The impact of biomedical knowledge accumulation on mortality: a bibliometric analysis of cancer data

Frank R. Lichtenberg

frank.lichtenberg@columbia.edu

Columbia University

and

National Bureau of Economic Research

27 September 2013

The impact of biomedical knowledge accumulation on mortality: a bibliometric analysis of cancer data

Abstract

I examine the relationship across diseases between the long-run growth in the number of publications about a disease and the change in the mortality rate from the disease. The diseases analyzed are almost all the different forms of cancer, i.e. cancer at different sites in the body (lung, colon, breast, etc.). The National Cancer Institute publishes annual data on cancer incidence as well as on cancer mortality, by cancer site. Failure to control for the growth in incidence (which it is not feasible to do for non-cancer diseases) may bias estimates of the effect of publication growth towards zero, because growth in the number of publications is positively correlated across diseases with growth in incidence.

Time-series data on the number of publications pertaining to each cancer site were obtained from PubMed. For articles published since 1975, it is possible to distinguish between publications indicating and not indicating any research funding support.

My estimates indicated that mortality rates: (1) are unrelated to the (current or lagged) stock of publications that had not received research funding; (2) are only weakly inversely related to the contemporaneous stock of published articles that received research funding; and (3) are strongly inversely related to the stock of articles that had received research funding and been published 5 and 10 years earlier. The effect after 10 years is 66% larger than the contemporaneous effect. The strong inverse correlation between mortality growth and growth in the lagged number of publications that were supported by research funding is not driven by a small number of outliers.

Frank R. Lichtenberg
Columbia University
504 Uris Hall
3022 Broadway
New York, NY 10027
frank.lichtenberg@columbia.edu

I. Introduction

Many people and organizations have expressed the view that biomedical research has yielded substantial improvements in longevity and health. Nabel (2009) said that "biomedical research provides the basis for progress in health and health care." Moses and Martin (2011) said that "since 1945, biomedical research has been viewed as the essential contributor to improving the health of individuals and populations, in both the developed and developing world." Cutler, Deaton and Lleras-Muney (2006) "tentatively identif[ied] the application of scientific advance and technical progress (some of which is induced by income and facilitated by education) as the ultimate determinant of health." The Federation of American Societies for Experimental Biology (2013) said that "research in the biomedical sciences has generated a wealth of new discoveries that are improving our health, extending our lives and raising our standard of living." The National Institutes of Health (NIH) said that "in the last 25 years, NIH-supported biomedical research has directly led to human health benefits that both extend lifespan and reduce illnesses" (NIH (2013a)). The Australian Government (2013) said that "the purpose of health and medical research (HMR) is to achieve better health for all Australians. Better health encompasses increased life expectancy, as well as social goals such as equity, affordability and quality of life."

The hypothesis that biomedical research has yielded substantial improvements in longevity and health has been examined using two kinds of evidence. The first type of evidence consists of qualitative "case studies" of specific diseases. NIH (2013b, 2013c) describes the impacts of its long-term efforts to understand, treat, and prevent chronic diseases (including cardiovascular disease, cancer, diabetes, depression), and how it has worked to combat infectious diseases such as HIV/AIDS and influenza by helping to develop new therapies, vaccines, diagnostic tests, and other technologies.

The second kind of evidence is indirect, (partially) econometric evidence. This evidence is indirect because it is based on evidence about two links in the following causal chain:

biomedical research ➔ new drugs, devices, and procedures ➔ longevity and health

Regarding the first link: the National Cancer Institute (NCI) says that "approximately one half of the chemotherapeutic drugs currently used by oncologists for cancer treatment were discovered and/or developed at NCI" (NCI (2013)), and Sampat and Lichtenberg (2011) demonstrated that

new drugs often build on upstream government research. Regarding the second link: a number of studies have examined the impact of the introduction and use of new drugs, devices, and procedures on longevity and health.[1] For example, Lichtenberg (2011) analyzed the impact of new drugs and imaging procedures on longevity in the U.S. using longitudinal state-level data; Lichtenberg (2013a) analyzed the impact of new drugs on longevity in France using longitudinal disease-level data; and Lichtenberg (2013b) analyzed the impact of therapeutic procedure innovation on hospital patient longevity in Western Australia using patient-level data.

In this paper, I will use a different econometric approach to assess the impact that biomedical research has had on longevity: a direct examination of the relationship across diseases between the long-run growth in the number of research publications and the change in the mortality rate (in most cases controlling for the disease incidence rate). I hypothesize that the growth in the number of research publications about a disease is a useful indicator of the growth in knowledge about the disease. As the National Science Foundation (NSF) says, "Research produces new knowledge, products, or processes. Research publications reflect contributions to knowledge" (NSF (2013)). In his model of endogenous technological change, Romer (1990) hypothesized an aggregate production function such that an economy's output depends on the "stock of ideas" that have previously been developed, as well as on the economy's endowments of labor and capital. The mortality model that I will estimate may be considered a health production function, in which the mortality rate is an (inverse) indicator of health output or outcomes, and the cumulative number of publications is analogous to the stock of ideas.

Previous research on the agricultural and manufacturing sectors of the economy has found that counts of publications are useful indicators of the stock of knowledge. Evenson and Kislev (1973) used the publication of crop-specific scientific papers as a measure of agricultural research output in 75 wheat- and maize- growing countries to explain increases in yield per unit land in these crops over the period 1948-68. They observed a strong and persistent relationship between agricultural research and biological productivity-yield in wheat and maize. This relationship existed both "between" countries and "within" countries over time. Adams (1990) utilized article count data in each science as measures of knowledge in his analysis of productivity growth in two-digit manufacturing industries during the period 1949-83.

[1] Fuchs (2010) stated that "since World War II…biomedical innovations (new drugs, devices, and procedures) have been the primary source of increases in longevity."

The diseases we will analyze are almost all the different forms of cancer, i.e. cancer at different sites in the body (lung, colon, breast, etc.). About one-fourth of U.S. deaths during the period 1999-2010 were due to cancer. The main reason we focus on cancer is that the NCI publishes annual data on cancer incidence[2] as well as on cancer mortality, by cancer site. Incidence data aren't available for most other diseases. A less important reason is that the NCI uses a uniform cancer site classification scheme for data covering the entire period 1975-present. There were significant changes in the disease classification scheme for other diseases between 1998 and 1999, when the system used to classify underlying cause of death was changed from the International Classification of Diseases (ICD) Ninth Revision to the ICD Tenth Revision. As the CDC (2013) notes, the two classification schemes are different enough to make direct comparisons of cause-of-death difficult.

In the next section, I will briefly describe the biomedical publications data I will use. In Section III, I develop the econometric model I will use to investigate the impact of contributions to knowledge (as measured by publication counts) on cancer mortality rates. Descriptive statistics will be presented in Section IV. Estimates of the econometric model will be presented in Section V. Section VI provides a summary and conclusions.

II. Biomedical publications data

Time-series data on the number of publications pertaining to each cancer site were obtained from PubMed, a database developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine (NLM), one of the institutes of the National Institutes of Health (NIH). The database was designed to provide access to citations (with abstracts) from biomedical journals. PubMed's primary data resource is MEDLINE, the NLM's premier bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences, such as molecular biology. MEDLINE contains bibliographic citations and author abstracts from about 4,600 biomedical journals published in the United States and 70 other countries. The database contains about 12 million citations dating back to the mid-1960s. Coverage is worldwide, but most

[2] A cancer incidence rate is the number of new cancers of a specific site/type occurring in a specified population during a year, usually expressed as the number of cancers per 100,000 population at risk.

records are from English-language sources or have English abstracts. In addition to MEDLINE citations, PubMed provides access to non-MEDLINE resources, such as out-of-scope citations, citations that precede MEDLINE selection, and PubMed Central (PMC) citations.[3]

A controlled vocabulary of biomedical terms, the NLM Medical Subject Headings (MeSH), is used to describe the subject of each journal article in MEDLINE. MeSH contains approximately 26 thousand terms and is updated annually to reflect changes in medicine and medical terminology. MeSH terms are arranged hierarchically by subject categories with more specific terms arranged beneath broader terms.[4] PubMed allows one to view this hierarchy and select terms for searching in the MeSH Database. Skilled subject analysts examine journal articles and assign to each the most specific MeSH terms applicable—typically ten to twelve. Applying the MeSH vocabulary ensures that articles are uniformly indexed by subject, whatever the author's words (National Center for Biotechnology Information (2013)). Figure 1 shows an abridged sample of a PubMed bibliographic citation. I use three attributes (search fields) in the citation: the date of publication (line 8), the MeSH headings (lines 27-36), and the Publication Type (lines 18-20).

For articles published since 1975, the Publication Types identify U.S. government and non-U.S. government[5] financial support of the research that resulted in the published papers when that support is mentioned in the articles (National Library of Medicine (2013b)). Figure 2 shows data on the number of PubMed publications pertaining to cancer that were published during the period 1975-2009, by extent and source of research support. Cancer was one of the main topics discussed (i.e., cancer was a "MeSH Major Topic") in about 1.5 million articles published during this period. About 30% of these articles mentioned either U.S. government support, non-U.S. government support, or both. 20% of the articles indicating any research funding support mentioned only U.S. government support; 63% of the articles indicating any research funding support mentioned only non-U.S. government support; and 17% of the articles

[3] Together, these are often referred to as "PubMed-only citations." Out-of-scope citations are primarily from general science and chemistry journals that contain life sciences articles indexed for MEDLINE, e.g., the plate tectonics or astrophysics articles from *Science* magazine. Publishers can also submit citations with publication dates that precede the journal's selection for MEDLINE indexing, usually because they want to create links to older content. PMC citations are taken from life sciences journals (MEDLINE or non-MEDLINE) that submit full-text articles to PMC.
[4] The MeSH Tree Structure can be browsed online; see National Library of Medicine (2013a).
[5] Non-U.S. government financial support includes support by American societies, institutes, state governments, universities, and private organizations, and by foreign sources (national, departmental, provincial, academic & private organizations).

indicating any research funding support mentioned both U.S. government and non-U.S. government support. This distribution of funding support by source is quite consistent with data compiled by Research!America (shown in Figure 3) on the distribution of 2011 U.S. biomedical and health R&D spending, by source of funding. The Research!America data indicate that the Federal government accounted for 29% of 2011 U.S. biomedical and health R&D spending. If we assume that the U.S. government deserves "half the credit" for articles that mentioned both U.S. government and non-U.S. government support, we can say that the U.S. government support accounted for 28.5% (= 20% + (17% / 2)) of the funding support for articles that received any funding support.

