Mortality across Generations

Saul Lach,[1] Yaacov Ritov,[2] Avi Simhon [3*]

[1]Department of Economics, The Hebrew University of Jerusalem
91905 Jerusalem, Israel
[2]Department of Statistics, The Hebrew University of Jerusalem
91905 Jerusalem, Israel
[3]Department of Agricultural Economics, Faculty of Agricultute, The Hebrew University of Jerusalem
Rehovot 76100, P.O.Box 12, Israel
[*]To whom correspondence should be addressed ; E-mail: asimhon@pluto.mscc.huji.ac.il.

We use a large sample of over half a million individuals to study the relationship between the longevity of fathers and their children. We find that when a father dies between 45 and 65 years of age, his age at death has no effect on his sons' longevity. For daughters, there is no effect when a father dies between the ages 45 and 85. When a father dies between 65 and 85 years of age, an additional year of life is associated with almost 2 additional months of life for his sons. Death after 85 years of age has an even stronger effect reaching over 3 additional months for sons and daughters alike. These correlations are a result of hereditary factors as well as socioeconomic conditions. Thus, our findings set an upper bound to the hereditary effect.

The common, almost instinctive, response to the question on whether parents' and children's longevity are correlated is something akin to "of course". But when we further ask about the strength of this relationship, responses vary and there is no clear agreement on how important this relationship is nor on what exactly it means. Consider the following hypothetical question: if your parents and my parents were born at the same time but yours died 10 years later, what would be the difference between our expected longevity?

In the first half of the 20^{th} century several attempts were made to answer this question but they were all marred by severe statistical problems (e.g., the use of small and non-representative samples), to the point that, in her comprehensive survey on this topic, (*10*) concludes that the "idea that heredity plays an important role in the determination of life span... has been more taken for granted than supported by exact scientific investigation." Later on, several studies found that the chance of children and siblings of centenarians surviving into their nineties is significantly higher than average (*14, 15*). Other studies used pairs of twins, e.g., (*13*) or adoptees (*17*) to elicit information on the genetic transmission of longevity. (*12*) studied the genealogy of European high nobility over hundreds of years and found a significant relationship between the longevity of the parents and that of their offspring. They find that this relationship is especially pronounced when the father died after the age of 70 and is stronger with daughters than with sons. (*18*) followed a group of 2370 middle-aged civil servants and their spouses for 25 years. They found a significant difference in mortality rates between those who had two parents alive and those who had none. Again, this relationship is stronger among women than among men. Thus, most studies confirm the prior belief on the existence of an intergenerational correlation in mortality, but they do not have the required data to *quantify the relationship in a statistically reliable way*.

This state of affairs is not that surprising. To quantify this relationship requires data on the birth and death dates of two generations, and the ability to link children to their parents. This means, for example, that someone dying in 1990 at the age of 80 is to be linked to his or her parents who were born in the 19^{th} century. These data are difficult to obtain in a form amenable to statistical research. Thus, it is not surprising that estimates of the relationship between parents' longevity and their children's life expectancy based on large, representative, samples are simply not available (*1*). Furthermore, (*19*) argues, on purely theoretical grounds, that finding a significant correlation between children's and their parents' longevity is unlikely because of individual heterogeneity (frailty) in the hazard rate of dying.

This paper is the first study that quantifies the correlation between individuals' life expectancy and their parents' longevity using a large and representative data set. The data, assembled in cooperation with the Central Bureau of Statistics of Israel, consist of 3.4 million observations of individuals matched to their parents, of which over 0.5 million correspond to dead fathers. Our goal is to provide a reliable answer to the questions posed at the beginning of this introduction.

We find a significantly strong longevity effect from fathers to their children. On average, the hazard rate of dying when the father dies 10 years later is reduced by 8% for sons and 5% percent for daughters. Both effects are significant and quantitatively important. We also find that this effect varies with the age at death of the father: for sons, life expectancy is not significantly related to their father's longevity when fathers die before reaching 65 years of age; for daughters, there is no relationship until age 85. After this age, we find a strong and significant positive effect. An additional year of a father's life between the ages of 65 and 85 is associated with an increase in his sons' lifetime of almost 2 months, but none for his daughters, this effect increases to 3 additional months for both sons and daughters when the father dies after age 85. That is, after age 65, the relationship between fathers' and their children's life expectancy appears to be convex in the father's age at death.

These findings cannot be solely interpreted as an hereditary effect transmitted through genetic material shared across generations. There are alternative socioeconomic paths through which a father's death at a certain age may affect his children's life expectancy. For example, the education of the child may be interrupted because of the early death of a parent. Since education is negatively correlated with mortality, not accounting for education will lead to a positive correlation in lifetime duration between fathers and their children. Thus, in our analysis we control for individual characteristics in two steps. First, we control for cohort of birth and immigration and for country of birth. Second, for a fraction of the sample, we also have income and schooling data and we examine the robustness of our estimates to these controls. To what degree the use of these controls enables us to isolate the hereditary component is something we simply cannot tell. Nevertheless, the finding of strong correlation effects is important because it allows us to quantify the nature of the unconditional relationship between the mortality of fathers and their children. If one adopts the view that the socioeconomic and hereditary channels all work in the same direction, then our quantitative analysis also allows us to set an upper bound to the importance of the hereditary transmission.

