DRAFT – PLEASE DO NOT QUOTE OR CITE

Health Inequality, Education and Medical Innovation

Sherry Glied
Department of Health Policy and Management
Mailman School of Public Health

Adriana Lleras-Muney
Department of Economics and Woodrow Wilson School
Princeton University

October 21, 2002

Corresponding author:
Sherry Glied
sag1@columbia.edu

DRAFT – PLEASE DO NOT QUOTE OR CITE

Abstract

Recent studies suggest that health inequalities across socio-economic groups in the US are large and have been growing. We hypothesize that, as in other, non-health contexts, this pattern occurs because more educated people benefit more than do the less educated from technological advances in medicine (see Nelson and Phelps, 1966 and Bartel and Lichtenberg, 1986). We test this hypothesis by examining the evolution of mortality differentials and medical innovation over time. We focus on cancer mortality and examine both the incidence of cancer and survival conditional on disease incidence. Although there have not been great improvements in cancer survival overall, there has been substantial progress in the treatment of some forms of cancer. We find evidence supporting the hypothesis that more educated people are better able to take advantage of new medical innovations.

In 15th century Florence, merchants deposited dowries into an investment fund on behalf of their daughters. In studying this investment fund, historians find that the annual death rate among women whose dowries exceeded 100 florins was just under half as great as those of women whose dowries were below 50 florins (Morrison, Kirshner, and Molha, 1977). Similar differences in health outcomes by socioeconomic status have been documented in settings as diverse as Civil War era Rhode Island and Victorian Scotland (Chapin, 1924; Davey Smith et al., 1992) and for many countries (including the US, New Zealand, Israel, Canada, Russia and most European countries[1]). Many factors can generate socioeconomic differences in health status at a point in time. For example, in 15th century Florence, people of higher socioeconomic status likely had better living conditions than those with fewer resources. Today, less educated Americans are far more likely to smoke cigarettes than are those with higher education.

What is more difficult to explain is that these differentials persist – and in some cases have even increased – during the past century (Carroll, Davey Smith, and Bennett, 1996). In the six centuries since the Florentine dowry fund existed, the average health of the population has improved considerably. Moreover, as the smoking example above indicates, the principal causes of death and disability today are quite different from those even a century ago (McKeown, 1976; Cutler and Meara, 2002). Housing, nutrition, and sanitation have improved; the infectious diseases that were the prime causes of death before World War I account for little mortality today; and access to effective medical care for those diseases that remain has become more widespread. Indeed, the elimination of many of these risk factors led some writers in the middle of the past century to speculate that social gradients in health would disappear (Kadushin, 1964). Kadushin argued "in modern Western countries, the relationship between social class and the

[1] For evidence on the US see for example Elo and Preston (1996) or Christenson et al. (1995); for New Zealand see Pearce et al. (1983); for Israel see Manor et al (1999); for Canada see Mustard et al. (1997), for Russia see Shkolnikov et al. (1998), for European countries see Kunst et al (1994).

prevalence of illness is certainly decreasing and most probably no longer even exists." But Kadushin was wrong. The risk factors for disease have changed and the gradients have remained.

Disparities in health by socioeconomic status have widened over the past century. For example, in the UK, the Black Report (1980) documented that socio-economic differences in mortality increased after World War II in spite of the introduction of the National Health Service in 1946. Some evidence of widening gradients also exists for Europe (see Kunst et al. 2001 or Shkolnikov, et al. 1998). Feldman et al. (1989), and Pappas et al. (1993), and Elo and Preston (1994) show that in the US disparities in health increased in the second half of the 20th century. Pappas et al. (1993) report that age-adjusted mortality rates for white men with some college dropped from 5.7 to 2.8 per 1000 from 1960 to 1986, whereas they declined only from 9 to 7.6 for those with less than a high school degree. The life expectancy of high school graduates at age 65 is now approximately one year longer than that of high school dropouts.[2]

Many recent studies of differentials focus on gradients associated with education and have found that education is closely correlated with health status (even controlling for income). Economists have devoted considerable attention to the question of why educational gradients in health exist. Numerous studies have shown that increases in education – such as those induced by compulsory schooling laws – lead causally to improvements in health status (for example see Lleras-Muney, 2002). Grossman (1972) suggested that education leads to better health by improving the technology for health production. This might include having access to more information about health risks, making better use of that information or more effectively searching for high quality health providers (Rosenzweig, 1995). Education may also have indirect effects on health by increasing income and improving access to the resources needed to improve health, a possibility we consider briefly below. These roles of education in the production of health provide an explanation for the existence of gradients in health. They do not,

[2] See Richards and Berry (1998)

however, offer an explanation for why gradients might increase or why the gradient moves from one disease to another.

We propose one explanation -- the gradient exists, and moves among diseases, because more educated people are better able to exploit technological advances in the medicine. The most educated make the best initial use of new information about different aspects of health. Over time, the information diffuses along the education gradient. Thus, if technological change ceased, we would expect the gradient to flatten as well. In the steady state, education might still have an effect on health, because better-educated people might be better able to use certain health care technologies (Goldman and Lackdawalla, 2002).

This hypothesis is an extension to health of Nelson and Phelps' (1966) theory that "the return to education is greater the faster the theoretical level of technology has been advancing (p. 72)." Bartel and Lichtenberg (1986) find evidence supporting this hypothesis in the context of the labor market, as does Wozniak (1984) in the agricultural sector.

Our hypothesis is closely related to the sociological conjecture that socioeconomic status is a "fundamental social cause" of gradients in health (Link and Phelan 1, Link, Northridge, Phelan, 2). A fundamental cause is one that involves access to resources that can be used to avoid or minimize risks, influences multiple risk factors, and affects multiple disease outcomes. Fundamental causes, however, cannot generate socioeconomic gradients in health except in the presence of change – change in knowledge about the risks of disease or about treatments (Link, Phelan). In this view, higher socioeconomic status enables people to better exploit new information and resources.

In order to find evidence for this hypothesis, we look at whether health differentials across education groups are greater for those diseases where there has been medical progress and smaller for those diseases for which there has been little innovation. We focus here on gradients in cancer incidence, diagnosis, and survival (conditional on stage of diagnosis). Cancer provides a good area to study because progress in the incidence, diagnosis, and treatment of cancer varies

substantially among cancer subtypes. Although overall progress in the "war on cancer" has been slow, substantial improvements have occurred for certain cancers. For example, there have been substantial improvements in survival after diagnosis of early stage colon cancer (DHHS, 2001). Good long-term epidemiological data – the SEER registries -- exist linking cancer diagnosis to cancer outcomes in a consistent way across subtypes. These data allow us to examine how the incidence of cancer, stage of diagnosis, age at diagnosis, and 5-year survival after diagnosis vary with education. In the case of other diseases, surveillance data on risk factors cannot be as readily linked to outcomes.[3]

This paper is organized as follows. The data and our measures of progress are discussed in section II. We next show how education affects cancer mortality in aggregate and by subtype (Section III). In section IV we present a simple model of the relationship between education health and the rate of innovation. Then using our data we relate our estimated education gradients to our measures of progress. In section V we consider different explanations of our results. Section VII concludes.

II- Data

Our data come from the SEER Cancer Incidence Public Use Database collected by the National Cancer Institute. The data contain information on every person diagnosed with cancer from 1973 to 1998 in 9 SEER registries. The SEER registries are composed of several counties located in San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah and Atlanta.[4] Information on vital status was recorded for all individuals in the sample as of 1998. These data allow us to look at mortality rates conditional on cancer diagnosis. Following the epidemiology literature we look at 5-year mortality rates, by analyzing the effect of covariates on

[3] For example, people with high cholesterol or high blood pressure are at risk of mortality from multiple diseases, making it difficult to link mortality outcomes with progress.

[4] The Seer data include two more registries (San Jose and LA) but we exclude them since data are only available from 1992 to 1998 for these registries.

the probability that an individual will die within 5 years of being diagnosed with cancer. To avoid censoring, we limit our sample to those diagnosed with cancer no later than 1993. The SEER data contain a large number of observations (N=2,556,432) so we can perform analysis of death rates within detailed disease category.

