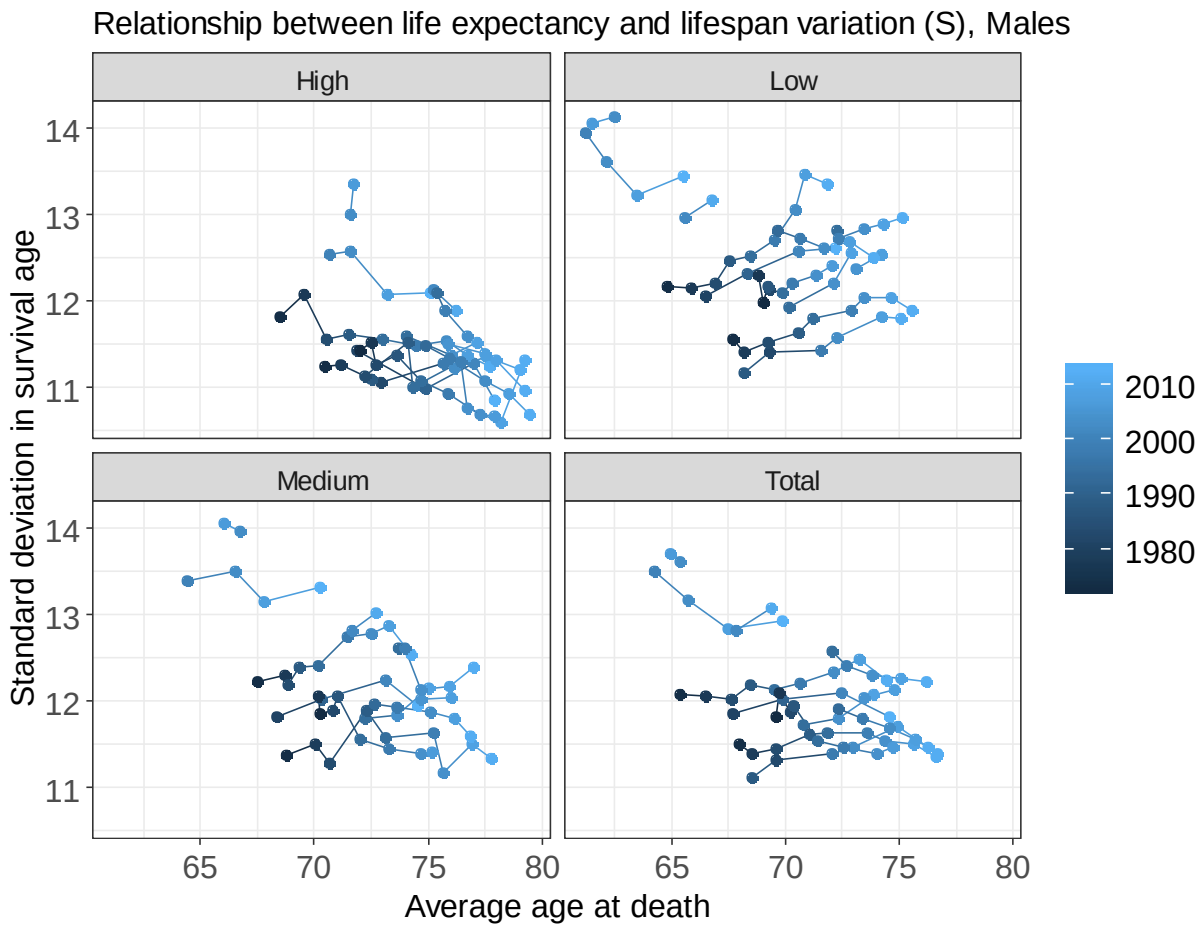

Figure 6: The association between the standard deviation and the average age at death, males