Our ability to distinguish between publications indicating and not indicating any research funding support will allow us to test the hypothesis that an increase in the number of publications indicating any research funding support has a larger (more negative) effect on mortality than an increase in the number of publications not indicating any research funding support; the latter may even have no effect. In principle, our ability to also distinguish between publications indicating U.S. government and non-U.S. government funding support could also allow us to separately examine the effects of both kinds of research funding on mortality. However, since almost half of the articles acknowledging U.S. government support also acknowledged non-U.S. government support, disentangling the effects of the two kinds of research funding on mortality may be difficult.

The PubMed database indicates the year of publication of each article, but not the year(s) in which research funding occurred (for articles that acknowledged research funding). However the NIH Reporter database (NIH (2013)) enables us to determine the start dates of NIH projects that yielded PubMed articles, as well as the publication dates of those articles. Hence, we can analyze the frequency distribution of the lag between project start date and the publication date of articles. The distribution of NIH-supported articles, by lag between project start date and publication date, is shown in Figure 4.[6] The median lag from project start to article publication is about six years. However, since this figure is based on right-censored data—articles that were or will be published after 2011 are excluded—six years should be considered a lower-bound estimate of the median lag from project start to article publication.

[6] Figure 4 is based on data on almost all NIH-supported articles published during 1985-2011 (N = 323,196), not just articles about cancer.

When former NIH Director Harold Varmus testified before Congress in 1998, he said that "the benefits of research are unpredictable…Although basic research projects initially may appear to be unrelated to any specific disease, findings from this research ultimately may prove to be a critical turning point in a long chain of discoveries leading to improved health" (Varmus (1998)). Determining whether or not a research project is applicable to a specific disease is therefore likely to be far easier six or more years after the project began (and articles are published) than it was when the project started.

III. Econometric model

Two types of statistics are often used to assess progress in the "war on cancer": survival rates and mortality rates**.** Survival rates are typically expressed as the proportion of patients alive at some point subsequent to the diagnosis of their cancer. For example, the observed 5-year survival rate is defined as follows:

5-year Survival Rate = Number of people diagnosed with cancer at time t alive at time t+5 / Number of people diagnosed with cancer at time t

= 1 – (Number of people diagnosed with cancer at time t dead at time t+5 / Number of people diagnosed with cancer at time t)

Hence, the survival rate is based on a *conditional* (upon previous diagnosis) mortality rate. The second type of statistic is the *unconditional* cancer mortality rate: the number of deaths, with cancer as the underlying cause of death, occurring during a year per 100,000 population.

The 5-year relative survival rate from cancer has increased steadily since the mid-1970s, from 49.1% for people diagnosed during 1975-1977 to 67.6% for people diagnosed during 2001-2008. Although this increase suggests that there has been significant progress in the war against cancer, it might simply be a reflection of (increasing) lead-time bias. Lead time bias is the bias that occurs when two tests for a disease are compared, and one test (the new, experimental one) diagnoses the disease earlier, but there is no effect on the outcome of the disease--it may appear that the test prolonged survival, when in fact it only resulted in earlier diagnosis when compared to traditional methods. Welch et al (2000) argued that "improving 5-year survival over time…should not be taken as evidence of improved prevention, screening, or therapy." They

argued that "while 5-year survival is a perfectly valid measure to compare cancer therapies in a randomized trial, comparisons of 5-year survival rates across time (or place) may be extremely misleading. If cancer patients in the past always had palpable tumors at the time of diagnosis while current cancer patients include those diagnosed with microscopic abnormalities, then 5-year survival would be expected to increase over time even if new screening and treatment strategies are ineffective. To avoid the problems introduced by changing patterns of diagnosis, observers have argued that progress against cancer be assessed using population-based mortality rates." Therefore, the dependent variable I will analyze will be the unconditional cancer mortality rate, rather than a variable based on the survival rate.[7]

The unconditional cancer mortality rate is essentially the unconditional probability of death from cancer (P(death from cancer)). The law of total probability implies the following:

$$P(\text{death from cancer}) = P(\text{death from cancer} \mid \text{cancer diagnosis}) * P(\text{cancer diagnosis}) + P(\text{death from cancer} \mid \text{no cancer diagnosis}) * (1 - P(\text{cancer diagnosis})) \quad (1)$$

The probability of dying from cancer is much lower than the probability of being diagnosed with cancer: in 2006, the cancer incidence rate was 2.5 times as high as the cancer mortality rate.[8] This suggests that the probability that a person who has never been diagnosed with cancer dies from cancer is quite small: $P(\text{death from cancer} \mid \text{no cancer diagnosis}) \approx 0$. In this case, eq. (1) reduces to

$$P(\text{death from cancer}) \approx P(\text{death from cancer} \mid \text{cancer diagnosis}) * P(\text{cancer diagnosis}) \quad (2)$$

Hence

$$\ln P(\text{death from cancer}) \approx \ln P(\text{death from cancer} \mid \text{cancer diagnosis}) + \ln P(\text{cancer diagnosis}) \quad (3)$$

[7] I will control for cancer incidence (by including it in the mortality equation), but in a completely unrestrictive manner. If changes in incidence are merely due to lead-time bias, the coefficient on incidence should be zero.
[8] 2006 U.S. age-adjusted incidence and mortality rates were 456.2 and 181.1, respectively.

I hypothesize that the conditional mortality rate (P(death from cancer | cancer diagnosis)) is inversely related to the (current or lagged) stock of useful knowledge about cancer.[9] The stock of knowledge is not directly observable, but I also hypothesize that the cumulative number of scientific publications is a meaningful indicator of the stock of knowledge.

$$\ln P(\text{death from cancer} \mid \text{cancer diagnosis}) = \beta \ln(\text{cum_pubs}_{t-k}) \quad (4)$$

Substituting (4) into (3),

$$\ln P(\text{death from cancer}) \approx \beta \ln(\text{cum_pubs}_{t-k}) + \ln P(\text{cancer diagnosis}) \quad (5)$$

To assess the impact of biomedical research on cancer mortality, I will estimate the following difference-in-differences version of eq. (5), based on longitudinal, cancer-site-level data on about 45 cancer sites:[10]

$$\ln(\text{mort_rate}_{st}) = \beta \ln(\text{cum_pubs}_{s,t-k}) + \gamma \ln(\text{inc_rate}_{st}) + \alpha_s + \delta_t + \varepsilon_{st} \quad (6)$$

mort_rate_{st} = the age-adjusted mortality rate from cancer at site s (s = 1,…, 47) in year t (t=1995,…,2009)

$\text{cum_pubs}_{s,t-k}$ = the number of PubMed articles published by the end of year t – k that were about cancer at site s

inc_rate_{st} = the age-adjusted incidence rate of cancer at site s in year t

α_s = a fixed effect for cancer site s

δ_t = a fixed effect for year t

ε_{st} = a disturbance

The fixed year effects control for time-varying factors that influence cancer mortality rates in general.

[9] The stock of useful knowledge may also affect the probability of diagnosis.

[10] Galiani et al (2005) used a difference-in-differences model to assess the impact of privatization of water services on child mortality in Argentina. They estimated their model using data classified by region and year, whereas the data I will use are classified by disease and year. Their "treatment variable" (whether water services were publicly or privately provided) was discrete, whereas my treatment variable (stocks of publications) are continuous.

I will estimate models based on eq. (6) using three alternative values of k: 0, 5, and 10.[11] For concreteness, suppose that k = 10. Now, let's write specific versions of eq. (6) for the first and last years of the sample period (t = 1995 and t = 2009):

$$\ln(\text{mort_rate}_{s,1995}) = \beta \ln(\text{cum_pubs}_{s,1985}) + \gamma \ln(\text{inc_rate}_{s,1995}) + \alpha_s + \delta_{1995} + \varepsilon_{s,1995} \quad (7)$$

$$\ln(\text{mort_rate}_{s,2009}) = \beta \ln(\text{cum_pubs}_{s,1999}) + \gamma \ln(\text{inc_rate}_{s,2009}) + \alpha_s + \delta_{2009} + \varepsilon_{s,2009} \quad (8)$$

Subtracting eq. (7) from eq. (8),

$$\ln(\text{mort_rate}_{s,2009} / \text{mort_rate}_{s,1995}) = \beta \ln(\text{cum_pubs}_{s,1999} / \text{cum_pubs}_{s,1985}) +$$
$$\gamma \ln(\text{inc_rate}_{s,2009} / \text{inc_rate}_{s,1995}) + (\delta_{2009} - \delta_{1995}) + (\varepsilon_{s,2009} - \varepsilon_{s,1995}) \quad (9)$$

or

$$\Delta\ln(\text{mort_rate}_s) = \beta\, \Delta\ln(\text{cum_pubs}_s) + \gamma\, \Delta\ln(\text{inc_rate}_s) + \delta' + \varepsilon_s' \quad (10)$$

where

$$\begin{aligned} \Delta\ln(\text{mort_rate}_s) &= \ln(\text{mort_rate}_{s,2009} / \text{mort_rate}_{s,1995}) \\ \Delta\ln(\text{cum_pubs}_s) &= \ln(\text{cum_pubs}_{s,1999} / \text{cum_pubs}_{s,1985}) \\ \Delta\ln(\text{inc_rate}_s) &= \ln(\text{inc_rate}_{s,2009} / \text{inc_rate}_{s,1995}) \\ \delta' &= (\delta_{2009} - \delta_{1995}) \end{aligned}$$

The cancer-site fixed effects that were included in the "within" model (eq. (6)) are no longer present in the "long-difference" model (eq. (10)); the intercept of eq. (10) is the difference between the initial- and end-year year fixed effects. In this simple model, the long-run growth of the age-adjusted cancer mortality rate depends on the long-run growth of the (lagged) cumulative number of publications, the long-run growth of the age-adjusted cancer incidence rate, and a constant.