Our data is based on records from the official Population Registry of the State of Israel. A record in the Registry has the individual's name and identity number as well his or her parents' names and, in some cases, their identity numbers. The parents' identity number, in turn, is used to access the parents' records at the Registry. The Registry includes information on the dates of birth and death if died before March 31, 2004, the last available update of the Registry. In this fashion, we can match children to parents and obtain their dates of birth and death. For a detailed description of the data please refer to the supporting material. Because of limitation in the data, we restrict our attention to the Jewish population, and link each individual only to his father. Table 1 explains the manner in which the sample used in the statistical analysis was assembled. Although the total number of matched observations is 3,421,545 (column (1)), the sample used in the statistical analysis is limited by several factors. (*2*) The survival analysis is based on 552,019 individuals with dead fathers, out of which 36,064 observations are not censored, i.e., the individual also died by March 31, 2004. The censoring, as expected, is large: about 93.5% of all observations matched to dead fathers are censored. Also, as expected, censoring is small in the first two cohorts (until 1919) and increases monotonically over the century (column (6)).

Besides hereditary determinants, a variety of socioeconomic, demographic, and environmental factors contribute to the longevity of a given individual. These factors comprise variables such as gender and ethnic origin, which are given at birth and do not change with age, the type of family and school in which the person was raised during childhood, and other variables that evolve during the individual's lifetime such as education, occupation and wealth. Several channels through which socioeconomic factors can affect a person's health have been explored in the literature. (*3*) Health status reflects the cumulative effects of an individual's characteristics and decisions taken over his or her lifetime. It is not very difficult to come up with a list of potential determinants of health status; the difficulties arise in trying to measure them.

We make a **proportional hazard assumption.** The Cox proportional hazard model is a convenient formulation because it does not specify a parametric form for the effect of age t on longevity. Thus, the hazard rate of dying at age t (*4*) is given by

$$\lambda(t|x_i) = \lambda_0(t)e^{x_i\beta},$$

where x_i is the vector of age-invariant covariates of individual i, and β is a vector of parameters. To be clear, a significant effect of a father's longevity on his children's hazard of dying cannot be traced back uniquely to any particular source, hereditary or socioeconomic/environmental sources. There are several plausible mechanisms through which parents' age at death can be related to their children's age at death that could be classified as either environmental, hereditary or socioeconomic. We will not be able to distinguish between them.

Our data is highly censored – most sons and daughters in the registry were still alive by 2004. The Cox model is the most convenient model to deal with censored data. In our sample, there is another type of censoring: to be part of the Population Registry individuals must have survived until after 1948 and, in the case of new immigrants, they had to survive until their year of immigration. We take care of both types of censoring in estimation. See the supporting material for other potential problems with the data and analysis. We obtain maximum (partial) likelihood estimates of the parameters.

An observation is an individual, dead or alive, linked to a dead father. Our first set of results are based on the sample with demographic covariates only. There are 552,019 observations in the sample that are matched to 237,131 fathers. Using the Cox model we estimated models with father's age at death as the main covariate while controlling for region of birth, cohort of birth and immigration year. We run these regressions separately for women (daughters) and for men (sons). The baseline group are Israeli-born individuals born during the 1930-39 cohort.

Table 2 reports estimates of the percentage change in the hazard of dying due to the father living 10 additional years. That is, $100 \times \left(e^{(365.25 \times 10)\beta_a} - 1\right)$, where β_a is the coefficient of the father's age at death appearing in $e^{x_i \beta}$. The hazard ratio e^{β_a} is raised to the power of 3652.5 because life duration is measured in days. Hazard ratios less than one indicate that a father's age at death has a positive effect on his children's survival probability (i.e., $\beta_a < 0$), while hazard ratios equal to one signify no effect (i.e., $\beta_a = 0$). The estimates of β_a, although numerically close to one, are always below one. In column (1), the effect of a father living 10 additional years is to decrease the chance of a son dying at any age by 6% and of a daughter by 2%. In parentheses are the 95% confidence interval for the effect of a 10-year increase in the father's age at death. As controls are added to the basic specification in column (1), the estimated effect of father's age at death increases to 8 percent and 5 percent, for sons and daughters respectively (column (4)), without affecting its precision. (*5*).

Next, we are interested in examining the possibility of non-linear effects in the relationship between father's age at death and children survival. In Table 3, we allow for the coefficient of father's age at death to vary across four age intervals. We did this by using splines at the specified knots in order to avoid discontinuities (*6*). For sons, we find a positive, but not strongly significant, relationship between father's age at death and children survival until age 45, no effect between 45 and 65 years of age, and then a positive, and significant, effect which becomes stronger after age 85. For daughters, we find essentially the same pattern except that father's age at death has virtually no effect if he dies between the 45 and 85 years of age.

A convenient way of presenting the estimated effect of any covariate is to trace the relationship between that covariate and life expectancy. This is particularly true for the effect of father's age at death which appears to vary with the age at death. Given the hazard function we can compute any moments of the distribution of longevity, T, conditional on x, where x includes father's age at death. Life expectancy at birth is,

$$E(T|x) = \int_0^\infty tf(t)dt = \int_0^\infty \left(e^{-\Lambda_0(t)}\right)^{e^{x\beta}} dt$$

where $f(t)$ is the density of t and $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ is the cumulative baseline hazard rate.

Given estimates of β and $\Lambda_0(t)$, $E(T|x)$ can be computed for any x. We use the estimates in column (4) of Table 3 to compute $E(T|x)$ for the baseline group – Israeli-born individuals born during the 1930-39 cohort – and at values of father's age at death ranging between 40 and 100. Figure 1 graphs $E(T|x)$ as a function of father's age at death for sons and daughters.