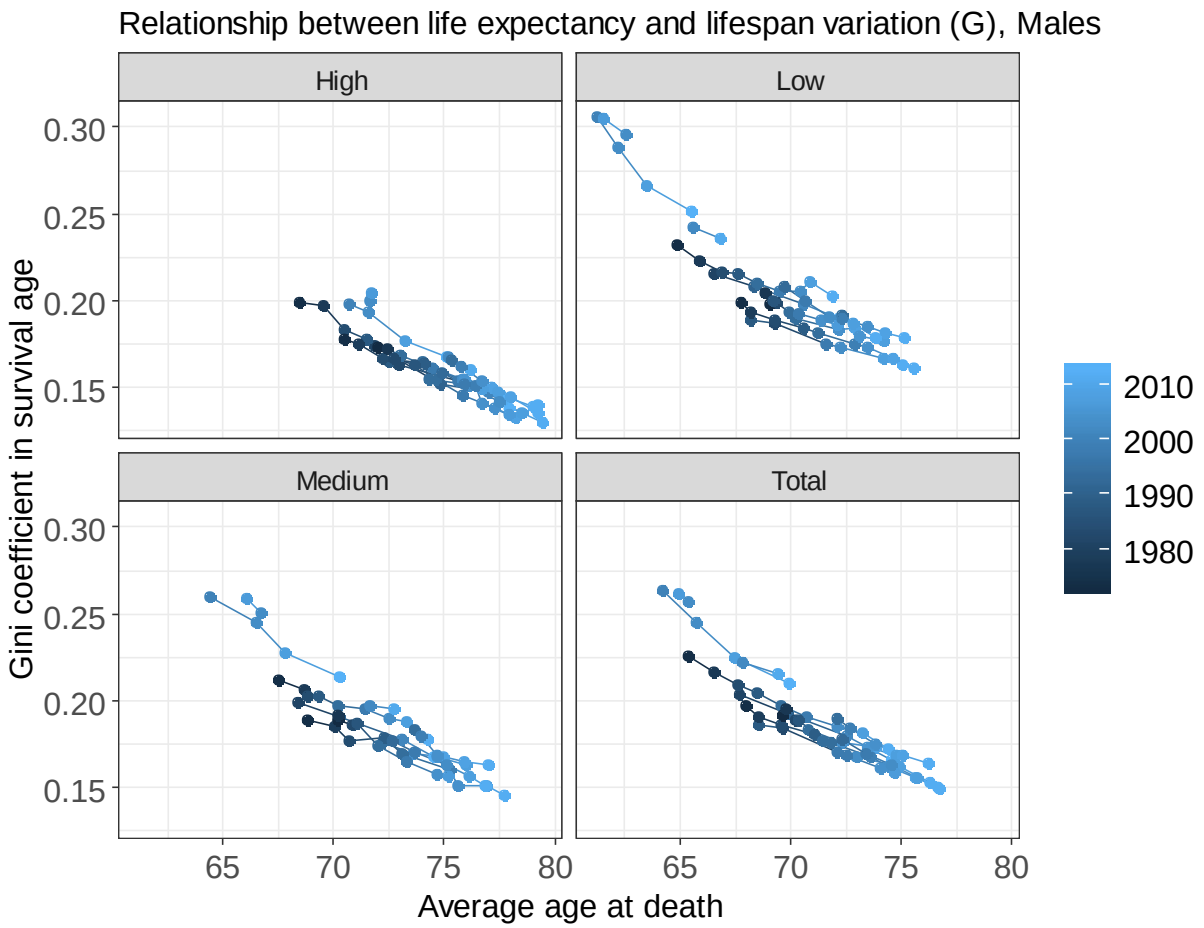

Figure 7: The association between the Gini coefficient and the average age at death, males