Education Measures

Unfortunately, the SEER registry data do not include information on educational status. Instead, we use two distinct proxies for educational status – compulsory schooling laws and average education level by cohort and registry.

The SEER data contain information about state of birth, year of birth, gender and race. We can therefore match individuals to compulsory attendance and child labor laws in place in their state of birth when they were 14 years of age. These laws, which implicitly specified the number of years that a child had to attend school, serve to identify the effect of education. The data on compulsory attendance and child labor laws were collected from multiple sources (See Lleras-Muney, 2001, for details). We use only two laws: the age at which a child had to enter school and the age at which he could get a work permit and leave school. The difference between these two variables measures the implicit number of years a child had to attend school. This variable ranges from 0 to 10 for the cohorts we study. Because compulsory schooling laws were most effective in the first half of the 20th century and they only affected whites (see Lleras-Muney 2001), we restrict our attention to white cohorts born between 1901 and 1925.

Several papers have shown that these laws had an impact on educational attainment, especially in the first half of the century (see Acemoglu and Angrist, 1999, Angrist and Krueger, 1991, Lleras-Muney, 2001, Margo and Finnegan, 1996, and Schmidt, 1996). By including them in place of education in a model of mortality/health, we are estimating a reduced form equation. Using the 1960 census we can estimate the first stage equation of the effect of compulsory schooling on education. Since the model is exactly identified and provided that we include the

same covariates in both estimations, the Two-Sample IV estimate of the effect of education on mortality can be calculated as the ratio of the reduced form equation estimate and the first stage estimate.[5] The advantage of this method is that we can argue (as other have) that the effects we measure are causal effects of education (direct or indirect).

As an alternative we also match individuals with average education levels in their cohort, gender, and registry. This measure of education can be calculated from the census in 1970, 1980 and 1990. We match individuals to education by decade, i.e. individuals diagnosed in the 1970s are matched to the average education in their cohort, gender, and registry, calculated from the 1970 census. Unfortunately mean education and income are not available for all possible cells. Therefore we must further restrict our sample to those individuals for whom average education and income can be imputed.

An advantage of the registry-level average education proxy is that we can also calculate total family income for the same cells. We can, therefore, include this income control in the regressions. A problem with this proxy, however, is that it may also capture average characteristics of the registry. For example, registries with more educated older people may have better cancer doctors. Furthermore, average education may be correlated with unobservable characteristics, such as rates of time preference (Fuchs, 1982).

Throughout the paper, we provide results using each of these two proxies. The two proxies both predict income (about equally well), but they are not highly correlated with each other. The simple correlation between them is 0.12. Using the Census data, we estimate individual education levels as a function of registry-cohort level average education and compulsory schooling laws. The result is:

[5] This method was used by Dee and Evans (1999).

$$(1)\quad \text{education}_i = \underset{(0.004)}{0.030}\ \text{registry mean education} + \underset{(0.010)}{0.070}\ \text{compulsory schooling} + \text{controls}$$

where controls include female, age, age^2, cohort, state of birth, registry, census year.

Outcome Measures

We examine four types of outcome measures: 5-year survival rates, cancer incidence rates, stage of diagnosis, and age at diagnosis. Survival rates are computed from individual level data. We focus most of our attention on gradients in survival rates (conditional on stage of diagnosis), because these gradients are largely independent of trends in the incidence of other diseases.

We calculate annual incidence rates by cancer site, interview year, gender cohort and state of birth. In the SEER data we count the number of people diagnosed with cancer by interview year, gender, cohort, state of birth and cancer site. We then divided by the population in that group. Population figures are only available from decennial censuses. We use the 1970, 1980 and 1990 censuses to generate estimates of the population for each gender/cohort/state-of-birth/site cell using a linear time trend. We match incidence rates calculated in this manner with compulsory schooling level by cohort and state-of-birth level. Alternatively, we calculate rates by cancer site, census year, gender and registry (where population was also imputed from the censuses for each cell) and match them to average education at the census year, gender and registry level. Incidence rates provide a measure of the effectiveness of cancer prevention. However, changes in cancer incidence (both overall, by education level, and by cancer subtype) also reflect progress in the incidence and treatment of other diseases, particularly cardiovascular disease. Increases in the age-adjusted incidence of cancer in a group may simply reflect a decline in cardiovascular disease mortality for that group.

Age at diagnosis provides information on both prevention (which may lead cancer to appear later) and diagnostic technology. Stage at diagnosis largely measures diagnostic improvements. We use a dichotomous variable (localized/not localized) for stage of diagnosis.

Summary statistics for the final sample used in this paper are in Table 1. Average age at diagnosis for this sample is around 70. About 2/3 of the population died within 5 years of diagnosis. The most common cancers are cancers of the digestive system, of the respiratory system and of the genital system. Note that our sample is relatively old because we exclude people born after 1925. [6]

Measures of progress

There is no consensus about how to measure either progress or the relative importance of progress. Our hypothesis is based on a measure of progress in knowledge and information about risk factors and treatment, but we do not have straightforward measures of knowledge. Instead, we compute multiple measures of innovation for each of the 81 cancer sites in the SEER data. Table 2 describes these measures.

The first measure looks at whether age-adjusted mortality rates have declined from 1969 to 1999. The National Cancer Institute provides a measure of the trend in age-adjusted mortality: the estimated annual percent change (EAPC), which is the coefficient from a log-linear regression of mortality rates on calendar year. Note that the EAPC is positive if age-adjusted mortality increased and negative if age-adjusted mortality decreased. A negative value for EAPC constitutes progress. Across all 81-cancer sites there has been progress in cancer mortality over this period.

Second, we look at changes in age-adjusted incidence rates for each disease. Declines in incidence rates will affect mortality but they reflect advances in prevention rather than in

[6] The average age at diagnosis is around 62 in the full SEER data. Our sample is older but not much more.

treatment. The National Cancer Institute provides a measure of the trend in age-adjusted mortality: the estimated annual percentage change in age-adjusted incidence (EAPC). As was the case with mortality, negative EAPC values constitute progress. Overall, there has been negative progress in incidence – rates have increased. As noted, because this increase in incidence also reflects progress in other diseases, it could be coincident with progress in cancer prevention (that is not well-reflected in this measure).

Using the SEER data, we compute a third progress measure -- the change in the 5-year survival rate conditional on diagnosis. This measure is related to changes in mortality rates but it is also affected by innovation in diagnosis technology. In the absence of any progress in diagnostic technology, this measure should capture innovation in treatment. Over this period, survival conditional on diagnosis has increased.

Finally, we use data on the number of drugs approved by the FDA to treat a particular cancer as a measure of the rate of pharmaceutical innovation for each particular site. As in Lleras-Muney and Lichenberg (2002) or Lichtenberg (2002), we use the number of new active ingredients approved by the FDA rather than the number of new drugs, which we consider a better measure of innovation in drug treatments (the FDA also approves generic equivalents and new dosages of the same drug for example). Although new drugs are an imperfect measure of medical innovation, they are our only direct measure of innovation in treatment. For each cancer site we construct two measures: the number of drugs that exist in the market as of 1999; and the number of drugs approved between 1973 and 1993. Since year of FDA approval is not known for all the drugs in our sample, the latter is measured with error but it has the advantage of being calculated exactly over the time period of interest. These data were constructed using several sources: First Data Bank provided a list of the drugs that are used to treat cancers, and the date of FDA approval of the active ingredient in each drug was kindly provided by Frank Lichtenberg.

While these measures are related to one another, the correlation among them is not very high (Table 3). Both improvements in incidence and in survival are correlated with mortality, but

the correlation of incidence is greater than that of survival. Survival and incidence are not correlated across cancer sites. Incidence is positively correlated with new drugs, which may indicate that pharmaceutical manufacturers target diseases with rising incidence. Drugs are relatively highly positively correlated with changes in survival.

