Pharmaceutical innovation and U.S. cancer survival in the 1990s: evidence from linked SEER-MEDSTAT data

Frank R. Lichtenberg

Columbia University
and
National Bureau of Economic Research

frank.lichtenberg@columbia.edu

13 April 2007

Pharmaceutical innovation and U.S. cancer survival in the 1990s: evidence from linked SEER-MEDSTAT data

Abstract

This study examines the impact of pharmaceutical innovation and other factors on the survival of U.S. cancer patients during the 1990s. In particular, it investigates whether cancer survival rates increased more for those cancer sites that had the largest increases in the proportion of drug treatments that were "new" treatments. We control for "expected survival," i.e. the survival of a comparable set of people that did not have cancer, thereby measuring the excess mortality that is associated with a cancer diagnosis. We also control for other types of medical innovation, i.e. innovation in surgical procedures, diagnostic radiology procedures, and radiation oncology procedures.

Data on observed and expected survival rates, the number of people diagnosed, mean age at diagnosis, and stage distribution are obtained from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) 1973-2003 Public-use Data. Estimates of rates of innovation in drugs and other treatment and diagnostic procedures were constructed from the MEDSTAT Marketscan database and other data sources.

We compute weighted least-squares estimates of 12 versions of a survival model, based on different survival intervals, functional forms, and sets of weights.

The drug vintage coefficient is positive and significant in almost every model. This indicates that the cancer sites whose drug vintage (measured by the share of post-1990 treatments) increased the most during the 1990s tended to have larger increases in observed survival rates, ceteris paribus. Estimates of the fraction of the 1992-1999 change in the observed survival rate that is attributable to the increased utilization of post-1990 drugs range from 12% to 121%. The estimated fraction is higher for shorter survival intervals, when observations are weighted by the number of MEDSTAT drug treatments, and for the logarithmic specification. The mean of the 12 estimates of the fraction of the 1992-1999 change in the observed survival rate that is attributable to the increased utilization of post-1990 drugs is 44%. Due to sampling and other measurement errors, these estimates may be conservative.

The coefficients on measures of other types of medical innovation (in radiation oncology, diagnostic radiology, and surgery innovation) are generally not significant. However these measures may be less reliable than the drug innovation measure: they are based upon the year in which the AMA established a new procedure code, which may be a far less meaningful indicator of innovation than the year in which the FDA first approved a drug. This topic warrants further research.

Frank R. Lichtenberg
Graduate School of Business
Columbia University
614 Uris Hall, 3022 Broadway
New York, NY 10027
Phone: (212) 854-4408
frank.lichtenberg@columbia.edu

Many clinical studies have compared the effects of newer and older drugs on cancer survival rates.[1] The findings of these studies have been mixed. Some studies have found that use of newer cancer drugs increased survival rates. For example, Richardson et al (2005) compared bortezomib (FDA approved May 2003) with high-dose dexamethasone (FDA approved October 1958) in patients with relapsed multiple myeloma who had received one to three previous therapies. They found that patients treated with bortezomib had a longer survival than patients treated with dexamethasone: the one-year survival rate was 80 percent among patients taking bortezomib and 66 percent among patients taking dexamethasone (P=0.003), and the hazard ratio for overall survival with bortezomib was 0.57 (P=0.001). Similarly, Kantarjian et al (2005) concluded that imatinib mesylate (FDA approved May 2000) improved survival compared with other therapies in patients with accelerated-phase chronic myelogenous leukemia.

Other studies have found that use of newer cancer drugs did not increase survival rates. For example, von der Maase et al (2005) compared long-term survival in patients with locally advanced or metastatic transitional cell carcinoma of the urothelium treated with cisplatin and either gemcitabine (FDA approved May 1996) or methotrexate/vinblastine/doxorubicin (all of which were approved before 1975). A total of 405 patients were randomly assigned: 203 to the gemcitabine/cisplatin arm and 202 to the methotrexate/vinblastine/doxorubicin/cisplatin arm. Overall survival was similar in both arms.

This paper will seek to determine the effect of pharmaceutical innovation—the use of newer drugs—*in general* on cancer survival rates. A reliable estimate of this effect can't be obtained by simply surveying previous clinical studies of specific drugs and cancer sites, for two reasons. First, there is considerable variation in the methodology and metrics used in these studies, rendering comparison and aggregation difficult.

[1] A PubMed search for ((“Survival Rate”) AND (“Antineoplastic Agents”)) AND (“Comparative Study”)) yields 387 items.

Second, these studies may not provide a complete or representative picture; there may be little or no published evidence about the survival impact of some drugs.[2]

We will investigate whether cancer survival rates increased more for those cancer sites that had the largest increases in the proportion of drug treatments that were "new" treatments. We will control for "expected survival," i.e. the survival of a comparable set of people that did not have cancer, thereby measuring the excess mortality that is associated with a cancer diagnosis. We will also control (imperfectly) for other types of medical innovation, i.e. innovation in surgical procedures, diagnostic radiology procedures, and radiation oncology procedures.

Section I of this paper sketches a simple theory of cancer survival. Section II presents an econometric specification based on this theory. Section III describes the construction of data used to estimate this model. Estimation issues are discussed in Section IV. Empirical results are presented in Section V. Section VI contains a summary and discussion.

I. A simple theory of cancer survival

We will use the following notation:

S = observed survival rate
E = expected survival rate
R = S / E = relative survival rate[3]
Q = treatment quality
P = disease progression at time of diagnosis
V = treatment vintage

We postulate the following simple theory of cancer survival:

$$R = S / E = f(Q, P) \quad (1)$$

where $f'_Q > 0$ and $f'_P < 0$, or, more generally,

$$S = f(E, Q, P) \quad (2)$$

where $f'_E > 0$, $f'_Q > 0$ and $f'_P < 0$.

---

[2] Johnson et al (2003) reported that only one-fourth of the oncology drug marketing applications approved by the FDA during the period January 1, 1990 to November 1, 2002 were based on direct evidence of survival benefits; 75% of approvals were based on surrogate end points (e.g. reduction in tumor size).
[3] Ederer et al (1961).

The observed survival rate is hypothesized to be an increasing function of expected survival and the quality of treatment, and a decreasing function of disease progression at time of diagnosis. Moreover, we hypothesize that treatment quality is an increasing function of treatment vintage:[4]

$$Q = f(V) \qquad (3)$$

where $f'_V > 0$. Substituting (3) into (2),

$$S = f(E, V, P) \qquad (4)$$

where $f'_E > 0$, $f'_V > 0$ and $f'_P < 0$. The observed survival rate is hypothesized to be an increasing function of expected survival and treatment vintage, and a decreasing function of disease progression at time of diagnosis.

Our primary objective is to estimate the effect of treatment vintage (V) on survival (S). Equation (4) indicates that if P is correlated with V, it is necessary to control adequately for P to obtain an unbiased estimate of the effect of treatment vintage on survival.

Measuring progression (or severity) of disease is often challenging in health economics. We will include five variables (or groups of variables) postulated to be indicators or determinants of the mean progression of disease:

(1) *Time dummies ("year effects")*: control for changes in mean disease progression that are invariant across cancer sites
(2) *Stage distribution of disease*: the fraction of patients diagnosed with in situ (stage 0), localized/regional (stages 1 and 2),[5] and distant (stage 4) cancer.[6]
(3) *Vintage of diagnostic radiology procedures*. Use of newer diagnostic radiology procedures may result in earlier detection, i.e. a reduction in P.
(4) *Number of people diagnosed and mean age at diagnosis*. Improvements in diagnostic technology are likely to lead to earlier detection. This would result in an increase in the number of people diagnosed and a reduction (or below-average increase) in their mean age.

