NBER WORKING PAPER SERIES

HEALTH INEQUALITY, EDUCATION AND MEDICAL INNOVATION

Sherry Glied
Adriana Lleras-Muney

Working Paper 9738
http://www.nber.org/papers/w9738

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2003

We would like to thank Angus Deaton, Joshua Graff Zivin, Bo Honoré , Darren Lubotsky, Christina Paxson, Jonathan Skinner , the seminar participants at the Princeton OPR seminar and the participants of the NBER Health Care meetings for useful comments. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

Health Inequality, Education and Medical Innovation
Sherry Glied and Adriana Lleras-Muney
NBER Working Paper No. 9738
May 2003
JEL No. I12, I20

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 are better able than to take advantage of technological advances in medicine than are the less educated. We test this hypothesis by relating education gradients in mortality with measures medical innovation. We focus on overall mortality and cancer mortality, examining both the incidence of cancer and survival conditional on disease incidence. We find evidence supporting the hypothesis that education gradients are steeper for diseases with more innovation.

Sherry Glied
Department of Health Policy and Management
Mailman School of Public Health
600 West 168th Street, Room 610
New York, NY 10032
and NBER
sag1@columbia.edu

Adriana Lleras-Muney
Department of Economics and Woodrow Wilson School
Princeton University
320 Wallace Hall
Princeton, NJ 08544
and NBER
alleras@princeton.edu

Socioeconomic disparities in health have increased over the past century in the U.S., the U.K., and continental Europe (Feldman et al., 1989, Pappas et al., 1993, Prston and Elo, 1994, Black Report, 1980, Kunst et al., 2001, Shkolnikov et al., 1998). In the U.S., between 1960 and 1986, the age-adjusted mortality rate for white men who had attended college declined from 5.7 to 2.8 per 1000, while the rate declined only from 9 to 7.6 for those who had not graduated high school (Pappas et al., 1993). Moreover, the principal causes of death and disability generating socioeconomic differentials today are quite different from those a century ago (McKeown, 1976; Cutler and Meara, 2002). Over this period, 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. Today, gradients are largest for chronic diseases.

Many factors can generate socioeconomic differences in health status at a point in time. What is more difficult to explain is that these differentials have increased and have shifted among diseases during the past century (Carroll, Davey, Smith, and Bennett, 1996). To our knowledge there is no existing research that has systematically investigated the question of why gradients move among diseases and widen over time.

In this paper we propose and test one explanation for this pattern: the gradient moves among diseases because more educated people are able to exploit technological advances in medicine more rapidly. For this reason, the gradient increases where and when technological change occurs. The most educated make the best initial use of new information about different aspects of health. Over time, this information diffuses along the education gradient. Thus, all else equal, if technological change ceased, we would expect the gradient to flatten as well.

Our 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)." A substantial literature examines this pattern in the labor market (see, for example, Bartel and Sicherman, 1999; Allen, 2001) and in the agricultural sector (Wozniak, 1984). Our hypothesis is also closely related to the sociological conjecture that socioeconomic status is a "fundamental social cause" of gradients in health (Link and Phelan 1995, Link, et al., 1998). 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. In this view, higher socioeconomic status enables people to better exploit new information and resources.

A recurring difficulty in the literature associating gradients in wages to technological progress is the difficulty of measuring progress (Griliches, 1994). This problem also arises in our context and we follow the labor literature in addressing it by considering several different measures of technological progress. We examine the effects of these measures in two different datasets that have distinctive strengths and weaknesses for this analysis. Although our data are limited in several dimensions, we find evidence that is generally (though not uniformly) supportive of our hypothesis: gradients appear to be largest for diseases where there has been more progress.

This paper is organized as follows. We first present a simple formalization of our proposed explanation of the relationship between education, health, and the rate of innovation (Section II). We then present a case study of the HIV epidemic to illustrate how such gradients may evolve.

In Section III, we describe our measures of progress and our empirical approach. In section IV we look at gradients in 5-year mortality across all diseases using data from the National Health Interview Survey (NHIS). We then relate our estimated education gradients to disease categories with divergent rates of progress.

In Section IV, we focus on gradients in cancer incidence and 5-year survival (conditional on stage of diagnosis) using data from the Surveillance, Epidemiology and End Results (SEER) database. Among diseases, cancer is second only to cardiovascular diseases as a contributor to educational gradients in mortality. Cancer provides a good area to study because improvements (albeit small) have occurred in the incidence, diagnosis, and treatment of several cancers, and these improvements vary substantially among cancer subtypes.[1] Additionally, the SEER data links cancer diagnosis to cancer outcomes in a consistent way across subtypes, whereas for other diseases, surveillance data on risk factors cannot be as readily linked to outcomes.[2]

1 For example, there have been substantial improvements in survival after diagnosis of early stage colon cancer (DHHS, 2001). Progress has been made in diagnosing breast and prostate cancer.
2 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 for each of these diseases with progress.

In Section V, we examine how changes in progress affect changes in the education gradient. In Section VI, we evaluate different causal explanations of our results. In Section VII, we relate gradients to technological progress prospectively. Section VIII concludes.

II. – Relating Outcomes to Progress

Many recent studies of health status differentials focus on gradients associated with education. These studies generally find that education is closely correlated with health status (even controlling for income), and suggest that the relationship is causal.[3] Several explanations have been proposed. 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. The extent to which education affects health may also depend on the available health care technology, because better-educated people might be better able to use certain health care technologies (Goldman and Lackdawalla, 2002). Rather than understanding the specific mechanism than generate the education gradient for particular disease at a given point in time, this paper's objective is to understand what makes education gradients increase over time or shift from one disease to another. To do this, we focus on the role of education in the diffusion of technological progress.

A formalization of the relationship between health, education and the rate of innovation

The following formalization 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:

3 For example see Lleras-Muney (2002) shows that increases in education induced by compulsory schooling laws lead causally to improvements in health status.

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

The frontier 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 and or treatments. This feature can be generated from maximization principles simply by assuming differential costs of technology adoption. Note that this model assumes that individuals have chosen education in previous periods and that technological changes are unanticipated. 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:

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

Since w'(e) is negative, the model 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 can test this prediction 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. Note that we think of technology here in very broad terms: it includes all innovations that affect the manner and the rate in which we can transform inputs into health. So in this view, new knowledge is considered innovation.

This simple model predicts that in the absence of technological change (λ = 0), there should be no difference between the educated and the uneducated. We do not believe that technology adoption is the only reason why education and health might be related. For example, even in the absence of progress we can expect education gradients if the more educated are better at utilizing complex technologies (e.g. Rosenzweig, 1995 Goldman and Smith, 2001). A more elaborate model could incorporate these other mechanisms. As the literature review in the introduction showed, there is a substantial amount of work that has investigated these. We will therefore focus only on whether higher rates of technological progress are associated with larger gradients in our empirical work.[4]

A Case Study

HIV disease provides an interesting case study of the relationship between education gradients and medical progress because the disease is new. At the beginning of the HIV epidemic, in the early 1980s, before information about transmission or treatment was available, cases were concentrated among gay men, a group with substantially higher than average educational attainment: 67% have at least a college degree (Bozzette, 1998).

Since the early 1980s, however, there have been markedly different trends in the HIV/AIDS epidemic for groups with different levels of education. Behavioral changes among gay men began immediately after the method of transmission of the human immunodeficiency virus was identified in 1982. In the largest study of transmissions in that period, new infections in this group fell from 20.8% of a susceptible cohort in 1982 to 2.1% of that cohort in 1983 (Centers for Disease Control, 1987). By the late 1980s, new cases among gay men were well below

4 Note that there is no sense in which there exist general equilibrium effects in the health model: if everyone obtains more schooling everyone's health improves. 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. In this respect our theory differs from the "fundamental causes" theory in that it is not a zero sum game.

projected rates, while rates among IV drug users were at or above projected levels, suggesting that there had been little behavioral change in this group (Bloom and Glied, 1992). As early as 1983, new AIDS cases began to be concentrated in low-income areas (Fordyce, 1998).