Note that our estimates of life expectancy vary between 74 and 82 for men and between 78 and 85 for women. These estimates are consistent with a life expectancy of 77.9 for men and 81.9 for women in 2002 as reported in the official statistics (*7*). We interpret this evidence as corroborating the validity of the statistical model.

If socio-economic factors are correlated across generations and if, in addition, they are positively correlated with longevity within generations, then the observed correlation between fathers and children's longevity may partly reflect the intergenerational transmission of socio-economic status. Thus, we would like to control for socio-economic factors in order to check the robustness of our estimates. In addition, there is independent interest in the effect of socioeconomic variables on mortality.

The 1983 Census provides us with data on the years of schooling and monthly wages for 20% of the population. Detailed description of the analysis is given in the supporting material. In Table 4, we find that the inclusion of schooling does not change the estimated effects of father's age at death, although the individual's education is a very significant determinant of the mortality hazard: for sons, an increase in one year of schooling reduces the hazard of dying by 5.3%. This is a significant effect but it cannot be given a causal interpretation. We repeat the same exercise using monthly wages. As with education, the inclusion of the predicted salary at age 50 does not affect the estimated intergenerational effects. Its effect, however, is very strong for sons and not significant for daughters: a 10% increase in the wage reduces the hazard of dying by 3%. These results suggest that the positive correlation between fathers and children's longevity cannot all be explained by a positive correlation in socio-economic status across generations.

References and Notes

1. An exception is the Icelandic data base which includes 270,000 living Icelanders in addition to most of their ancestors since the ninth century (see Gudmundsson H. D. F., A. K. H. Gudbjartsson, M. F Gudbjartsson, J. R. Gulcher, and K. Stefansson. Inheritance of human longevity in iceland. *European Journal of Human Genetics*, 8:733–749, 2000.

2. First, 4,682 individuals dying before January 1, 1949 or before their year of immigration were deleted from the sample; the new figures are in column (2). To be part of the Population Registry individuals must have survived until January 1, 1949 and, in the case of new immigrants arriving after 1949, they had to survive to their year of immigration. Observations with dates of death preceding this logical threshold were treated as error and deleted.

Second, observations with fathers that did not die by March 31, 2004 are also excluded from the analysis. This is a binding constraint since only about 16 percent of the matched observations belong to individuals with dead fathers. As expected, the percentage of dead fathers is very high in the first cohorts but declines rapidly. Some of these observations belong to individuals (children) that have died by March 31, 2004 (column (5)); the rest are censored at this date.

3. Richer people may have access to better health care; the level of stress appears to be related to status level (*16*). Behavior towards risky activities, as well as eating and drinking habits, may be related to income (*8*). See also the work of (*9, 11*) where mortality is related to the individual's income (or wealth) and to the ranking of the individual in the income distribution.

4. The hazard rate, $\lambda(t|x_i)dt$ is the probability of dying at interval of length dt shortly after individual with covariate value x_i arrived to time t .

5. Likelihood ratio tests indicate that the controls are important: the null hypothesis that cohort of birth and immigration, and region of birth have no effects is always rejected.

6. We experimented with different spline specifications: knots every 15 and 20 years starting at a father's age at death of 25, and at father's age at death of 30. The final specification appearing in Table 3 had the largest value of the likelihood function. In any case, the pattern of coefficients in the various specifications were very similar.

7. See Table 3.20 in the Statistical Abstract of Israel, 2004.

8. J. Adda and V. Lechene. Smoking and endogenous mortality: Does heterogeneity in life expectancy explain differences in smoking behavior? October 2001.

9. O. Attanazio and C. Emmerson. Mortality in the uk. Technical Report 8241, NBER, 2001.

10. Cohen H. B. Family patterns of mortality and life span. *The quarterly review of Biology*, 39:130–181, 1964.

11. A. Deaton and C. Paxson. Mortality education, income, and inequality among american cohorts. Technical Report 7140, NBER, 1999.

12. L. A. Gavrilov and N.S. Gavrilova. Biodemographic study of familial determinants of human longevity. *Population: An English Selection*, 13:197–222, 2001.

13. A. M. Herskind, M. McGue, N. V. Holm, T. I. A. Sorensen, and J.W. Vaupel B. Harvald. The heritability of human longevity: A population-based study of 2872 danish twin pairs born. *Human Genetics*, 97:319–323, 1996.

14. T. Perls, E. Bubrick, J. Vijg C. Wager, and L. Kruglyak. Siblings of centanerians live longer. *The Lancet*, 351:1560, 1998.

15. T. Perls, M. Shea-Drinkwater, J. Bowen-Flynn, S. Kang S. Ridge, E. Joyce, M. Daly, S. Brewster, and A. Puca L. Kunkel. Exceptional familial clustering for extreme longevity in humans. *JAGS*, 48:1483–1485, 2000.

16. J. P. Smith. Healthy bodies and thick wallets: the dual relation between health and economic status. *Journal of Economic Perspectives*, 13:145–166, 1999.

17. T.I.A. Sorensen, G. Nielsen, P. Andersen, and T. Teasdale. Genetic and environmental influences on premature death in adult adoptees. *New England Journal of Medicine*, 318:727–732, 9999.

18. J. Vandenbroucke, A. Matroos, C. van der Heide-Wessel, and R. van der Heide. Parental survival, an independent predictor of longevity in middle-aged persons. *American Journal of Epidemiology*, 119:742–750, 1984.