[4] The vintage of a treatment is the year in which the treatment was first used. For example, the vintage of a drug is the year that the drug's active ingredient was first approved by the FDA.
[5] We combine stages 1 and 2 because these two stages are merged in the case of prostate cancer in SEER data.
[6] The omitted stage category is "unstaged" (SEER Historic Stage A 9). All lymphomas and leukemias are considered unstaged

II. Econometric specification

Based on this theory, we propose the following econometric model of observed survival:

$$f(S_{it}) = \beta_1 \text{drug\_new\%}_{it} + \beta_2 \text{rad\_onc\_new\%}_{it} + \beta_3 \text{rad\_diag\_new\%}_{it} + \beta_4 \text{surg\_new\%}_{it} + \beta_5 f(E_{it}) + \beta_6 \ln(N_{it}) + \beta_7 \text{age}_{it} + \beta_8 \text{in\_situ\%}_{it} + \beta_9 \text{loc\_reg\%}_{it} + \beta_{10} \text{distant\%}_{it} + \alpha_i + \delta_t + \varepsilon_{it} \quad (5)$$

S_{it} = the observed survival rate of people diagnosed with cancer originating at site i in year t. The observed survival rate is the probability of surviving all causes of death for a specified time interval. Observed survival does not consider cause of death, it simply looks at who is alive and who is not.

$\text{drug\_new\%}_{it}$ = % of drug treatments administered in year t associated with cancer originating at site i that used drugs approved by the FDA after 1990

$\text{rad\_onc\_new\%}_{it}$ = % of radiation oncology procedures performed in year t associated with cancer originating at site i whose CPT codes were established by the American Medical Association (AMA) after 1990

$\text{rad\_diag\_new\%}_{it}$ = % of diagnostic radiation procedures performed in year t associated with cancer originating at site i whose CPT codes were established by the AMA after 1990

$\text{surg\_new\%}_{it}$ = % of surgical procedures performed in year t associated with cancer originating at site i in year t whose CPT codes were established by the AMA after 1990

E_{it} = the expected survival rate of people diagnosed with cancer originating at site i in year t. The expected survival rate is the observed survival rate of a comparable (in terms of race, sex, and age) set of people who do not have cancer.

age_{it} = the mean age of people diagnosed with cancer originating at site i in year t

N_{it} = the number of people diagnosed with cancer originating at site i in year t

$\text{in\_situ\%}_{it}$ = the fraction of cancers originating at site i in year t that were diagnosed in situ (stage 0)

$\text{loc\_reg\%}_{it}$ = the fraction of cancers originating at site i in year t that were diagnosed as localized or regional (stage 1 or 2)

$\text{distant\%}_{it}$ = the fraction of cancers originating at site i in year t that were diagnosed as distant (stage 4)

α_i = fixed cancer-site effects

δ_t = fixed year effects

Due to the presence of fixed cancer-site effects and year effects, this is a difference-in-differences model. A positive and significant estimate of β_1 would signify that there were above-average increases in observed survival rates of cancer sites with above average increases in drug_new%, *ceteris paribus*.

Since the expected survival rate is based on the age- (as well as race- and sex-) distribution of a comparable set of people who do not have cancer, controlling for mean age as well as expected survival may be redundant.

III. Data construction

Survival data. Data on observed and expected survival rates, the number of people diagnosed, mean age at diagnosis, and stage distribution were obtained from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) 1973-2003 Public-use Data. I used data from SEER 9 registries, which are Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. In this data set, cases diagnosed from 1973 through 2003 are available for all registries except Seattle-Puget Sound (1974+) and Atlanta (1975+). The database contains one record for each of 3,260,176 tumors. However, the treatment innovation measures can only be constructed for the period 1992-2003.

Cancer cases were classified using the SEER Cancer Causes of Death Recode 1969+ (9/17/2004).[7] This classification includes 68 non-overlapping types of cancer.

Survival rates may be calculated for a variety of time intervals (1-year, 2-year, etc.). We will estimate models of 1-year, 2-year, and 3-year survival rates. (The longer the time interval, the shorter the available time series.) Table 1 shows 1992 and 2000 2-year survival data for the top 30 (ranked by 1992 incidence) cancer sites, or groups of cancer sites. The 2-year relative survival rate for all cancer sites combined increased from 72% in 1992 to 75% in 2000. The 1992-2002 change in survival rates varied considerably across cancer sites. For example, the relative survival rate for Cervix Uteri declined from 86% to 81%, while the relative survival rate for Skin excluding Basal and Squamous increased from 83% to 97%. The relative survival rate for Leukemia declined

[7] http://seer.cancer.gov/codrecode/1969+_d09172004/index.html

from 59% to 57%, while the relative survival rate for Non-Hodgkin Lymphoma increased from 63% to 72%.

Treatment vintage data. First I will describe the construction of the drug innovation measure ($drug\_new\%_{it}$). A similar approach was used to construct the other innovation measures. The drug innovation measure was defined as follows:

$$drug\_new\%_{it} = \sum_p FREQ_{pit}\, POST1990_p / \sum_p FREQ_{pit} \qquad (6)$$

where

$FREQ_{pit}$ = the number of times drug p was used to treat cancer originating at site i in year t (t = 1992, 1993,…, 2000)

$POST1990_p$ = 1 if drug p was approved by the FDA after 1990
= 0 otherwise

Data on utilization of medical procedures and products, by diagnosis and year ($FREQ_{pit}$), were obtained from the MEDSTAT Marketscan database. MEDSTAT contains data on outpatient and inpatient services (procedures) and outpatient prescriptions of hundreds of thousands, or even millions, of individuals.

It is worth distinguishing between two types of drugs: self-administered drugs, and drugs administered by physicians and other medical providers (e.g., chemotherapy). Utilization of self-administered drugs is reported in outpatient prescription records (claims). These records generally don't include any information about the patient's diagnosis. In contrast, drugs that are administered by physicians and other medical providers are reported as outpatient and inpatient services (procedures). These records include information about the patient's diagnosis.

For most diseases other than cancer, the vast majority of drugs are self-administered, and determining the diagnosis associated with a particular drug's use (hence measuring $FREQ_{pit}$) can be difficult. But an important fraction of drug treatments for cancer are administered by physicians and other medical providers. We will use data on provider-administered drugs only, since the number of times provider-administered drug p was used to treat cancer originating at site i in year t can be measured precisely.

Each MEDSTAT outpatient and inpatient service record contains one procedure code and one or more ICD-9 diagnosis codes. Codes for drugs administered by providers

are Healthcare Common Procedure Coding System (HCPCS) Level II Codes.[8] Table 2 shows the top 40 (ranked by frequency) provider-administered drugs associated with all cancer diagnoses in 2003.

Only about a third of the drug treatments administered to cancer patients involve cancer drugs (antineoplastic agents). We will report estimates of two versions of eq. (5): one does not distinguish between cancer drugs and other drugs, and the other does. Table 3 shows a comparison of the top 40 provider-administered drugs associated with two major cancer sites (colon and breast) in 2003.