The effects of new information on the gradient in HIV incidence were later magnified by the effects of new treatment technologies on HIV mortality. We use data from the Centers for Disease Control WONDER system to map annual death rates among the population living with AIDS by exposure category (Figure 1). There were two major treatment advances in HIV care over this period. After a short period of clinical trials, the FDA approved the first effective AIDS drug, AZT, in 1987 (Brown, 1987). A second, more effective, group of drugs, the highly active antiretroviral therapies (HAART) based on protease inhibitors, was introduced beginning in the late 1990s. The FDA approved the first HAART-related protease inhibitor, invirase, in December 1995 and the development of combination therapies that made use of these drugs followed over the subsequent 18 months (McGinley and Womack, 1995; Goldman and Lakdawalla, 2001). As Figure 1 shows, the introduction of AZT and of HAART drugs was associated with a divergence in the death rate between gay men and other transmission groups.[5] This pattern suggests that the more educated group was quicker in taking advantage of the new treatment technology.

Case studies of specific populations support this finding (see Appendix Table 1). Crystal, Sambamoorthi, and Merzel report that there were significant socioeconomic differences (measured by race and exposure group) in receipt of AZT in a cohort studied in 1987-1988, but that these differences had largely disappeared by 1989-1990. Two studies examine changing socioeconomic status disparities in HAART between 1996 and 1998 (Cunningham et al, 2001; Sambamoorthi et al, 2001). Cunningham et al. find that the gap between the percentage of college educated HIV-infected people who had ever used HAART compared to the percentage among those with less than a high school degree shrank form 27 points in 1996 (49:22) to 14 points in 1998 (79:65). Sambamoorthi et al, 2001 find a similar pattern in receipt of HAART over time by race and exposure group.

5 Note that these death rates are conditional on incidence of AIDS. Our data do not capture the extent to which people with HIV began taking these drugs before developing AIDS symptoms and never progressed to AIDS.

Today, lower educational attainment is highly correlated with mortality among AIDS patients (Schechter, 1994). More highly educated patients are more likely to adhere to therapy (Goldman and Smith, 2002), have greater access to antiretroviral therapy and protease inhibitors (Sorvillo, 1999), are more likely to participate in clinical trials (Seltzer, 1989), and have far more knowledge about AIDS (Sorenson, 1999). The HIV case study provides an interesting example of how education may interact with new treatment technologies. Several other case studies similarly document more rapid diffusion of new health innovations among more highly educated people relative to less educated people (see, for example, Link et al. (1998)). In the next sections, we attempt to find more systematic evidence for this pattern.

III. –Progress measures and empirical approach

Measures of progress

There is no consensus about how to measure either progress or the relative importance of progress (Allen, 2001; Bartel and Sicherman, 1999). Instead, we compute multiple measures of innovation for each of the 55 diseases in the NHIS-MCD and the 81 cancer sites in the SEER data. Table 1 describes these measures. Not all progress measures are available in both datasets.

Measurement of progress by disease is particularly problematic because progress in the prevention or treatment of one disease can leave a larger population susceptible to another disease. Thus, the most straightforward measures of technological progress are those that describe innovation in survival conditional on diagnosis of a disease.[6]

1. Number of drugs

The SEER data contain information on survival conditional on diagnosis. We can link these data to information on the number of drugs approved by the FDA to treat a particular cancer (SEER), a direct measure of the rate of pharmaceutical innovation for each particular cancer site.[7] We cannot assign drugs to disease

6 Note that progress in the treatment or prevention of other diseases may still affect outcomes conditional on diagnosis if the newly susceptible population is different (for example, more or less fragile) from the original population.

7 As in Lleras-Muney and Lichtenberg (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

categories in the NHIS because the match between drugs and causes of death is highly imperfect, since drugs are used for conditions that can lead to death from multiple causes. For example, drugs to control diabetes can reduce death rates from diabetes, heart disease, stroke, kidney failure, and other conditions.

2. Change in 5-year survival rates

Using the SEER data, we compute the change in the 5-year survival rate conditional on diagnosis. Like the drug measure, this measure is related to innovation that affects survival conditional on diagnosis. Unlike drugs, which only capture a specific type of innovation, this measure is more comprehensive: progress in surgical procedures, radiation and other aspects will be reflected in the survival rate. However note that this measure is also affected by innovations in diagnostic technology that lead to cancers being diagnosed earlier.[8] Over this period, survival conditional on diagnosis has increased.

3. Change in age-adjusted mortality

A broader measure of progress, which can be confounded by changes in disease incidence, is the change in the age-adjusted mortality rate. The National Cancer Institute provides a measure of the trend in age-adjusted mortality from each type of cancer: the estimated annual percent change (EAPC), which is the coefficient from a log-linear regression of mortality rates on calendar year. The EAPC is positive if age-adjusted mortality increased and negative if age-adjusted mortality decreased; therefore a *negative* value for EAPC constitutes progress. Across all 81-cancer sites there has been progress in cancer mortality over this period.

For the NHIS data we calculate the same measure (EAPC) of age-adjusted mortality using data from the compressed mortality files provided by the CDC. The compressed mortality files provide us with age-adjusted mortality rates for whites for each year from 1986 to 1995 that we use to calculate the EAPC using the same method that the National Cancer Institute uses.

4. Changes in age-adjusted incidence

---

for example). For each cancer site we compute the number of drugs that exist in the market as of 1999. 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.

8 We examine this directly by evaluating changes in the gradient in stage of diagnosis, discussed in footnote xx below, and by controlling for stage of diagnosis in the regression analyses.

A final measure of technological change is innovation in disease prevention. New knowledge associated with disease prevention should allow people to avoid getting a disease in the first place. Unfortunately, we do not have any direct measures of information about disease risk factors.

The efficiency with which epidemiologists are able to identify risk factors for disease increases as the proportion of all susceptible people who develop the disease approaches 0.5 (Rothman and Greenland, 1998). In our context, the prevalence of every disease or disease subtype in the population is well below 0.5. This suggests that the amount of available information about risk factors for a disease or disease subtype will be increasing in the incidence of that disease.[9] We conjecture that as disease incidence increases, information (and gradients) also increase. This pattern should hold even if increases in disease incidence are simply a consequence of improvements in the prevention and treatment of other diseases. We examine this interpretation of changes in incidence in Section VII below.

The National Cancer Institute provides a measure of the trend in age-adjusted incidence: the estimated annual percentage change in age-adjusted incidence (EAPCI). In light of our argument above, positive EAPCI values constitute proxies for progress in the identification of disease risk factors. Overall, incidence rates for most cancer subtypes have increased. Note that there are multiple reasons why incidence could increase including advances in diagnostic technology, increases in environmental risk or in progress in the prevention and treatment of other diseases (who leave more people alive and susceptible to cancer). Our conjecture that information about risk factors increases with increasing prevalence should hold regardless of the causes for increasing incidence.[10] Innovation measures by disease are listed in appendices A and B.

While these measures are related to one another, the correlation among them is not very high, suggesting that they all describe distinct components of progress (Table 2). Changes in incidence and mortality are positively and highly correlated. Increases in survival and in mortality are negatively correlated, but survival and incidence are not highly correlated. Both incidence and mortality are positively correlated with new

9 In a preliminary search of Medline, we found that publications on the etiology of diseases are increasing in disease prevalence.

10 Note that as information about risk factors diffuses incidences will fall, and so we would expect incidence to diminish

drugs, which may indicate that pharmaceutical manufacturers target diseases with rising incidence (as we discuss below). Drugs are positively correlated with changes in survival.

Empirical implementation Across Diseases and Disease Subtypes

Our hypothesis suggests that gradients in education should be greatest where medical progress has been greatest. Our empirical strategy to evaluate this hypothesis consists in estimating education gradients (by disease or disease subtype) and relating these gradients to measures of medical progress. Although there exist different measures of the gradient, in this study we always define gradients as the difference (rather than the ratio) in health outcomes between educated and uneducated individuals. We consider three empirical specifications.