19. J. Vaupel. Inherited frailty and longevity. *Demography*, 25, 1988.

Mortality across Generations
Supporting material

Saul Lach,[1] Yaacov Ritov,[2] Avi Simhon [3*]

[1]Department of Economics, The Hebrew University of Jerusalem
91905 Jerusalem, Israel
[2]Department of Statistics, The Hebrew University of Jerusalem
91905 Jerusalem, Israel
[3]Department of Agricultural Economics, Faculty of Agricultute, The Hebrew University of Jerusalem
Rehovot 76100, P.O.Box 12, Israel
[*]To whom correspondence should be addressed ; E-mail: asimhon@pluto.mscc.huji.ac.il.

The Data Set: linking children to parents

Our data is based on records from the official Population Registry of the State of Israel. A record in the Registry has the individual's name and identity number as well his or her parents' names and, in some cases, their identity numbers. The parents' identity number, in turn, is used to access the parents' records at the Registry. The Registry includes information on the dates of birth and death if died before March 31, 2004, the last update of the Registry, but does not record the cause of death. In this fashion, we can match children to parents and obtain their dates of birth and death.

Column (1) in Table 5 gives the total number of records on individuals by cohort of birth. After deleting emigrants, individuals born before or during 1900 and individuals with obvious errors in their vital statistics, the total number of records in the Registry by March 31, 2004 is 5,798,066 (*1*). Using the available identity numbers we matched individuals to their fathers and the number of matches appears in column (2). The matching to mothers was problematic and is not used in this paper for reasons to be discussed below. Overall, about 54% of the individual records can be matched to their fathers but the distribution of these matches across cohorts varies considerably. The percentage of matched records, column (3), increases monotonically from about less than a tenth of a percentage point in the 1901-1909 cohort to 96% in the 2000-2004 cohort. Until 1939 less than 1% of the records can be matched in the Registry. This is problematic because we need matched data precisely for the earlier cohorts where death of both children **and** their father is prevalent. Fortunately , the Israeli Population Registry has a unique feature that allows us to generate many more matches than those available directly from the Registry.

The Registry was established in 1948 with the creation of the State of Israel. Surveyors went house by house and recorded demographic data of each person in the household (children and adults). In particular, for each individual they recorded the names (first and last) of their parents and assigned **consecutive identity numbers** to all household members. Furthermore, families who immigrated to Israel following its establishment in 1948 were also assigned consecutive identity numbers. Thus, the identity numbers of most of the Jewish (*2*) population in Israel, who were born in Israel before 1948 or immigrated to Israel after 1948, are bundled together by families. We exploit this feature to generate additional matches between children and their parents. The algorithm that generates these matches (to fathers) works as follows:

1. Sort the Population Registry by ascending id number.

2. For each record i in the Registry, select the 10 records preceding and the 10 records following record i.

3. Among these 20 records, select those that are males.

4. Among these records, select those with a first and last name equal to the father's name in record i.

5. Among these, select those whose age is older than record $i's$ age by at least 15 years.

6. Among these records, select the one whose id number is closest to the id number of record i. This record is the **father** and his id number is added to record i.

Column (4) in Table 5 provides the number of matches found by the algorithm only. Overall, the algorithm generates about 150,000 additional matches. Notice the importance of the algorithm in generating matches in the earlier cohorts. Thus, the total number of individuals matched to their fathers (in the Registry or by the algorithm) is the sum of columns (2), (4) ,(5) and (6) which gives 3,421,545, representing 59% of the total number of records

Some of the matches that were created by the algorithm already exist in the Registry. In most of these cases the matches identified by the algorithm accord with the Registry (column (6)), but in some cases the algorithm assigned a father to an individual who has a different father in the Registry (column (5)). The percentage of mismatches is minimal: it averages to about a tenth of a percentage point, reaching 2.6% in one cohort only. Thus, we are quite confident that the matches generated by the algorithm are reliable.

A similar algorithm was used to match individuals to their mothers but the results were far from satisfactory. The main reason for this negative result is that when the Registry was first computerized in the 1960's the father's name was recorded, whereas the mother's name was omitted in order to save computer resources. Only in 1980 when a new computer was purchased, the mother's name was gradually added, a process that was completed in the mid nineties. Thus, by 1980 only 20% of the individuals had their mother's name recorded in the Registry; this proportion reached 98% by 1996. Unfortunately, this update was carried out only for the records of the living individuals. Consequently, the earlier a person died the lower the probability that his/her mother name is recorded in the Registry. Thus, the matching to mothers in the earlier cohorts is likely to induce a selectivity bias in the sample. We therefore focus our analysis on the effects of fathers' longevity only.

A possible source of bias in our sample is that the procedure we used to match children to their parents cannot be applied to individuals born in Israel after 1948. In the Registry, only a fraction of those born in Israel in the early 50s are matched to their parents, while our procedure matched a larger fraction of that cohort who immigrated to Israel. This should not affect our analysis as long as the reason for not matching children to their parents is unrelated to their health status or other uncontrolled factors. Our sample of matched observations would not be a random sample of the total population if it contained disproportionately more people who were in contact with the authorities due to health problems or because they were welfare recipients.

Table 1 explains the manner in which the sample used in the statistical analysis was assembled. Although the total number of matched observations is 3,421,545 (column (1)), the sample used in the statistical analysis is limited by several factors (*1*). The survival analysis is based on 552,019 observations, of which 36,064 observations are not censored. The censoring, as expected, is large: about 93.5% of all observations matched to dead fathers are censored. Also, as expected, censoring is small in the first two cohorts (until 1919) and increases monotonically over the century (column (6)).