III. – The Effect of Education on Cancer Outcomes

We first examine the overall effect of education on cancer survival. These results are reported in Table 4. The effect of education on overall survival is negative and significant, using either the compulsory schooling or mean education specifications. The effect of education is greater for cancers affecting men than for those affecting women. This finding is consistent with other studies (e.g., Elo and Preston, 1994), which also show that the effect of education on health is greater for men than for women.

Using the compulsory schooling laws, we estimate two-stage IV estimates of the effect of education. We estimate the first stage using the 1960 census and find that the effect of years of compulsory schooling on education is 0.051. Using this information, we calculate that the TSIV estimate of the effect of education on mortality is somewhere between –0.06 and –0.04. This estimate is somewhat higher than the OLS estimates, which is often the case when the instruments affect only those at the lower end of the distribution of education. The TSIV estimates, however, are close to those in the existing empirical literature on education effects. Elo and Preston (1996) find that the effect of education on 5-year mortality rates is between [-0.02;-0.05]. Since our objective in these paper is not to provide accurate estimates of the effect of education on cancer, but rather to look at whether the education gradient is related to progress, in the remainder of the paper we will present reduced form estimates of the effect of compulsory schooling as proxies for the effect of schooling.

Next, we look at cancers according to a primary classification of 16 types and estimate the effect of education separately for each type of cancer (Table 5). We find that the effect of

education on cancer survival differs by type of cancer. For example, the effect of education on survival with urinary system cancers is about 5 times greater than the effect on survival with respiratory cancers. We also find that significant gradients by education exist for several common cancer types. The results are quite similar for the two measures of education.

The next two tables report the effect of education on intermediate cancer outcome measures – age at diagnosis and stage at diagnosis (Tables 6 and 7). We find that higher mean education is associated with later age at diagnosis (although we do not find this result using the compulsory schooling measure). This result suggests that the apparent survival advantage of more educated people is not simply due to earlier diagnosis of conditions that might never have led to mortality. More educated people are likely to be older, not younger, when they are diagnosed with cancer. Higher mean education or compulsory schooling is also correlated with a greater probability that cancer is localized when diagnosed. Localized cancers are likely to be more treatable than cancers found after they have spread.

We next examine the effects of education on the incidence of cancer (Table 8). Changes in cancer incidence are likely to be associated with behavioral changes – or changes in the incidence of other diseases -- rather than changes in medical treatment. We find that increases in compulsory schooling lead to significant reductions in cancer incidence for all cancers and for several specific types of cancer. By contrast, we find that increases in mean education at the registry level have little impact on incidence. The difference between these results may be an artifact of the aggregation process. In computing incidence we rates, we use data that are aggregated. The aggregation is much greater using mean education (N=20,348) than using compulsory schooling laws (N=336,509). Results using compulsory schooling include lots of cells that each contain very few people, while results that use mean education have fewer cells but more people in each cell.

To summarize, our results suggest that education has significant effects on the age of incidence of cancer, stage of cancer diagnosis, and survival after cancer diagnosis and (by some

measures) on the incidence of cancer. More educated people are likely to be older when they are diagnosed, they are more likely to be initially diagnosed with localized cancer, they are more likely to survive for 5 years after diagnosis, and they may be les likely to get cancer at all. Most of these results are similar regardless of which measure of education is used.

IV – Relating Outcomes to Progress

A simple model of health, education and the rate of innovation

The following model closely follows Nelson and Phelps (1966). This model is only illustrative. It captures the basic features of our hypothesis in a simple fashion and provides guidelines for our empirical approach.

Suppose that the health H of an individual can be modeled as a function of the level of technology A that the individual has access to, and other inputs C:

$$H = H(A, C)$$

The theoretical level of technology is given by T(t), where

$$T(t) = T_0 e^{\lambda t}$$

T(t) is the level of technology if technology is instantaneously diffused and λ is the exogenous rate of technological progress. Suppose now that the level of technology available to any individual depends on how rapidly individuals adopt new technologies, and that the lag between innovation and adoption is a decreasing function of education, so that

$$A(t) = T(t - w(e)) = T_0 e^{\lambda(t - w(e))}$$

where $w'(e)<0$. This key assumption captures the ideas that were presented in the introduction -- that is that the more educated "adopt" new technologies at a faster rate because of better access to information, better use of information, and better capacity to search for better providers/treatments. This feature can be generated from maximization principles simply by assuming differential costs of technology adoption.

In this context, we can express the health of the individual as:

$$H = H(T_0 e^{\lambda(t-w(e))}, C)$$

The derivative of the health production function with respect to education gives us what is known as "the education gradient" in health. It gives the marginal gain in health induced by an additional unit of schooling. In this model it can be expressed as:

$$\frac{\partial H}{\partial e} = -\lambda w'(e) A \frac{\partial H}{\partial A} > 0$$
$$\frac{\partial H}{\partial e \partial \lambda} = -w'(e) A \frac{\partial H}{\partial A} > 0$$

This model predicts that in the absence of technological change, there should be no difference between the educated and the uneducated. Since w'(e) is negative, the model also predicts that health is an increasing function of education and that the rate of return of education is larger the higher the rate of technological change. We test these two predictions in the data, first by estimating the disease-specific education gradient, and then by relating the size of the gradient to measures of innovations that proxy for the parameter λ. Again note that in this model technology is a broad concept that captures all innovations that affect the manner and the rate in which we can transform inputs into health.

Although this model is simple, it is a good description in the context of health. While the income returns to education are determined by the labor market, the returns to education in health are only determined by the individual health production function. There is no sense in which there exist general equilibrium effects in the health model: if everyone obtains more schooling everyone's health improves. Individuals are assumed to have chosen education in previous periods and we are assuming that technological changes are unanticipated.

Empirical evidence on the education gradient and innovation

Our hypothesis suggests that gradients in education should be greatest where medical progress has been greatest. To test this, we next relate education gradients to measures of

medical progress. In this section, we compute gradients for the 81 sites of cancer as classified in the SEER registry and examine how they relate to each of the progress measures. We implement three econometric specifications.

In the first specification, we estimate a linear probability model of the probability of dying within 5 years after diagnosis (conditional on stage at diagnosis):

$$(2)\quad P(died=1)=\beta_0+\beta_1 education+\beta_2 education*progress+X\gamma+e\,.$$

Where X includes 47 state of birth, 24 cohort, 8 registry, 2 decade, and 4 stage of diagnosis dummies), dummies for cancer sites. Our model suggests β_2 should be negative. In the strictest interpretation of the model we should further expect that β_1 be 0. However since our measures of progress are not perfect and because there might be additional reasons why education is related to health that are not captured by the model, we might observe negative estimates of β_1. Note that in this specification we constrain the coefficients on all variables to be the same across all cancer sites, except that we allow education to vary with our measure of medical progress.

In a second specification, we free up the functional form by separating the two stages. Initially, we run a single individual level regression, including all the controls above, and interacting education with cancer site dummies:

$$P(died_{ij}=1)=\beta_0+\sum_{j=1}^{81}\beta_j education_i*cancer_j+X_{ij}\gamma+e_{ij}$$

where i indexes individuals, and j indexes the 81 different cancers. The resulting 81 coefficients (labeled β_j) from the interaction terms (education*cancer site) become the dependent variable in a second stage regression where medical progress related to that cancer site is the independent variable:

$$\beta_j=\delta_0+\delta_1 progress_j+\varepsilon_j,\ j=1...81$$

This second stage regression is weighted by the variance of β, the effect of education. In this specification, the effect of education is allowed to vary by cancer site, but all other variables

are constrained to have the same effect across sites. A further advantage of this specification is that measures of progress are not included in the estimates of education gradients. Since our measures of progress are all based on the SEER data, this two-stage method eliminates a possible endogeneity of the progress measures.

In a third specification, we run separate regressions for each cancer site, including all the controls. We use the coefficient on education from each regression as the dependent variable in the second stage regression. In this fully flexible form, none of the variables are constrained to be the same across sites. Unfortunately, because of the very large number of dummy variables in our specification, there is little statistical power available to identify results in this specification. In principle, the dummy variables alone fully identify 73,782 observations, and for many cancer sites we have far fewer observations available (see Appendix 1).