Supplementary Figures for Females, corresponding to Figs 1-7

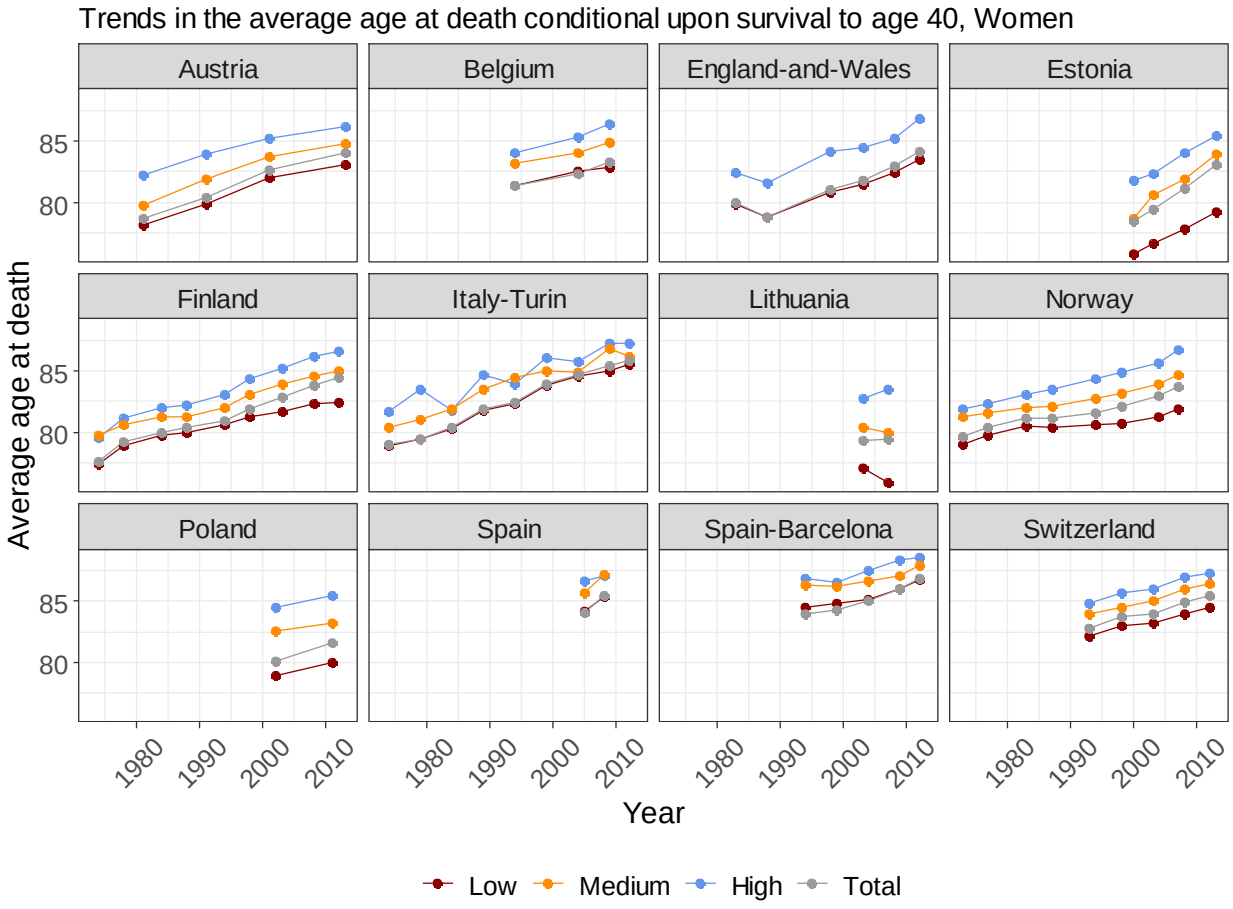

Figure S1.

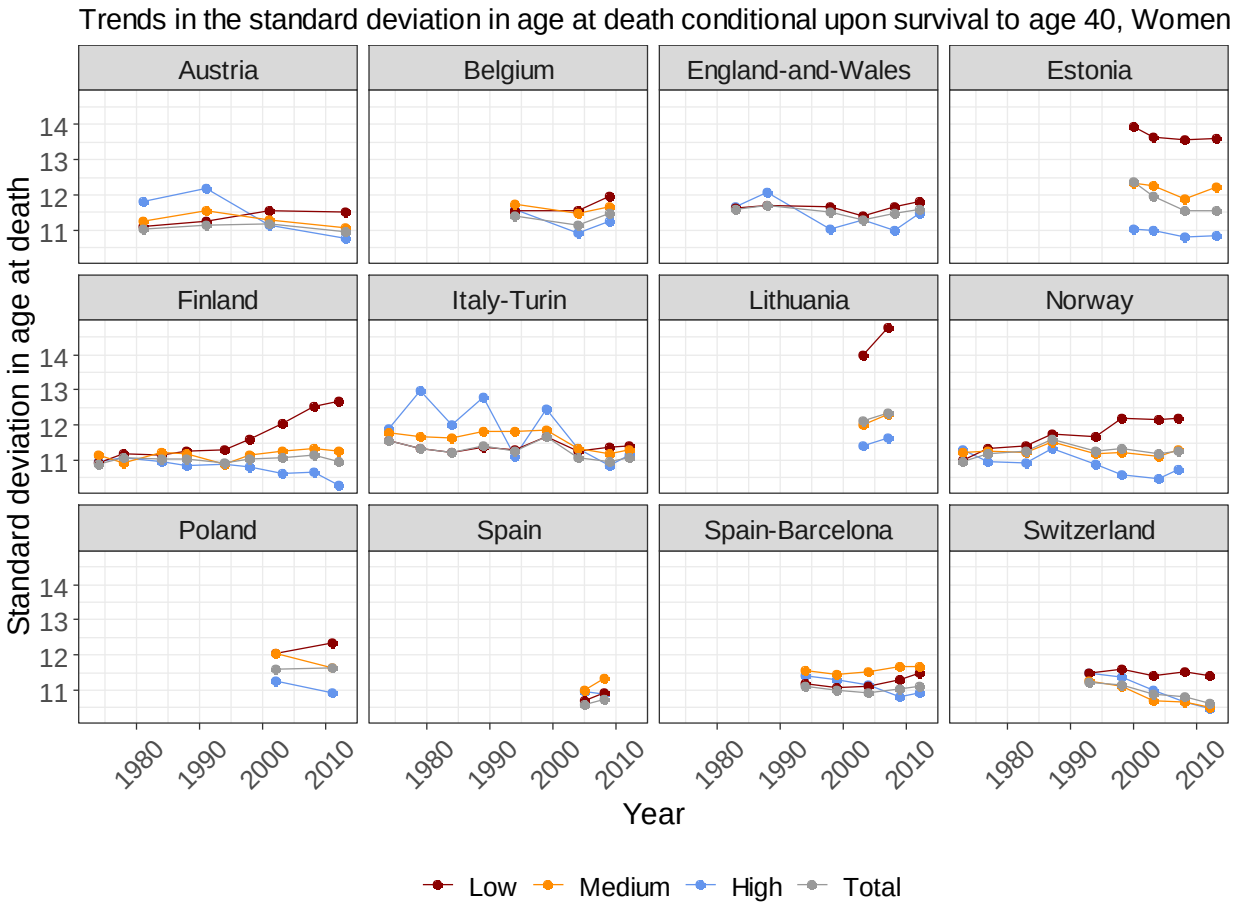
Trends in the standard deviation in age at death conditional upon survival to age 40, Women
Austria
Belgium
England-and-Wales
Estonia
Finland
Italy-Turin
Lithuania
Norway
Poland
Spain
Spain-Barcelona
Switzerland
Standard deviation in age at death
11
12
13
14
1980
1990
2000
2010
Year
Low
Medium
High
Total