Eq. (10) can easily be generalized to allow for two or three different stocks of publications:

$$\Delta\ln(\text{mort_rate}_s) = \beta_{\text{RESEARCH}}\, \Delta\ln(\text{cum_research_pubs}_s)$$
$$+ \beta_{\text{NON-RESEARCH}}\, \Delta\ln(\text{cum_non_research_pubs}_s) + \gamma\, \Delta\, \ln(\text{inc_rate}_s) + \delta' + \varepsilon_s' \quad (11)$$

$$\Delta\, \ln(\text{mort_rate}_s) = \beta_{\text{RESEARCH_US_GOV}}\, \Delta\ln(\text{cum_US_gov_research_pubs}_s)$$

[11] Since data on financial support of research that resulted in published papers begin in 1975, it is not practical to specify longer lags (k > 10).

$$+ \beta_{RESEARCH_OTHER}\,\Delta\ln(cum_other_research_pubs_s)$$

$$+ \beta_{NON\text{-}RESEARCH}\,\Delta\ln(cum_non_research_pubs_s) + \gamma\,\Delta\ln(inc_rate_s) + \delta' + \varepsilon_s' \qquad (12)$$

where

$cum_research_pubs_{s,t-k}$	= the number of PubMed articles indicating any research funding support published by the end of year t–k that were about cancer at site s
$cum_non_research_pubs_{s,t-k}$	= the number of PubMed articles not indicating any research funding support published by the end of year t–k that were about cancer at site s
$cum_US_gov_research_pubs_{s,t-k}$	= the number of PubMed articles indicating US government research funding support published by the end of year t–k that were about cancer at site s
$cum_other_research_pubs_{s,t-k}$	= the number of PubMed articles indicating non-US government research funding support published by the end of year t–k that were about cancer at site s

I will estimate eqs. 10-12 for three different values of k (0, 5, and 10). These equations will be estimated via weighted least-squares, weighting by the mean mortality rate of cancer site s during the period 1985-2009. Since the dependent variable is the log of the mortality rate, I am analyzing percentage changes in the mortality rate. As shown in Figure 5, the data exhibit heteroskedasticity: cancer sites with low average mortality rates exhibit much larger positive and negative percentage changes in mortality rates than cancer sites with high average mortality rates. Weighted least squares is appropriate in the presence of heteroskedasticity.

IV. Descriptive statistics

Data on age-adjusted incidence and mortality rates were obtained from SEER Cancer Query Systems (National Cancer Institute (2013b)). Incidence and mortality rates of all malignant cancers combined during the period 1973-2009 are shown in Figure 6. Incidence and mortality both increased between the mid-1970s and the early 1990s, when both began to decline. Between 1992 and 2009, the incidence rate declined 9% and the mortality rate declined 19%.

Age-adjusted mortality and incidence rates in 1995 and 2009 and PubMed publication counts 10 years earlier (in 1985 and 1999) for the top 18 cancer sites (ranked by mean mortality

rate) are shown in Table 1.[12] Lung cancer had the largest mean mortality rate by far; it accounted for more than one in four cancer deaths. Between 1995 and 2009, the lung cancer incidence rate declined 12% and the lung cancer mortality rate declined 17%. The cumulative number of PubMed publications about lung cancer (cum_pubs) approximately doubled between 1985 and 1999; the cumulative number of PubMed publications about lung cancer that cited any research support (cum_research_pubs) more than tripled.

The second largest cancer (ranked by mean mortality rate) was colon cancer. The incidence and mortality rates of colon cancer declined about twice as much as the incidence and mortality rates of lung cancer: by 23% and 34%, respectively. But lagged cum_pubs and cum_research_pubs increased more slowly for colon cancer than they did for lung cancer: by 77% and 139%, respectively.

The third largest cancer (ranked by mean mortality rate) was breast cancer. The breast cancer incidence rate declined just 4%, while the breast cancer mortality rate declined by 29%. Lagged cum_pubs and cum_research_pubs increased more for breast cancer than they did for lung cancer: by 144% and 294%, respectively.

Weighted means, standard deviations, and correlation coefficients across 47 cancer sites of 1995-2009 growth in mortality, incidence, and cumulative number of publications 10 years earlier are shown in Table 2. Observations are weighted by mean mortality rate. The weighted mean declines in mortality and incidence are consistent with the data shown in Figure 6. The mean log change in publications acknowledging research funding (cum_research_pubs) was almost twice as large as the mean log change in total publications (cum_pubs); this is at least partly due to the fact that only articles published after 1974 include information about research funding. The mean log change in publications acknowledging non-U.S. government research funding (cum_other_research_pubs) was 81% larger than the mean log change in publications acknowledging U.S. government research funding (cum_gov_research_pubs). This is consistent with data compiled by Research!America, which indicate that the Federal government's share of U.S. biomedical R&D has been declining; it fell from 34% in 2002 to 29% in 2011.

As shown in the first row of correlation coefficients in Table 2, there is a significant positive correlation across cancer sites between the growth in incidence and the growth in

[12] Age-adjusted mortality and incidence rates and PubMed publication counts for the other 29 cancer sites not included in Table 1 are shown in Appendix Table 1.

mortality: cancer sites with larger declines in incidence had larger declines in mortality. The correlation between mortality growth and growth in non-research publications is insignificant, but the correlations between mortality growth and growth in cum_research_pubs, cum_gov_research_pubs, and cum_other_research_pubs are negative and significant. The correlation between the growth of government and other research publications is quite high (r = 0.856), suggesting that disentangling the effects of the two kinds of research funding on mortality may be difficult.

V. Estimates of models of 1995-2009 growth of the age-adjusted cancer mortality rate

Weighted least-squares estimates of models of 1995-2009 growth of the age-adjusted cancer mortality rate (eqs. 10-12) are shown in Table 3. The equations were estimated using three alternative assumed values of the lag (k) from cumulative publications to the mortality rate: 0, 5, and 10 years. k = 0 in models 1-5; k = 5 in models 6-10; and k = 10 in models 11-15.

Model 1 is a simple regression of the growth in the mortality rate on the growth in cum_pubs, i.e. the growth in the incidence rate is excluded. The coefficient on the growth in cum_pubs is insignificant. Model 2 includes the growth in the incidence rate as well as the growth in cum_pubs. In this model, the coefficient on the growth in cum_pubs is negative and highly significant (and the coefficient on the growth in the incidence rate (γ) is positive and significant). This indicates that failure to control for the growth in incidence (which it is not feasible to do for non-cancer diseases) may bias estimates of the coefficient on the growth in cum_pubs (β) towards zero, because growth in the number of publications is positively correlated across diseases with growth in incidence.[13] In model 3, the growth in cum_pubs is replaced by the growth in cum_research_pubs. The coefficient on the growth in cum_research_pubs is also negative and highly significant. However, when we control (in model 4) for the growth in cum_non_research_pubs, the estimate of $\beta_{RESEARCH}$ is only marginally significant (p-value = 0.092).[14] Model 5 is an estimate of eq. (12), in which cum_research_pubs

[13] The coefficient on incidence growth is positive, but (contrary to eq. (5) above) significantly less than one: a 10% rise in incidence is associated with a 7.3% rise in mortality. This may be at least partly due to the fact that measured incidence is a noisy indicator of true incidence, e.g. due to changing patterns of diagnosis and a changing degree of lead-time bias.

[14] As shown in Table 2, the correlation across cancer sites between growth in cum_ research_pubs and growth in cum_non_research_pubs is quite high (0.647).

is disaggregated into cum_gov_research_pubs and cum_other_research_pubs. Neither $\beta_{RESEARCH_US_GOV}$ nor $\beta_{RESEARCH_OTHER}$ is significant, which is not surprising given the high correlation across cancer sites between the growth of government and other research publications.

Models 6-10 are identical to models 1-5, except the assumed lag from cumulative publications to the mortality rate is five years rather than zero years. The estimates of models 6-8 are similar to the estimates of models 1-3, but the contrast between models 9 and 4 (which include both cum_research_pubs and cum_non_research_pubs) is interesting. Although $\beta_{RESEARCH}$ is only marginally significant (p-value = 0.092) in model 4, it is highly significant (p-value = 0.012) in model 9. This means that although the mortality rate is only weakly inversely related to the contemporaneous stock of publications that had received research funding (controlling for the contemporaneous stock of publications that had not received research funding), it is strongly inversely related to the stock of publications that had received research funding 5 years earlier. Moreover, the magnitude of the point estimate of $\beta_{RESEARCH}$ is 46% larger in model 9 than it is model 4.

In models 11-15, the assumed lag from cumulative publications to the mortality rate is ten years. As shown in Figure 7, the magnitude of the point estimate of $\beta_{RESEARCH}$ in model 14 is 14% larger than it is in model 9, and 66% larger than it is in model 4.

Figure 8 shows the partial correlation across cancer sites between the 1985-1999 log change in the number of research publications and the 1995-2009 log change in the mortality rate, controlling for the 1995-2009 log change in the incidence rate. The figure is a plot of the residuals from the weighted simple regression of $\Delta\ln(mort_rate_s)$ on $\Delta\ln(inc_rate_s)$ against the residuals from the weighted simple regression of $\Delta\ln(cum_research_pubs_s)$ on $\Delta\ln(inc_rate_s)$, where we assume a 10-year lag from cumulative publications to the mortality rate.[15] The figure suggests that the strong inverse correlation between mortality growth and growth in the lagged number of publications that were supported by research funding is not being driven by a small number of outliers. If we exclude lung cancer, which receives the greatest weight by far, from the sample, the estimate of $\beta_{RESEARCH}$ in eq. (13) hardly changes: $\beta_{RESEARCH} = -0.285$ (T = -3.22; p-value = 0.003).

[15] Figure 8 is a partial regression plot of model 13 in Table 3.

The magnitude of $\beta_{RESEARCH}$ in eq. (13) is quite large. As shown in Table 2, the weighted mean value of Δln(cum_research_pubs$_s$) is 1.538. The average annual rate of increase in lagged cum_research_pubs during 1995-2009 was 11.0% (= 1.538 / 14). Eq. (13) implies that, during the period 1995-2009, the growth in the lagged number of publications that were supported by research funding reduced the age-adjusted cancer mortality rate by 3.5% (= -0.319 * 11.0%) per year. During that period, the age-adjusted cancer mortality rate declined at an average annual rate of 1.5%.[16] This means that, in the absence of *any* growth in the lagged number of publications that were supported by research funding, the age-adjusted cancer mortality rate would have *increased* at an average annual rate of 2.0%. However, since there was such rapid growth in the number of publications, estimating what would have happened in the absence of any growth requires substantial out-of-sample prediction, which is certainly subject to great uncertainty.