The results from these 3 specifications are shown in Tables 9-11.

The first panel of Table 9 provides the results of the most constrained specification where progress is measured by reductions in age-adjusted mortality. Education improves survival in both education proxy specifications, but the results for progress are contradictory – the mean education measure suggests that the education gradient is steeper where progress has been greater, but the compulsory schooling measure shows the opposite. The 2nd panel shows the results where progress is measured as age-adjusted incidence. Here, using either measure of education, we find that progress reduces the gradient in education. This implies that the cancer survival gradient in education is steepest for those cancers whose incidence is increasing. One reason for this may be that the more educated are more likely to be vigilant about obtaining screening and treatment for those cancers that are most common.

The third panel, which uses our preferred measure of progress, survival after diagnosis, shows that the education gradient in survival is steepest for those diseases where survival has been improving. In fact, the direct effect of education is zero or positive in this specification. The final panel shows results for drugs, which also have their effect primarily on survival after

diagnosis. Here too, we find that the education gradient is steepest for those diseases where there has bee n the most progress.

Table 10 shows results for the more flexible specification. The results in this specification for progress measured as incidence, survival, or drugs are similar in direction and magnitude to those in the constrained specification, although significance levels are lower.

Table 11 shows the results for the fully flexible specification. Again, the results are largely in the same direction as those in the more constrained specification but the estimates are much smaller in magnitude and are mainly statistically insignificant.

We next examine the effects of progress on education gradients in incidence and stage of diagnosis. We report results for the partially flexible specification only. Table 12 shows results for the education gradient in incidence. We find that gradients in incidence are greatest for those cancers that are becoming more prevalent. We also find that gradients in incidence are greater for those cancers where progress in survival is greatest. Progress in survival may reflect greater understanding of cancer, which in turn generates gradients in incidence.

Table 13 shows results for stage of diagnosis. We find that gradients in stage of diagnosis are greatest for those diseases where there has been little progress in overall mortality (which combines effects of incidence and survival) and in those diseases where there has been progress in drugs. Again, this finding is consistent with educated people seeking out early diagnosis for those diseases where there are more treatment options post-diagnosis.

V—Mechanisms by Which Education May Affect Outcomes

Our results support our hypothesis, suggesting that education enables people to make more effective use of technological progress in cancer treatment. As suggested earlier, this effect may be a consequence of the indirect impact of education – through generating higher income – or the direct, causal impact of education itself. We do not have strong tests that allow us to distinguish between these hypotheses, but we consider three here.

First, we compare the effects of education with and without controls for average family income. These results are reported in the first 2 columns of Table 14. We find that average family income has an independent effect on cancer survival. However, we also find that the relationship between the education gradient in survival and measures of progress is virtually unaffected by adding controls for family.

Next, we compare the effects of education for those diagnosed before and after Medicare eligibility (age 65). For the population below Medicare eligibility age, education may be related to differences in health insurance and access to medical care, but this should be less true of the population 65 and over. Note that because our sample is quite old, the sample of people diagnosed before age 65 is relatively small. Nonetheless, we find effects that go in the same direction for both samples and effects for progress measured in survival are significant for both subsamples. Overall, however, the correlations between the gradient and progress appear, if anything, greater for the population with Medicare than for the population below age 65.

Finally, we review the results in the prior panels that show the effects of education measured as compulsory schooling and as mean education and note that the results yielded by the two specifications are quite similar in almost every case. The compulsory schooling measure can reasonably viewed as showing a causal effect of education here, particularly in examining the effects of survival after diagnosis of cancer. These results suggest that education itself, rather than some other characteristic of those who choose to become educated, has an effect on cancer survival.

VI. Conclusions and Limitations

This study provides evidence that education gradients in health are greatest where technological progress has been greatest. It is, thus, a test of the Nelson and Phelps hypothesis in a novel context. An alternative approach to this problem would be to examine changes in gradients over time and associate them with changes in progress over time. We cannot do this

with our data because our sample is restricted to 25 cohorts. The limited scope of our data makes it difficult to measure discrete education gradients at different points in time. We did attempt to measure changes in gradients by cancer disease subtype but our estimates were extremely imprecise. Furthermore, our cohorts are aging, and the effect of education on survival is diminishing with age. Thus, these gradients will also be biased by the general aging of our limited sample.

Comparing gradients to progress is made even more difficult because we have no hypotheses about the lags between innovation and diffusion across education groups. These lags are likely to vary with the particular technological innovation considered. In some cases, the less educated may find it very difficult to catch up to the more educated (for example, smoking differentials persist to the present). In other cases, the less educated may catch up faster.

A final problem in conducting our analysis is that rates of technological progress may themselves be endogenous. We observe some evidence of endogenous technological progress in the positive correlation between the change in incidence and the number of drugs that exist for treatment of a given disease. General correlations between incidence and progress will not bias our results. It seems somewhat unlikely that technological progress would be endogenous to the education gradient in incidence or survival with a particular disease.

MORE CONCLUSIONS TO COME

References

Adler, Nancy E. et al, "Socioeconomic Status and Health, the Challenge of the Gradient," American Psychologist, vol 49, No. 1 January 1994

Acemoglu, Daron and Joshua Angrist, "How Large are the Social Returns to Education? Evidence from Compulsory schooling Laws," NBER Working Paper No. W7444, December 1999

Angrist, Joshua D. and Alan B. Krueger, "Does Compulsory School Attendance Affect Schooling and Earnings?," Quarterly Journal of Economics, November 1991

Bartel, Ann P. and Frank R. Lichtenberg, "The Comparative Advantage of Educated Workers in implementing new technology," the Review of Economics and Statistics, February 1987

Chapin, C.V. (1924) Deaths among taxpayers and non-taxpayers of income tax, Providence, 1865. American Journal of Public Health, 14, 647-51.

Christenson, Bruce A., and Nan E. Johnson, "Educational Inequality in Adult Mortality: An Assessment with Death Certificate Data from Michigan," Demography 32. May 1995

Elo, Irma T. and Samuel H. Preston, "Educational Differentials in Mortality: United States, 1979-85," Social Science and Medicine 42(1), 1996

Feldman J. D. Makuc, J. Kleinman and J. Cornoni-Huntley, "National Trends in Educational Differences in Mortality," American Journal of Epidemiology, 1989

Fuchs, Victor R. and Harold C. Sox Jr., "Physicians' View of the Relative Importance of Thirty Medical innovations," Health Affaris, Volume 20, Number 5, September/October 2001

Grossman, Michael, ``The Human Capital Model of the Demand for Health," in the Handbook of Health Economics, edited by Culyer et al, Elsevier 2000

Kitagawa and Hauser, Differential Mortality in the United States: a Study in Socioeconomic Epidemiology. Cambridge, MA: Harvard University Press, 1973.

Lichtenberg, Frank, "the Benefits and Costs of Newer Drugs: Evidence from the 1996 Medical Expenditure Panel Survey," mimeo, Columbia University 2000

Link, Bruce G. et at, "Social Epidemiology and the Fundamental Cause Concept: on the Structuring of effective Cancer Screens by Socioeconomic Status," The Milkbank Quarterly Vol. 76, No. 3, 1998

Lleras-Muney, Adriana "The Relationship between Education and Adult Mortality in the U.S." mimeo, Columbia University 2001

Lleras-Muney, Adriana, "Were State Laws on Compulsory Education Effective? An analysis from 1915 to 1939," mimeo, Columbia University, 2001

Margo, Robert A. and T. Aldrich Finegan, "Compulsory Schooling Legislation and School Attendance in Turn-of-The Century America: A 'Natural Experiment'," Economics Letters, October 1996

Marmot, Michael J., M. J. Shipley, and G. Rose, 1984, "Inequalities in death—specific explanations of a general pattern?" Lancet, 1, 1003–6.