In the most flexible, specification, which we employ in both the NHIS and SEER data, we estimate separate regressions for each disease (NHIS) or cancer site (SEER), including a full set of controls,

$$(1)\quad P(died=1)_{ij} = \beta_0 + \beta_{1j} education_{ij} + X_{ij}\gamma + e_{ij}, j = 1...n.$$

where i indexes individuals, and j indexes the n different diseases. We use the coefficient on education from each regression as the dependent variable in the second stage regression that contains as many observations as disease gradients (n),

$$(2)\quad \beta_j = \delta_0 + \delta_1 progress_j + \varepsilon_j, j = 1...n$$

According to our hypothesis, we expect δ_1 to be negative: larger progress translates into more negative gradients. 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 SEER specification, there is little statistical power available to identify results using this specification in these data so for these data, we begin with two less flexible specifications.[11] In the least flexible specification, we estimate a linear probability model of the probability of dying within 5 years after diagnosis (conditional on stage at diagnosis), where education is interacted with progress:

11 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).

$$(3)\quad P(died=1)=\beta_0+\beta_1 education+\beta_2 education*progress+X\gamma+e\,.$$

Our model suggests β_2 should be negative. 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, preferred specification, we free up the functional form by separating the two stages. Initially, we run a single individual level regression, including a full set of controls and interacting education with cancer site dummies and stage of diagnosis dummies:

$$(4)\ P(died_{ij}=1)=\beta_0+\sum_{j=1}^{81}\beta_j education_i*cancer_j+\sum_{j=1}^{81}\theta_j stage_j*cancer_j+X_{ij}\gamma+e_{ij}$$

where i indexes individuals, and j indexes the 81 different cancers. As with the fully flexible form, 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:

$$(5)\quad \beta_j=\delta_0+\delta_1 progress_j+\varepsilon_j,\ j=1...81$$

In this specification, the effects of education and stage are allowed to vary by cancer site, but all other variables are constrained to have the same effect across sites. As in the fully flexible specification, measures of progress are not included when estimating education gradients, eliminating the possible endogeneity of the progress measures with respect to the probability of dying.

IV- Gradients by Disease – Results from the NHIS

Data

We first measure education gradients in mortality using the NHIS-MCD files. The NHIS is an annual cross sectional survey of the U.S. population. All respondents 18 years of age or older in the 1986-1994 surveys were subsequently matched to the Mortality Cause of Death (MCD) files from 1986 to 1995. Following the epidemiological literature, we focus on five-year mortality and, to avoid censoring, use only observations from the 1986-1990 NHIS interviews. The data contain several socio-economic variables including years of education and family income, and the

mortality data contain information on all causes of death. For purposes of comparison with the cancer data described below, we restrict the NHIS sample to whites ages 40 and above. In all, our sample from the NHIS-MCD files contains 164,373 observations. Summary statistics for the NHIS-MCD data are reported in Table 3.

The Effect of Education on Mortality

We first examine the overall effect of education on the probability of surviving 5 years. We estimate the model in (1) above where the X includes Hispanic status, gender, marital status, interview year and single years-of-age dummies.[12] Table 4 documents the gradient in education for all-cause mortality. The effect of education is negative and significant, slightly larger for males than for females but this difference is not statistically significant. At the mean, these coefficients imply that one more year of education reduces 5-year mortality by about 5%. Table 5 shows the results by disease category for broad categories. We find a negative effect of education on mortality from cancer, respiratory system diseases, cardiovascular diseases, digestive system diseases and other diseases. The effect of education on mortality from diabetes or infectious diseases is negative but not statistically significant.

The Effect of Progress on the Gradient

We next relate the education gradients by disease with progress in mortality for that disease. Figure 2 shows the relationship for broad disease categories. The figure shows (excluding digestive and genitourinary diseases), consistent with our hypothesis, that the gradient is largest for diseases where mortality has decreased most.

We next examine the relationship using 55 detailed categories of death.[13] For each cause of death we estimate the effect of education on the probability of dying in the next five years as in tables 4 and 5. We then match coefficients from these regressions to changes in disease-specific mortality. Table 6 presents results from regressions that are weighted by the number of deaths from that disease in our sample (column 1) and by the inverse of the variance of the first stage coefficient (column 2).

12 Cox proportional hazard models yielded similar results but given that the data are large and that we include many variables, these estimations take a very long times to converge. We therefore present linear probability models instead.

13 The NHIS-MCD recodes ICD9 causes of death into 72 categories. We use this classifications We excluded deaths from external causes, deaths from birth complications, deaths from ill-defined causes, and causes of death for which we did not observe any deaths. See appendix B for a detailed list of the causes of death we use.

When weighting by the number of deaths, we find that the gradient in education is larger and statistically significant for diseases where there has been more progress. Medical progress, measured as changes in mortality, explains about 13.6% of the gradient in education (evaluated at the mean for progress, i.e., 0.0000612*1.067/0.0048). When weighting by the standard error of the coefficient, however, the result is negative but much smaller and not statistically significant.

In order to investigate why the different weights strongly affect the results, we plotted education gradients and their standard errors against the number of deaths in that disease (Figure 3). Figure 3 shows that both gradients and standard errors are larger for more common diseases. That the standard errors increases with the number of deaths is not surprising: the variance of a binary random variable is (1-p)p, and since p (the probability of dying, i.e. the number of deaths divided by population) is always less than 0.5, the variance increases with the number of deaths. But note that gradients increase with the number of deaths. Therefore weighting by the standard error leads us to place the most weight on rare causes of death. We conclude that the regressions that are weighted with the number of deaths are more accurate.

Note that this exercise has uncovered an interesting pattern that is consistent with our hypothesis about progress in incidence. Education gradients appear to increase with disease incidence. We discuss this pattern more extensively below.

V. Gradients by Disease Subtype – Results from the SEER Data

Our cancer 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.[14] 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. To avoid censoring in our analysis of 5-year mortality, we limit our sample to those diagnosed with cancer no

14 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.

later than 1993. The SEER data contain a large number of observations (N=2,556,432) so we can perform analyses of death rates within detailed disease category (sites).

We focus on two outcome measures in the cancer data: 5-year mortality rates and cancer incidence rates. We focus most of our attention on gradients in mortality rates (conditional on stage of diagnosis), because these gradients are relatively independent of trends in the incidence of other diseases.

Summary statistics for the final SEER sample used in this paper are in Table 7. 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.[15]

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. Several papers have shown that these laws had an impact on educational attainment.[16] These laws, which implicitly specified the number of years that a child had to attend school, serve to identify the effect of education.[17] The implicit number of compulsory years ranges from 0 to 10 for the cohorts we study.

By including laws in place of education in a model of mortality/health, we are estimating a reduced form equation. The advantage of this method is that we can argue that the effects we measure are causal effects of education. Note however that we are possibly identifying the effect of education only for those affected by these laws, i.e. those at the lower end of the distribution of education. Because compulsory schooling

15 The average age at diagnosis is around 62 in the full SEER data. Our sample is older but not much more.
16 See Acemoglu and Angrist (1999), Angrist and Krueger (1991), Lleras-Muney (2002b), Margo and Finnegan, (1996) and Schmidt (1996).
17 The data on compulsory attendance and child labor laws were collected from multiple sources (See Lleras-Muney, 2002, 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.

laws were most effective in the first half of the 20th century and they only affected whites (see Lleras-Muney 2002a) we restrict our attention to white cohorts born between 1901 and 1925.

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 is available.

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 or it may be correlated with unobservable characteristics, such as rates of time preference (Fuchs, 1982). Finally when using average education, the other coefficients in the regression will most surely be biased since the error term now contains the difference between individual and mean education, which is most likely correlated with other covariates in our model.

Throughout the cancer analyses, we provide results using each of these two proxies. Although there are many reasons as we just discussed why the two measures might provide different answers, we will feel more confident about our results to the extent that both measures provide similar estimates. 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:

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

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

The Effect of Education on Outcomes

In the SEER data, we examine the overall effect of education on the probability of dying 5 years conditional on being diagnosed with cancer. We estimate the model in (1) above where X includes 47 state of birth, 24 cohort dummies, 8 registry dummies, 2 decade dummies, 4 stage of diagnosis dummies, and 80 dummies for cancer sites.[18] Since we include state of birth and cohort dummies, the effect of compulsory schooling laws is identified from variations in the laws within states over time.

These results are reported in Table 8. The effect of education on overall mortality 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. The effect of education on the probability of dying from cancer that we measure here is greater than that suggested by the NHIS all-cause data, but notice that in the NHIS we are estimating unconditional probabilities, whereas everyone in the SEER data has cancer.

Using the compulsory schooling laws, we estimate two-stage IV estimates of the effect of education in Table 8. Using the 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.[19] We estimate the first stage using the 1960, 1970 and 1980 censuses and find that the effect of compulsory schooling on education is 0.079. Using this information we calculate that the TSIV estimate of the effect of education on mortality is somewhere between –0.02 and –0.05. At the means, this coefficient suggests that one more year of schooling reduces the probability of dying post-diagnosis by 3-7%. The TSIV estimates are close (somewhat larger) to those found by Elo and Preston (1996) who report that the effect of one more years of schooling on 5-year (all cause) mortality is also between 0.02 and 0.05. Since our objective in this paper is not to provide accurate estimates of the effect of education on cancer, but

18 Cox proportional hazard models yielded similar results but given that the data are large and that we include many variables, these estimations take a very long times to converge. We therefore present linear probability models instead.

