American Economic Review: Papers & Proceedings 2009, 99:2, 133–138
http://www.aeaweb.org/articles.php?doi=10.1257/aer.99.2.133

Average Earnings and Long-Term Mortality: Evidence from Administrative Data

By Daniel Sullivan and Till von Wachter*

In this paper we exploit a unique database that merges longitudinal earnings data on Pennsylvanian workers with national death records to study the detailed nature of the correlation between earnings and mortality. We find that the estimates typically reported in the literature, which are based on single years of earnings data, are likely to understate substantially the strength of the association between income and mortality. In particular, relative to a single year of earnings, the average of earnings over a six-year period predicts a 70 percent greater impact of income on mortality. In addition, controlling for the mean level of earnings over a period, we find that greater earnings volatility is associated with higher mortality.

We also examine the lag structure of the relationship between earnings and mortality. We find that conditional on a small number of years of recent earning levels, there is little or no correlation between earnings levels in earlier years and current mortality. This runs counter to the interpretation of the earnings-mortality correlation that views workers as "buying health" by allocating a steady fraction of their earnings to the accumulation of a health stock. It is also inconsistent with interpretations of the earnings-mortality link emphasizing the correlation of both variables with an unobserved, time-invariant trait, such as the rate at which workers discount the future. We find that explaining the patterns of correlation appears to require a relatively complex theory that we leave to future research.

*Sullivan: Federal Reserve Bank of Chicago, 230 South LaSalle Street, Chicago, IL 60604 (e-mail: Daniel.Sullivan@chi.frb.org); von Wachter: Columbia University, NBER, CEPR, and IZA, 420 West 118th St. 1022 IAB, New York, NY 10027 (e-mail: vw2112@columbia.edu).

I. Relationship to the Literature

The existing literature documents a robust correlation of income with mortality and other health outcomes. Typical estimates relate mortality over various horizons to basic measures of income, such as annual earnings from survey data (e.g., Angus Deaton and Christina Paxson 1999). However, because surveys measure earnings with significant error (e.g., John Bound and Alan Krueger 1991), such estimates may be biased toward zero. Similarly, earnings are subject to substantial short-term variation (e.g., John Abowd and David Card 1989), which implies that a single year of data will fail to capture the full effect of earnings on mortality over a period of several years. The administrative data we analyze allow us to overcome these limitations, because they are measured with much less error than survey data and are longitudinal in nature. They also allow us to examine the relationship of other career outcomes, such as earnings variability, to mortality.

In addition to capturing direct effects of income on health, typical estimates may also reflect reverse causality or the effects of omitted variables affecting both health and income (e.g., James Smith 1999; David Cutler, Deaton, and Adriana Lleras-Muney 2006). Determining the correct interpretation of the earnings-mortality correlation is difficult, in part because most data sources with information on the health of large samples of workers contain only limited information on earnings. Our use of longitudinal data allows us to examine the correlation in ways that may shed light on the plausibility of alternative interpretations.

For example, if the correlation between income and mortality arises because workers invest a constant fraction, a, of their income, y_{it}, in a health stock, H_{it}, which otherwise depreciates at a constant rate, δ, then that stock would evolve as $H_{it} = ay_{it} + (1 - \delta) H_{it-1}$. This implies

that the stock is a function of current and previous earnings of the form

$$(1) \quad H_{it} = \sum_{s=0}^{t-1} a y_{it-s} (1-\delta)^s + (1-\delta)^t H_{i0}.$$

This expression suggests that mortality should depend on the whole history of workers' earnings, with weights that decline smoothly with time since the earnings were obtained. In particular, if the depreciation rate is relatively low, then even conditional on recent earnings, higher earnings in earlier years should be associated with lower current mortality.

Higher earnings could also be associated with lower mortality, because in the United States workers with higher earnings are more likely to have health insurance. As having health insurance lowers the cost of preventative care, it may imply additional health stock investments in a manner consistent with (1). In addition, however, having health insurance may allow a worker to obtain better care in the event of a random health shock. Having health insurance in earlier periods is not protective in this manner. Thus, this link from income to mortality suggests less dependence on lagged income, controlling for current income.

Mortality may also be correlated with earnings because both variables are influenced by some other unobserved variable, such as the rate at which workers discount the future. That is, the correlation could arise simply because more patient workers invest more in obtaining skills, thus have high earnings and invest more in their health, and thus have low mortality. This interpretation also has implications for a distributed lag specification relating mortality to earnings histories. In particular, it suggests that earnings matter for mortality because they are a noisy indicator of discount rates. But this applies to earnings levels in all years. Indeed, if patience is the only driver of mortality, is constant over time, and is related to earnings in all years equally closely, then in a regression of health on a vector of lagged earnings levels, one would expect to see equal coefficients for all years' earnings. This is an implication we can investigate with longitudinal earnings data.