We used Multum's Lexicon database (http://www.multum.com/Lexicon.htm) to determine the active ingredients of the drugs corresponding to each of these HCPCS Level II Codes. We used data from the Drugs@FDA database (http://www.fda.gov/cder/drugsatfda/datafiles/default.htm) to determine the year in which each active ingredient was first approved by the FDA.

The following table shows the mean value of drug_new%, and the number of provider-administered drugs upon which that statistic is based, for all cancer sites combined during the period 1992-2003.

Year	drug_new%	Number of provider-administered drugs	Number of firms covered by MEDSTAT data
1992	9%	17,731	45
1993	12%	20,134	45
1994	18%	22,516	45
1995	17%	29,798	45
1996	18%	55,190	92
1997	20%	62,235	92
1998	25%	101,221	92
1999	29%	187,838	95
2000	32%	216,476	98
2001	33%	230,066	103

[8] Level II of the HCPCS is a standardized coding system that is used primarily to identify products, supplies, and services not included in the CPT codes, such as ambulance services and durable medical equipment, prosthetics, orthotics, and supplies (DMEPOS) when used outside a physician's office. Because Medicare and other insurers cover a variety of services, supplies, and equipment that are not identified by CPT codes, the level II HCPCS codes were established for submitting claims for these items. The development and use of level II of the HCPCS began in the 1980's. Level II codes are also referred to as alpha-numeric codes because they consist of a single alphabetical letter followed by 4 numeric digits, while CPT codes are identified using 5 numeric digits. See http://www.cms.hhs.gov/medicare/hcpcs/codpayproc.asp

2002	32%	328,234	200
2003	33%	486,409	200

The fraction of post-1990 drugs increased from 9% in 1992 to 33% in 2003. The 1992 figure is based on 17,731 observations (drug treatments), and the 2003 figure is based on 486,409 treatments. The increase in sample size is partly due to the fact that the number of firms covered by the MEDSTAT data increased from 45 in 1992 to 200 in 2003.

A similar procedure was used to construct the radiology and surgery innovation measures (rad_onc_new%$_{it}$, rad_diag_new%$_{it}$, and surg_new%$_{it}$). However, unlike drugs, radiology and surgical procedures are not subject to FDA approval,[9] so FDA approval dates can't be used to measure the vintage of these procedures.

Radiology and surgical procedures are coded using Current Procedural Terminology (CPT) codes that are established and maintained by the American Medical Association.[10] The AMA publishes a database (*CPT Assistant Archives 1990-2003)* that provides information about the year in which each CPT code was first established. The data are left-censored: if a code was established prior to 1990, we know only that it is a pre-1990 code. To construct the radiology and surgery innovation measures, we defined POST1990$_p$ as follows:

POST1990$_p$ = 1 if the CPT code for procedure p was established by the AMA after 1990
= 0 otherwise

The radiology and surgery innovation measures are probably less reliable than the drug innovation measure, because FDA approval of a drug is more meaningful indicator than AMA establishment of a new CPT code. For example, measuring surgical innovation using CPT code changes may be problematic. Closer inspection of the data on surgical procedures reveals that some "new" procedures are probably just relabeled or reclassified old procedures, rather than true innovations. For example, the three procedures whose codes were added in 1997 which were most frequently performed in

[9] Some new procedures may be closely related to medical device innovations, which are subject to FDA approval, but linking procedure innovations to FDA approvals of medical devices is difficult.
[10] For a description of how CPT codes are maintained, the committees involved, and the entire CPT process, including the evolution of CPT, see http://www.ama-assn.org/ama/pub/category/3112.html.

1997 were 98940, 98941, and 98942, which correspond to different types of chiropractic manipulative treatment of the spine. Undoubtedly, this type of treatment was performed well before 1997. A new CPT code should therefore be considered a necessary condition for a medical innovation, but not a sufficient condition: all innovations have new CPT codes, but some new CPT codes are not innovations. The fraction of procedures with new CPT codes exceeds the fraction of truly innovative procedures, perhaps by a significant amount, and the degree of overstatement varies across diseases. In the future, I hope to develop improved measures of radiology and surgery innovation.

Table 4 presents data on innovation measures in 2003, by cancer site, ranked by number of drug treatments.

IV. Estimation issues

Several issues regarding the estimation of eq. (5) should be considered before we present the empirical results. These issues are: (1) functional form; (2) measurement error; (3) weighting; and (4) estimation by cancer stage vs. overall estimation.

Functional form. The dependent variable of eq. (5) is specified to be an arbitrary function of the observed survival rate, f(S). Because the survival rate is bounded between zero and one, a linear function (e.g., f(S) = S) would not be an appropriate choice. We will estimate the model using two alternative functional forms. The first is the probit, i.e. $f(S) = F^{-1}(S)$, where $F^{-1}( )$ is the inverse of the standard normal cumulative distribution. The second is the logarithmic, i.e. f(S) = ln(S).[11] As shown in eq. (5), the same transformation is applied to the expected survival rate.

Measurement error. As described above, the survival data and the treatment vintage data were obtained from different data sources and are based on different populations. The survival data were obtained from SEER 9 public-use data, which primarily covers elderly people in certain regions of the U.S.[12] The treatment vintage data were obtained from the MEDSTAT Marketscan database, which primarily covers nonelderly people in other regions of the U.S.[13] As rich as the MEDSTAT database is, it provides data on only a

[11] $F^{-1}(S)$ is similar to ln(S / (1 – S)).
[12] About two-thirds of cancer patients are 65 or over.
[13] MEDSTAT has data on some patients in Medicare health plans, but these data were not available for this study.

small proportion of all cancer treatments provided in the U.S. This may be illustrated with utilization data for a specific drug treatment, an ondansetron HCL injection (HCPCS code J2405). In 2003 MEDSTAT data, this procedure was performed in connection with a cancer diagnosis 11,845 times. According to Medicare Part B Physician/Supplier National Data (http://www.cms.hhs.gov/MedicareFeeforSvcPartsAB/Downloads/LEVEL2SERV03.pdf) there were 6,381,294 allowed services of ondansetron HCL injection in 2003. The MEDSTAT frequency is only 0.2% of the Medicare frequency.

We assume that treatment innovation indicators based on MEDSTAT data are useful, albeit noisy, indicators of the treatment innovation experienced by patients in SEER 9 registries. This sampling error is likely to bias the coefficients on the treatment innovation measures towards zero.

Weighting. Eq. (5) is to be estimated using grouped data, where groups are defined by cancer site and year of diagnosis. These groups are very heterogeneous in terms of size, where size is measured either by number of SEER 9 patients or number of MEDSTAT treatments. For example, as shown in Table 4, in 2003 there were 140,122 drug treatments for breast cancer, and only 32 for cancer of the eye and orbit.

We will estimate eq. (5) via weighted least-squares (WLS), where the weight is a measure of size. Consider two different measures of size: the number of SEER 9 patients, and the number of MEDSTAT drug treatments. In a given year, the correlation across cancer sites between these two measures is quite high. For example, in 2003 the correlation between the number of SEER 9 patients and the number of MEDSTAT drug treatments is about .85. This suggests that the choice between these two weights wouldn't make much difference. As noted above, however, the MEDSTAT sample size increased dramatically over time, suggesting that the more recent innovation measures are far more reliable, and therefore deserve much greater weight.[14]

We will estimate eq. (5) with two different sets of weights--the number of SEER 9 patients, and the number of MEDSTAT drug treatments. We believe that the estimates based on the latter set of weights are more credible.