Figure S2

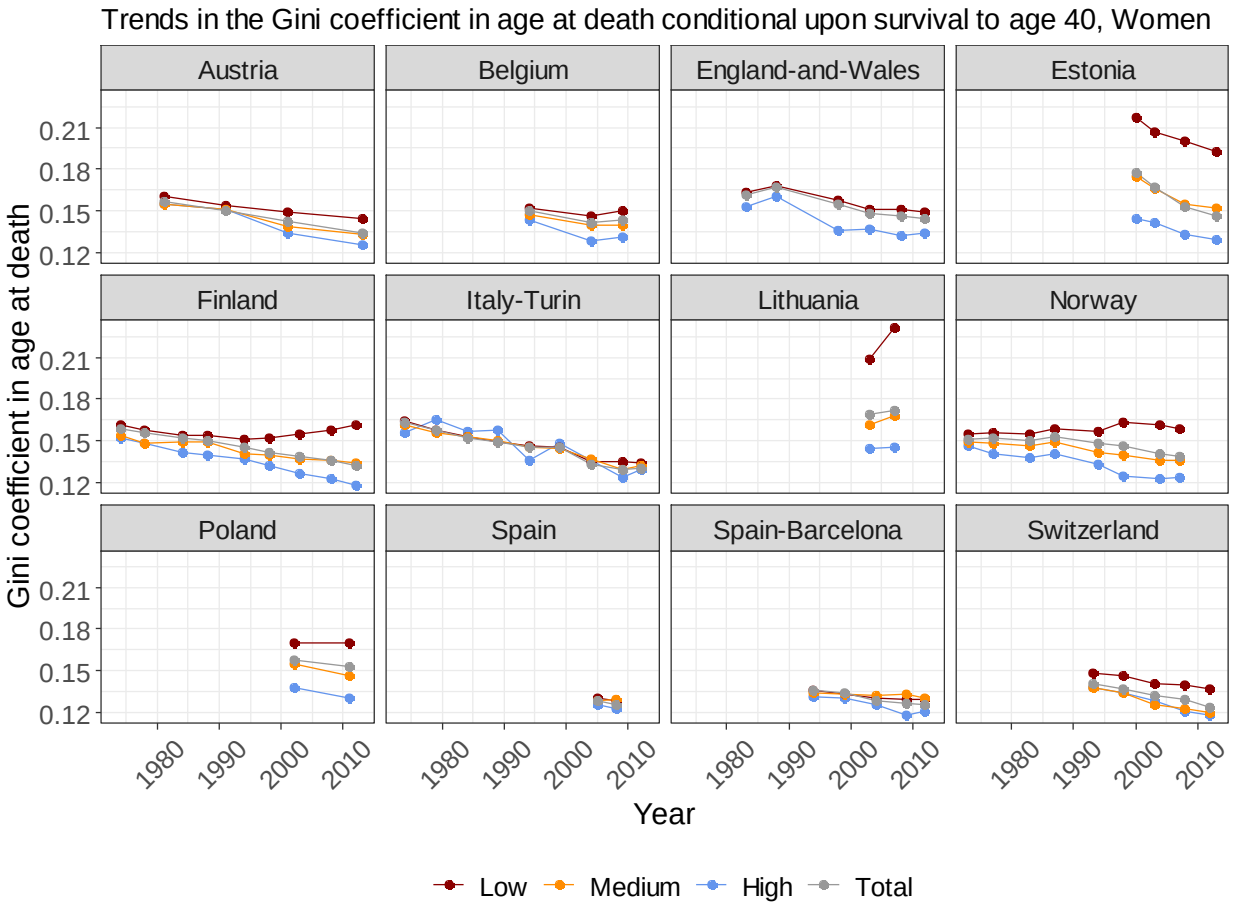
Trends in the Gini coefficient in age at death conditional upon survival to age 40, Women
Austria
Belgium
England-and-Wales
Estonia
Finland
Italy-Turin
Lithuania
Norway
Poland
Spain
Spain-Barcelona
Switzerland
Gini coefficient in age at death
0.21
0.18
0.15
0.12
1980
1990
2000
2010
Year
Low
Medium
High
Total