Model limitation and verification

As mentioned in the main text, the analysis is restricted to the longevity of individuals with dead fathers. This could generate a bias, since the chance of entering the sample is higher the shorter is the father's life span. To assess the magnitude of this bias we repeat our estimation for a sample consisting of individuals who where born prior to 1939 (column 5) because in this subsample more than 98% of the parents died by 2004 (column (4) in Table 1). Column (5) in Table 2 gives the estimated effects from this subsample. The parameters' estimates do not differ much; they are even more negative. Thus, we do not find evidence that the selection of individuals with dead fathers is generating substantial biases in the estimated effects of interest. Another problem that may arise is that siblings in the sample may not constitute independent observations. Adjusting the standard errors for the presence of siblings was not possible due to computer limitations. When the sample was restricted to one sibling per family (the oldest), the results in all specifications were essentially the same in terms of the coefficients and their significance. This suggests that correcting for the bias in the standard errors would not substantially change the estimated parameters.

Socio-economic factors

The 1983 Census provides us with data on the years of schooling and monthly wages. For each individual, these variables are observed at age "1983 – year of birth". We restricted the sample to those individuals older than 21 years of age at the time of the Census (i.e., born after 1962). Column (1) in Table 4 presents the results of adding education to the basic equation. Because socioeconomic data from the 1983 census are available for about 20% of the overall population, column (2) reestimates the model in Table 3, column (4), on the restricted sample. The inclusion of schooling in column (1) does not change the estimated effects of father's age at death, compared to column (2). As shown in column (1), the individual's education is a very significant determinant of the mortality hazard: for sons, an increase in one year of schooling reduces the hazard of dying by 5.3%. This is a significant effect but it cannot be given a causal interpretation. This effect is robust: omitting father's age at death from the regression does not change the estimated effect of education on mortality (column (3)). Thus, it seems that father's age at death is not just picking up socio-economic effects.

We repeat the same exercise using monthly wages. Since wages are observed at different ages for different individuals we constructed a predicted wage at age 50 as follows. We used the cross-section data in 1983 to estimate an OLS regression of log monthly salary on a cubic in age, schooling and a vector of demographic covariates (cohort and country of birth and year of immigration) for men and women separately. We then used the estimated parameters to estimate the percentage growth in predicted salary between the person's age in 1983 and age 50 and applied this growth rate to his or her observed salary in 1983. Column (4) reports the estimates obtained when using this predicted log wage at age 50 (*3*). As with education, its inclusion does not affect the estimated intergenerational effects. Its effect, however, is very strong for sons and not significant for daughters: a 10% increase in the wage reduces the hazard of dying by 3%.

References and Notes

1. Before the deletions the Registry has 6,862,849 observations. We first deleted 658,522 observations of individuals that are marked in the Registry as "emigrants." In effect, the Central Bureau of Statistics assigns an emigrant status to individuals who could not be contacted in the decennial Censuses. According to their calculations, over 80 percent of them emigrated to other countries (since the establishment of the Registry approximately 6% of the Jewish population emigrated). These individuals are usually registered as "alive" in the Registry even if they died abroad. For the same reason we deleted 134,296 observations of individuals whose fathers have emigrant status and missing date of death. Next, we deleted 255,111 observations of individuals born before or during 1900 (of course, we allow parents who were born before 1900 and have children who were born after 1900). We further deleted 9,084 individuals with missing gender information, and 315 observations where the individual was born within 13 year of the father. In addition, 2,746 observations with negative life durations, and 4,343 observations with missing information on the country of origin were

also deleted. Finally, 68 observations whose fathers died after the age of 110, and 298 observations whose fathers died before reaching 13 years of age were also deleted. The total number of observations deleted is 1,064,783.

2. We restric the analysis to the Jewish population mainly because the frequent multiple appearance of the same names in the same neighborhood of the Arabic population made our matching algorithms inappropriate to this population.

3. The analysis is restricted to those between 21 and 65 years of age for men, and 21 and 60 years of age for women. In fact, the estimates are very similar to those obtained using the age-varying wage.

Table 1 . Sample Selection and Censored Observations

Cohort	Number of Matched Obs. (1)	Number of Matched Obs. after Deletions (2)	Number of Matched Obs. with Dead Fathers (3)	Percent with Dead Fathers (4) = (3):(2)	Number of Matches with Dead Fathers & Children (5)	Percent Uncensored (6) = (5):(3)
1901-1909	828	828	823	99.4	807	98.1
1910-1919	1,858	1,851	1,839	99.4	1,511	82.2
1920-1929	5,937	5,931	5,891	99.3	2,916	49.5
1930-1939	36,189	36,182	35,346	97.7	7,635	21.6
1940-1949	107,572	107,566	94,094	87.5	9,221	9.8
1950-1959	296,831	296,820	171,015	57.6	7,613	4.5
1960-1969	460,452	460,269	142,584	31.0	4,123	2.9
1970-1979	670,004	668,407	68,198	10.2	1,725	2.5
1980-1989	772,342	770,643	24,703	3.2	436	1.8
1990-1999	792,945	792,021	7,011	0.9	71	1.0
2000-2004	276,587	276,345	515	0.2	6	1.2
Total	**3,421,545**	**3,416,863**	**552,019**	**16.2**	**36,064**	**6.5**