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 on outcomes.[20]

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 9). 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 cancers of the respiratory system, genital system, urinary system and for buccal cavity and pharynx cancers. We do not find statistically significant effects for other cancers. The results are similar for the two measures of education.

We next examine the effects of education on the incidence of cancer (Table 10). 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. Alternatively, we calculate rates by cancer site, census year, gender and registry. We then divided by the population in that group obtained from the 1970, 1980 and 1990 censuses. We generate estimates of the population for each gender/cohort/state-of-birth/site cell using a linear time trend. We then match these incidence rates with our education measures.

Incidence rates provide a measure of the progress in knowledge about cancer risk factors. 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. For example, increases in the age-adjusted incidence of cancer in a group may simply reflect a decline in cardiovascular disease mortality for that group. We hypothesize that 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

19 This method was used by Dee and Evans (1999).
20 The inclusion of state-of-birth and cohort specific variables (such as infant mortality, number of hospitals per mile and number of doctors per capita in state-of-birth at age 14) did not affect the results. Results available upon request.

and for several specific types of cancer. These effects are large (one more year of education decreases the incidence rate by about 40%). By contrast, we find that increases in mean education at the registry level have little impact on incidence.[21]

The difference between the compulsory schooling and mean education results may be an artifact of the aggregation process. In computing incidence 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.

We also find that higher mean education is associated with later age at diagnosis, although we do not find this result using the compulsory schooling measure (not shown). These coefficients are small: the largest effect of mean schooling indicates that a one year increase in mean schooling induces a .1% reduction in the age at diagnosis, the implied effect of education is larger – about 1.3% – since mean schooling is only partially related to education.

This result may at first appear surprising since more educated people are likely to be diagnosed early.[22] However more educated people are likely to be older, not younger, when they get cancer. Given that the more educated are diagnosed at older ages, we conclude that the apparent survival advantage of more educated people is not simply due to earlier diagnosis.

To summarize, our results suggest that education has significant effects on all cause morality, 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. In the SEER data, more educated people are likely to be older when they are diagnosed with cancer, they are more likely to survive for 5 years after diagnosis, and they may be less likely to get cancer at all. Most of these results are similar regardless of which measure of education is used.

The Effect of Progress on the Gradient in Mortality

21 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. Using the coefficient on compulsory schooling, we can calculate that one more year of schooling results in an 11% increase in the probability that the cancer is localized at time of diagnosis.

22 The calculation is not reliable however since, unlike the case of compulsory school, we cannot argue that mean education is really an instrument for education.

As we had done for the NHIS, we next relate the education gradients in morality by disease with measures of progress for that disease. In this case, we use the three specifications described above (see Tables 11-13).

The first panel of Table 11 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 mortality 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 information about risk factors, proxied by age-adjusted incidence. Here, using either measure of education, we find that progress increases the gradient in education. This implies that the cancer survival gradient in education is steepest for those cancers whose incidence is increasing. This result is consistent with our earlier finding in the NHIS that gradients were larger for more prevalent diseases. We comment again on this result below.

The third panel, which uses our preferred measure of progress, increases in survival after diagnosis, shows that the education gradient in survival is steepest for those diseases where survival is improving. 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 been the most progress.

Table 12 shows results for our preferred 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 somewhat lower. Table 13 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, as we had expected based on the large number of dummy variables, the estimates are much smaller in magnitude and are mainly statistically insignificant.

All these results are weighted by the inverse of the variance of the estimated education coefficient. Unlike in the NHIS, results weighting by the number of individuals with the disease give similar results (See Appendix Table 4). Intuitively, this is because in the SEER data the variance of the estimated coefficient of education

is estimated using a different sample for each disease: diseases with larger incidence have more precisely estimated coefficients. Additionally, note that the probability of death is in general larger than 0.5 so that again if we weight by the number of deaths, we place more weight on larger diseases. Consequently in the SEER, any weighting scheme places relatively more weight on cancers that are more common.

We next examine the effect of progress on incidence (Table 14). Using our preferred, flexible specification, we find that again that gradients in incidence become larger (more negative) when there is progress in information about risk factors, proxied by increasing disease incidence.

We also find that gradients in incidence become larger (more negative) when survival rates improve (or when more drugs are introduced). If pharmaceutical companies targeted their efforts toward diseases that particularly affected more educated people, we would expect to see a positive correlation between drugs and the gradient. Instead, we find that highly educated people are less likely to get those cancers where we observe the largest improvements in survival. This result provides important evidence suggesting that technological progress drives the gradient, rather than the gradient driving technological progress.

VI – Changes in Progress and in the Gradient

Our model suggests that progress affects the gradient because more educated people are quicker to take advantage of progress. New technologies later diffuse to less educated people. This theory has implications about the relationship between the gradient and the timing of progress. It suggests that more recent progress should lead to widening in the gradient, while progress years earlier should lead to a narrowing of the gradient. The HIV case study is consistent with this pattern. It also suggests that changes in the gradient should be related to changes in progress over time.

We have limited data to test these secondary hypotheses. Because of the construction of the SEER education measures, they vary only as cohorts age and there is much less variation in predicted education among older cohorts than among younger cohorts, confounding efforts to examine changes in the gradient over time. The NHIS data cover only a very short time span (1986-1994), but we can conduct preliminary analyses of these data.

We estimate two regressions using the NHIS gradients we estimated earlier (Table 15). First, we examine whether more recent or older progress has greater effects on the gradient. We examine how recent progress and older progress affect the gradient measured in 1990. We find that only recent progress leads to a widening in the gradient. Earlier progress appears to have only a very small, negative effect on the size of the gradient. Second, we examine whether the change in the education gradient by disease is related to the change in progress for that disease. We find that gradients widened most for those diseases where progress was greatest. These results are both consistent with our theory.

VII. Incidence, gradients and technology: are they related?

We conjectured that information about risk factors would be most efficiently obtained for diseases with higher prevalence. Thus, we expect that progress in the discovery of disease risk factors is likely to be increasing in disease incidence. Similarly, prior research has suggested that technological progress in treatment is related to the burden of a disease (Lichtenberg and Waldfogel, 2001). Thus, another way to examine our hypothesis is to see whether factors that predict future rates of technological progress also predict future gradients. An observation that future progress and future gradients are both associated with the initial prevalence of a disease would provide indirect evidence of our hypothesis.

We construct this indirect test by examining the relationship between the number of deaths in 1980 (calculated from the Mortality detail files in 1980) and the change in age-adjusted mortality for that disease from 1985 to 1995, or alternatively from 1990 to 1999 (using CDC data). We likewise look at the relationship between the number of deaths in 1980 by disease and the size of the education gradient for that disease in the NHIS from 1986 to 1990.

We present the results using un-weighted regressions, and regressions that weight by the inverse of the variance of the estimate of the rate of progress EAPC (recall that this is a regression of mortality rates on time).[23] Here, the weighted regressions place more emphasis on those diseases where the mortality trend is more precisely measured.

23 We calculated the variance of the EAPC using the Delta method.

Results of these analyses are reported in Table 16. The results suggest that the higher the number for deaths in a given disease in 1980 the greater the percentage decline in age-adjusted mortality from 1985 to 1995. The effect is negative and, when weighting by the inverse of the variance, statistically significant. On the other hand, the number of deaths appears to be unrelated to progress in the later 1990-1998 period. Most interestingly, the education gradient 1986-1990 is always larger for diseases with more deaths in 1980 (results are significant irrespective of weighing scheme). Finally, the number of drugs 1986-1996 is positively correlated with number of deaths in 1980 but the coefficient is not significant.

Overall these results do suggest that education gradients and progress are driven by the same factors, in particular incidence. This suggests that information about diseases is increasing in disease incidence. Innovation in treatment may also be occurring more for diseases that are common or are becoming common.

VIII—Mechanisms by Which Education May Affect Outcomes

Our results are largely (though not uniformly) consistent with our hypothesis, suggesting that education may enable people to make more effective use of technological progress in reducing mortality or in surviving cancer. They do not, however, explain the mechanisms through which this might occur.