VI. Summary and conclusions

Previous research on the agricultural and manufacturing sectors of the economy has found that counts of publications are useful indicators of the stock of knowledge: they are strongly positively correlated with productivity. In this paper, I have examined the relationship across diseases between the long-run growth in the number of publications about a disease and the change in the mortality rate from the disease.

The diseases I analyzed are almost all the different forms of cancer, i.e. cancer at different sites in the body (lung, colon, breast, etc.). About one-fourth of U.S. deaths during the period 1999-2010 were due to cancer. The main reason I focused on cancer is that the National Cancer Institute publishes annual data on cancer incidence as well as on cancer mortality, by cancer site. Failure to control for the growth in incidence (which it is not feasible to do for non-cancer diseases) may bias estimates of the effect of publication growth towards zero, because growth in the number of publications is positively correlated across diseases with growth in incidence.

[16] Eq. (13) implies that declining incidence accounted for about 1/6 of the decline in mortality.

Time-series data on the number of publications pertaining to each cancer site were obtained from PubMed. For articles published since 1975, it is possible to distinguish between publications indicating and not indicating any research funding support.

My estimates indicated that mortality rates: (1) are unrelated to the (current or lagged) stock of publications that had not received research funding; (2) are only weakly inversely related to the contemporaneous stock of published articles that received research funding; and (3) are strongly inversely related to the stock of articles that had received research funding and been published 5 and 10 years earlier. The effect after 10 years is 66% larger than the contemporaneous effect. The strong inverse correlation between mortality growth and growth in the lagged number of publications that were supported by research funding is not driven by a small number of outliers.

Research!America (2013) estimates that U.S. biomedical and health R&D spending (from all sources) declined by more than 3% in fiscal year 2011, and that this is the first drop in overall spending since 2002. While most of that decrease reflects the end of American Recovery and Reinvestment Act (ARRA) funding, which allocated $10.4 billion to the National Institutes of Health over two fiscal years (2009-2010), federal funding declined beyond the drop attributable to ARRA. In subsequent years, across-the-board cuts could cut billions more out of the federal research budget. The White House Office of Management and Budget estimated that the NIH alone could lose $2.53 billion in funding in fiscal year 2013. The evidence in this paper strongly suggests that reductions in biomedical and health R&D spending will ultimately have an adverse effect on U.S. longevity growth.

References

Adams JD (1990), "Fundamental Stocks of Knowledge and Productivity Growth," *Journal of Political Economy* 98 (4): 673-702, August.

Australian Government (2013), National Health and Medical Research Council, Meeting the Challenges of the Next Decade to Improve Health through Research https://www.nhmrc.gov.au/_files_nhmrc/file/research/nhmrc_submission_background_mckeon_120417.pdf

Canese K, Jentsch J, Myers C, PubMed: The Bibliographic Database, Chapter 2 of *The NCBI Handbook* [Internet]. McEntyre J, Ostell J, editors. Bethesda (MD): National Center for Biotechnology Information (US); 2002-, http://www.ncbi.nlm.nih.gov/books/NBK21094/

Centers for Disease Control and Prevention (2013), Compressed Mortality File 1968-2010, http://wonder.cdc.gov/wonder/help/cmf.html#Compressed%20Mortality%20File:%20ICD%20Revision

Cutler D, Deaton A, Lleras-Muney A (2006). "The Determinants Of Mortality," *Journal of Economic Perspectives* 20(3,Summer), 97-120.

Evenson RE, Kislev Y (1973). "Research and Productivity in Wheat and Maize." *Journal of Political Economy* 81(6): 1309-29.

Federation of American Societies for Experimental Biology (2013), Benefits of Biomedical Research, http://www.faseb.org/Policy-and-Government-Affairs/Data-Compilations/Benefits-of-Biomedical-Research.aspx

Fuchs, VR (2010), "New Priorities for Future Biomedical Innovations," *N Engl J Med* 363:704-706 August 19, http://www.nejm.org/doi/full/10.1056/NEJMp0906597

Galiani S, Gertler P, Schargrodsky E, (2005). "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy* 113 (1), February, 83-120.

Lichtenberg FR (2011), "The quality of medical care, behavioral risk factors, and longevity growth**,"** *International Journal of Health Care Finance and Economics* 11(1): 1-34, March.

Lichtenberg FR (2013a), "The impact of pharmaceutical innovation on longevity and medical expenditure in France, 2000–2009," *Economics and Human Biology*.

Lichtenberg FR (2013b), "The impact of therapeutic procedure innovation on hospital patient longevity: Evidence from Western Australia, 2000-2007," *Social Science and Medicine* 77: 50-9, January 2013.

Moses H, Martin JB (2011), Biomedical Research and Health Advances. *N Engl J Med* 364:567-571, February 10. http://www.nejm.org/doi/full/10.1056/NEJMsb1007634

Nabel EG (2009). Linking biomedical research to health care. *J Clin Invest.* 119(10):2858–2858, Oct. 1.

National Cancer Institute (2013a), Drug Discovery at the National Cancer Institute, http://www.cancer.gov/cancertopics/factsheet/NCI/drugdiscovery

National Cancer Institute (2013b), SEER Cancer Query Systems (CanQues), http://seer.cancer.gov/canques/

National Center for Biotechnology Information (2013), PubMed Help [Internet], http://www.ncbi.nlm.nih.gov/books/NBK3827/

National Institutes of Health (2013a), NIH Research's Impact on Health, http://www.nih.gov/about/impact/impact_health.pdf

National Institutes of Health (2013b), NIH…Turning Discovery into Health, http://www.nih.gov/about/discovery/viewbook_2011.pdf ,

National Institutes of Health (2013c), NIH Research's Impact on Scientific Knowledge, http://www.nih.gov/about/impact/impact_knowledge.pdf

National Institutes of Health (2012), ExPORTER - NIH RePORTER Database Download, http://exporter.nih.gov/

National Library of Medicine (2013a), MeSH Tree Structure, http://www.nlm.nih.gov/cgi/mesh/2013/MB_cgi

National Library of Medicine (2013b), Funding Support (Grant) Information in MEDLINE/PubMed – 2013, http://www.nlm.nih.gov/bsd/funding_support.html

National Science Foundation (2013), Key Science and Engineering Indicators 2012 Digest, Research Outputs: Publications and Patents http://www.nsf.gov/statistics/digest12/outputs.cfm

Research!America (2013), U.S. Investment in Health Research, http://www.researchamerica.org/research_investment

Romer, P., 1990. Endogenous Technological Change. *Journal of Political Economy* 98 (5, Part 2), S71-S102.

Sampat, B. N., Lichtenberg, F. R., 2011. What are the Respective Roles of the Public and Private Sectors in Pharmaceutical Innovation? *Health Affairs* 30(2), 332-9.

Varmus H (1998), New Developments in Medical Research: NIH and Patient Groups, Testimony Before the House Commerce Committee, Subcommittee on Health and Environment, March 26, http://www.hhs.gov/asl/testify/t980326a.html

Welch, H. Gilbert, Lisa M. Schwartz, and Steven Woloshin (2000), "Are Increasing 5-Year Survival Rates Evidence of Success Against Cancer?," *JAMA* 283(22): 2975-2978.