Figure S3

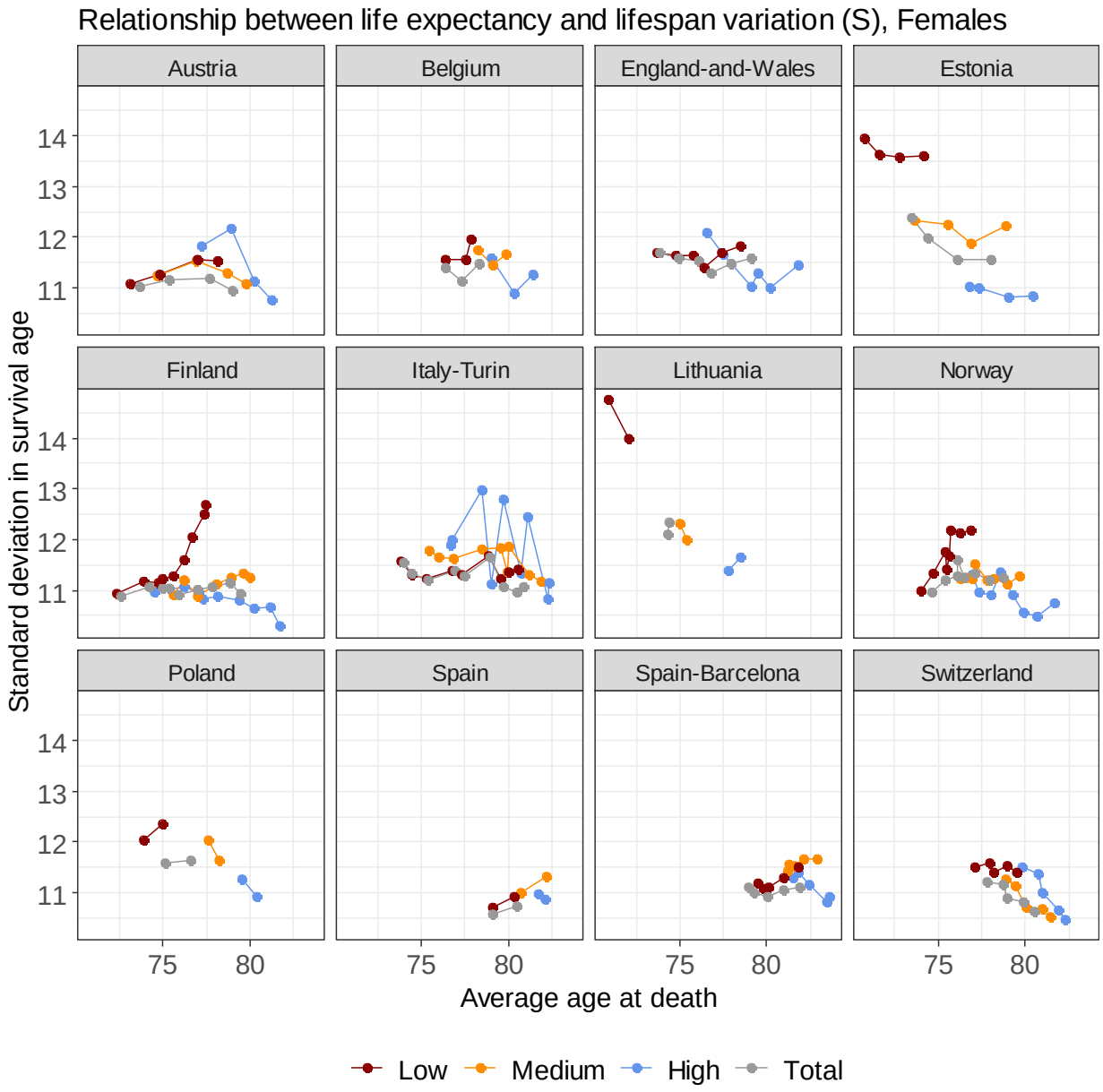

Figure S4: Note: the high educated Turin women are a very small population

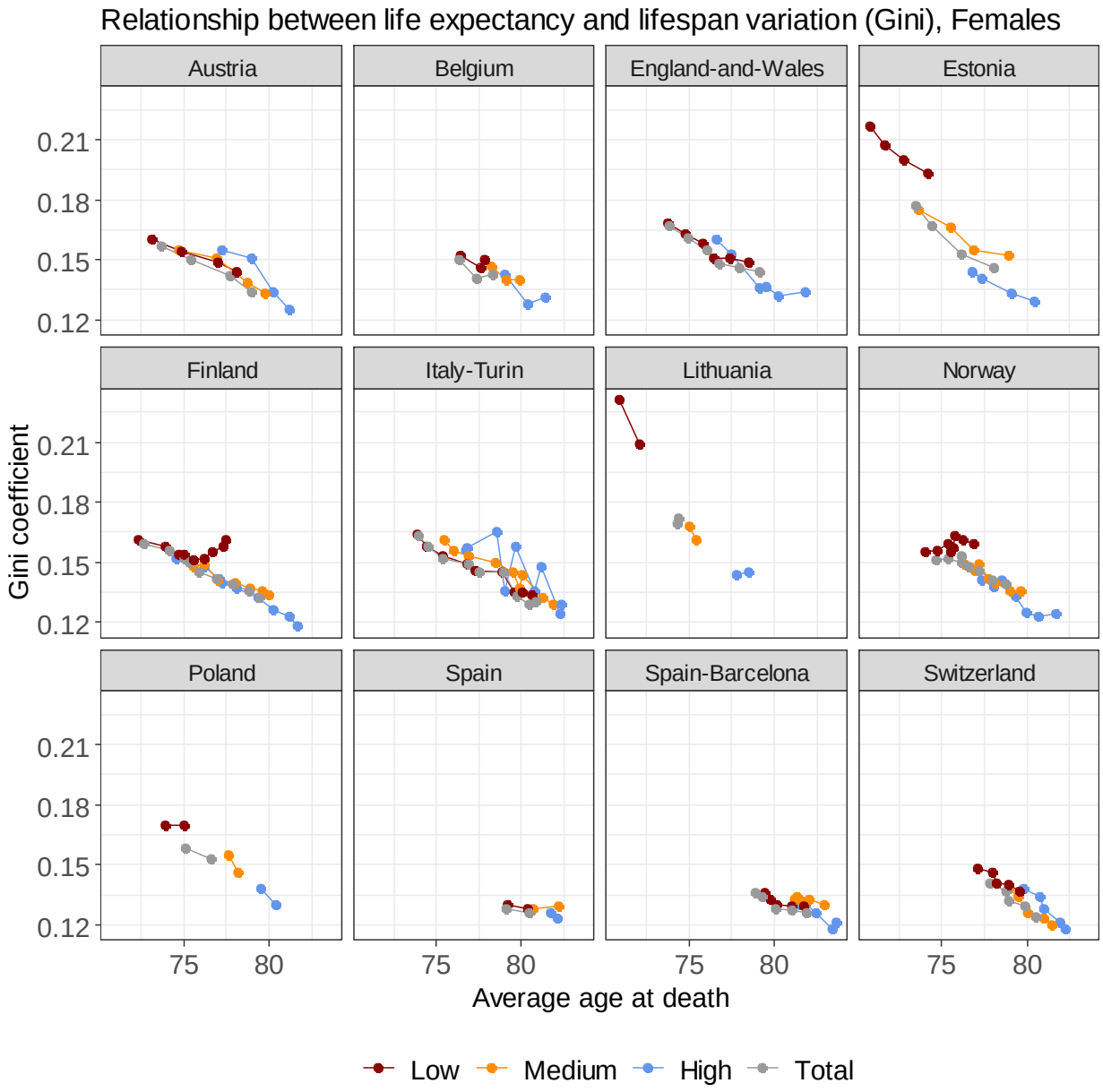
Relationship between life expectancy and lifespan variation (Gini), Females
Austria
Belgium
England-and-Wales
Estonia
Finland
Italy-Turin
Lithuania
Norway
Poland
Spain
Spain-Barcelona
Switzerland
Gini coefficient
0.21
0.18
0.15
0.12
75
80
Average age at death
Low
Medium
High
Total