The existing literature on disparities in cancer treatment between whites and blacks and among education groups suggests a broad array of mechanisms that might generate the relationship we observe (Shavers and Brown, 2002). It describes differences between groups in receipt of radiation treatment following surgery, staging of cancer, nature of radiation therapy received, receipt of adjuvant chemotherapy, receipt of surgery, and aggressiveness of treatment. Groups also differ in the rate of referral to oncologists (Earle et al., 2002), the rate at which they followed up on screening mammograms (Strzelcyk and Dignan, 2002), and the rate at which they participate in clinical trials (Svensson CK, 1989). There are many ways that education could generate these differences. Education might have direct effects (for example, by making it less difficult for people to understand consent procedures in clinical trials or to follow new health information) or indirect effects (for example, because poorly educated people are less likely to have health insurance or more likely

to live in low income areas where oncologists are unavailable and environmental risk factors are greater).

We do not have strong tests that allow us to distinguish between these hypotheses, but we consider two here. First, we compare the effects of education with and without controls for average family income. The results for both the SEER and the NHIS are reported in the first 2 columns of Table 17. 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 persists even when adding controls for family income.

Next, we compare the effects of education for those diagnosed before and after Medicare eligibility (age 65) in the SEER data. 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 are significant for both sub samples. 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.

Another possible mechanism is that it is not education per se but some other characteristic of people who become educated that drives our results. In this respect, we note that the results in the SEER that show the effects of education measured as compulsory schooling and as mean education 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.

IX. Conclusions and Limitations

This study finds some evidence to support our hypothesis: gradients appear to be larger for diseases where there has been more progress. While we do not find that

all measures of progress are correlated with education gradients, the bulk of the evidence is quite suggestive.

Our results do not explain the mechanisms through which gradients arise, although our findings suggest that the relationship between the gradient and technological progress is not explained away by family income. Mechanisms are likely to vary among diseases and to change over time. For any specific condition at a specific point in time, understanding the mechanisms is critical to reducing the gradients. Over time, however, there is unlikely to be a single mechanism that can account for gradients.

In exploring the relationship between education gradients and rates of innovation, we find that gradients are largest for diseases that afflict many, and that gradients increase for diseases where incidence is increasing. We also find evidence that incidence predicts future innovation. Our interpretation is therefore that innovation occurs for diseases that are common or are becoming common and that this is a reason why education gradients to appear for those diseases.

Our data have some important limitations. In the SEER data, we do not have direct measures of education. Use of proxy measures limits our ability to examine changes in gradients over time. In the NHIS data, our samples are relatively small, particularly for some causes of death. Moreover, the causes of death recorded in the NHIS are quite broad, and progress may be very different for some subtypes of disease than for others. Additionally we do not have a unique measure of progress, but rather a number of proxies. Our results suggest that progress explains 14% - 31% of education gradient, but the precision of this estimate is low given the limited variation in education or diseases across our sample.

Gradients in health outcomes that arise because of technological progress make the distribution of health less equitable. The existence of a gradient suggests that there exists the technological potential for improving the health of the less well off. But gradients that arise due to improvements in the health of the most well off, rather than a diminution in the health of the least well off may be a Pareto efficient outcome of technological progress in medicine.

References

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

Allen, Steven G., "Technology and the Wage Structure." Journal of Labor Economics, Vo. 19, No. 2: 440-483.

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 Nachum Sicherman, "Technological Change and Wages: An Interindustry Analysis," The Journal of Political Economy, April 1999: 285-325.

Bloom, D.E and S. Glied. 1992. "Projecting the Number of AIDS Cases in the United States." International Journal of Forecasting 8: 339-365.

Bozzette, S.A. et al. 1998. "The Care of HIV-infected adults in the United States. HIV Cost and Services Utilization Study Consortium." New England Journal of Medicine 339(26):1897-904.

Brown, J. 1987. "Approval of AZT." Food and Drug Administration, U.S. Department of Health and Human Services. < http://www.aegis.com/news/fda/1987/FD870301.html> (11 February 2003).

Centers for Disease Control (1987). "Human Immunodeficiency Virus Infection in the United States: A review of current knowledge." Morbidity and Mortality Weekly Report, 36, 18 Dec., S-6.

Crystal, Stephen; Sambamoorthi, Usha;Merzel, Cheryl. The diffusion of innovation of AIDS treatment: Zidovudine use in two New Jersey cohorts. Health Services Research; Chicago; Oct 1995: 30(4):593-615.

Cunningham WE, Markson LE, Andersen RM, Crystal SH, Fleishman JA, Golin C, Gifford A, Liu HH, Nakazono TT, Morton SC, Bozzette SA, Shapiro MF, Wenger NS. Prevalence and Predictors of Highly Active Antiretroviral Therapy Use in Persons with HIV Infection in the U.S. Journal of Acquired Immune Deficiency Syndromes, Vol. 25, No. 2, October 2000, pp. 115-123

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

Fordyce, E.J. et al. 1998. "Economic and Geographic Diversity in AIDS Incidence Among HIV Exposure Groups in New York City: 1983-1995." AIDS and Public Policy Journal 13(3): 103-114.

Fuchs, Victor R., ``Time Preference and Health: An Exploratory Study,: in Victor Fuchs, Ed., *Economic Aspects of Health*, Chicago: The University of Chicago Press, 1982

Goldman, D.P and J.P. Smith. 2002. "Can Patient Self-Management Help Explain the SES Health Gradient?" Proceeding of the National Academy of Sciences of the United States of America 99(16): 10929-10934.

Griliches, Zvi. Productivity, R&D, and the Data Constraint," American Economic Review 84 (March 1994): 1-23.

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

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, "Were State Laws on Compulsory Education Effective? An analysis from 1915 to 1939," Journal of Law and Economics, Vol. XLV(2), October 2002a

Lleras-Muney, Adriana "The Relationship between Education and Adult Mortality in the U.S." NBER Working Paper #8986, May 2002b

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

McGinley, L. and A. Womack. 1995. “Technology & Health: An FDA Panel Urges Approval of AIDS Drug.” .” Food and Drug Administration, U.S. Department of Health and Human Services. <http://www.aegis.com/news/wsj/1995/WJ951101.html> (11 February 2003).

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

Rothman, Kenneth J. and Sander Greenland. Modern Epidemiology, Second Edition. Philadelphia: Lippincott-Raven , 1998.

Sambamoorthi, Usha, et al. 2001. "Use of Protease Inhibitors and Non-Nucleoside Reverse Transcriptase Inhibitors Among Medicaid Beneficiaries with AIDS." American Journal of Public Health 91(9): 1474-81.

Schechter, M.T, et al. 1994.“Higher Socioeconomic Status Is Associated With Slower Progression of HIV Infection Independent of Access to Health Care.” Journal of Clinical Epidemiology47(1): 59-67.

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

Seltzer, R. 1989. "Race, Age Education and Knowledge of AIDS." Sociology and Social Research 73(4): 189-193

Sorenson, W., Oar, H.S., and K. Corson. 1999. "The Relationship Between Educational Background and Decline to Death In a Sample of Louisiana Male AIDS Cases." International Journal of STD and AIDS 10(3): 195-198.

Sorvillo, F. et al. 1999. "Use of Protease Inhibitors Among Persons with Aids in Los Angeles County." AIDS Care 11(2): 147-155.

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: Summary Statistics on measures of progress

Variable	Obs	Mean	Std. Dev.	Min	Max
NHIS data					
Estimated Annual Percent Change in age-adjusted mortality [(1)]	55	-1.067	4.149	-9.064	15.043
SEER cancer data					
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

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.

Table 2: 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)
SEER cancer data				
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

Notes: See previous table for definitions and data sources.

TABLE 3: NHIS Summary statistics

	Obs	Mean	Std. Dev.	Min	Max
Died in 5 years=1	164373	0.095	0.293	0	1
Education	164373	12.113	3.239	0	18
Interview Year (1986=0)	164373	2.222	1.312	0	4
Hispanic	164373	0.049	0.216	0	1
Married	164373	0.734	0.442	0	1
Female	164373	0.540	0.498	0	1
Age	164373	58.081	12.482	40	90
Causes of death					
Infectious diseases	164373	0.0014	0.0378	0	1
Cancer	164373	0.0280	0.1649	0	1
Diabetes	164373	0.0021	0.0458	0	1
Cardiovascular Diseases	164373	0.0419	0.2004	0	1
Respiratory System Diseases	164373	0.0077	0.0871	0	1
Digestive System Diseases	164373	0.0018	0.0427	0	1
Urinary System	164373	0.0009	0.0300	0	1
Other Diseases	164373	0.0014	0.0375	0	1

Notes: Data: NHIS-MDC data.
Sample consists of whites ages 40 and above with no missing values.