McKeown, Thomas, The modern rise of Population, New York: Academic Press, 1976

Nelson, Richard R. and Edmund S. Phelps, "Investment in Humans, Technological Diffusion and Economics Growth," American Economic Review, 1966

Pappas, Gregory, Susan Queen, Wolber Hadden and Grail Fisher, "The Increasing Disparity in Mortality Between Socioeconomic Groups in the United States, 1960 and 1986," The New England Journal of Medicine, 1993

Richards, H. and Berry, R. (1998). U.S. life tables for 1990 by sex, race, and education. Journal of Forensic Economics 11, 9–26.

Rosenzweig, Mark R., "Why are there returns to Schooling?," American Economic Review, Volume 85, Issue 2, May 1995

Rosenzweig, Mark R. and Paul T. Schultz "Schooling, Information and Nonmarket Productivity: Contraceptive Use and its effectiveness," International Economic Review, May 1989, 30920), pp. 457-77

Schmidt, Stefanie, "School Quality, Compulsory Education Laws, and the Growth of American High School Attendance, 1915-1935," MIT Ph.D. Dissertation 1996

Wozniak, Gregory D., "The Adoption of Interrelated Innovations: A Human Capital Approach," The Review of Economics and Statistics, Vol. 66, No. 1. (Feb., 1984), pp. 70-79.

Table 1: SEER Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Years of compulsory school	711450	6.93	1.05	0	10
Mean education in cohort, gender and registry	711450	10.93	1.06	4.944	16
Mean total family income in cohort, gender and registry	711450	30258.68	36101.75	891.277	527999.5
Female=1	711450	0.46	0.50	0	1
Birth year	711450	1913.68	6.67	1901	1925
Age at Diagnosis	711450	69.48	8.00	47	92
Hispanic=1	711450	0.02	0.14	0	1
Married=1	711450	0.64	0.48	0	1
Died within 5 year of diagnosis=1	711450	0.63	0.48	0	1
Year of diagnosis	711450	1983.65	5.74	1973	1993
Cancer Site (Broad categories)					
Bones and joints	711450	0.00	0.03	0	1
Brain and other nervous system	711450	0.01	0.11	0	1
Breast	711450	0.12	0.33	0	1
Digestive system	711450	0.23	0.42	0	1
Endocrine system	711450	0.00	0.07	0	1
Eye and orbit	711450	0.00	0.04	0	1
Genital system	711450	0.20	0.40	0	1
Leukimia	711450	0.03	0.16	0	1
Lymphomas	711450	0.03	0.18	0	1
Buccal cavity and pharynx	711450	0.03	0.17	0	1
Multiple Myeloma	711450	0.01	0.11	0	1
Ill-defined and unspecified sites	711450	0.03	0.17	0	1
Respiratory system	711450	0.20	0.40	0	1
Skin	711450	0.02	0.13	0	1
Soft tissue	711450	0.00	0.06	0	1
Urinary system	711450	0.07	0.26	0	1

Notes: Total family income was deflated using the CPI. The base year is 1989.

Table 2: Summary Statistics on measures of progress for 81 cancer sites.

Variable	Obs	Mean	Std. Dev.	Min	Max
Estimated Annual Percent Change in age-adjusted mortality [1]	81	-1.279	2.572	-9.1	8.5
Estimated Annual Percent Change in the age-adjusted incidence rate[2]	80	0.238	2.196	-7.3	9
Change in the 5-year survival rate, conditional on diagnosis[3]	81	0.079	0.101	-0.36	0.299
Number of drugs[4]	81	9.654	10.015	0	48
Number of drugs approved 1973-1993[5]	81	3.259	3.130	0	10

Notes:

(1) Estimated Annual Percent Change in age-adjusted mortality is calculated as follows:
$EAPC=(e^b-1)*100$,
where b is the coefficient from the following regression:
log(rate)= constant + b*(calendar year),
where rate refers to the age-adjusted mortality rate for whites, and the time period used to calculate the change is 1969 to 1999. This data are provided by the National Cancer Institute, mortality rates are calculated from Vital Statistics using the entire US. Age adjustments use the 2000 US population. This statistics is calculated for men and women jointly with the exception of diseases of the genital system which are calculated for each gender separately (site recodes 27010, 27020, 27030, 27040, 27050, 27060 and 27070 for women; site recodes 28020, 28030 and 28040 for men).

(2) Estimated Annual Percent Change in age-adjusted incidence is calculated as follows:
$EAPC=(e^b-1)*100$,
where b is the coefficient from the following regression:
log(rate)= constant + b*(calendar year),
where rate refers to the age-adjusted incidence rate for whites, and the time period used to calculate the change is 1973 to 1999. This data are provided by the National Cancer Institute, age-adjusted incidence rates are calculated using 9 registries in the SEER data base. Age adjustments use the 2000 US population. This statistics is calculated for men and women jointly with the exception of diseases of the genital system which are calculated for each gender separately (site recodes 27010, 27020, 27030, 27040, 27050, 27060 and 27070 for women; site recodes 28020, 28030 and 28040 for men). This statistic is not provided for Other Monocytic Leukimia (site recode 35033).

(3) Change in the 5-year survival rate conditional on diagnosis is calculated as follows:
(% diagnosed in 1973,1974,1975 who died in 5 years)-(% diagnosed in 1991, 1992, 1993 who died in 5 years), where only whites are used to calculate the survival rates. This statistics is calculated by the authors using the SEER mortality data.

(4) The number of drugs by cancer site is calculated only using the number of distinct active ingredients approved by the FDA. In other words, we do not simply calculate the number of drugs in the market, we calculate the number of chemically distinct compounds available, which results in a much smaller number of drugs available. Note that it is not always straightforward to assign drugs to cancer sites. Therefore there is some measurement error. A list of all cancer drugs, the conditions they are used for and their year of approval is available from the authors upon request.

(5) The number of drugs approved by the FDA between 1973 and 1993 was calculated using information on the year of approval by the FDA. Note that year of approval could not be calculated for some drugs. Cancer drugs for which the year of approval was not available were excluded from the calculation.

Table 3: Correlation between the different measures of progress

	Estimated Annual Percent in the age-adjusted mortality rate	Change in the 5-year survival rate conditional on diagnosis	Estimated Annual Percent change in the age-adjusted incidence rate	Number of drugs (active ingredients)	Number of drugs (active ingredients) approved 1973-1993
Estimated Annual percent change in age-adjusted mortality rate	1				
Change in the 5-year survival rate, conditional on diagnosis	-0.20	1			
Estimated Annual Percent Change in the age-adjusted incidence rate	0.56	-0.05	1		
Number of drugs	0.17	0.30	0.17	1	
Number of drugs approved 1973-1993	0.08	0.25	0.08	0.81	1

Notes: See previous table for definitions and data sources.

Table 4: The effect of education on the probability of dying in the next 5 years conditional on cancer diagnosis (all cancers)

	Effect of education measured using compulsory schooling laws			Effect of education measured using mean education in cohort, gender and registry	
	Effect of compulsory school on the probability of dying in 5 years	Years of compulsory schooling on education 1960 Census	TSIV Effect of education on the probability of dying in 5 years	No controls for family income	Controls for mean family income
All	-0.002***	0.051**	-0.039*	-0.005***	-0.004***
	(0.001)	(0.009)	(0.021)	(0.001)	(0.001)
Males	-0.002*	0.035**	-0.057	-0.003**	-0.002*
	(0.001)	(0.011)	(0.034)	(0.001)	(0.001)
Females	-0.003***	0.067**	-0.044**	-0.000	0.000
	(0.001)	(0.011)	(0.017)	(0.002)	(0.002)

Notes: Standard errors in parentheses. Regressions include age at diagnosis, age at diagnosis squared, 47 state of birth dummies, 24 cohort dummies, 80 cancer site dummies, 2 decade dummies, 8 registry dummies and 4 stage of cancer at diagnosis dummies. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed. Standard errors for the TSIV estimates were calculated using the Delta method.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: The effect of education on 5-year death rates by cancer type