[14] As shown in Table 1, the SEER 9 sample size barely increased over time.

Estimation: overall vs. by stage. The SEER microdata contain information about the stage of cancer at time of diagnosis. Thus, it is feasible to calculate observed and expected survival rates by cancer site, year, *and stage*. However, we believe that our approach (controlling for stage distribution rather than analysis by stage) is preferable, for two reasons. First, there is no information about cancer stage in MEDSTAT. Hence, we can't construct stage-specific treatment innovation measures.

Even if we could construct such measures, due to a phenomenon known as *stage migration*, analysis by cancer stage is probably inappropriate. Changes in stage-specific survival may provide a distorted view of true survival change. In particular, the survival rate for every stage may improve even when overall survival does not change.

The assignment of a given stage to a particular cancer may change over time due to advances in diagnostic technology. Stage migration occurs when diagnostic procedures change over time, resulting in an increase in the probability that a given cancer will be diagnosed in a *more advanced* stage. For example, certain distant metastases that would have been undetectable a few years ago can now be diagnosed by a computer tomography (CT) scan or by magnetic resonance imaging (MRI). Therefore, some patients who would have been diagnosed previously as having cancer in a *localized* or *regional* stage are now diagnosed as having cancer in a *distant* stage. The likely result would be to remove the worst survivors — those with previously undetected distant metastases — from the localized and regional categories and put them into the distant category. As a result, the stage-at-diagnosis distribution for a cancer may become less favorable over time, but the survival rates for each stage may improve: the early stage will *lose* cases that will survive *shorter* than those remaining in that category, while the advanced stage will *gain* cases that will survive *longer* than those already in that category. However, *overall survival would not change* (Feinstein et al., 1985). Stage migration is an important concept to understand when examining temporal trends in survival by stage at diagnosis as well as temporal trends in stage distributions; it could affect the analysis of virtually all solid tumors.[15]

[15] *SEER Cancer Statistics Review 1973-1999 Overview*, p. 12.

V. Empirical results

We computed weighted least-squares estimates of 12 versions of eq. (5): three survival intervals (1-year, 2-year, and 3-year survival), two functional forms (probit and logarithmic), and two sets of weights (the number of MEDSTAT drug treatments, and the number of SEER 9 patients diagnosed); 3 * 2 * 2 = 12. Table 5 shows estimates of eq. (5) for all three survival intervals, based on the probit functional form and weighting by the number of MEDSTAT drug treatments.

For all three survival intervals, the coefficient on drug_new% is positive and significant (p-value < .03). This indicates that the cancer sites whose drug vintage (measured by the share of post-1990 treatments) increased the most during the 1990s tended to have larger increases in observed survival rates, ceteris paribus.

None of the coefficients on the diagnostic radiology innovation measure or the surgery innovation measure are significant. The coefficient on the radiology oncology innovation measure is positive and significant in the 2-year survival equation, but not in the other two equations.

The coefficients on the expected survival variable $F^{-1}(E)$ are all positive and significant. This indicates that part of the increase in the observed survival rate of cancer patients can be attributed to factors that also increased the survival of people who did not have cancer.

The coefficients on the log of the number of SEER 9 patients diagnosed are also all positive and significant. This may be capturing the impact of improved (earlier) detection: one would expect above-average increases in the number of people diagnosed for cancer sites with the greatest improvements in detection.

The age coefficient is negative and significant in the 1-year survival equation but insignificant in the other two equations. However, as noted above, the expected survival rate is based on the age- (as well as race- and sex-) distribution of a comparable set of people who do not have cancer, so controlling for mean age as well as expected survival may be redundant.

Now let's consider the stage distribution coefficients. Since survival is inversely related to disease progression, one might expect the in_situ% coefficient to exceed the loc_reg% coefficient, and the loc_reg% coefficient to exceed the distant% coefficient.

This is the case in the 2-year and 3-year survival equations, but not in the 1-year equation. Due to stage migration, however, this ordering is not necessarily to be expected. Suppose that there was no change in the true stage distribution of any cancer site, but that some cancer sites had improved detection. These cancer sites would have the largest increase in distant%, and might also have above-average increases in survival.

The estimates shown in Table 5 do not distinguish between cancer drugs and other drugs. Table 6 shows the effect of distinguishing between these two types of drugs. Line 1 of Table 6 shows the estimates of the drug coefficients when cancer drugs and other drugs are pooled. Lines 2 and 3 show the estimates of the drug coefficients when cancer drugs and other drugs are disaggregated. In the 1-year and 2-year survival models, the coefficient on cancer_drug_new% is positive and highly significant, whereas the coefficient on other_drug_new% is insignificant. This suggests that the gains in cancer survival are primarily attributable to cancer drugs as opposed to other drugs. In the 3-year survival model, neither of the coefficients in lines 2 and 3 are significant, although the coefficient on drug_new% (the utilization-weighted average of cancer_drug_new% and other_drug_new%) is positive and significant. This may be attributable to the fact that the 3-year estimates are based on 47% fewer observations (drug treatments) than the 1-year estimates and 24% fewer than the 2-year estimates. In the remainder of the paper we will consider models in which cancer drugs and other drugs are pooled.

As noted above, we also estimated models using an alternative (logarithmic) functional form, and an alternative set of weights (the number of SEER 9 patients diagnosed). To conserve space, we will not present the full estimates of these other nine models. But to enable assessment of the robustness of the estimates, we present in Table 7 the t-statistics (indicating both the sign and statistical significance) on the treatment innovation measures, $F^{-1}(E)$, ln(N), and age from all 12 models.

The drug vintage coefficient is positive and significant in 11 of the 12 models, and positive and marginally significant (p-value = .07) in the other model. The only other variable whose coefficient is generally significant with a consistent sign is the log of the number of SEER 9 patients diagnosed; this coefficient is positive and significant in 9 of the 12 models. The coefficient on the expected survival term is positive and

significant in only 4 models. The radiation oncology, diagnostic radiology, and surgery innovation coefficients are positive and significant in 2, 1, and 0 models, respectively.

From our estimates, we can calculate the fraction of the 1992-1999 change in the observed survival rate that is attributable to the increased utilization of post-1990 drugs, ceteris paribus. The fraction is equal to

$$(\beta_1 * (\text{drug\_new\%}_{1999} - \text{drug\_new\%}_{1992})) / (f(S_{1999}) - f(S_{1992}))$$

As noted above, the overall value of drug_new% increased from 9% in 1992 to 29% in 1999, so the above expression reduces to $(\beta_1 * 20\%) / (f(S_{1999}) - f(S_{1992}))$. The following table shows the observed survival rates for all cancer sites combined in the years 1992 and 1999:

Survival interval	1-year	2-year	3-year
S_{1992}	76.4%	67.1%	61.6%
S_{1999}	78.1%	69.9%	65.0%
$S_{1999} - S_{1992}$	1.7%	2.8%	3.4%

The following table shows estimates from each of the 12 models of the fraction of the 1992-1999 change in the observed survival rate that is attributable to the increased utilization of post-1990 drugs:

Functional form	probit	probit	log	log
Weight	No. of MEDSTAT drug treatments	No. of SEER 9 patients diagnosed	No. of MEDSTAT drug treatments	No. of SEER 9 patients diagnosed
1-year	39%	26%	121%	77%
2-year	37%	14%	68%	40%
3-year	20%	12%	41%	29%

Estimates of the fraction of the 1992-1999 change in the observed survival rate that is attributable to the increased utilization of post-1990 drugs range from 12% to 121%. The estimated fraction is higher for shorter survival intervals, when observations are weighted by the number of MEDSTAT drug treatments, and for the logarithmic specification. The

mean of the 12 estimates of the fraction of the 1992-1999 change in the observed survival rate that is attributable to the increased utilization of post-1990 drugs is 44%.