TABLE 4: Results with the NHIS
The effect of education on the probability of dying in five years
(all causes of death)

Dependent Variable: dead=1 if died within 5 year	All	Males	Females
Education	-0.0048***	-0.0055***	-0.0035***
	(0.0003)	(0.0004)	(0.0004)
Interview Year (1986=0)	-0.0008	-0.0014	-0.0004
	(0.0006)	(0.0009)	(0.0007)
Hispanic	-0.0204**	-0.0207**	-0.0194**
	(0.0030)	(0.0048)	(0.0036)
Married	-0.0137***	-0.0354***	-0.0153**
	(0.0019)	(0.0035)	(0.0022)
Female	-0.0489***		
	(0.0015)		
N	164,710	75,770	88,940

Notes: Data: NHIS-MDC data. Sample consist of whites ages 40 and above with no missing values. Regressions also include single year of age dummies and use person weights provided by NHIS.

TABLE 5: The effect of education by disease type in the NHIS

Dependent Variable: dead=1 if died in 5 year	Education[(1)]
Infectious diseases	-0.000040
	(0.000040)
Cancer	-0.001249***
	(0.000158)
Diabetes	-0.000051
	(0.000050)
Cardiovascular Diseases	-0.002355***
	(0.000183)
Respiratory System Diseases	-0.000628***
	(0.000083)
Digestive System Diseases	-0.000068*
	(0.000034)
Urinary System	-0.000058**
	(0.000027)
Other Diseases	-0.000074**
	(0.000038)

Notes: Data: NHIS-MDC data.
Sample consist of whites ages 40 and above with no missing values. Regressions also include single year of age dummies and use person weights provided by NHIS.

Table 6: Results using the NHIS: Is the Effect of education on mortality larger for diseases where more progress has occurred between 1985 and 1995? Fully Flexible specification

Dependent variable: Effect of education on the probability of dying within 5 years of diagnosis (separate regression by cause of death)	Education	Education
WEIGHT	Inverse of variance of beta	Number of deaths by disease in 1986
Progress measured by decreases in age-adjusted mortality		
(-Estimated annual percent change in age-adjusted mortality)	-1.64e-07 2.26e-07	-0.0000612*** (0.0000216)

Notes: Standard errors in parentheses. N=55. Each coefficient comes from a separate regression, where the effect of education for each cause of death 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 in 5 years after the interview, which includes single age dummies, family income, female dummy, Hispanic dummy and interview year. We obtained 55 different coefficients (and their standard errors) by running a regression for each cause of death. Sample consists of whites ages 40 and above with no missing data.

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

Table 7: 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 localized or in situ when diagnosed	711450	0.38	0.48	0	1
Incidence rate*		0.0028	0.007	0.00006	0.1
<u>Cancer Site (Broad categories)</u>					
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.
*incidence rates are calculated using aggregated data (see text)

Table 8: 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 1970/80/90 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***	0.079***	-0.025*	-0.005***	-0.004***
	(0.001)	(0.015)	(0.014)	(0.001)	(0.001)
Males	-0.002*	0.041*	-0.049	-0.003**	-0.002*
	(0.001)	(0.024)	(0.038)	(0.001)	(0.001)
Females	-0.003***	0.107***	-0.028**	-0.000	0.000
	(0.001)	(0.019)	(0.011)	(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 9: 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)
Leukemia	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 10: The Effect of education on Incidence rates

	Compulsory school [(1)]	Mean education in cohort gender and registry[(2)]
Sample		
All cancers	-0.00008***	0.0000
	(0.00001)	(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 11: 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 increases in age-adjusted incidence rates		
Education	0.001**	0.000
	(0.001)	(0.001)
Education*(Estimated annual percent change 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	-0.00039***
	(0.000)	(0.00003)

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 12: Is the Effect of education on mortality larger for diseases where more progress has occurred between 1973 and 1998? Flexible specification-Including interaction between site and stage

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)	0.00069	-0.00125
	(0.00051)	(0.00088)
Progress measured by increases in incidence rates		
(Estimated annual percent change in age adjusted incidence rates)	-0.00119**	0.00026
	(0.00055)	(0.00099)
Progress measured by increases in the 5 year-survival rates after diagnosis		
change in 5-year survival rate conditional on diagnosis	-0.02706**	-0.0596***
	(0.01257)	(0.0218)
Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)		
number of drugs	-0.00015	-0.00039**
	(0.00010)	(0.00017)

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 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
<u>Progress measured by decreases in age-adjusted mortality</u>		
(-Estimated annual percent change in age-adjusted mortality)	-0.00003 (0.00001)	-0.00073 (0.00056)
<u>Progress measured by increases in incidence rates</u>		
(Estimated annual percent change in age adjusted incidence rates)	0.00002 (0.00001)	-0.00017 (0.00050)
<u>Progress measured by increases in the 5 year-survival rates after diagnosis</u>		
change in 5-year survival rate conditional on diagnosis	-0.0007 (0.0077)	-0.01780** (0.00818)
<u>Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)</u>		
number of drugs	0.00001 (0.00005)	-0.00001 (0.00006)

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 14: 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
<u>Progress measured by decreases in age-adjusted mortality</u>		
Estimated annual percent change in age-adjusted mortality	0.00001 (0.00001)	-0.00000 (0.00001)
<u>Progress measured by increases in incidence rates</u>		
Estimated annual percent change in age adjusted incidence rates	-0.00003*** (0.00001)	-0.00002* (0.00001)
<u>Progress measured by increases in the 5 year-survival rates after diagnosis</u>		
change in 5-year survival rate conditional on diagnosis	-0.00078*** (0.00015)	-0.00087*** (0.00019)
<u>Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)</u>		
number of drugs	-0.00002*** (0.00000)	-0.00000 (0.00000)

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 15: Changes in the 4-Year Mortality Education Gradient over Time (NHIS)

	Change in the Effect of Education on the Probability of Dying 1990 vs. 1986	Effect of Education on the Probability of Dying 1990
Change in the Rate of Technological Progress 1990-1993 vs. 1983-1986	-0.000019**	
	(0.000007)	
Tech Progress 1983-1985		-0.000015
		(0.000014)
Tech Progress 1986-1989		0.000006
		(0.000022)
Tech Progress 1990-1993		-0.000055**
		(0.000027)

Notes: Standard errors in parentheses. N=55. Each coefficient comes from a separate regression, where the effect of education for each cause of death 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 in 5 years after the interview, which includes single age dummies, family income, female dummy, Hispanic dummy and interview year. We obtained 55 different coefficients (and their standard errors) by running a regression for each cause of death. Sample consists of whites ages 40 and above with no missing data.

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

Table 16: Does incidence predict progress and gradients?

Dependent variable:	Estimated Annual Percentage Change in age-adjusted mortality 1985-1995	Estimated Annual Percentage Change in age-adjusted mortality 1985-1995	Estimated Annual Percentage Change in age-adjusted mortality 1990-1998	Estimated Annual Percentage Change in age-adjusted mortality 1990-1998	Gradient in education 1986-1990 (NHIS)	Gradient in education 1986-1990 (NHIS)	Number of drugs approved 1986-1996
WEIGHT (inverse of)	--	Variance of EAPC 1985-1995	--	Variance of EAPC	--	Variance of estimated gradient	--
Number of deaths in 1980	-1.93e-06	-5.83e-06**	9.52e-07	-4.15e-07	-2.41e-09**	-1.70e-09**	4.37e-06
	(9.32e-06)	(1.53e-06)	(9.40e-06)	(3.37e-07)	(2.24e-10)	(2.76e-10)	(0.0000252)

Notes: Standard errors in parentheses. Each coefficient comes from a separate regression, where the effect of education for each cause of death 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 in 5 years after the interview, which includes single age dummies, family income, female dummy, Hispanic dummy and interview year. We obtained 55 different coefficients (and their standard errors) by running a regression for each cause of death. Sample consists of whites ages 40 and above with no missing data.