Table 2. Proportional Hazard Model for Life Duration

Percentage change in hazard of dying when father lives 10 additional years

	Sons					**Daughters**				
	Hazard Ratio					**Hazard Ratio**				
	(1)	(2)	(3)	(4)	(5) Born ≤ 1939	(1)	(2)	(3)	(4)	(5) Born ≤ 1939
Father's Age at Death	-6	-8	-9	-8	-12	-2	-5	-5	-5	-10
	(-7, -5)	(-9, -7)	(-10, -8)	(-9, -7)	(-14, -11)	(-3, 0)	(-7, -3)	(-7, -3)	(-6, -3)	(-14, -6)
Region of birth	no	no	no	yes	yes	no	no	no	yes	yes
Cohort of birth	no	yes	yes	yes	yes	no	yes	yes	yes	yes
Cohort of Immigration	no	no	yes	yes	yes	no	no	yes	yes	yes
Number of Observations	309,818	309,818	309,818	309,818	35,840	242,201	242,201	242,201	242,201	8,059
% Censored	91.3	91.3	91.3	91.3	70.2	96.2	96.2	96.2	96.2	72.7
Log-likelihood	-305,947.3	-305,697.1	-305,619.1	-305,614.1	-102,714.6	-102,036.9	-101,725.8	-101,697.3	-101,695.2	-17,599.1

Notes: 95% confidence intervals in parentheses.

The coefficients of the dummies for region of birth (Asia, Africa, Europe and America, USSR) are not reported.

The coefficients of the dummies for 10-year cohorts of birth and immigration (1901-1909, 1910-1919,…1990-1999,2000-2004) are not reported.

Table 3. Proportional Hazard Model for Life Duration - Non-linear Effects

Percentage change in hazard of dying when father lives 10 additional years

	Sons					**Daughters**				
	Hazard Ratio					**Hazard Ratio**				
	(1)	(2)	(3)	(4)	(5) Born ≤ 1939	(1)	(2)	(3)	(4)	(5) Born ≤ 1939
Father's Age at Death ≤ 45	-10	-8	-9	-9	-60	-15	-11	-11	-11	3155
	(-18, 0)	(-17, 1)	(-17, 1)	(-17, 1)	(-85, 5)	(-25, -3)	(-23, 1)	(-22, 1)	(-22, 1)	(-97, 316523)
Father's Age at Death in (45,65]	2	0	0	0	0	-5	-4	-4	-4	-18
	(-2, 5)	(-3, 3)	(-3, 4)	(-3, 4)	(-8, 10)	(-10, 0)	(-9, 2)	(-9, 1)	(-9, 1)	(-34, 2)
Father's Age at Death in (65,85]	-7	-10	-10	-10	-13	6	-1	-1	-1	-5
	(-9, -5)	(-12, -8)	(-12, -8)	(-12, -8)	(-16, -12)	(2, 10)	(-5, 3)	(-5, 3)	(-5, 3)	(-12, 3)
Father's Age at Death > 85	-19	-20	-20	-20	-20	-21	-25	-24	-24	-23
	(-24, -14)	(-25, -15)	(-25, -15)	(-25, -15)	(-26, -13)	(-30, -12)	(-33, -16)	(-33, -15)	(-32, -15)	(-35, -10)
Region of birth	no	no	no	yes	yes	no	no	no	yes	yes
Cohort of birth	no	yes	yes	yes	yes	no	yes	yes	yes	yes
Cohort of Immigration	no	no	yes	yes	yes	no	no	yes	yes	yes
Number of Observations	309,818	309,818	309,818	309,818	35,840	242,201	242,201	242,201	242,201	8,059
% Censored	91.3	91.3	91.3	91.3	70.2	96.2	96.2	96.2	96.2	72.7
Log-likelihood	-305,920	305,667	-305,591.9	-305,587.2	-102,705.6	-102,023.1	-101,716.0	-101,688.2	-101,686.3	-17,596.2

Notes: 95% confidence intervals in parentheses.

The coefficients of the dummies for region of birth (Asia, Africa, Europe and America, USSR) are not reported.

The coefficients of the dummies for 10-year cohorts of birth and immigration (1901-1909, 1910-1919,…1990-1999,2000-2004) are not reported.

Table 4. Proportional Hazard Model for Life Duration - Non-linear Effects and Socioeconomic Controls

Percentage change in hazard of dying when father lives 10 additional years / schooling increases by 1 year / wages increase by 10%