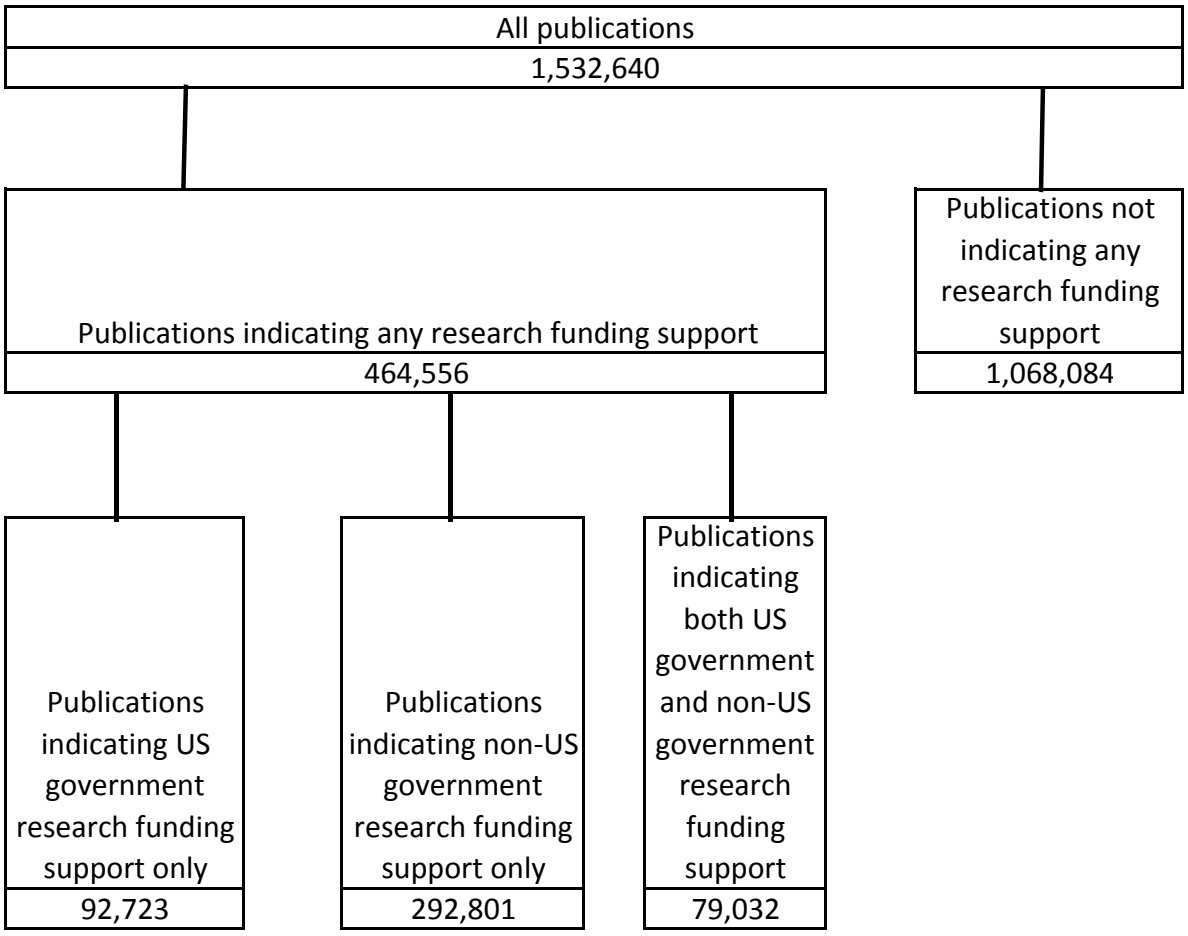
Figure 1

Abridged sample of a PubMed bibliographic citation

Line	
1	PMID- 20425429
2	OWN - NLM
3	STAT- MEDLINE
4	DA - 20100428
5	DCOM- 20100810
6	VI - 4
7	IP - 3
8	**DP - 2009 Jul**
9	TI - Application of immunotherapy in pediatric leukemia.
10	PG - 159-66
11	LID - 10.1007/s11899-009-0022-5 [doi]
12	AD - Center for Cancer Research, National Cancer Institute, National Institutes of
13	Health, Building 10, Room 1W-3750, 9000 Rockville Pike, MSC-1104, Bethesda, MD
14	20892, USA. waynea@mail.nih.gov
15	FAU - Wayne, Alan S
16	AU - Wayne AS
17	LA - eng
18	PT - Journal Article
19	**PT - Research Support, N.I.H., Intramural**
20	PT - Review
21	PL - United States
22	TA - Curr Hematol Malig Rep
23	JT - Current hematologic malignancy reports
24	JID - 101262565
25	RN - 0 (Immunotoxins)
26	SB - IM
27	MH - Child
28	MH - Graft vs Leukemia Effect/immunology
29	MH - Hematopoietic Stem Cell Transplantation/methods
30	MH - Humans
31	MH - Immunotherapy/*methods
32	MH - Immunotherapy, Adoptive/methods
33	MH - Immunotoxins/immunology/therapeutic use
34	**MH - Leukemia/immunology/pathology/*therapy**
35	MH - Models, Immunological
36	MH - Transplantation, Homologous
37	RF - 50
38	EDAT- 2010/04/29 06:00
39	MHDA- 2010/08/11 06:00
40	CRDT- 2010/04/29 06:00
41	AID - 10.1007/s11899-009-0022-5 [doi]
42	PST - ppublish
43	SO - Curr Hematol Malig Rep. 2009 Jul;4(3):159-66. doi: 10.1007/s11899-009-0022-5

Figure 2

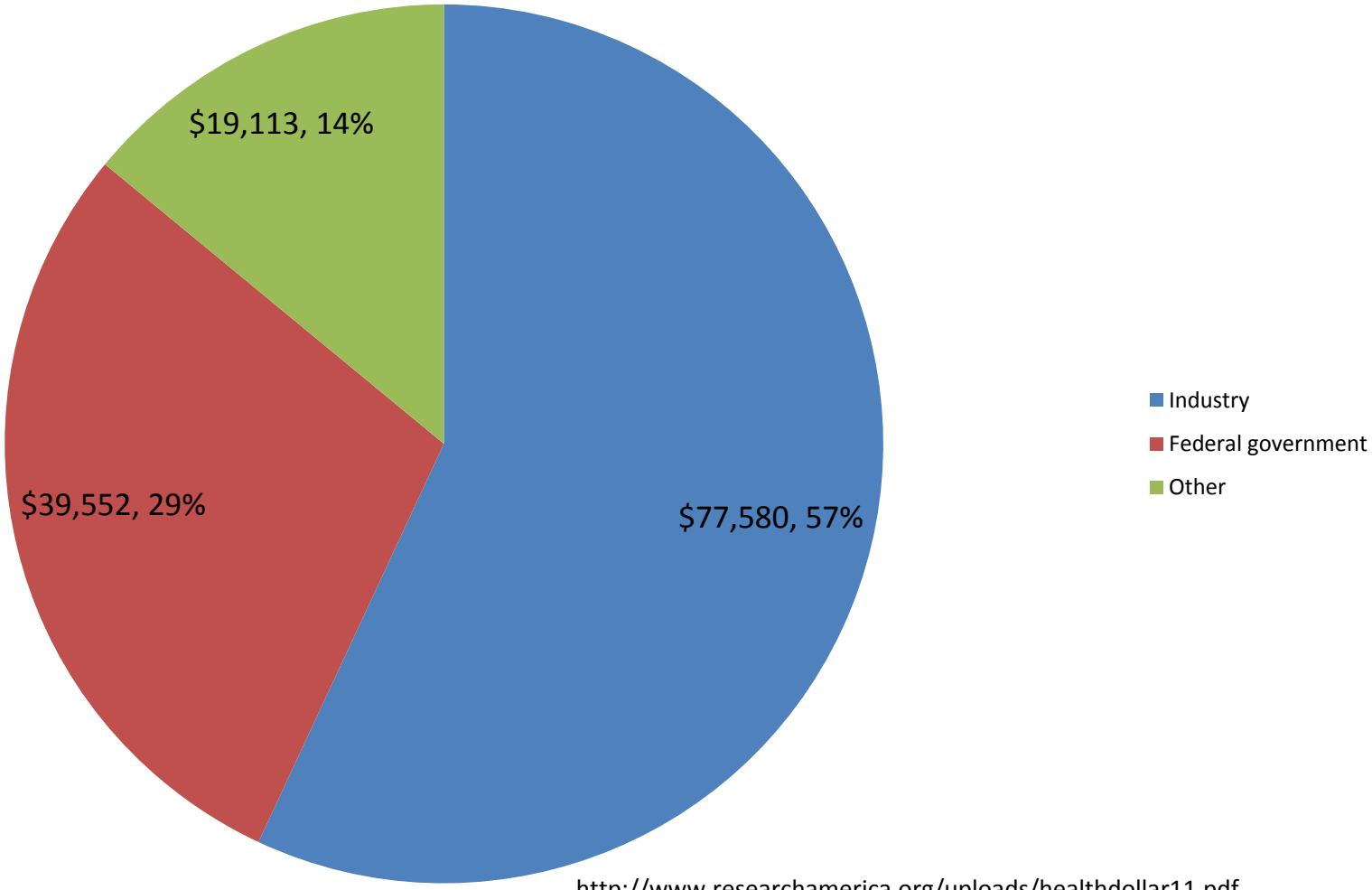
Number of PubMed publications pertaining to cancer* that were published during the period 1975-2009, by extent and source of research support



* PubMed publications pertaining to cancer are those identified by the search "neoplasms[MeSH Major Topic]"

Figure 3

2011 U.S. Biomedical and Health R&D Spending (millions of dollars)



http://www.researchamerica.org/uploads/healthdollar11.pdf

Figure 4
Distribution of NIH-supported articles,
by lag between project start date and publication date

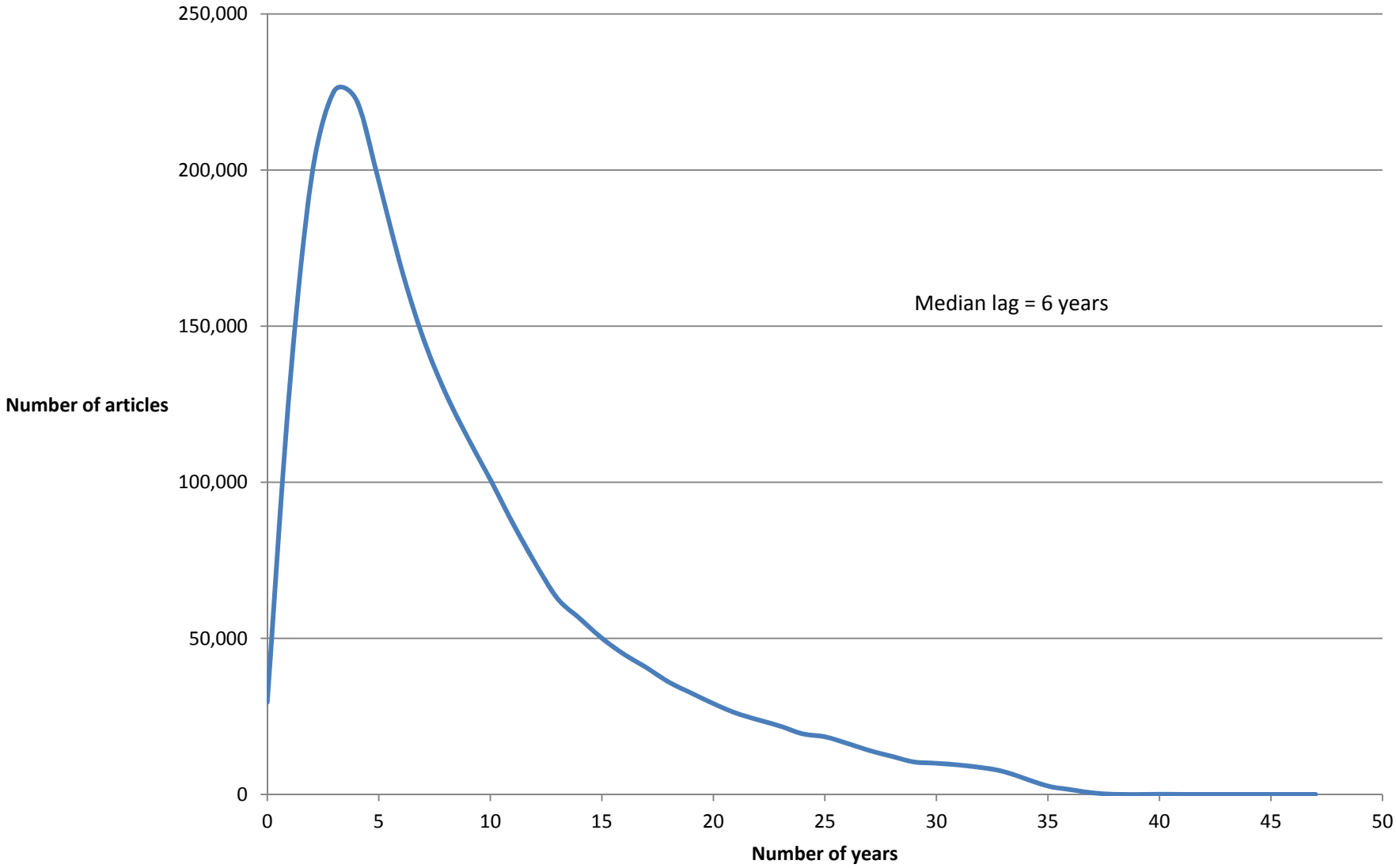


Figure 5
Heteroskedasticity: relationship across cancer sites between m
mean mortality rate and log change in mortality rate