Figure S5

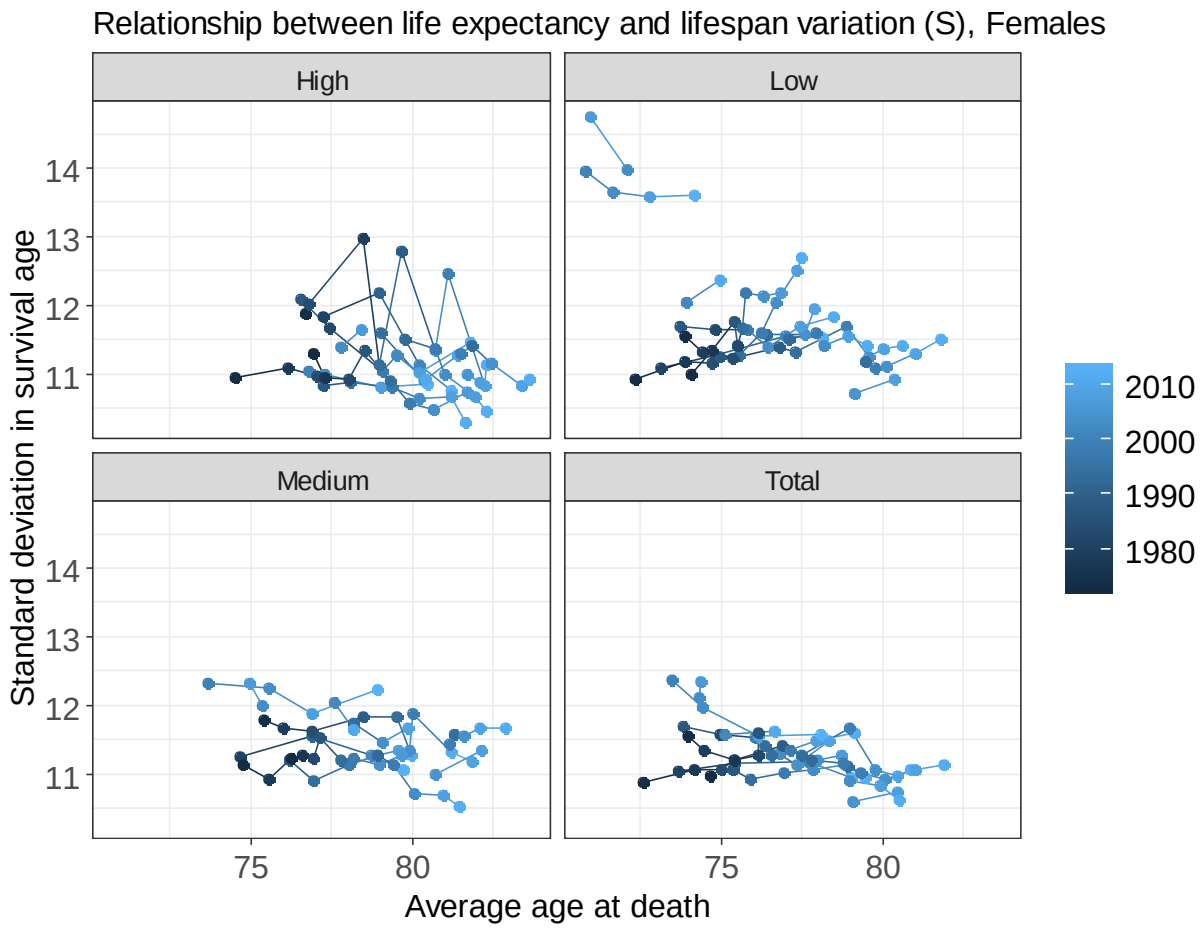
Relationship between life expectancy and lifespan variation (S), Females
High
Low
Medium
Total
Standard deviation in survival age
14
13
12
11
75
80
Average age at death
2010
2000
1990
1980