VI. Summary and discussion

Previous studies (Brenner (2002)) have shown that long-term survival rates for many types of cancer have substantially improved in past decades because of advances in early detection and treatment. This study has examined the impact of pharmaceutical innovation and other factors on the survival of U.S. cancer patients during the 1990s. In particular, it investigated whether cancer survival rates increased more for those cancer sites that had the largest increases in the proportion of drug treatments that were "new" treatments. By controlling for "expected survival," i.e. the survival of a comparable set of people that did not have cancer, we measured the excess mortality that is associated with a cancer diagnosis. We also controlled for other types of medical innovation, i.e. innovation in surgical procedures, diagnostic radiology procedures, and radiation oncology procedures.

Data on observed and expected survival rates, the number of people diagnosed, mean age at diagnosis, and stage distribution were obtained from the National Cancer Institute's SEER public-use data. Estimates of rates of innovation in drugs and other treatment and diagnostic procedures were constructed from the MEDSTAT Marketscan database and other data sources. Treatment innovation indicators based on MEDSTAT data are likely to be useful, albeit noisy, indicators of the treatment innovation experienced by patients in SEER registries. This sampling error is likely to bias the coefficients on the treatment innovation measures towards zero.

We computed weighted least-squares estimates of 12 versions of a survival model, based on different survival intervals, functional forms, and sets of weights. The drug vintage coefficient was positive and significant in almost every model. This indicates that the cancer sites whose drug vintage (measured by the share of post-1990 treatments) increased the most during the 1990s tended to have larger increases in observed survival rates, ceteris paribus.

Estimates of the fraction of the 1992-1999 change in the observed survival rate that is attributable to the increased utilization of post-1990 drugs ranged from 12% to

121%. The estimated fraction is higher for shorter survival intervals, when observations are weighted by the number of MEDSTAT drug treatments, and for the logarithmic specification. The mean of the 12 estimates of the fraction of the 1992-1999 change in the observed survival rate that is attributable to the increased utilization of post-1990 drugs is 44%. Due to sampling and other measurement errors, these estimates may be conservative.

The coefficients on measures of other types of medical innovation (in radiation oncology, diagnostic radiology, and surgery innovation) were generally not significant. However these measures may be less reliable than the drug innovation measure: they were based upon the year in which the AMA established a new procedure code, which may be a far less meaningful indicator of innovation than the year in which the FDA first approved a drug. This topic warrants further research.

References

Axtell LM, Cutler SJ. (1961), "The relative survival rate: A statistical methodology," *J Natl Cancer Inst Monogr* 6, pp. 101-121.

Brenner H. (2002), "Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis," *Lancet* Oct 12; 360(9340):1131-5.

Feinstein AR, Sosin DM, Wells CK (1985), "The Will Rogers phenomenon: Stage migration and new diagnostic techniques as a source of misleading statistics for survival of cancer," *New England Journal of Medicine* 312, pp.1604-1608.

Johnson, John, Grant Williams, and Richard Pazdur, "End points and United States Food and Drug Administration Approval of Oncology Drugs," *Journal of Clinical Oncology* 21 (7), 1 April 2003, pp. 1404-11.

Kantarjian H, Talpaz M, O'Brien S, Giles F, Faderl S, Verstovsek S, Garcia-Manero G, Shan J, Rios MB, Champlin R, de Lima M, Cortes J. (2005), "Survival benefit with imatinib mesylate therapy in patients with accelerated-phase chronic myelogenous leukemia--comparison with historic experience." *Cancer*. May 15; 103(10): 2099-108.

Mosby (2004), Mosby's Drug Consult 2004, 14th Edition, http://www.mosbysdrugconsult.com/

Richardson PG, et al (2005), "Bortezomib or high-dose dexamethasone for relapsed multiple myeloma." *N Engl J Med.* Jun 16;352(24):2487-98.

von der Maase H, Sengelov L, Roberts JT, Ricci S, Dogliotti L, Oliver T, Moore MJ, Zimmermann A, Arning M (2005), “Long-term survival results of a randomized trial comparing gemcitabine plus cisplatin, with methotrexate, vinblastine, doxorubicin, plus cisplatin in patients with bladder cancer,” *J Clin Oncol.* Jul 20; 23(21): 4602-8.

Table 1
1992 and 2000 survival data, top 30 (ranked by 1992 incidence) cancer sites

Site	Number diagnosed		Two-year survival rates					
			Observed		Expected		Relative	
	1992	2000	1992	2000	1992	2000	1992	2000
Prostate	19832	17449	89%	92%	90%	92%	99%	100%
Breast	15660	19450	91%	93%	96%	96%	95%	97%
Lung and Bronchus	12710	12455	23%	25%	93%	93%	25%	27%
Colon excluding Rectum	7935	8020	68%	68%	91%	91%	74%	75%
Lymphoma	4231	4785	64%	70%	95%	94%	67%	75%
Melanoma of the Skin	3901	6199	92%	94%	96%	96%	96%	99%
Urinary Bladder	3856	4082	78%	79%	91%	91%	86%	88%
Non-Hodgkin Lymphoma	3559	4095	59%	67%	94%	94%	63%	72%
Rectum and Rectosigmoid Junction	3160	3275	71%	74%	93%	93%	77%	80%
Sigmoid Colon	2809	2459	75%	76%	92%	92%	81%	82%
Corpus Uteri	2727	2889	87%	87%	95%	96%	91%	91%
Leukemia	2501	2661	54%	53%	93%	93%	59%	57%
NHL - Nodal	2394	2725	59%	65%	94%	94%	62%	70%
Rectum	2094	2324	71%	74%	93%	93%	77%	80%
Pancreas	2078	2267	8%	10%	92%	92%	9%	11%
Kidney and Renal Pelvis	1948	2460	66%	70%	94%	94%	70%	75%
Ovary	1860	1962	64%	67%	96%	96%	67%	70%
Cecum	1753	1839	63%	65%	90%	90%	69%	73%
Cervix Uteri	1731	943	85%	79%	98%	98%	86%	81%
Stomach	1693	1690	30%	31%	92%	91%	33%	34%
Brain and Other Nervous System	1464	1532	38%	40%	96%	96%	40%	41%
Brain	1376	1410	36%	37%	96%	96%	38%	38%
Thyroid	1263	1779	93%	96%	98%	98%	95%	97%
Lymphocytic Leukemia	1253	1209	77%	77%	93%	93%	83%	82%
Other Non-Epithelial Skin	1199	537	42%	79%	98%	94%	43%	84%
NHL - Extranodal	1165	1370	60%	70%	94%	94%	64%	75%
Myeloma	1117	1205	52%	55%	92%	92%	57%	59%
Myeloid and Monocytic Leukemia	1067	1293	32%	34%	93%	93%	35%	36%
Rectosigmoid Junction	1066	951	70%	75%	93%	93%	76%	80%
Ascending Colon	1043	1397	67%	68%	91%	90%	74%	76%

Table 2

Top 40 (ranked by frequency) provider-administered drugs associated with cancer diagnoses in 2003