Dependent Variable:		Died within 5 years of diagnosis	
	N	Compulsory schooling	Mean education
Buccal cavity and pharynx	21356	0.004	-0.010*
		(0.004)	(0.006)
Digestive system	165944	-0.002	-0.003*
		(0.001)	(0.002)
Respiratory system	140033	-0.002*	-0.001
		(0.001)	(0.002)
Bones and joints	675	0.022	-0.008
		(0.023)	(0.031)
Soft tissue	2472	0.002	0.003
		(0.012)	(0.016)
Skin	12338	0.001	-0.002
		(0.005)	(0.007)
Breast	87729	-0.002	0.002
		(0.002)	(0.003)
Genital system	140671	-0.003**	-0.010***
		(0.001)	(0.002)
Urinary system	52514	-0.010***	-0.006*
		(0.003)	(0.003)
Eye and orbit	1233	0.009	-0.02
		(0.018)	(0.023)
Brain and other nervous system	9008	0.001	0.006
		(0.004)	(0.005)
Endocrine system	3516	-0.002	0.002
		(0.009)	(0.014)
Lymphomas	24162	-0.001	0.001
		(0.004)	(0.005)
Multiple Myeloma	9017	0.001	0.004
		(0.005)	(0.007)
Leukimia	18561	-0.004	-0.002
		(0.004)	(0.005)
Ill-defined and unspecified sites	22221	0.001	-0.003
		(0.003)	(0.003)

Notes: Each coefficient reported is estimated using a separate regression. Standard errors in parentheses. Regressions include age at diagnosis, age at diagnosis squared, diagnosis year, 47 state of birth dummies, 24 cohort dummies, cancer site dummies, 2 decade dummies, 8 registry dummies and 4 stage of cancer at diagnosis dummies. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: The effect of education on age at diagnosis by cancer type

Dependent Variable:		Age at diagnosis	
	N	Compulsory schooling	Mean education
All cancer sites	711450	-0.003	0.027***
		(0.004)	(0.005)
Buccal cavity and pharynx	21356	0.013	0.010
		(0.023)	(0.030)
Digestive system	165944	-0.002	0.017*
		(0.008)	(0.010)
Respiratory system	140033	-0.011	0.015
		(0.01)	(0.012)
Bones and joints	675	-0.015	-0.008
		(0.125)	(0.165)
Soft tissue	2472	0.011	0.111
		(0.061)	(0.083)
Skin	12338	0.033	-0.018
		(0.029)	(0.035)
Breast	87729	-0.001	0.064***
		(0.012)	(0.018)
Genital system	140671	0.000	0.033***
		(0.008)	(0.010)
Urinary system	52514	-0.001	0.005
		(0.015)	(0.018)
Eye and orbit	1233	-0.099	0.038
		(0.091)	(0.116)
Brain and other nervous system	9008	-0.055	0.101**
		(0.037)	(0.049)
Endocrine system	3516	-0.036	-0.167**
		(0.054)	(0.083)
Lymphomas	24162	-0.002	-0.019
		(0.018)	(0.024)
Multiple Myeloma	9017	0.010	-0.029
		(0.033)	(0.045)
Leukimia	18561	-0.013	0.021
		(0.023)	(0.030)
Ill-defined and unspecified sites	22221	-0.013	0.072***
		(0.020)	(0.028)

Notes: Each coefficient reported is estimated using a separate regression. Standard errors in parentheses. Regressions include diagnosis year, age, age squared, 47 state of birth dummies, 24 cohort dummies, cancer site dummies, 2 decade dummies, 8 registry dummies and 4 stage of cancer at diagnosis dummies. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 7: The effect of education on probability that cancer is in situ or localized at time of diagnosis- by cancer type

Dependent Variable:		Cancer stage in situ or localized	
	N	Compulsory schooling	Mean education
All cancer sites	711450	0.000	0.001
		(0.001)	(0.001)
Buccal cavity and pharynx	21356	-0.004	0.002
		(0.004)	(0.006)
Digestive system	165944	-0.001	0.003
		(0.001)	(0.002)
Respiratory system	140033	-0.002	0.000
		(0.002)	(0.002)
Bones and joints	675	0.029	-0.013
		(0.025)	(0.033)
Soft tissue	2472	-0.012	-0.008
		(0.013)	(0.018)
Skin	12338	0.007	-0.004
		(0.005)	(0.006)
Breast	87729	0.004	0.006
		(0.002)	(0.004)
Genital system	140671	0.003*	-0.003
		(0.002)	(0.002)
Urinary system	52514	0.002	0.007**
		(0.003)	(0.003)
Eye and orbit	1233	-0.023	0.003
		(0.016)	(0.021)
Brain and other nervous system	9008	0.001	0.003*
		(0.001)	(0.002)
Endocrine system	3516	-0.018*	-0.012
		(0.011)	(0.017)
Lymphomas	24162	-0.001	0.000
		(0.003)	(0.004)
Multiple Myeloma	9017	-	-
Leukimia	18561	-	-
Ill-defined and unspecified sites	22221	-	-

Notes: Each coefficient reported is estimated using a separate regression. Standard errors in parentheses. Regressions include 47 state of birth dummies, 24 cohort dummies, cancer site dummies, 2 decade dummies, 8 registry dummies. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 8: The Effect of education (measured by compulsory schooling years) on Incidence rates

	Compulsory school [1]	Mean education in cohort gender and registry[2]
Sample		
All cancers	-0.0001***	0.0000
	(0.000)	(0.0000)
Buccal cavity and pharynx	-0.0001	0.0000
	(0.000)	(0.0001)
Digestive system	-0.0001***	0.0000
	(0.000)	(0.0000)
Respiratory system	-0.0001	-0.0002
	(0.0001)	(0.0002)
Bones and joints	0.000	-0.0002*
	(0.000)	(0.0001)
Soft tissue	-0.0001**	0.0000
	(0.0001)	(0.0001)
Skin	-0.0001***	0.0001
	(0.000)	(0.0001)
Breast	0.0000	0.0001
	(0.0001)	(0.0002)
Genital system	-0.0001	-0.0003
	(0.0001)	(0.0004)
Urinary system	-0.0001**	0.0001
	(0.0000)	(0.0001)
Eye and orbit	0.0000	0.0000
	(0.0001)	(0.0000)
Brain and other nervous system	-0.0001	0.0000
	(0.0000)	(0.0001
Endocrine system	0.0000	0.0000
	(0.0000)	(0.0000
Lymphomas	0.0000	0.0002**
	(0.0000)	(0.0001)
Multiple Myeloma	-0.0002**	0.0000
	(0.0001)	(0.0000)
Leukimia	-0.0001	0.0000
	(0.0001)	(0.0000)
Ill-defined and unspecified sites	-0.0001**	0.0000
	(0.0001)	(0.0001)

Notes: standard errors in parenthesis.
(1)Data that has been aggregated by cancer site, diagnosis year, gender, cohort and state-of-birth. N=336,509. Regressions include age, age squared, diagnosis year, state-of-birth dummies, cohort dummies, cancer site dummies and census year dummies.
(2) Data that has been aggregated by cancer site, diagnosis year, gender, cohort and registry of residence. N=20,348. Regressions include age, age squared, diagnosis year, registry dummies, cohort dummies, cancer site dummies and census year dummies.

Table 9: Is the Effect of Education on mortality larger for diseases where more progress has occurred between 1973 and 1998?