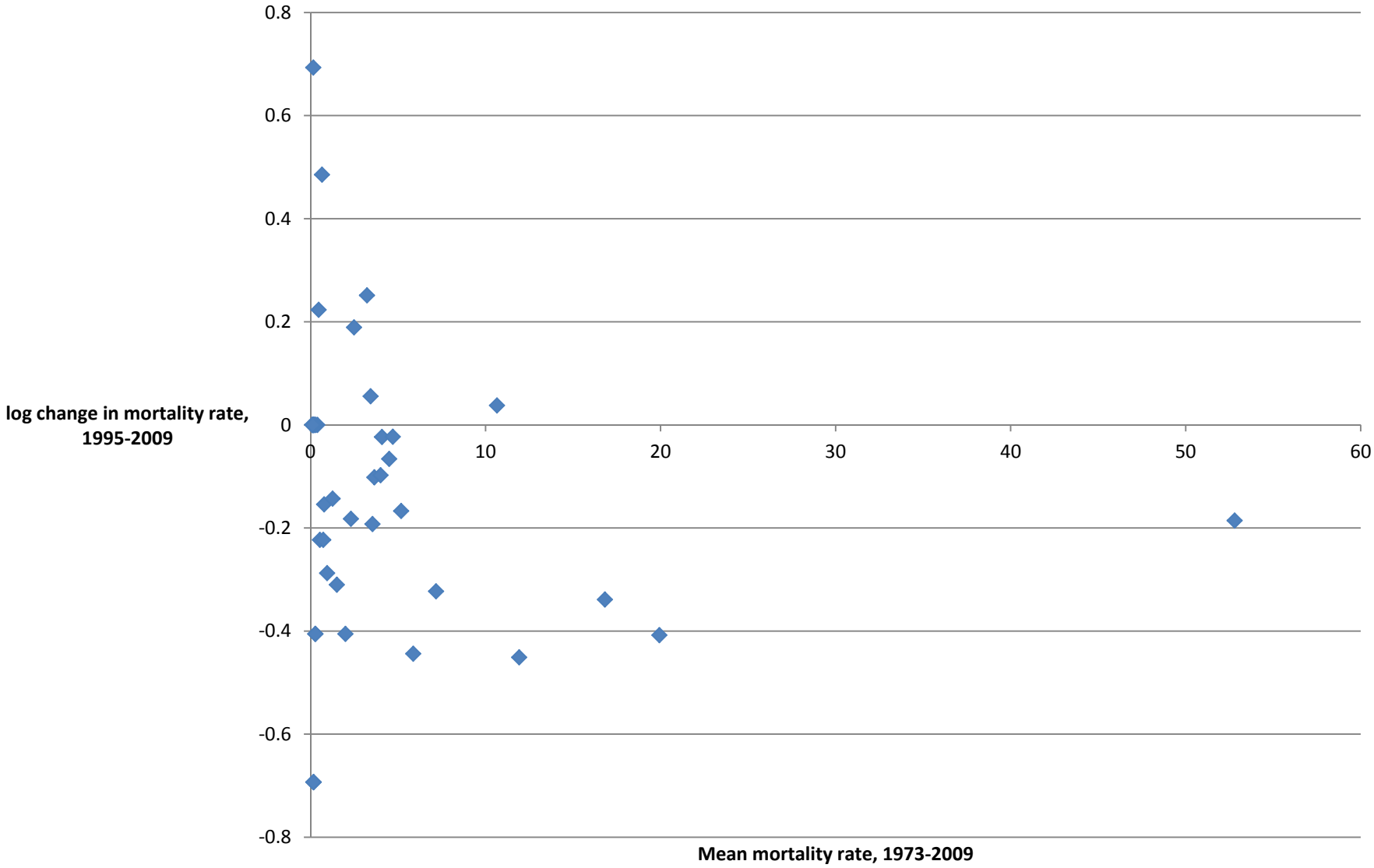


Figure 6
Incidence and mortality rates per 100,000 population:
all malignant cancers, 1973-2009

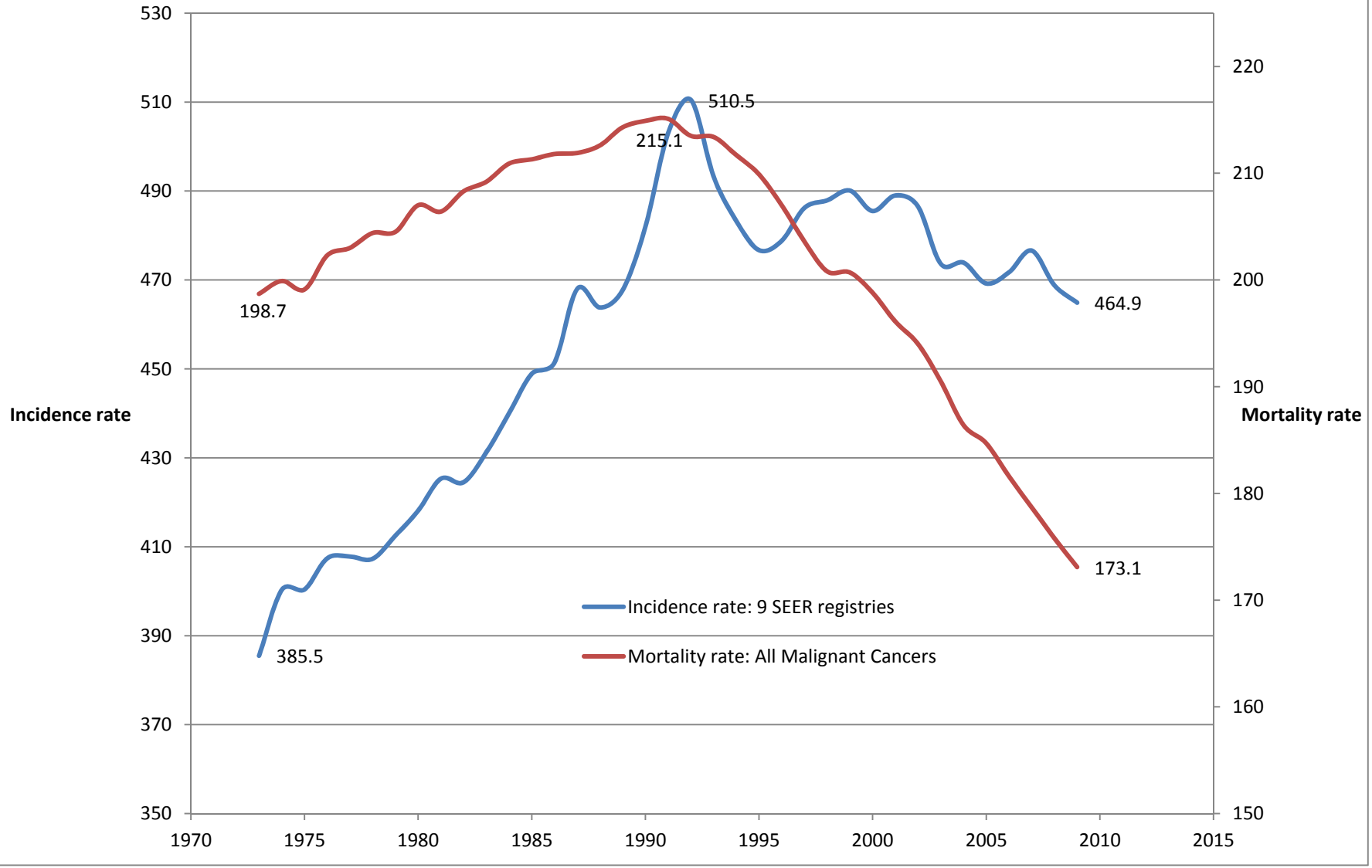


Figure 7

Estimates of $-\beta_{RESEARCH}$ in eq. (11) based on three alternative assumed values of the lag (k) from cumulative publications to the mortality rate