Drug	COUNT	PERCENT
J1100-Dexamethasone Sodium Phos	47702	9.8
J1642-Inj Heparin Sodium Per 10 U	35259	7.2
J1260-Dolasetron Mesylate	27981	5.8
J9190-Fluorouracil Injection	25578	5.3
J7051-Sterile Saline/Water	21459	4.4
J7040-Normal Saline Solution Infus	20002	4.1
J1626-Granisetron HCl Injection	17075	3.5
J0640-Leucovorin Calcium Injection	16925	3.5
J1200-Diphenhydramine HCl Injectio	16876	3.5
J1644-Inj Heparin Sodium Per 1000u	16481	3.4
J2405-Ondansetron HCl Injection	11845	2.4
J9265-Paclitaxel Injection	11587	2.4
J9000-Doxorubic HCl 10 Mg Vl Chemo	10445	2.1
J9045-Carboplatin Injection	10082	2.1
J9170-Docetaxel	9982	2.1
J2912-Sodium Chloride Injection	9729	2.0
J9355-Trastuzumab	9175	1.9
J7030-Normal Saline Solution Infus	8756	1.8
J3487-Zoledronic Acid	8422	1.7
J9201-Gemcitabine HCl	8118	1.7
J0880-Darbepoetin Alfa Injection	7407	1.5
J2060-Lorazepam Injection	6136	1.3
J9206-Irinotecan Injection	6071	1.2
J1441-Filgrastim 480 Mcg Injection	5886	1.2
J9310-Rituximab Cancer Treatment	5275	1.1
J9390-Vinorelbine Tartrate/10 Mg	4998	1.0
J9214-Interferon Alfa-2b Inj	4472	0.9
J9093-Cyclophosphamide Lyophilized	3997	0.8
J2430-Pamidronate Disodium /30 Mg	3724	0.8
J9096-Cyclophosphamide Lyophilized	3373	0.7
J9060-Cisplatin 10 Mg Injection	3340	0.7
J9217-Leuprolide Acetate Suspnsion	3303	0.7
J9182-Etoposide 100 Mg Inj	3288	0.7
J9185-Fludarabine Phosphate Inj	3275	0.7
J1440-Filgrastim 300 Mcg Injection	3144	0.6
J2780-Ranitidine Hydrochloride Inj	3032	0.6
J9181-Etoposide 10 Mg Inj	2891	0.6
J3475-Inj Magnesium Sulfate	2708	0.6
J3480-Inj Potassium Chloride	2503	0.5
J2820-Sargramostim Injection	2490	0.5

Table 3

Comparison of top 20 (ranked by frequency) provider-administered drugs associated with two major cancer sites in 2003

Colon excluding Rectum

Drug	COUNT	PERCENT
J9190-Fluorouracil Injection	12222	22.6
J0640-Leucovorin Calcium Injection	10460	19.3
J1100-Dexamethasone Sodium Phos	4819	8.9
J1642-Inj Heparin Sodium Per 10 U	3885	7.2
J1260-Dolasetron Mesylate	3224	6.0
J9206-Irinotecan Injection	2975	5.5
J7051-Sterile Saline/Water	2501	4.6
J1626-Granisetron HCl Injection	2058	3.8
J7040-Normal Saline Solution Infus	1825	3.4
J1644-Inj Heparin Sodium Per 1000u	1766	3.3
J2405-Ondansetron HCl Injection	1287	2.4
J0460-Atropine Sulfate Injection	1250	2.3
J2912-Sodium Chloride Injection	1050	1.9
J2060-Lorazepam Injection	736	1.4
J1200-Diphenhydramine HCl Injectio	518	1.0
J0880-Darbepoetin Alfa Injection	500	0.9
J7030-Normal Saline Solution Infus	256	0.5
J7042-5% Dextrose/Normal Saline	242	0.4
J1441-Filgrastim 480 Mcg Injection	235	0.4
J7070-D5w Infusion	173	0.3

Breast

Drug	COUNT	PERCENT
J1100-Dexamethasone Sodium Phos	15110	10.8
J1642-Inj Heparin Sodium Per 10 U	11852	8.5
J9355-Trastuzumab	8763	6.3
J1260-Dolasetron Mesylate	7957	5.7
J7051-Sterile Saline/Water	6717	4.8
J9000-Doxorubic HCl 10 Mg Vl Chemo	6607	4.7
J1644-Inj Heparin Sodium Per 1000u	6317	4.5
J9170-Docetaxel	5649	4.0
J7040-Normal Saline Solution Infus	5576	4.0
J1200-Diphenhydramine HCl Injectio	5075	3.6
J1626-Granisetron HCl Injection	5037	3.6
J9265-Paclitaxel Injection	4067	2.9
J9190-Fluorouracil Injection	3548	2.5
J2405-Ondansetron HCl Injection	3449	2.5
J2912-Sodium Chloride Injection	3349	2.4
J9390-Vinorelbine Tartrate/10 Mg	3314	2.4
J3487-Zoledronic Acid	3186	2.3
J9093-Cyclophosphamide Lyophilized	2814	2.0
J9096-Cyclophosphamide Lyophilized	2562	1.8
J0880-Darbepoetin Alfa Injection	2447	1.7

Table 4
Innovation measures in 2003, by cancer site, ranked by number of drug treatments

	Post-1990 procedures/total procedures				Number of procedures			
Cancer site	drugs	radiation oncology	diagnostic radiology	surgery	drugs	radiation oncology	diagnostic radiology	surgery
Breast	37%	58%	7%	20%	140122	122050	40188	74214
Lung and Bronchus	40%	53%	5%	16%	62313	31442	20300	18635
Colon excluding Rectum	20%	57%	2%	17%	54159	2716	8656	17067
Non-Hodgkin Lymphoma	33%	53%	3%	22%	37597	6649	20290	16863
Ovary	41%	55%	2%	16%	25487	1135	5405	9300
Rectum and Rectosigmoid Junction	20%	57%	2%	19%	25404	9632	4674	8897
Miscellaneous Malignant Cancer	42%	52%	14%	21%	18083	33948	21616	13649
Myeloma	44%	51%	12%	29%	11139	1761	2975	6286
Hodgkin Lymphoma	21%	54%	2%	24%	9699	2997	4937	3555
Prostate	26%	54%	5%	12%	8812	43453	7279	19765
Pancreas	41%	55%	4%	22%	8738	3497	2466	3446
Urinary Bladder	24%	60%	3%	4%	8059	1164	2902	11918
Melanoma of the Skin	16%	48%	6%	10%	7070	947	4853	9497
Chronic Lymphocytic Leukemia	48%	52%	1%	15%	6691	124	894	3660
Esophagus	30%	58%	2%	25%	5302	3672	2330	2420
Testis	21%	58%	1%	6%	5194	1629	3453	1740
Acute Lymphocytic Leukemia	12%	44%	10%	43%	4101	268	626	3072
Cervix Uteri	28%	46%	4%	19%	3656	5027	1676	1927
Soft Tissue including Heart	27%	55%	9%	18%	3509	3218	2683	2109
Stomach	22%	58%	2%	25%	3497	1439	1129	1821
Kidney and Renal Pelvis	35%	56%	5%	18%	3113	928	5527	3572
Corpus Uteri	30%	44%	3%	17%	2931	5120	1903	2556
Acute myeloid	26%	15%	4%	38%	2498	161	717	3039
Anus, Anal Canal and Anorectum	22%	51%	3%	19%	2348	2491	440	1074
Brain and Other Nervous System	35%	53%	63%	14%	2109	12183	5117	4597
Small Intestine	20%	54%	2%	16%	1977	59	350	578
Liver	25%	56%	9%	26%	1885	183	1444	1280
Larynx	37%	58%	5%	7%	1836	5957	801	1934
Bones and Joints	29%	53%	12%	19%	1825	1079	1998	1489
Tongue	32%	50%	7%	7%	1770	4478	609	1180
Other Non-Epithelial Skin	18%	62%	6%	21%	1210	3861	569	78555

