- © ¬ ¹ ¥º ¸ © ¸¡ ¡ ²¤¤ · © ± ¸ ¡ © ° ¥ ¤
The Hebrew University of Jerusalem

º © ¬ ¤ § ¬ « ¬ ¸ « § ® ¦ « ¸ ® ¤ ¬ ¤ ° ® © ¥ ¬ ¤ § ¬ « ¬ ¤ « · ¬ § ® ¤
The Center for Agricultural Economic Research
The Department of Agricultural Economics and Management

Discussion Paper No. 4.06

LONGEVITY ACROSS GENERATIONS

by

Saul Lach, Yaacov Rtion and Avi Simhon

Papers by members of the Department can be found in their home sites:

מאמרים של חברי המחלקה נמצאים גם באתרי הבית שלהם:

http://departments.agri.huji.ac.il/economics/indexe.html

P.O. Box 12, Rehovot 76100

ת.ד. 12, רחובות 76100

Longevity across Generations

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

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. Attempts to quantify it were marred by severe statistical problems such as the use of small and non-representative samples. We use a sample of over half a million individuals in Israel to quantify the relationship between the longevity of fathers and their children. This is the first study to empirically address the correlation in longevity across generations using a large and representative data set. Our findings are summarized in Figure 1. When a father dies between 45 and 65 years of age, his age at death has no effect on his sons' longevity. However, when he 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. Interestingly, for daughters, there is no effect when a father dies between the ages 45 and 85. These correlations are a result of hereditary factors as well as socio-economic conditions. As explained below, our findings set an upper bound to the hereditary effect.

Figure 1: Children's expected longevity and father's age at death

In her 1964 survey Cohen concluded 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

[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 Agriculture, The Hebrew University of Jerusalem, Rehovot 76100, P.O.Box 12, Israel.

scientific investigation"[1]. Later on, several studies found that the chance of children and siblings of centenarians surviving into their late nineties is significantly higher than average[2,3]. Other studies used pairs of twins[4], adoptees[5], and even the genealogy of European high nobility over several centuries to elicit information on the genetic transmission of longevity[6]. Another study followed a group of 2370 middle-aged civil servants and their spouses for 25 years[7]. While most studies confirm the prior belief on the existence of an intergenerational correlation in mortality, they do not have the required data to quantify the relationship in a statistically reliable way, let alone study how this relationship changes with the father's age at death.

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. An exception is the Icelandic data base which includes 270,000 living Icelanders in addition to most of their ancestors since the ninth century[8]. Furthermore, some researchers argue, 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[9].

The data, assembled in cooperation with the Central Bureau of Statistics of Israel, are 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 as 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 the person 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. Because of limitation in the data, we restrict our attention to the Jewish population, and link each individual only to his father. Although the total number of matched observations is 3,421,545, the survival analysis is based on 552,019 individuals linked to 237,131 dead fathers. Our data are highly censored – most sons and daughters in the registry were still alive by 2004 (about 93.5% of all the observations are censored). Also, as expected, censoring is small in the first two cohorts (individuals born before 1919) and increases monotonically over the century (For a detailed description of the data please refer to the supplementary information).

We make a proportional hazard assumption. The Cox proportional hazard model is a convenient formulation because it allows us to assess the effect of father's age at death without specifying a parametric form for the effect of (own) age on longevity. Thus, the hazard rate of dying at age t is given by

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

where x_i is a vector of age-invariant covariates of individual i, β is a vector of parameters and $\lambda_0(t)$ is called the "baseline hazard". The vector x includes the age at death of the father as well as available socio-economic variables (see below). The hazard rate $\lambda(t|x_i)$ is the probability of dying during an interval of length dt shortly after individual i (with covariate value x_i) arrived to time t. 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 supplementary information 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 a specification of the vector of covariates x that includes demographic covariates

only (region of birth, cohort of birth and immigration year) and father's age at death as the main covariate of interest 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 1 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. Likelihood ratio tests indicate that the controls are important: the null hypothesis that cohort of birth and immigration, and country of birth have no effects is always rejected.

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 2, 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. 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 column 4 had the largest value of the likelihood function. In any case, the pattern of coefficients in the various specifications were very similar. 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 2 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[10]. We interpret this evidence as corroborating the validity of the statistical model.