II. Data and Statistical Methodology

Our data on workers' employment and earnings histories are derived from the unemployment insurance records of the state of Pennsylvania over the period from 1974 to 1991. For a 5 percent sample of workers who held jobs covered by unemployment insurance, we observe quarterly earnings from each Pennsylvania (PA) employer.[1] Our data on mortality are derived from a database compiled by the Social Security Administration and cover deaths occurring anywhere in the United States between 1974 and 2006. The accuracy of the death information has been found to be good for the sample of mature and older male workers we consider.[2]

We are interested in measuring the association of mortality with measures of average earnings over several years. Such averages are most informative for workers who are continuously in the labor force. In addition, if a worker has zero recorded earnings in a particular year, we do not know if he is working outside the covered Pennsylvania work force or truly has zero earnings. Thus, we limit our samples to men with significant attachment to the Pennsylvania labor force over relevant periods. For example, our base sample is limited to male workers born between 1920 and 1959 with some earnings in each year from 1974 to 1979. We then relate average earnings over this period to mortality from 1980 to 2006.

We estimate the effects of average earnings and other career outcomes on mortality over various follow-up periods using a standard logistic regression framework. Specifically, we estimate a number of logistic regression models of the form

$$(2) \qquad \ln\left(\frac{p_{it}}{1-p_{it}}\right) = \mathbf{x}_{it}\beta + \chi_{a(i,t)} + \varphi_t,$$

where $p_{it} \equiv \Pr\{Death_{it} = 1 \mid Death_{it-1} = 0\}$ is the hazard of worker i dying in year t given survival through year $t-1$, and $\mathbf{x}_{it}$ is a vector of career outcomes, such as the average or the standard

[1] See Louis Jacobson, Robert LaLonde, and Daniel Sullivan (1993) for a detailed description of the data and its advantages and disadvantages. An "employer" in our data refers to a firm, which may operate multiple establishments as long as they are in Pennsylvania.

[2] The Social Security Administration's Death Master File (DMF) is described and evaluated in Mark E. Hill and Ira Rosenwaike (2001). Coverage of the death data is better in the 1990s, for older workers, and for men.

TABLE 1—EFFECT OF EARNINGS ON 1980–2006 MORTALITY: WORKERS WITH POSITIVE EARNINGS EACH YEAR 1974–1979 AND BORN 1920–1959

	Males			
Estimation sample	(1)	(2)	(3)	(4)
Log 1979 earnings	−0.226			
	(0.010)			
Log 1974–1979 average annual earnings		−0.384	−0.328	−0.381
		(0.016)	(0.018)	(0.030)
Log standard deviation 1974–1979 quarterly earnings			0.112	0.110
			(0.012)	(0.012)
Number quarters zero earnings 1974–1979			−0.014	−0.014
			(0.010)	(0.010)
Earnings growth 1974–1979			−0.061	−0.061
			(0.013)	(0.013)
Log average earnings × 1979 nonmanufacturing job				0.076
				(0.033)
Observations	1,658,374	1,658,374	1,658,374	1,693,763

Notes: Entries are coefficients of logistic regression models of the mortality hazard. Models also include year effects and a quartic in age.

deviation of quarterly earnings.[3] Because the probability of death in a particular year is typically quite low, the increase in the log-odds ratio associated with change in a component of $\mathbf{x}_{it}$ is approximately equal to the percentage increase in the death rate itself. All the specifications we report below include year dummies (φ_t), which among other things may control for variation over time in the completeness of the Social Security Administration's death records. They also include a fourth-order polynomial in age $(\chi_{a(i,t)})$.

III. The Correlation of Average Earnings and Mortality

Table 1 shows coefficient estimates from logistic regression models based on alternative measures of earnings and other career outcomes over the 1974–1979 period. We start with a specification similar to many in the literature in its use of a single year of earnings data. As shown in column 1, log earnings in 1979 is highly associated with mortality over the 1980–2006 period. The coefficient estimate of −0.226 indicates that a 10 percent increase in 1979 earnings is associated with an approximately 2.3 percent lower hazard of death over the follow-up period. If, instead of 1979 earnings, we use the average earnings over the 1974–1979 period, as in column 2, the coefficient rises to −0.384. This 70 percent increase in the strength of the effect suggests that by averaging, we eliminate a substantial amount of random fluctuations of annual earnings around an underlying trend. This is consistent with classic studies suggesting that earnings exhibit significant short-term fluctuations around underlying trends (e.g., Abowd and Card 1989). These short-term fluctuations are equivalent to measurement error if mortality depends primarily on underlying trends in earnings.