	Post-1990 procedures/total procedures				Number of procedures			
Cancer site	drugs	radiation oncology	diagnostic radiology	surgery	drugs	radiation oncology	diagnostic radiology	surgery
Peritoneum, Omentum and Mesentery	36%	62%	1%	29%	1115	21	182	340
Other Endocrine including Thymus$	29%	54%	25%	34%	1027	787	861	698
Nasopharynx	29%	51%	23%	19%	738	1064	321	536
Tonsil	35%	53%	3%	4%	682	2823	423	603
Trachea, Mediastinum and Other Respiratory Organs	34%	59%	2%	18%	672	657	256	288
Other Female Genital Organs	33%	41%	3%	22%	663	228	171	339
Other Myeloid/Monocytic Leukemia	11%	25%	8%	29%	640	4	40	356
Uterus, NOS	26%	50%	3%	22%	627	1047	669	356
Intrahepatic Bile Duct	49%	64%	3%	32%	613	204	260	287
Oropharynx	36%	58%	1%	9%	584	908	80	334
Salivary Gland	29%	54%	13%	10%	573	1557	288	427
Thyroid	15%	46%	10%	5%	558	780	1215	4447
Hypopharynx	26%	50%	4%	22%	551	644	68	212
Gallbladder	31%	42%	10%	31%	500	198	263	310
Other Urinary Organs	27%	63%	6%	9%	493	68	160	280
Other Biliary	37%	58%	8%	24%	445	511	322	336
Gum and Other Mouth	34%	56%	12%	11%	421	870	257	494
Aleukemic, subleukemic and NOS	18%	34%	11%	33%	418	212	632	775
Nose, Nasal Cavity and Middle Ear	35%	56%	27%	45%	377	1099	296	398
Vagina	24%	52%	11%	24%	336	474	161	140
Other Oral Cavity and Pharynx	33%	38%	6%	8%	333	546	172	357
Chronic Myeloid Leukemia	23%	18%	6%	27%	300	39	155	1626
Other Acute Leukemia	14%	6%	4%	40%	296	17	139	290
Other Lymphocytic Leukemia	48%	6%	3%	20%	280	17	168	668
Vulva	32%	67%	3%	31%	221	404	160	419
Other Digestive Organs	53%	37%	0%	8%	203	19	110	157
Pleura	36%	65%	3%	32%	117	72	97	132
Floor of Mouth	41%	64%	2%	9%	111	1062	61	221
Retroperitoneum	56%	37%	3%	14%	108	79	140	191
Penis	11%	58%	4%	11%	81	80	80	83
Ureter	27%	57%	1%	19%	63	93	104	152
Lip	2%	48%	4%	3%	53	137	26	208
Acute Monocytic Leukemia	23%	.	7%	60%	43	.	14	53
Eye and Orbit	25%	38%	24%	10%	32	578	242	384

Table 5
Weighted least-squares estimates of eq. (5),
based on probit functional form and weighting by number of MEDSTAT drug treatments

	1-year survival			2-year survival			3-year survival		
Parameter	**Estimate**	**t-Value**	**Prob > t**	**Estimate**	**t-Value**	**Prob > t**	**Estimate**	**t-Value**	**Prob > t**
drug_new%	0.110	2.62	0.0092	0.145	3.86	0.0001	0.089	2.19	0.0289
rad_diag_new%	0.022	0.18	0.8564	-0.107	-0.93	0.3519	0.054	0.41	0.6839
rad_onc_new%	-0.044	-0.97	0.3343	0.092	2.02	0.0442	0.000	0.01	0.9945
surg_new%	-0.006	-0.04	0.9687	0.003	0.02	0.9834	-0.126	-0.73	0.4678
$F^{-1}(E)$	0.582	4.69	<.0001	0.252	2.24	0.0259	0.410	3.34	0.0009
ln(N)	0.182	5.05	<.0001	0.082	2.18	0.0296	0.205	5.17	<.0001
age	-0.009	-2.03	0.0425	-0.006	-1.25	0.2131	-0.003	-0.51	0.607
in_situ%	1.240	7.23	<.0001	1.108	6.45	<.0001	1.469	8.30	<.0001
loc_reg%	1.439	11.87	<.0001	0.964	6.32	<.0001	1.381	8.23	<.0001
distant%	1.174	20.83	<.0001	-0.078	-0.38	0.7031	0.280	1.21	0.2276
δ_{1992}	0.030	1.33	0.1845	-0.020	-1.04	0.2974	-0.022	-1.28	0.1998
δ_{1993}	0.015	0.77	0.4402	-0.023	-1.43	0.1534	-0.020	-1.37	0.1708
δ_{1994}	0.019	1.17	0.244	-0.027	-2.02	0.0438	-0.022	-1.89	0.0597
δ_{1995}	0.009	0.57	0.5685	-0.021	-1.57	0.1165	-0.015	-1.42	0.1567
δ_{1996}	0.011	0.83	0.4079	-0.015	-1.35	0.1762	-0.008	-0.93	0.351
δ_{1997}	0.030	2.50	0.0129	-0.003	-0.37	0.7147	0.008	1.03	0.3028
δ_{1998}	0.024	2.42	0.016	-0.004	-0.59	0.5548	0.008	1.65	0.1007
δ_{1999}	0.009	1.15	0.2523	-0.007	-1.26	0.2085	0.000	.	.
δ_{2000}	0.015	2.41	0.0162	0.000	.	.			
δ_{2001}	0.000	.	.						
Degrees of freedom									
Model	80			79			78		
Error	453			395			338		
Corrected total	533			474			416		
R^2	0.999			0.999			0.999		

The dependent variable is $F^{-1}(S_{it})$, where S_{it} is the observed survival rate of people diagnosed with cancer originating at site i in year t, and $F^{-1}( )$ is the inverse of the standard normal cumulative distribution. All models include cancer-site fixed effects.