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

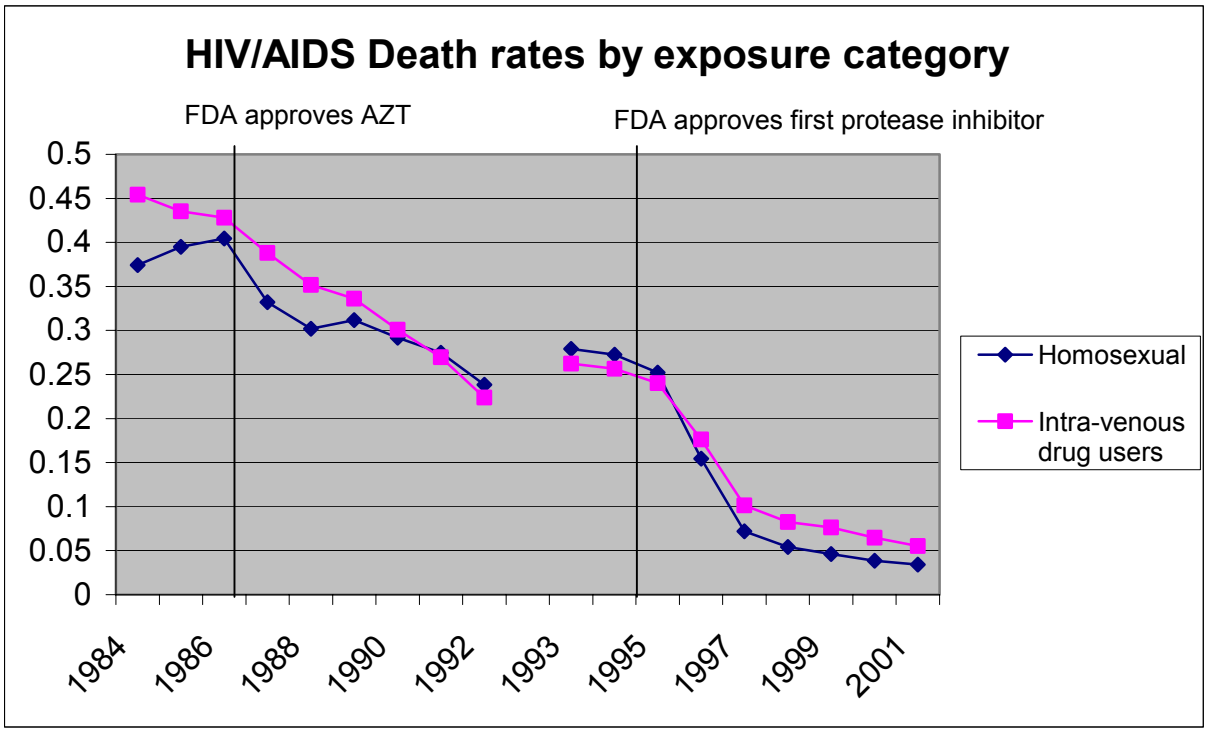
Table 17: Is the education-mortality gradient only due to access?

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
NHIS data				
Progress measured by decreases in age-adjusted mortality				
(-Estimated annual percent change in age-adjusted mortality)	-0.0000612*** (0.0000216)	-0.0000402** (0.0000192)		
number of drugs approved 1973-1993	4.64e-06 (4.93e-06)	5.44e-06 (4.18e-06)		
SEER cancer data				
Progress measured by decreases in age-adjusted mortality				
(-Estimated annual percent change in age- adjusted mortality)	-0.00076 (0.00129)	-0.00079 (0.00128)	0.00062 (0.00116)	-0.00110 (0.00095)
Progress measured by decreases in incidence rates				
(-Estimated annual percent change in age adjusted incidence rates)	0.00429*** (0.00109)	0.00422*** (0.00109)	0.00490*** (0.00089)	0.00103 (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*** (0.01575)	-0.11160*** (0.01580)	-0.11919*** (0.01124)	-0.05217*** (0.01685)
Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)				
number of drugs	-0.00040*** (0.00014)	-0.00039*** (0.00014)	-0.00053*** (0.00011)	-0.00011 (0.00011)

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%

Figure 1: HIV/AIDS Case study



Source: 1984-1992 from the AIDS Microfiche Data on CDC Wonder. We divided the number of deaths per year by the cumulative population (number diagnosed total minus the death from previous years). 1993-2001 is from the CDC HIV/AIDS Surveillance Report Vol. 13(2). We divided the death rate by the total number of people living with HIV/AIDS.

Figure 2: Relating education gradients and changes in mortality rates: NHIS

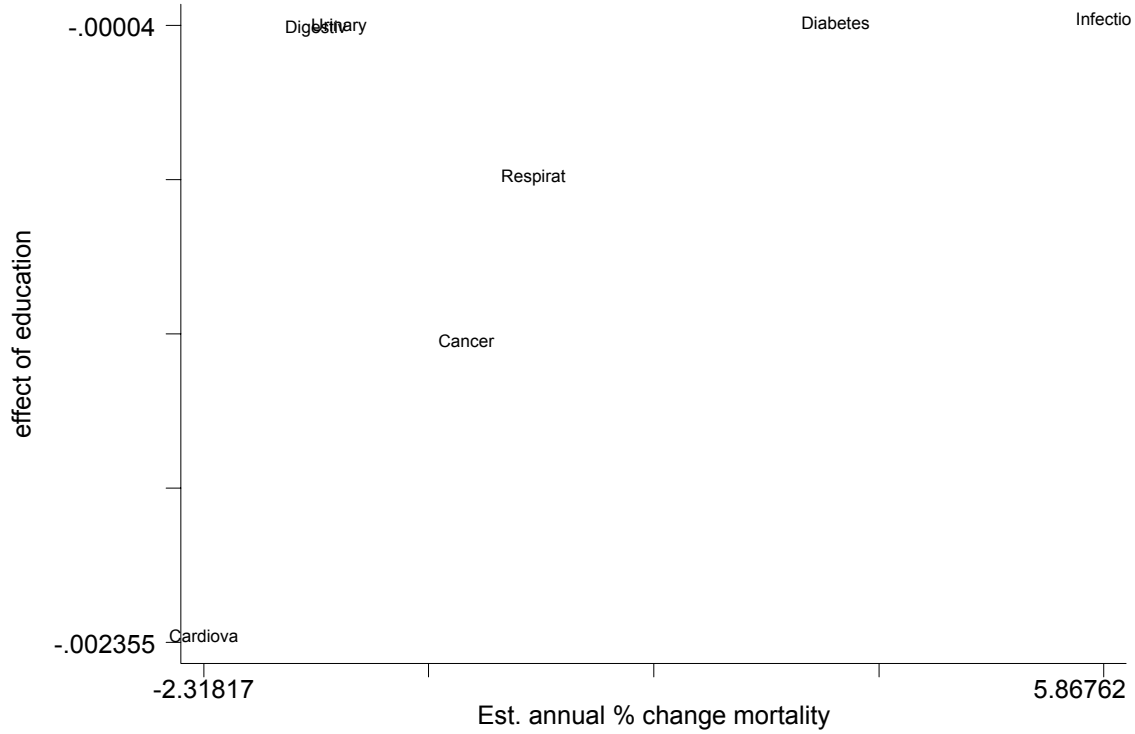
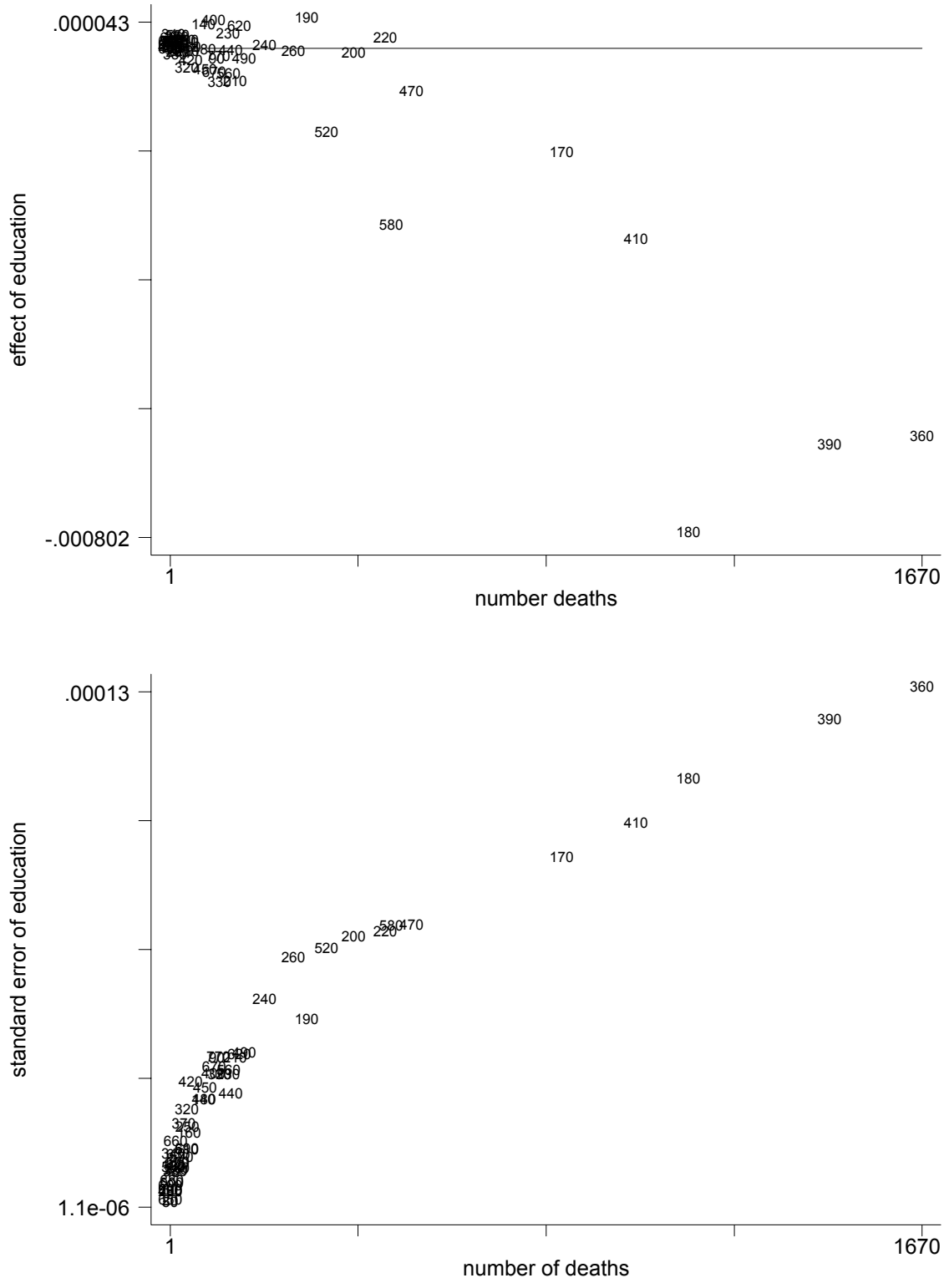