Figure S6

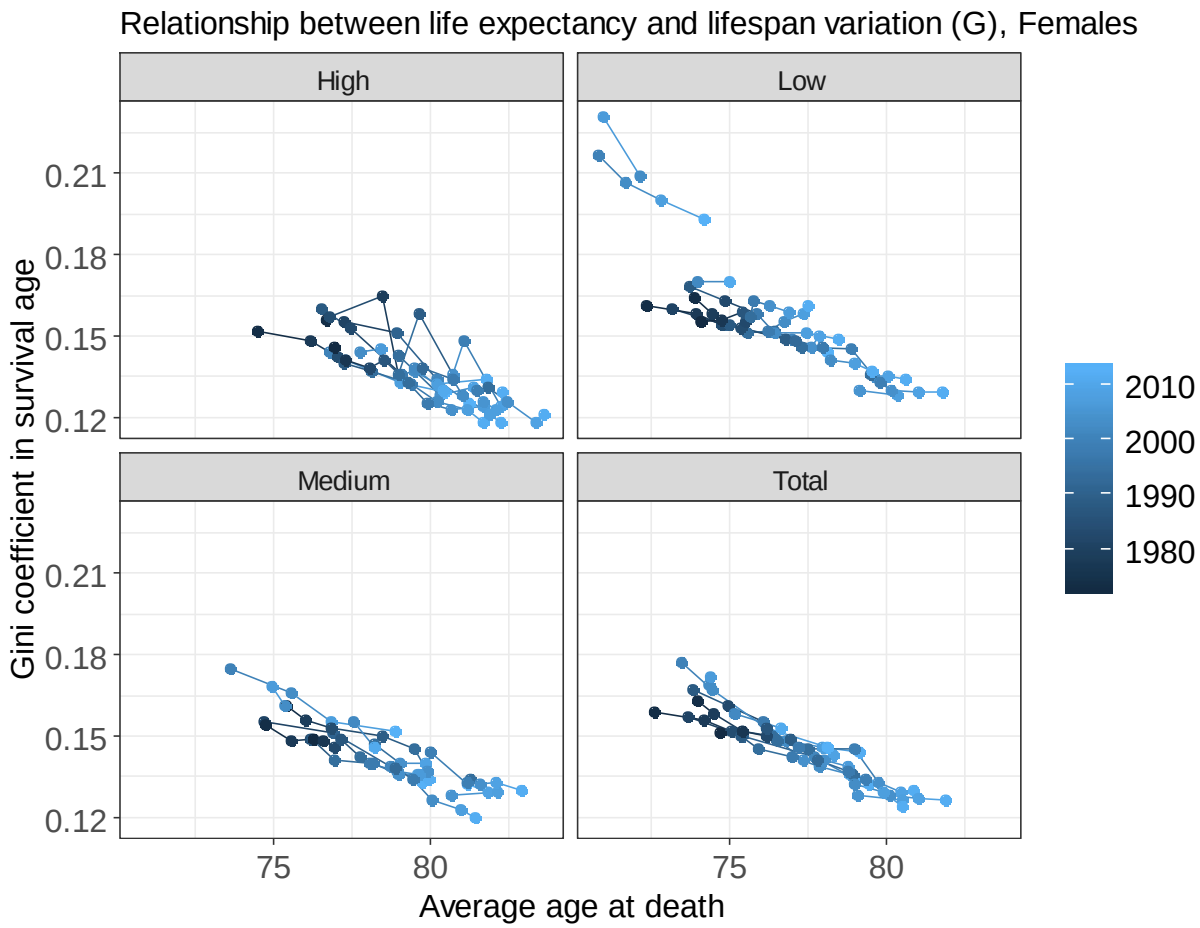
Relationship between life expectancy and lifespan variation (G), Females
High
Low
Medium
Total
Gini coefficient in survival age
0.21
0.18
0.15
0.12
75
80
Average age at death
2010
2000
1990
1980

Figure S7