Table 6
Distinguishing between cancer drugs and other drugs

		1-year survival			2-year survival			3-year survival		
Line	Parameter	Estimate	t-Value	Prob > t	Estimate	t-Value	Prob > t	Estimate	t-Value	Prob > t
1	drug_new%	0.110	2.62	0.0092	0.145	3.86	0.0001	0.089	2.19	0.0289
2	cancer_drug_new%	0.086	3.64	0.0003	0.078	3.41	0.0007	0.008	0.28	0.7810
3	other_drug_new%	0.019	0.55	0.5819	0.024	0.79	0.4279	0.018	0.62	0.5388

Table 7
t-statistics (indicating signs and statistical significance) of estimated coefficients of 12 survival models

Survival interval	1-year	2-year	3-year	1-year	2-year	3-year	1-year	2-year	3-year	1-year	2-year	3-year			
Functional form	probit	probit	probit	probit	probit	probit	log	log	log	log	log	log			
Weight	No. of MEDSTAT drug treatments	No. of MEDSTAT drug treatments	No. of MEDSTAT drug treatments	No. of SEER 9 patients diagnosed	No. of SEER 9 patients diagnosed	No. of SEER 9 patients diagnosed	No. of MEDSTAT drug treatments	No. of MEDSTAT drug treatments	No. of MEDSTAT drug treatments	No. of SEER 9 patients diagnosed	No. of SEER 9 patients diagnosed	No. of SEER 9 patients diagnosed			
														No. of positive and signif. coeffs.	No. of negative and signif. coeffs.
Coefficient															
drug_new%	**2.62**	**3.86**	**2.19**	**2.79**	**1.99**	1.80	**4.21**	**3.85**	**2.65**	**4.87**	**3.25**	**2.41**		11	0
rad_diag_new%	0.18	-0.93	0.41	1.94	-0.30	-0.86	1.46	0.34	-0.11	**2.10**	0.00	-0.85		1	0
rad_onc_new%	-0.97	**2.02**	0.01	-1.24	-0.68	-0.61	-0.60	**2.44**	1.39	0.46	1.23	1.57		2	0
surg_new%	-0.04	0.02	-0.73	0.50	0.07	0.52	-0.41	-0.95	**-2.32**	-0.40	-0.27	0.44		0	1
f(E)	**4.69**	**2.24**	**3.34**	**4.09**	1.88	1.47	0.48	-0.08	0.14	-1.89	**-2.54**	-1.65		4	1
ln(N)	**5.05**	**2.18**	**5.17**	**7.30**	**3.03**	**3.32**	**3.41**	**-2.68**	1.03	**6.18**	1.83	**2.36**		9	1
age	**-2.03**	-1.25	-0.51	**4.02**	**4.22**	**3.45**	**-3.67**	**-3.19**	**-2.64**	-1.70	0.11	0.80		3	4

Appendix Table 1
SEER Cancer Causes of Death Recode 1969+ (9/17/2004)

Cancer Causes of Death	Recode	ICD-9 (1979-1998) #	ICD-10 (1999+) #
All Malignant Cancers	--	140-208, 238.6	C00-C97
Oral Cavity and Pharynx			
Lip	20010	140	C00
Tongue	20020	141	C01-C02
Salivary Gland	20030	142	C07-C08
Floor of Mouth	20040	144	C04
Gum and Other Mouth	20050	143, 145	C03, C05-C06
Nasopharynx	20060	147	C11
Tonsil	20070	146.0-146.2	C09
Oropharynx	20080	146.3-146.9	C10
Hypopharynx	20090	148	C12-C13
Other Oral Cavity and Pharynx	20100	149	C14
Digestive System			
Esophagus	21010	150	C15
Stomach	21020	151	C16
Small Intestine	21030	152	C17
Colon and Rectum			
Colon excluding Rectum	21040	153, 159.0	C18, C26.0
Rectum and Rectosigmoid Junction	21050	154.0-154.1	C19-C20
Anus, Anal Canal and Anorectum	21060	154.2-154.3, 154.8	C21
Liver and Intrahepatic Bile Duct			
Liver	21071	155.0, 155.2	C22.0, C22.2-C22.4, C22.7, C22.9
Intrahepatic Bile Duct	21072	155.1	C22.1
Gallbladder	21080	156	C23
Other Biliary	21090	156.1-156.2, 156.8-156.9	C24
Pancreas	21100	157	C25
Retroperitoneum	21110	158	C48.0
Peritoneum, Omentum and Mesentery	21120	158.8-158.9	C45.1+, C48.1-C48.2
Other Digestive Organs	21130	159.8-159.9	C26.8-C26.9, C48.8
Respiratory System			
Nose, Nasal Cavity and Middle Ear	22010	160	C30-C31
Larynx	22020	161	C32
Lung and Bronchus	22030	162.2-162.5, 162.8-162.9	C34
Pleura	22050	163	C38.4, C45.0+
Trachea, Mediastinum and Other Respiratory Organs	22060	162.0, 164.2-164.3, 164.8-164.9, 165	C33, C38.1-C38.3, C38.8, C39
Bones and Joints	23000	170	C40-C41
Soft Tissue including Heart$	24000	164.1, 171	C47, C49, C38.0, C45.2+
Skin excluding Basal and Squamous			
Melanoma of the Skin	25010	172	C43
Other Non-Epithelial Skin	25020	173	C44, C46+
Breast	26000	174-175	C50
Female Genital System			
Cervix Uteri	27010	180	C53
Corpus and Uterus, NOS			

Corpus Uteri	27020	182	C54
Uterus, NOS	27030	179	C55
Ovary	27040	183	C56
Vagina	27050	184	C52
Vulva	27060	184.1-184.4	C51
Other Female Genital Organs	27070	181, 183.2-183.5, 183.8-183.9, 184.8-184.9	C57-C58
Male Genital System			
Prostate	28010	185	C61
Testis	28020	186	C62
Penis	28030	187.1-187.4	C60
Other Male Genital Organs	28040	187.5-187.9	C63
Urinary System			
Urinary Bladder	29010	188	C67
Kidney and Renal Pelvis	29020	189.0-189.1	C64-C65
Ureter	29030	189.2	C66
Other Urinary Organs	29040	189.3-189.4, 189.8-189.9	C68
Eye and Orbit	30000	190	C69
Brain and Other Nervous System	31010	191, 192	C70, C71, C72
Endocrine System			
Thyroid	32010	193	C73
Other Endocrine including Thymus$	32020	164.0, 194	C37, C74-C75
Lymphoma			
Hodgkin Lymphoma	33010	201	C81
Non-Hodgkin Lymphoma	33040	200, 202.0-202.2, 202.8-202.9	C82-C85, C96.3
Myeloma	34000	203.0, 238.6	C90.0, C90.2
Leukemia			
Lymphocytic Leukemia			
Acute Lymphocytic Leukemia	35011	204	C91.0
Chronic Lymphocytic Leukemia	35012	204.1	C91.1
Other Lymphocytic Leukemia	35013	202.4, 204.2, 204.8-204.9	C91.2-C91.4, C91.7, C91.9
Myeloid and Monocytic Leukemia			
Acute myeloid	35021	205.0, 207.0, 207.2	C92.0, C92.4-C92.5, C94.0, C94.2
Acute Monocytic Leukemia	35031	206	C93.0
Chronic Myeloid Leukemia	35022	205.1	C92.1
Other Myeloid/Monocytic Leukemia	35023	205.2-205.3, 205.8-205.9, 206.1-206.2, 206.8-206.9	C92.2-C92.3, C92.7, C92.9, C93.1-C93.2, C93.7, C93.9
Other Leukemia			
Other Acute Leukemia	35041	208	C94.4, C94.5, C95.0
Aleukemic, subleukemic and NOS	35043	203.1, 207.1, 207.8, 208.1-208.2, 208.8-208.9	C90.1, C91.5, C94.1, C94.3, C94.7, C95.1, C95.2, C95.7, C95.9
Mesothelioma (ICD-10 only)+	36010	N/A	C45+
Kaposi Sarcoma (ICD-10 only)+	36020	N/A	C46+
Miscellaneous Malignant Cancer	37000	159.1, 195-199, 202.3, 202.5 202.6, 203.8	C26.1, C45.7+, C45.9+, C76-C80, C88, C96.0-C96.2, C96.7, C96.9, C97

http://seer.cancer.gov/codrecode/1969+_d09172004/index.html