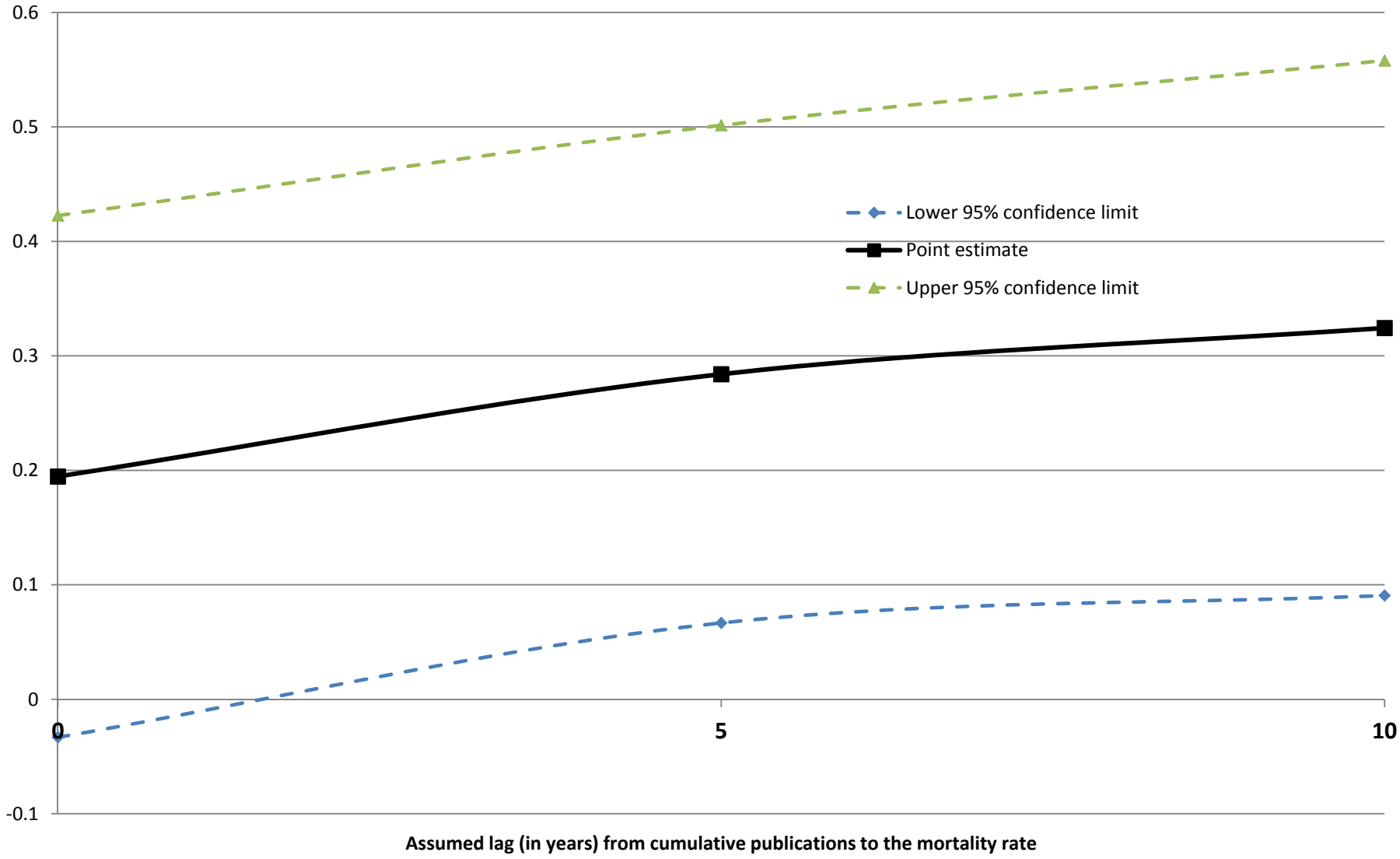
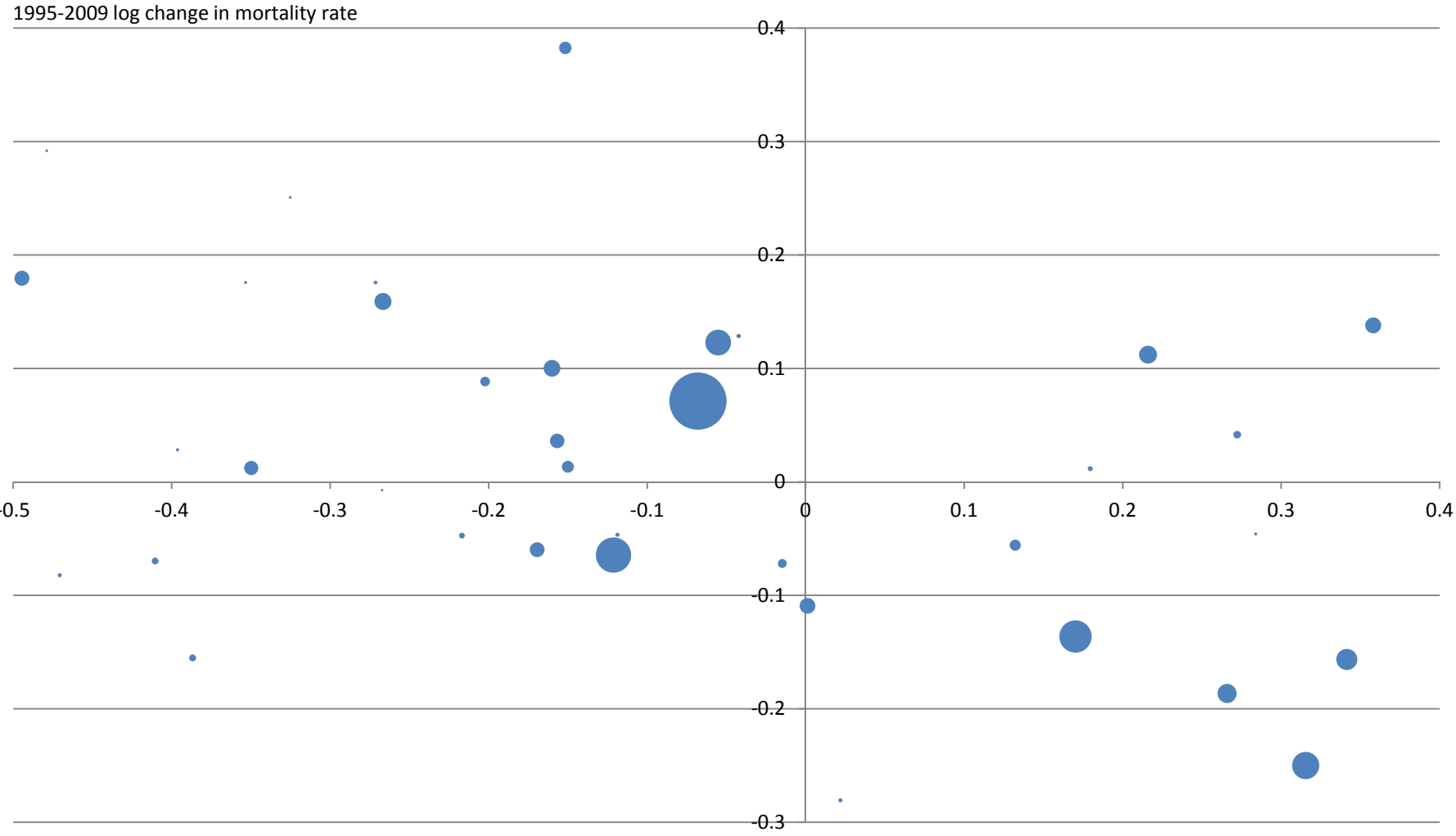


Figure 8

Partial correlation across cancer sites between 1985-1999 log change in number of research publications and 1995-2009 log change in mortality rate,

controlling for 1995-2009 log change in incidence rate



Note: bubble sizes are proportional to mean age-adjusted mortality rate during 1973-2009

Table 1

Mortality and incidence rates in 1985 and 2009 and PubMed publication counts ten years earlier (in 1985 and 1999), top 18 cancer sites (ranked by mean mortality rate)

Site	mean mort_rate	mean inc_rate	year	mort_rate	inc_rate	cum_pubs	cum_research_pubs	cum_non_research_pubs	cum_US_gov_research_pubs	cum_other_research_pubs
lung	52.8	62.9	1995	58.4	66.8	53,044	8,171	44,873	3,515	5,781
			2009	48.5	58.8	102,847	24,704	78,143	8,623	19,825
colonic	19.9	41.2	1995	19.4	39.7	23,420	5,757	17,663	2,771	4,030
			2009	12.9	30.5	41,498	13,776	27,722	5,414	10,770
breast	16.8	67.7	1995	17.4	72.8	56,987	11,766	45,221	4,941	8,554
			2009	12.4	69.8	138,938	46,391	92,547	18,331	36,643
prostatic	11.9	61.6	1995	13.5	72.2	18,053	3,677	14,376	1,826	2,381
			2009	8.6	69.4	58,195	21,043	37,152	10,400	15,453
pancreatic	10.7	11.7	1995	10.4	11.1	15,068	2,764	12,304	1,274	1,948
			2009	10.8	12.8	33,384	8,582	24,802	3,199	7,081
lymphoma non hodgkin	7.2	17.0	1995	8.7	19.9	24,535	4,979	19,556	2,165	3,789
			2009	6.3	20.2	53,953	14,397	39,556	4,830	12,061
stomach	5.9	9.6	1995	5.3	8.3	28,427	2,309	26,118	498	1,957
			2009	3.4	7.3	50,347	9,076	41,271	1,205	8,380
ovarian	5.2	8.2	1995	5.2	8.0	19,902	3,398	16,504	1,337	2,629
			2009	4.4	6.8	41,689	11,493	30,196	4,220	9,473
urinary bladder	4.7	20.8	1995	4.4	20.6	17,011	2,866	14,145	1,313	1,966
			2009	4.3	20.4	30,209	6,773	23,436	2,356	5,292
brain	4.5	6.5	1995	4.7	6.5	43,989	6,557	37,432	3,147	4,651
			2009	4.4	6.6	78,921	17,739	61,182	6,821	14,331
esophageal	4.1	4.5	1995	4.3	4.4	12,573	953	11,620	314	742
			2009	4.2	4.5	25,312	4,464	20,848	1,071	3,860
kidney	4.0	10.7	1995	4.3	11.1	20,046	2,597	17,449	1,184	1,869
			2009	3.9	14.9	37,923	6,775	31,148	2,332	5,456
rectal	3.6	16.1	1995	3.1	14.4	16,615	2,037	14,578	752	1,493
			2009	2.8	12.2	26,336	3,769	22,567	1,056	3,052
multiple myeloma	3.5	5.5	1995	4.0	5.7	10,753	1,727	9,026	681	1,247
			2009	3.3	6.1	19,888	5,449	14,439	1,801	4,569
skin	3.4	16.2	1995	3.5	18.1	34,051	5,061	28,990	2,330	3,581
			2009	3.7	24.6	67,141	13,234	53,907	4,930	10,464
liver	3.2	3.9	1995	3.5	3.7	36,499	8,241	28,258	3,824	5,678
			2009	4.5	7.1	73,272	20,603	52,669	5,951	16,988
leukemia myeloid acute	2.5	3.4	1995	2.4	3.7	14,685	5,363	9,322	2,299	4,167
			2009	2.9	3.6	25,170	10,049	15,121	3,507	8,492
leukemia lymphoid	2.3	6.4	1995	2.4	6.5	22,164	6,512	15,652	2,733	5,112
			2009	2.0	6.3	39,587	14,531	25,056	4,848	12,447

Table 2

Weighted means, standard deviations, and correlation coefficients across 47 cancer sites of 1995-2009 growth in mortality, incidence, and number of publications ten years earlier

	$\Delta\ln(mort_rate_s)$	$\Delta\ln(inc_rate_s)$	$\Delta\ln(cum_pubs_s)$	$\Delta\ln(cum_research_pubs_s)$	$\Delta\ln(cum_non_research_pubs_s)$	$\Delta\ln(cum_US_gov_research_pubs_s)$	$\Delta\ln(cum_other_research_pubs_s)$
Mean	-0.210	-0.053	0.813	1.538	0.694	1.081	1.957
Std. dev.	0.356	0.348	0.364	0.434	0.345	0.410	0.481
	Correlation coefficients						
$\Delta\ln(mort_rate_s)$	1.000	**0.631**	-0.119	**-0.348**	-0.032	**-0.411**	**-0.348**
$\Delta\ln(inc_rate_s)$		1.000	0.246	0.058	**0.304**	0.026	0.120
$\Delta\ln(cum_pubs_s)$			1.000	**0.667**	**0.981**	**0.741**	**0.680**
$\Delta\ln(cum_research_pubs_s)$				1.000	**0.647**	**0.908**	**0.939**
$\Delta\ln(cum_non_research_pubs_s)$					1.000	**0.690**	**0.643**
$\Delta\ln(cum_US_gov_research_pubs_s)$						1.000	**0.856**
$\Delta\ln(cum_other_research_pubs_s)$							1.000

Observations are weighted by mean mortality rate.
Correlation coefficients in bold are statistically significant (p-value < 0.05).