Besides hereditary determinants, a variety of socioeconomic, demographic, and environmental factors contribute to the longevity of a given individual. If socio-economic factors are correlated across generations and if, in addition, they are positively correlated with longevity within generations[11–14], then the observed correlation between fathers and children's longevity may reflect in part the intergenerational transmission of socio-economic status. In addition,

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 also lead to a positive correlation in lifetime duration between fathers and their children. Thus, we would like to control for socio-economic factors in order to sharpen the interpretation, and check the robustness, of our estimates. Moreover, 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 supplementary information. In Table 3, 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 but 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.

We also have data on the underlying cause of death of individuals who died between 1968 and 2000. Overall, neoplasms and diseases of the circulatory systems (including heart diseases and strokes) are the two most prevalent causes of death accounting for about half the deaths in our sample. However, external causes (including vehicle and other accidents) are the major cause of death for individuals dying before age 39.

Table 4 presents the effect of father's longevity on his children's hazard rate by the father's

cause of death. Excluding individuals whose father died of external causes (column 1) results in the same pattern of coefficients obtained for the entire population. In columns (2) and (3) we use only observations where the father's death was caused by a neoplasm or by a disease of the circulatory system, respectively. In the case of neoplasms, the pattern is similar to the one in Table 2 except that now the effect when the father dies after the age of 85 is not significant. This could be attributed to the very small number of observations in this category. In the case of circulatory diseases, the pattern is similar to the one in Table 2.

Our findings cannot be solely interpreted as an hereditary effect transmitted through genetic material shared across generations but we showed that they are robust to the inclusion of demographic and socio-economic 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. Moreover, the estimated coefficients do not vary much with the father's cause of death suggesting that for predicting longevity, the father's age at death appears to be more important than his cause of death.

[1] Cohen, H.B., Family patterns of mortality and life span, *The Quarterly Review of Biology*, **39**, 1964, 130–181.

[2] Perls T., E. Bubrick, C. Wager, J. Vijg, L. Kruglyak, "Siblings of centanerians live longer", *The Lancet*, May **23**,1998, v351 n9115 p1560.

[3] Perls, T., M. Shea-Drinkwater, J. Bowen-Flynn, S. Ridge, S. Kang, E. Joyce, M. Daly, S. Brewster, Louis Kunkel, A. Puca, "Exceptional familial clustering for extreme longevity in humans", *JAGS* 48, 1483-1485, 2000.

[4] Herskind A. M., M. McGue, N. V. Holm, T. I. A. Sorensen, B. Harvald, J.W. Vaupel, "The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900", *Human Genetics*, **97**, 1996, 319-323.

[5] Sorensen T.I.A., G. Nielsen, P. Andersen, T. Teasdale, "Genetic and environmental influences on premature death in adult adoptees", *New England Journal of Medicine*, **318**, 727-732.

[6] Gavrilov L.A. and N.S. Gavrilova, "Biodemographic study of familial determinants of human longevity", *Population: An English Selection*, **13(1)**, 2001, 197-222.

[7] Vandenbroucke J., 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(5)**, 1984, 742-750.

[8] Gudmundsson H., D. F. Gudbjartsson, A. Kong, H. Gudbjartsson, M. Frigge, J.R. Gulcher, and K. Stefansson, "Inheritance of human longevity in iceland", *European Journal of Human Genetics*, **8**, 2000, 743-749.

[9] Vaupel J., "Inherited frailty and longevity", *Demography*, **25(2)**, May 1988.

[10] Central Bureau of Statistics, *Statistical Abstract of Israel*, Table 20.3, Jerusalem, 2004.

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

[12] Adda, J. and V. Lechene,"Smoking and endogenous mortality: does heterogeneity in life expectancy explain differences in smoking behavior?", October 2001, Mimeo.

[13] Attanazio, O. and C. Emmerson, "Mortality in the UK", NBER WP 8241, 2001.

[14] Deaton, A. and C. Paxson, "Mortality, education, income, and inequality among american cohorts", NBER WP 7140, 1999.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Central Bureau of Statistics of Israel for preparing and making the data available for research, and the Israel Science Foundation and The Morris Falk

Institute of Economic Research for funding this project.

Author Contributions Saul Lach and Avi Simhon came up with the original question, assembled the data set, analyzed it and jointly wrote the paper. Yaacov Ritov assisted with the statistical analysis and its interpretation. All authors discussed the results and commented on the manuscript.

Author Information Correspondence should be addressed to

Yaacov Ritov (yaacov.ritov@huji.ac.il)

Table 1. 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.