The remaining columns of Table 1 show how mortality relates to some other career outcomes that we are able to measure using our longitudinal earnings data. Column 3 shows that conditional on mean earnings, higher earnings variability, as measured by the log of the standard deviation of quarterly earnings, is associated with increased

[3] This is a standard logistic regression model, and we obtain our parameter estimates by maximum likelihood. Workers contribute one observation for each year they are alive during the follow-up period. The risk set evolves over time as workers die. Bradley Efron (1988) shows that the logistic model we estimate approximates standard continuous parametric models of the survival hazard. Note that in some specifications, the components of χ_{it} are calculated over a base period that precedes the follow-up period. In those cases, it is constant over the estimation period. However, when we examine the lag structure of the relationship, its components are defined relative to the year of the observation, and thus change over time.

Table 2—Effect of Average Annual Earnings on Mortality at Different Lags: Workers with Earnings in Preceding 10 Years

	Males			
Estimation sample	(1)	(2)	(3)	(4)
Log average earnings 1–5 years before follow-up	−0.498 (0.051)		−0.628 (0.078)	
Log average earnings 6–10 years before follow-up		−0.358 (0.059)	0.199 (0.094)	
Log average earnings 1–3 years before follow-up				−0.461 (0.066)
Log average earnings 4–6 years before follow-up				−0.088 (0.102)
Log average earnings 7–10 years before follow-up				0.136 (0.098)
Observations	379,665	379,665	379,665	379,665

Notes: Entries are coefficients of logistic regression models of the mortality hazard. Models also include year effects and a quartic in age.

mortality. This may be because significant fluctuations in earnings increase stress as workers struggle to adapt to changing economic circumstances. It is also possible that high earnings variability during the 1974–1979 period reflects health problems during that period that show up as increased mortality during the follow-up. In the latter case, the standard deviation can be viewed as a control for health, so it is worth noting that its inclusion in the model lowers only modestly the coefficient associated with mean earnings.

Conditional on the mean and standard deviation of earnings, the number of quarters of zero earnings is not associated with variation in subsequent mortality. However, workers with a higher (logarithmic) growth rate in annual earnings from 1974 to 1979 tend to have lower mortality during the follow-up period, which may reflect better earnings prospects during that period. The last column of Table 1 shows that the negative association between earnings and mortality is stronger for workers whose 1979 job was in a manufacturing industry.

IV. The Effect of Lagged Earnings on Mortality

To examine the lag structure of the relationship between earnings and mortality, we limit our estimation sample to worker-year observations for which we observe positive earnings in each of the previous ten years. This implies a follow-up period of 1984–1992. Because we also want to focus in this section on workers in mid career, we further limit the sample to years in which workers were under age 60. Column 1 of Table 2 shows that a higher average earnings level in the preceding five years is highly associated with lower mortality. The coefficient of −0.498 suggests that a 10 percent increase in that average would lower the probability of death in the following year by approximately 5 percent. Column 2 shows that the effect is attenuated, but still strong, if we look at average earnings over the period six to ten years before the year in question. However, column 3 shows that conditional on earnings in the most recent five-year period, higher earnings in the previous five years are associated with higher, not lower, mortality. In addition, the coefficient on recent earnings increases noticeably. Column 4 splits the previous ten years into three, rather than two, subperiods. The positive association of lagged earnings with mortality is no longer statistically significant, but the results confirm that it is only recent earnings that are independently associated with reduced mortality.

For the results of Table 2 to be consistent with an interpretation along the lines of (1), in which workers buy health by allocating a steady fraction of income to maintain a health stock, the rate of depreciation of that stock would have to be quite high, perhaps implausibly high. Thus the results of Table 2 cast some doubt on that interpretation. They may fit somewhat better with the idea that higher earnings bring greater

TABLE 3—EFFECT OF AVERAGE PRERETIREMENT EARNINGS ON POSTRETIREMENT MORTALITY: WORKERS WITH EARNINGS AGE 55–59

	Males	
Estimation sample	(1)	(2)
Log avg. earnings 1–5 years before age 60	−0.353	
	(0.028)	
Log avg. earnings 1–5 years before retirement × 0–4 years after age 60		−0.362
		(0.029)
Log avg. earnings 1–5 years before retirement × 5–9 years after age 60		−0.350
		(0.028)
Log avg. earnings 1–5 years before retirement × 10–14 years after age 60		−0.345
		(0.028)
Log avg. earnings 1–5 years before retirement × 15 or more years after age 60		(0.370)
		(0.030)
Observations	188,744	188,744

Notes: Entries are coefficients of logistic regression models of the mortality hazard. Models also include year effects and a quartic in age.

access to health insurance, and insurance brings access to better medical care in the event of a health shock. That interpretation emphasizes current coverage, which would be related only to current earnings.