Table 3

Weighted least-squares estimates of models of 1995-2009 growth of the age-adjusted cancer mortality rate (eqs. 10-12)

Model	Publication lag (years)	Statistic	Δln(inc_rate$_s$)	Δln(cum_pubs$_s$)	Δln(cum_research_pubs$_s$)	Δln(cum_non_research_pubs$_s$)	Δln(cum_US_gov_research_pubs$_s$)	Δln(cum_other_research_pubs$_s$)	Intercept(δ')
1	0	Estimate	.	-0.249	.	.	.	.	-0.034
		T	.	-1.588	.	.	.	.	-0.295
		PVALUE	.	0.120	.	.	.	.	0.769
2	0	Estimate	**0.732**	**-0.426**	.	.	.	.	0.130
		T	**6.772**	**-3.801**	.	.	.	.	1.573
		PVALUE	**0.000**	**0.000**	.	.	.	.	0.124
3	0	Estimate	**0.653**	.	**-0.262**	.	.	.	0.120
		T	**6.102**	.	**-3.555**	.	.	.	1.404
		PVALUE	**0.000**	.	**0.001**	.	.	.	0.168
4	0	Estimate	0.698	.	-0.195	-0.172	.	.	0.149
		T	5.745	.	-1.726	-0.792	.	.	1.596
		PVALUE	0.000	.	0.092	0.433	.	.	0.118
5	0	Estimate	**0.721**	.	.	-0.217	0.024	-0.202	0.186
		T	**5.148**	.	.	-0.819	0.117	-0.937	1.110
		PVALUE	**0.000**	.	.	0.418	0.907	0.355	0.274
6	5	Estimate	.	-0.245	.	.	.	.	-0.029
		T	.	-1.619	.	.	.	.	-0.254
		PVALUE	.	0.113	.	.	.	.	0.801
7	5	Estimate	**0.738**	**-0.422**	.	.	.	.	0.140
		T	**6.869**	**-3.928**	.	.	.	.	1.697
		PVALUE	**0.000**	**0.000**	.	.	.	.	0.097
8	5	Estimate	**0.664**	.	**-0.300**	.	.	.	0.202
		T	**6.555**	.	**-4.366**	.	.	.	2.281
		PVALUE	**0.000**	.	**0.000**	.	.	.	0.028
9	5	Estimate	**0.673**	.	**-0.284**	-0.037	.	.	0.206
		T	**5.928**	.	**-2.641**	-0.191	.	.	2.239
		PVALUE	**0.000**	.	**0.012**	0.850	.	.	0.031
10	5	Estimate	**0.632**	.	.	0.074	-0.156	-0.157	0.156
		T	**4.781**	.	.	0.317	-0.823	-0.831	0.953
		PVALUE	**0.000**	.	.	0.753	0.416	0.411	0.346

Table 3

Weighted least-squares estimates of models of 1995-2009 growth of the age-adjusted cancer mortality rate (eqs. 10-12)

Model	Publication lag (years)	Statistic	Δln(inc_rate$_s$)	Δln(cum_pubs$_s$)	Δln(cum_research_pubs$_s$)	Δln(cum_non_research_pubs$_s$)	Δln(cum_US_gov_research_pubs$_s$)	Δln(cum_other_research_pubs$_s$)	Intercept(δ')
11	10	Estimate	.	-0.092	.	.	.	.	-0.134
		T	.	-0.587	.	.	.	.	-1.027
		PVALUE	.	0.561	.	.	.	.	0.310
12	10	Estimate	**0.718**	**-0.299**	**.**	**.**	.	.	0.071
		T	**5.967**	**-2.476**	**.**	**.**	.	.	0.701
		PVALUE	**0.000**	**0.018**	**.**	**.**	.	.	0.488
13	10	Estimate	**0.663**	.	**-0.319**	**.**	.	.	0.316
		T	**6.132**	.	**-3.578**	**.**	.	.	2.278
		PVALUE	**0.000**	.	**0.001**	**.**	.	.	0.028
14	10	Estimate	**0.660**	.	**-0.324**	0.011	.	.	0.316
		T	**5.551**	.	**-2.807**	0.067	.	.	2.248
		PVALUE	**0.000**	.	**0.008**	0.947	.	.	0.030
15	10	Estimate	**0.608**	.	.	0.189	-0.302	-0.162	0.336
		T	**5.090**	.	.	1.092	-1.546	-1.089	2.031
		PVALUE	**0.000**	.	.	0.282	0.130	0.283	0.049

Appendix Table 1
Mortality and incidence rates in 1985 and 2009 and PubMed publication counts ten years earlier (in 1985 and 1999), 29 cancer sites not included in Table 1 (ranked by mean mortality rate)

Site	mean mort_rate	mean inc_rate	year	mort_rate	inc_rate	cum_pubs	cum_research_pubs	cum_non_research_pubs	cum_US_gov_research_pubs	cum_other_research_pubs
uterine cervical	2.0	5.3	1995	1.8	4.6	22,427	2,696	19,731	951	2,033
			2009	1.2	3.5	40,186	8,743	31,443	2,440	7,209
laryngeal	1.5	4.5	1995	1.5	4.4	10,896	712	10,184	272	513
			2009	1.1	3.1	16,687	1,780	14,907	404	1,511
soft tissue	1.3	2.6	1995	1.5	2.8	5,568	851	4,717	365	581
			2009	1.3	3.3	14,062	1,967	12,095	626	1,552
gallbladder	1.0	1.4	1995	0.8	1.4	2,682	126	2,556	35	102
			2009	0.6	1.1	4,829	437	4,392	67	392
tongue	0.8	2.6	1995	0.7	2.5	3,066	172	2,894	60	124
			2009	0.6	3.3	4,998	641	4,357	120	563
hodgkin disease	0.7	2.9	1995	0.5	2.8	16,042	2,100	13,942	959	1,379
			2009	0.4	2.9	21,495	3,772	17,723	1,311	2,894
bile duct	0.7	0.5	1995	0.8	0.8	3,768	258	3,510	69	210
			2009	1.3	0.8	8,299	962	7,337	232	845
bone	0.5	0.9	1995	0.5	1.0	39,473	2,573	36,900	1,041	1,809
			2009	0.4	1.0	67,454	6,968	60,486	2,125	5,714
thyroid	0.5	6.8	1995	0.4	6.2	5,946	1,235	4,711	302	1,076
			2009	0.5	14.3	18,907	4,484	14,423	950	4,040
intestinal	0.4	1.5	1995	0.4	1.7	48,426	8,720	39,706	3,588	6,525
			2009	0.4	2.2	107,614	30,022	77,592	9,530	24,852
vulvar	0.3	1.3	1995	0.3	1.3	3,028	239	2,789	99	177
			2009	0.3	1.4	4,935	493	4,442	142	406
salivary	0.3	1.2	1995	0.3	1.2	5,633	467	5,166	121	370
			2009	0.2	1.3	9,873	1,155	8,718	217	1,014
nasopharyngeal	0.3	0.7	1995	0.3	0.8	4,381	628	3,753	232	461
			2009	0.2	0.6	8,289	2,167	6,122	348	1,946
tonsillar	0.3	1.3	1995	0.2	1.2	933	52	881	21	35
			2009	0.2	1.8	1,396	130	1,266	35	104
nose	0.2	0.7	1995	0.2	0.7	6,017	309	5,708	117	214
			2009	0.2	0.7	10,094	692	9,402	166	562
hypopharyngeal	0.2	1.0	1995	0.2	0.9	453	53	400	11	49
			2009	0.1	0.6	1,374	213	1,161	24	203
oropharyngeal	0.2	0.3	1995	0.2	0.3	1,488	141	1,347	43	113
			2009	0.2	0.3	3,254	540	2,714	132	459
pleural	0.2	0.0	1995	0.2	.	3,160	310	2,850	86	251
			2009	0.1	0.0	7,062	1,058	6,004	246	947

Site	mean mort_rate	mean inc_rate	year	mort_rate	inc_rate	cum_pubs	cum_research_pubs	cum_non_research_pubs	cum_US_gov_research_pubs	cum_other_research_pubs
testicular	0.2	2.4	1995	0.1	2.3	10,173	1,471	8,702	598	1,050
			2009	0.1	2.9	16,191	2,707	13,484	852	2,158
vaginal	0.2	0.4	1995	0.2	0.4	2,221	153	2,068	90	79
			2009	0.1	0.4	3,119	233	2,886	113	148
tracheal mediastinal	0.2	0.2	1995	0.1	0.2	7,340	371	6,969	215	213
			2009	0.1	0.2	10,607	604	10,003	261	420
peritoneal	0.2	0.4	1995	0.1	0.4	3,250	267	2,983	94	215
			2009	0.2	0.6	7,395	931	6,464	258	781
anus	0.1	1.2	1995	0.2	1.2	1,822	138	1,684	53	102
			2009	0.2	1.7	3,175	372	2,803	132	289
retroperitoneal	0.1	0.5	1995	0.1	0.4	3,436	128	3,308	55	89
			2009	0.1	0.4	5,445	255	5,190	85	198
eye orbital	0.1	0.8	1995	0.1	0.9	14,335	2,241	12,094	1,059	1,705
			2009	0.1	0.8	23,738	4,771	18,967	1,638	4,047
ureteral	0.1	0.6	1995	0.1	0.6	1,765	47	1,718	13	36
			2009	0.1	0.5	2,738	125	2,613	20	111
penile	0.1	0.4	1995	0.1	0.3	1,947	107	1,840	44	73
			2009	0.1	0.4	3,211	193	3,018	59	150
leukemia monocytic acute	0.1	0.2	1995	0.1	0.3	978	258	720	96	205
			2009	0.0	0.2	1,416	423	993	115	364
lip	0.0	1.5	1995	0.0	1.3	1,469	48	1,421	12	42
			2009	0.0	0.6	2,064	103	1,961	17	95