Figure 3: Gradients and the Number of Deaths



Appendix A: 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 A (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

Appendix B: Progress measures for 56 causes of death

Recode	Disease name	Number of deaths in sample	Estimated Annual Percent change in the age-adjusted mortality rate	Number of drugs approved 1986-1996
20	Shigellosis and amebiases	1	15.04	2
40	Tuberculosis of respiratory system	6	-4.70	0
50	Other tuberculosis	1	0.00	1
80	Meningococcal infection	1	0.00	0
90	Septicemia	105	-1.27	6
120	Viral Hepatitis	9	12.94	0
140	All other infections	75	-2.30	60
160	Neoplasms-lip, oral cavity and pharynx	43	-1.67	0
170	Neoplasms-digestive system	870	-0.74	2
180	Neoplasm-respiratory system	1151	0.74	0
190	Neoplasms-breast	304	-0.95	0
200	Neoplasms-genital organs	408	0.47	1
210	Neoplasms-urinary organs	145	0.39	0
220	Neoplasms-unspecified site	478	-0.26	6
230	Leukemia	130	0.17	5
240	Other malignant neoplasms of lymphatic tissues	210	1.68	10
250	Begin neoplasms	39	0.25	33
260	Diabetes	274	3.43	1
270	Nutritional deficiencies	16	-0.13	1
280	Anemias	18	-0.27	0
290	Meningitis	1	-7.31	6
320	Rheumatic fever and rheumatic heart disease	38	-3.85	0
330	Hypertensive heart disease	111	0.36	0
340	Hypertensive heart and renal disease	9	-2.70	0
360	Acute myocardial infarction	1670	-3.71	2
370	Other forms of ischemic heart disease	31	-4.40	0
380	Angina pectoris	12	-4.70	10
390	Old myocardial infection, chronic heart disease	1465	-2.03	4
400	Other diseases of endocardium	98	3.65	1
410	All other forms of heart disease	1035	-2.02	26
420	Hypertension with or without renal disease	47	2.94	24
440	Intracerebral and other intracranial hemorrhage	135	-0.90	0
450	Cerebral Thrombosis	79	-8.25	1
460	Cerebral embolism	2	-2.75	1
470	All other late effects of cerebrovascular disease	537	-0.82	2
480	Atheosclerosis	76	-6.10	0
490	Other disease of arteries, arterioles and capillaries	165	0.10	3
500	Acute bronchitis and bronchiolitis	5	-4.56	1
520	Pneumonia	348	-0.38	29
530	Influenza	8	-7.48	2
550	Bronchitis, chronic and unspecified	18	-2.17	14

Appendix B: Progress measures for 56 causes of death

Recode	Disease name	Number of deaths in sample	Estimated Annual Percent change in the age-adjusted mortality rate	Number of drugs approved 1986-1996
560	Emphysema	131	0.68	3
570	Asthma	26	1.99	5
580	other chronic obstructive pulmonary diseases	492	2.56	2
590	Ulcer of stomach and duodenum	38	-3.34	15
600	Apendicitis	2	-8.44	0
610	Hernia	38	-0.77	0
620	Chronic liver disease and cirrhosis	155	-1.61	2
630	Cholelithiasis and other disorders of the gallbladder	19	-3.18	1
650	Acute glomerulonephritis and nephrotic syndrome	2	0.00	0
660	Chronic glomerulonephritis and nephrotic syndrome	13	-0.99	0
670	Renal failure	99	-1.03	0
680	Infections of the kidney	5	-9.06	1
690	Hyperplasia of prostate	1	0.00	4
730	Congenital anomalies	14	-1.22	2

Appendix Table 1:
Case Studies of HIV/AIDS Treatment Disparities Over Time

Study	Treatment	Outcome			
Cunningham et al. (2000)	Highly Active Antiretroviral Therapy (HAART)	Characteristic	Percent with cumulative incident use by 1996		Percent with cumulative incident use by second follow-up (1998)
		Race:			
		Black	47		78
		White	20		59
		Hispanic	34		73
		Exposure Group			
		IVDU	32		65
		Homosexual	45		78
		Education:			
		Some HS	22		65
		HS Diploma	35		70
		Some College	41		71
		College degree	49		79
Sambamoorthi et al. (2001)	Protease Inhibitor/Non-Nucleoside Reverse Transcriptase Inhibitor	Characteristic	Odds Ratio 1996	Odds Ratio 1997	Odds Ratio 1998
		Race			
		White	-	-	-
		Black	.43	.58	.81
		Hispanic	.54	.56	.75
		Exposure Group			
		IVDU	.73	.98	1.13
		Non-IVDU	-	-	-
Crystal et al. (1995)	Zidovudine Treatment	Characteristic	Odds Ratio 1987-1988		Odds Ratio 1989-1990
		Race			
		White	.46		.79
		Black	-		-
		Exposure Group			
		IVDU	.84		1.09
		Non-IVDU			-

Appendix Table 2: 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%

Appendix Table 3: 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%

Appendix Table 4: Weighting SEER regression by number of individuals with the disease (alternative specification to Table 12)

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)	0.00022 (0.00048)	-0.00175 (0.00121)
Progress measured by increases in incidence rates		
(Estimated annual percent change in age adjusted incidence rates)	-0.00094** (0.00045)	-0.00323*** (0.00112)
Progress measured by increases in the 5 year-survival rates after diagnosis		
change in 5-year survival rate conditional on diagnosis	-0.01426* (0.00758)	-0.09576*** (0.01629)
Progress measured by the number of drugs available for treatment (Match by 3 digit icd9 code)		
number of drugs	-0.00003 (0.00005)	-0.00022* (0.00012)

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 for 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%