	Sons						**Daughters**					
	Hazard Ratio						**Hazard Ratio**					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Father's Age at Death ≤ 45	58	51	-	24	26	-	-55	-53	-	-78	-79	-
	(-33, 276)	(-36, 254)		(-62, 302)	(-61, 305)		(-79, 2)	(-79, 1)		(-91, -49)	(-91, -51)	
Father's Age at Death in (45,65]	-7	-7		-7	-7	-	3	2	-	-7	-7	-
	(-18, 5)	(-18, 5)		(-22, 12)	(-23, 11)		(-22, 21)	(-22, 22)		(-26, 54)	(-26, 55)	
Father's Age at Death in (65,85]	-13	-13		-13	-13		-3	-3	-	5	5	-
	(-19, -7)	(-19, -7)		(-21, -4)	(-21, -4)		(-9, 17)	(-9, 17)		(-15, 30)	(-15, 30)	
Father's Age at Death > 85	-31	-31	-	-34	-32	-	-30	-34	-	-74	-74	-
	(-43, -15)	(-44, -16)		(-52, -11)	(-50, -8)		(-52, 0)	(-54, -4)		(-90, -34)	(-90, -34)	
Schooling (years) in 1983	-5.3	-	-5.3	-	-	-	-5.1	-	-5.2	-	-	-
	(-6.3, -4.4)		(-6.3, -4.4)				(-6.9, -3.4)		(-7, -3.5)			
Predicted Net Monthly Wage at Age 50		-	-	-2.99	-	-2.98	-	-	-	0.00	-	0.00
				(-3.79, -2.17)		(-3.79, -2.17)				(-2.23, 1.51)		(-2.28, 1.45)
Number of Observations	34,439	34,439	34,439	18,765	18,765	18,765	24,451	24,451	24,451	10,143	10,143	10,143
% Censored	92.1	92.1	92.1	93.1	93.1	93.1	96.6	96.6	96.6	97.2	97.2	97.2
Log-likelihood	-24,284	-24,343	-24,320	-10,982.6	-11,005.4	-11,000.5	-6,984.3	-7,000.6	-6,989.1	-2,241.7	-2,241.8	-2,252.2

Notes: 95% confidence intervals in parentheses.

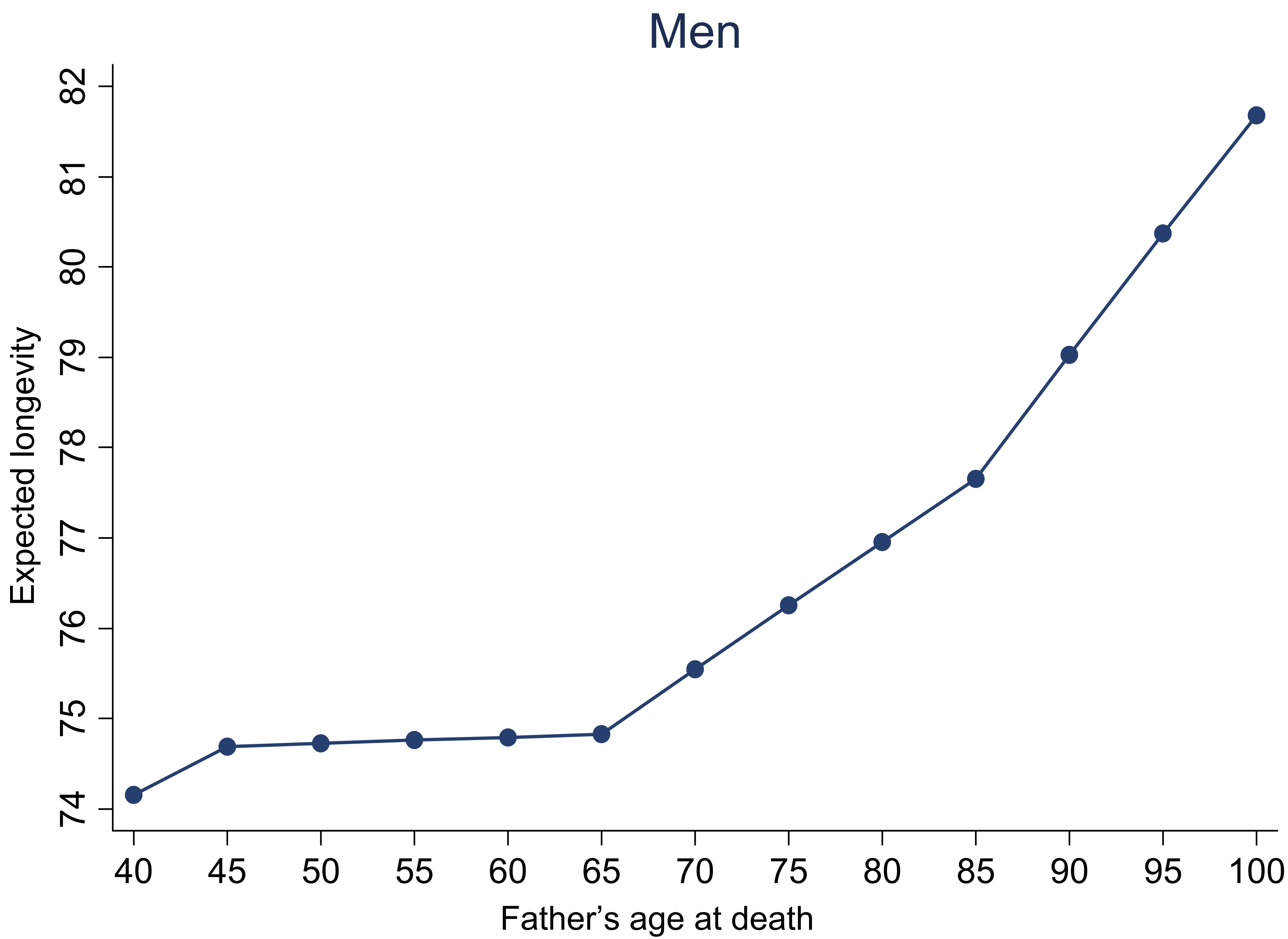
The coefficients of the dummies for region of birth (Asia, Africa, Europe and America, USSR) are not reported.