Dependent variable: Died within 5 years of diagnosis	Compulsory schooling law	Mean education in gender, cohort and registry
Progress measured by decreases in age-adjusted mortality		
Education	-0.002***	-0.005***
	(0.001)	(0.001)
Education*(-Estimated annual percent change in age-adjusted mortality)	0.001**	-0.001***
	(0.000)	(0.000)
Progress measured by decreases in age-adjusted incidence rates		
Education	-0.001**	0.000
	(0.001)	(0.001)
Education*(-Estimated annual percent in age-adjusted incidence rates)	0.001***	0.005***
	(0.000)	(0.000)
Progress measured by increases in 5 year-survival rates after diagnosis		
Education	-0.001	0.010***
	(0.001)	(0.001)
Education*change in 5-year survival rate conditional on diagnosis	-0.013***	-0.117***
	(0.005)	(0.005)
Progress measured by the number of drugs available (Match by 3 digit icd9 code)		
Education	-0.002*	0.004***
	(0.001)	(0.001)
Education*Number of drugs	-0.000	-0.000***
	(0.000)	(0.000)
Education	-0.003**	-0.001
	(0.001)	(0.001)
Education*Number of drugs approved 1973-1993	0.000	-0.001***
	(0.000)	(0.000)

Notes: Standard errors in parentheses. Regressions include diagnosis year, age at diagnosis, age at diagnosis squared, 47 state of birth dummies, 24 cohort dummies, 80 cancer site dummies, 2 decade dummies, 8 registry dummies and 4 stage of cancer at diagnosis dummies. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 10: Is the Effect of education on mortality larger for diseases where more progress has occurred between 1973 and 1998?
Flexible specification

Dependent variable: Effect of education on the probability of dying within 5 years of diagnosis (education*cancer site dummies)	Compulsory school	Mean education
Progress measured by decreases in age-adjusted mortality		
-Estimated annual percent change in age-adjusted mortality	0.00062	-0.00076
	(0.00050)	(0.00129)
Progress measured by decreases in incidence rates		
-Estimated annual percent change in age adjusted incidence rates	0.00116**	0.00429***
	(0.00045)	(0.00109)
Progress measured by increases in the 5 year-survival rates after diagnosis		
change in 5-year survival rate conditional on diagnosis	-0.01302*	-0.11264***
	(0.00781)	(0.01575)
Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)		
number of drugs	-0.00004	-0.00040***
	(0.00005)	(0.00014)
number of drugs approved 1973-1993	0.00003	-0.00057
	(0.00021)	(0.00054)

Notes: Standard errors in parentheses. N=81. Each coefficient comes from a separate regression, where the effect of education for each cancer is regressed on a constant and on the relevant measure of innovation, using the variance of the effect of education as weights. The effect of education is the coefficient on education in a regression of the probability of dying, which includes age at diagnosis, age at diagnosis squared, 47 state of birth dummies, 24 cohort dummies, 2 decade dummies, 8 registry dummies, 80 cancer site dummies and 4 stage of cancer at diagnosis dummies. We obtained 78 different coefficients (and their standard errors) by running a regression where education is interacted with a dummy fir each cancer site. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed.
* significant at 10%; ** significant at 5%

Table 11: Is the Effect of Education on mortality larger for diseases where more progress has occurred between 1973 and 1998?
Fully flexible specification

Dependent variable: Effect of education on the probability of dying within 5 years of diagnosis (separate regression by cancer site)	Compulsory school	Mean education
Progress measured by decreases in age-adjusted mortality		
-Estimated annual percent change in age-adjusted mortality	-0.00003 (0.00001)	-0.00073 (0.00056)
Progress measured by decreases in incidence rates		
-Estimated annual percent change in age adjusted incidence rates	-0.00002 (0.00001)	0.00017 (0.00050)
Progress measured by increases in the 5 year-survival rates after diagnosis		
change in 5-year survival rate conditional on diagnosis	-0.0007 (0.0077)	-0.01780** (0.00818)
Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)		
number of drugs	0.00001 (0.00005)	-0.00001 (0.00006)
number of drugs approved 1973-1993	0.0001 (0.0002)	-0.00006 (0.00023)

Notes: Standard errors in parentheses. N=78. Each coefficient comes from a separate regression, where the effect of education for each cancer is regressed on a constant and on the relevant measure of innovation, using the variance of the effect of education as weights. The effect of education is the coefficient on education in a regression of the probability of dying, which includes age at diagnosis, age at diagnosis squared, 47 state of birth dummies, 24 cohort dummies, 2 decade dummies, 8 registry dummies and 4 stage of cancer at diagnosis dummies. We obtained 78 different coefficients (and their standard errors) by running a regression for each cancer site. There are 3 cancers for which the regressions could not be estimated because of small sample sizes. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed.
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 12: Is the Effect of Education on incidence larger for diseases where more progress has occurred between 1973 and 1998?
Flexible specification

Dependent variable: Effect of education on the annual incidence rate (education*cancer site dummies)	Compulsory school	Mean education
Progress measured by decreases in age-adjusted mortality		
-Estimated annual percent change in age-adjusted mortality	0.00001 (0.00001)	-0.00000 (0.00001)
Progress measured by decreases in incidence rates		
-Estimated annual percent change in age adjusted incidence rates	0.00003*** (0.00001)	0.00002* (0.00001)
Progress measured by increases in the 5 year-survival rates after diagnosis		
change in 5-year survival rate conditional on diagnosis	-0.00078*** (0.00015)	-0.00087*** (0.00019)
Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)		
number of drugs	-0.00000*** (0.00000)	-0.00000 (0.00000)
number of drugs approved 1973-1993	-0.00002*** (0.00000)	-0.00001* (0.00001)

Notes: Standard errors in parentheses. N=81. Each coefficient comes from a separate regression, where the effect of education for each cancer is regressed on a constant and on the relevant measure of innovation, using the variance of the effect of education as weights. The effect of education is the coefficient on education in a regression of the probability of dying, which includes age at diagnosis, age at diagnosis squared, 47 state of birth dummies, 24 cohort dummies, 2 decade dummies, 8 registry dummies, 80 cancer site dummies and 4 stage of cancer at diagnosis dummies. We obtained 78 different coefficients (and their standard errors) by running a regression where education is interacted with a dummy fir each cancer site. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed.
* significant at 10%; ** significant at 5%

Table 13: Is the Effect of Education on stage of cancer at diagnosis larger for diseases where more progress has occurred between 1973 and 1998?
Flexible specification

Dependent variable: Effect of education on the stage of cancer at diagnosis (education*cancer site dummies)	Compulsory school	Mean education
Progress measured by decreases in age-adjusted mortality		
-Estimated annual percent change in age-adjusted mortality	0.00018 (0.00049)	0.00260** (0.00116)
Progress measured by decreases in incidence rates		
-Estimated annual percent change in age adjusted incidence rates	0.00022 (0.00046)	0.00174 (0.00109)
Progress measured by increases in the 5 year-survival rates after diagnosis		
change in 5-year survival rate conditional on diagnosis	0.00251 (0.00777)	-0.02786 (0.01841)
Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)		
number of drugs	-0.00016*** (0.00005)	0.00008 (0.00013)
number of drugs approved 1973-1993	-0.00041** (0.00020)	-0.00089* (0.00049)

Notes: Standard errors in parentheses. N=81. Each coefficient comes from a separate regression, where the effect of education for each cancer is regressed on a constant and on the relevant measure of innovation, using the variance of the effect of education as weights. The effect of education is the coefficient on education in a regression of the probability of dying, which includes age at diagnosis, age at diagnosis squared, 47 state of birth dummies, 24 cohort dummies, 2 decade dummies, 8 registry dummies, 80 cancer site dummies and 4 stage of cancer at diagnosis dummies. We obtained 78 different coefficients (and their standard errors) by running a regression where education is interacted with a dummy fir each cancer site. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed.
* significant at 10%; ** significant at 5%

Table 14: Is the education-mortality gradient only due to access? Flexible specification

Dependent variable: Effect of education on the probability of dying within 5 years of diagnosis (education*cancer site dummies)	Mean education No income	Mean education Control for total family income	Effect for those 65 and above	Effect for those below 65
Progress measured by decreases in age-adjusted mortality				
-Estimated annual percent change in age-adjusted mortality	-0.00076	-0.00079	0.00062	-0.00110
	(0.00129)	(0.00128)	(0.00116)	(0.00095)
Progress measured by decreases in incidence rates				
-Estimated annual percent change in age adjusted incidence rates	0.00429***	0.00422***	0.00490***	0.00103
	(0.00109)	(0.00109)	(0.00089)	(0.00099)
Progress measured by increases in the 5 year-survival rates after diagnosis				
change in 5-year survival rate conditional on diagnosis	-0.11264***	-0.11160***	-0.11919***	-0.05217***
	(0.01575)	(0.01580)	(0.01124)	(0.01685)
Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)				
number of drugs	-0.00040***	-0.00039***	-0.00053***	-0.00011
	(0.00014)	(0.00014)	(0.00011)	(0.00011)
number of drugs approved 1973-1993	-0.00057	-0.00055	-0.00138***	0.00017
	(0.00054)	(0.00054)	(0.00045)	(0.00042)

Notes: Standard errors in parentheses. N=81. Each coefficient comes from a separate regression, where the effect of education for each cancer is regressed on a constant and on the relevant measure of innovation, using the variance of the effect of education as weights. The effect of education is the coefficient on education in a regression of the probability of dying, which includes age at diagnosis, age at diagnosis squared, 47 state of birth dummies, 24 cohort dummies, 2 decade dummies, 8 registry dummies, 80 cancer site dummies and 4 stage of cancer at diagnosis dummies. We obtained 78 different coefficients (and their standard errors) by running a regression where education is interacted with a dummy fir each cancer site. Sample consists of whites born in the 48 states between 1901 and 1925 for whom education could be imputed.