Finally, the results in Table 2 are inconsistent with the idea that the correlation of earnings and mortality is entirely due to both variables being influenced by a common, time-invariant unobserved trait like patience. Average age-adjusted earnings over adjacent five-year periods are likely to be equally good indicators of such a trait. Thus, one can show that they should have approximately the same coefficients in a logistic regression for mortality if the trait is the only reason for the correlation of earnings with mortality. Of course, it is still possible that a time-varying omitted variable could explain the patterns observed, but such a story would have to be more complex.

V. The Effect of Pre-Retirement Earnings on Post-Retirement Mortality

One concern with the estimates shown above is that they may be affected by reverse causality: Deteriorating health may lower earnings in the near term and lead to earlier death in a follow-up period. The importance of such effects should be lower when mortality outcomes are observed well after earnings are recorded. However, a difficulty with results based on long follow-up periods is that base-period earnings may become a relatively poor proxy for earnings during the follow-up because shocks to the earnings process may occur over time. The likelihood of such shocks is lower, however, for those in retirement. Instead, such workers' incomes are likely to be determined primarily by their earnings in the last several years of their working careers. Thus, in this section, we analyze the association of preretirement earnings with postretirement mortality.

Rather than trying to determine exactly when retirement occurs for individual workers, we arbitrarily assume that it begins at age 60. Thus, in Table 3 we present results on the mortality of workers age 60 and above as a function of their earnings when they were in their late 50s. The samples are limited to workers who had positive earnings in each of the five years before they reached age 60. Column 1 shows the effect of average earnings from age 55 to 59 on mortality age 60 and later, assuming this effect is constant over time. The coefficient of −0.353 is comparable to the effects estimated in Table 1 for a similarly long follow-up period. Column 2 introduces interactions of preretirement earnings with periods after retirement to show how the effect of retirement earnings varies over time. The results in column 2 suggest that they are quite constant over time.[4]

[4] This is consistent with results in Table 2, which suggest that the direct effect of pre-retirement earnings on mortality fades relatively quickly. The remaining effect is likely to be due to a stable correlation of preretirement earnings with postretirement income.

VI. Summary and Conclusions

We have examined the structure of the relationship between earnings and mortality using a newly constructed dataset that links longitudinal earnings records to death records. By examining earnings measures that average over several years, instead of a single year, we find that the association between earnings and mortality is substantially stronger than has been reported in the literature. We also find that other characteristics of the earnings process are predictive of mortality. In particular, given the mean level of earnings, higher variability of earnings is associated with increased mortality.

During the prime working years, we find that it is only the most recent few years of earnings that are independently associated with mortality. Holding constant the average of earnings over the previous few years, earnings in earlier years are not significantly associated with mortality. These results call into question the notion of workers using earnings to build up a slowly depreciating health stock. They also rule out interpretations of the earning-mortality correlation as merely one of a common dependence on a time-invariant trait such as patience.

REFERENCES

Abowd, John M., and David Card. 1989. "On the Covariance Structure of Earnings and Hours Changes." *Econometrica*, 57(2): 411–45.

Bound, John, and Alan B. Krueger. 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, 9(1): 1–24.

Case, Ann, Darren Lubotsky, and Christine Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review*, 92(5): 1308–34.

Currie, Janet, and Mark Stabile. 2003. "Socioeconomic Status and Child Health: Why Is the Relationship Stronger for Older Children?" *American Economic Review*, 93(5): 1813–23.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney. 2006. "The Determinants of Mortality." *Journal of Economic Perspectives*, 20(3): 97–120.

Deaton, Angus, and Christina Paxson. 1999. "Mortality, Education, Income, and Inequality among American Cohorts." National Bureau of Economic Research Working Paper 7140.

Efron, Bradley. 1988. "Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve." *Journal of the American Statistical Association*, 83(402): 414–25.

Hill, Mark E., and Ira Rosenwaike. 2001. "The Social Security Administration's Death Master File: The Completeness of Death Reporting at Older Ages." *Social Security Bulletin*, 64(1): 45–51.

Jacobson, Louis S., Robert J. LaLonde, and Daniel Sullivan. 1993. "Earnings Losses of Displaced Workers." *American Economic Review*, 83(4): 685–709.

Smith, James P. 1999. "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status." *Journal of Economic Perspectives*, 13(2): 145–66.

Sullivan, Daniel, and Till von Wachter. Forthcoming. "Job Displacement and Mortality: An Analysis using Administrative Data." *Quarterly Journal of Economics*.