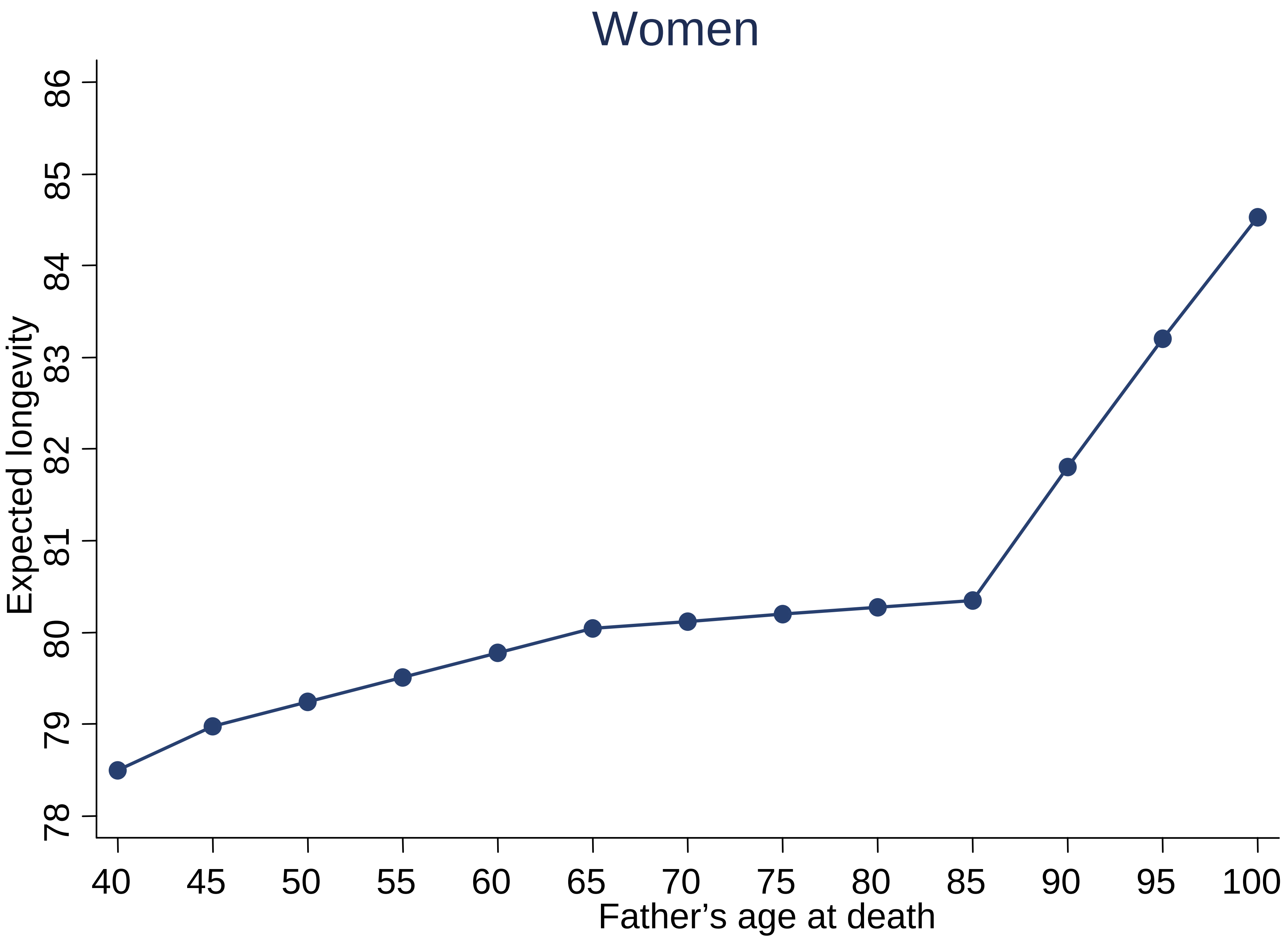
The coefficients of the dummies for 10-year cohorts of birth and immigration (1901-1909, 1910-1919,…1990-1999,2000-2004) are not reported.

Table 5 . Data Availability in Population Registry and Matches to Fathers

Cohort	Total Number of Records (1)	Matches in Registry Only (2)	Percent of Matches (3) = (2):(1)	Matches by Algorithm Only (4)	Matches in Registry and by Algorithm: Not Agree (5)	Matches in Registry and by Algorithm: Agree (6)	Percent of Missmatches (7) = (5):((5)+(6))
1901-1909	228,808	21	0.0	796	0	11	0.0
1910-1919	329,600	122	0.0	1,680	0	56	0.0
1920-1929	405,765	741	0.2	4,922	7	267	2.6
1930-1939	405,927	3,629	0.9	30,587	11	1,962	0.6
1940-1949	482,588	18,064	3.7	79,116	28	10,364	0.3
1950-1959	649,075	246,801	38.0	24,217	31	25,782	0.1
1960-1969	611,694	429,865	70.3	6,065	12	24,510	0.0
1970-1979	767,100	633,601	82.6	1,883	3	34,517	0.0
1980-1989	813,891	733,257	90.1	169	10	38,906	0.0
1990-1999	816,651	781,656	95.7	39	2	11,248	0.0
2000-2004	286,967	276,095	96.2	1	0	491	0.0
Total	**5,798,066**	**3,123,852**	**53.9**	**149,475**	**104**	**148,114**	**0.1**

Notes:
Jewish population only


Men
Expected longevity
74
75
76
77
78
79
80
81
82
40
45
50
55
60
65
70
75
80
85
90
95
100
Father's age at death



Women
Expected longevity
78
79
80
81
82
83
84
85
86
40
45
50
55
60
65
70
75
80
85
90
95
100
Father’s age at death