* significant at 10%; ** significant at 5%

Appendix: Progress measures for 81 cancer sites

Code	Cancer Site Name	# of white persons born in the 48 states between 1901 and 1925 in our sample	Estimated Annual Percent in the age-adjusted mortality rate	Change in the 5-year survival rate conditional on diagnosis	Estimated Annual Percent in the age-adjusted incidence rate	Number of drugs (active ingredients)	Number of drugs approved 1973-1993 (active ingredients)
20010	Lip	2976	-5.5	0.036	-3.2	12	7
20020	Tongue	4166	-1.7	0.164	0.5	12	7
20030	Salivary gland	1467	-1.5	-0.009	0.6	12	7
20040	Floor of mouth	2579	-4.5	0.034	-2	12	7
20050	Gum & other mouth	3631	-1.4	0.055	-0.4	12	7
20060	Nasopharynx	699	-1.5	0.225	-0.9	12	7
20070	Tonsil	2031	-2.6	0.208	0.2	12	7
20080	Oropharynx	646	1.3	0.109	-0.3	12	7
20090	Hypopharynx	2345	-2.5	0.090	-1	12	7
20100	Other buccal cavity and pharynx	816	-0.9	0.091	-0.1	12	7
21010	Esophagus	7415	1.2	0.079	1.2	0	0
21020	Stomach	14866	-2.7	0.046	-2	1	1
21030	Small intestine	1998	0.2	0.130	2.5	0	0
21041	Cecum	16445	-0.9	0.098	0.1	8	3
21042	Appendix	394	-0.9	-0.235	1.3	8	3
21043	Ascending colon	9755	-0.9	0.123	0.5	8	3
21044	Hepatic flexure	3197	-0.9	0.111	2.1	8	3
21045	Transverse colon	7421	-0.9	0.140	-1.3	8	3
21046	Splenic flexure	2825	-0.9	0.118	0.1	8	3
21047	Descending colon	5650	-0.9	0.147	-2	8	3
21048	Sigmoid colon	26684	-0.9	0.137	-1.1	8	3
21049	Large intestine, NOS	3418	-0.9	-0.001	-0.6	8	3
21051	Rectosigmoid junction	10896	-2.9	0.143	-1.1	8	3
21052	Rectum	20641	-2.9	0.145	-0.9	8	3
21060	Anus, anal canal & anorectum	1597	4.3	0.044	2.1	0	0
21071	Liver	3995	0.8	-0.038	2.4	0	0
21072	Intrahepatic bile duct	602	8.5	-0.078	9	0	0
21080	Gallbladder	2687	-3	0.025	-2.5	0	0
21090	Other biliary	2569	-2.2	0.054	-0.7	6	4
21100	Pancreas	21280	-0.1	-0.013	-0.4	12	8
21110	Retroperitoneum	618	-4.4	0.201	-0.6	0	0
21120	Peritoneum, omentum & mesentery	404	0.4	0.172	6.7	0	0
21130	Other digestive organs	587	-2.8	0.022	-0.2	0	0
22010	Nasal cavity, middle ear & accessory sinuses	1134	-2.6	0.060	-0.2	0	0
22020	Larynx	9818	-0.7	0.009	-1	4	1
22030	Lung and bronchus	127003	1.6	0.031	1	23	10
22050	Pleura	1718	0	-0.029	2.4	0	0
22060	Trachea, mediastinum & other respiratory organs	360	-4.5	0.073	-1.3	0	0
23000	Bones & joints	675	-3.2	0.173	0.6	11	6
24000	Soft tissue (including heart)	2472	1.8	0.071	0.9	20	5

Appendix (continued): Progress measures for 81 cancer sites

Code	Cancer Site Name	# of white persons born in the 48 states between 1901 and 1925 in our sample	Estimated Annual Percent in the age-adjusted mortality rate	Change in the 5-year survival rate conditional on diagnosis	Estimated Annual Percent in the age-adjusted incidence rate	Number of drugs (active ingredients)	Number of drugs approved 1973-1993 (active ingredients)
25010	Melanomas-skin	11289	1.4	0.134	3.6	22	8
25020	Other non-epithelial skin	1049	-0.3	-0.356	4.9	1	1
26000	Breast	87729	-0.4	0.140	1.3	48	10
27010	Cervix	7630	-2.9	0.059	-2	12	5
27020	Corpus	26216	-0.4	-0.055	-1.2	1	0
27030	Uterus, NOS	361	-2.3	-0.087	-2	8	3
27040	Ovary	13758	-0.4	0.154	0.5	25	7
27050	Vagina	834	-1.4	0.099	-1.2	0	0
27060	Vulva	2281	-0.8	0.096	0.9	1	1
27070	Other female genital organs	621	-0.5	0.173	-0.2	0	0
28010	Prostate	87592	0.5	0.298	3.2	34	10
28020	Testis	358	-4.7	0.201	2	14	7
28030	Penis	826	-1.9	0.106	-1.4	2	1
28040	Other male genital organs	194	-3	0.023	0.8	0	0
29010	Bladder	35240	-1	0.098	0.5	16	4
29020	Kidney and Renal pelvis	14873	0.7	0.102	1.8	15	4
29030	Ureter	1626	-0.3	0.090	-1	0	0
29040	Other urinary organs	775	-1.2	0.128	-0.6	0	0
30000	Eye & orbut	1233	-2.6	0.046	-0.6	1	0
31010	Brain	8712	1.5	0.092	0.8	8	3
31040	Other nervous system	296	-9.1	0.169	0.7	8	3
32010	Thyroid	2979	-1.3	0.036	2	7	2
32020	Other endocrine (include. Thymus)	537	0	0.160	1.3	13	6
33011	Hodgkin's Disease-Nodal	1899	-4.6	0.168	-0.2	27	6
33012	Extranodal	52	-4.6	0.160	1.4	2	0
33041	Non- Hodgkin's Lymphomas--Nodal	17122	1.9	0.078	1.9	38	9
33042	Extranodal	5089	1.9	0.054	4.7	40	9
34000	Multiple myeloma	9017	1.3	0.100	0.7	20	3
35011	Acute lymphocytic leukimia	490	-1.6	0.248	1.2	22	2
35012	Chronic lymphocytic	7328	0.6	0.083	-0.5	20	2
35013	Other lymphocytic	267	-5.8	0.070	-1.8	14	1
35021	Acute granulocytic	4509	0.2	0.077	0.5	15	4
35022	Chronic granulocytic	2507	-0.7	0.166	-0.2	15	2
35023	Other granulocytic	536	-4.5	0.159	-7.3	4	1
35031	Acute monocytic Leukimia	339	-5.3	0.127	0.5	3	0
35032	Chronic monocytic leukimia	38	-2.9	-0.118	-4	1	0
35033	Other monocytic leukimia	50	-8.9	0.119		1	0
35041	Other acute leukimia	1080	0.5	-0.058	-0.2	3	1
35042	Other chronic	65	-0.6	-0.141	-2	2	0
35043	Aleukemic, subacute, and NOS	1352	1.2	0.162	0.5	12	2
37000	Ill defined and unspecified sites	22221	0.8	0.050	-0.7	22	9