Do fixed patent terms distort innovation? Evidence from cancer clinical trials*

Eric Budish
University of Chicago

Benjamin N. Roin
Harvard Law School

Heidi Williams
MIT and NBER

January 12, 2013

Preliminary and incomplete: Please do not cite or circulate

Abstract

Patents award innovators a fixed period of market exclusivity. Yet, since in many industries firms file patents at the time of discovery ("invention") rather than first sale ("commercialization"), *effective* patent terms vary: inventions that commercialize at the time of invention receive a full patent term, whereas inventions that have a long time lag between invention and commercialization receive substantially reduced - or in extreme cases, zero - effective patent terms. We present a simple model formalizing how this variation may inefficiently distort research and development (R&D). We then explore this distortion empirically in the context of cancer R&D, where clinical trials are shorter - and hence, effective patent terms longer - for drugs targeting late-stage cancer patients, relative to drugs targeting early-stage cancer patients or cancer prevention. Using a newly constructed data set on cancer clinical trial investments, we provide several sources of evidence consistent with fixed patent terms distorting cancer R&D. We discuss several policy levers that could eliminate this distortion.

*Financial support from NIA Grant Number T32-AG000186 to the NBER, NSF Grant Number 1151497, the Chicago Booth Initiative on Global Markets, the NBER Innovation Policy and the Economy program, and the Petrie-Flom Center at Harvard Law School are gratefully acknowledged. We thank Meru Bhanot, David Burk, Toby Chaiken, Brian Cordonnier, Niels Joaquin, Megan McGrath, and Ana Medrano-Fernandez for excellent research assistance. David Autor, Joe Doyle, Dan Fetter, Amy Finkelstein, Amanda Kowalski, Ariel Pakes, Scott Stern, Glen Weyl, and seminar participants at the Cornell Empirical Patent Law Conference, Georgia Tech, Harvard Law School, MIT, the NBER productivity lunch, Toulouse, UC-Berkeley Haas, UIUC, and Wellesley provided very helpful comments. Contact: eric.budish@chicagobooth.edu, benroin@law.harvard.edu, heidiw@mit.edu.

1 Introduction

Patents aim to incentivize innovation by awarding inventors a temporary period of market exclusivity. In most countries, the current patent term is twenty years from the filing date of the application. It has long been recognized that heterogeneity across inventions - such as variation in risk-adjusted costs of development - implies that this type of uniform patent term will award "too much" market exclusivity to some inventions, and will be insufficient to motivate the development of others.[1] In this paper, we theoretically and empirically investigate the consequences of one particular form of heterogeneity. Because in many industries firms file patents at the time of discovery ("invention") rather than at the time of first sale ("commercialization"), fixed patent terms generate variable *effective* patent lengths: inventions that commercialize at the time of invention receive a full twenty-year patent term, whereas inventions that take longer to commercialize realize a shorter effective patent life. In the extreme, patents offer no incentive to develop technologies that would take longer than twenty years to commercialize. Under a fixed patent term, research and development (R&D) investments may thus be distorted away from technologies with long time lags between invention and commercialization.

We first present a simple theoretical model to characterize the conditions under which fixed patent terms will distort innovation. For this distortion to arise, firms must face a risk of losing the ability to patent if they wait until commercialization to file; the innovation process must involve a time lag between invention and commercialization; patents must be an effective means of generating monopoly profits; the technology must be imitable once patents expire; and the technology must have a long "total useful life" relative to the length of the patent term. Mechanically, these conditions together generate variation in the ratio of an invention's "monopoly life" to its total useful life. Since society cares about an invention's total useful life, but private firms care only about monopoly life, a distortion emerges: society might value invention A more highly than invention B, but private industry may choose to develop B but not A. It is important to point out that, other things equal, commercialization lag lowers both monopoly life and total useful life: both society and private firms prefer inventions to reach the market quickly. However, commercialization lag reduces monopoly life more rapidly than total useful life, hence the distortion away from inventions with both a long total useful life and a long commercialization lag.

The model clarifies that the fixed patent term distortion is unlikely to arise in many industries, but the conditions necessary for this distortion to exist naturally fit the pharmaceutical industry. Pharmaceutical firms face strong incentives to file patent applications at the time of invention. However, prior to marketing

[1] Optimal patent length and optimal patent breadth have long been topics of study in the economics and law literatures; see, *e.g.*, Machlup (1958), Nordhaus (1969), Scherer (1972), Nordhaus (1972), Kaplow (1984), Klemperer (1990), Gilbert and Shapiro (1990), and Scotchmer (1991).

a new drug, firms must submit clinical trial results to the US Food and Drug Administration (FDA) documenting that their product meets a set of safety and efficacy standards. These clinical trials generate a lag between the time of patenting and the time of commercialization, which reduces an invention's effective patent life - more so for drugs that require longer clinical trials. A key determinant of clinical trial length is patient survival time: because clinical trials generally must show evidence that the treatment improves mortality-related outcomes, clinical trials tend to be longer - and hence, effective patent terms shorter - when enrolling patients with longer survival times. All else equal, firms are thus awarded longer terms of market exclusivity for successfully developing drugs to treat patients with short survival times relative to drugs to treat patients with longer survival times.[2]

Consider two examples of clinical trials for prostate cancer treatments, both published in the *New England Journal of Medicine* in 2011. A first study, de Bono et al. (2011), analyzed a treatment for metastatic prostate cancer (an advanced stage of prostate cancer with a five-year survival rate on the order of 20 percent). The study tracked patient survival for a median time of 12.8 months, and estimated statistically significant improvements in survival (a gain of 3.9 months of life on average). A second study, Jones et al. (2011), analyzed a treatment for localized prostate cancer (an early stage of prostate cancer with a five-year survival rate on the order of 80 percent). The study tracked patient survival for a median time of 9.1 years, estimating statistically significant improvements in survival. As expected, this stark difference in patient follow-up times translates into large differences in clinical trial length: 3 years for the metastatic patient trial versus 18 years for the localized patient trial. All else equal, the drug treating localized patients would receive 15 fewer years of effective patent life relative to the drug treating metastatic patients. Given this difference in effective patent life, we expect private firms to provide "too few" clinical trials for patients with long expected survival times. Consistent with this idea, the study of metastatic cancer patients was funded by a private firm (Cougar Biotechnology) whereas the study of localized cancer patients was funded by the National Cancer Institute.

Our empirical work investigates the basic insight behind this example in the context of R&D investments on cancer treatments. Cancer drug development tends to be specific to a cancer type (*e.g.* prostate) and stage of disease (*e.g.* metastatic) - as in the examples above - providing a natural framework for estimating how survival time and R&D investments vary across different groups of patients. Aggregating survival information from patient-level cancer registry data, we document stark variation in survival times across patients of different cancer types and stages of disease. In order to measure R&D investments on cancer treatments relevant to each cancer type and stage of disease, we use newly-constructed data from a

[2] As we discuss in Section 2.5.1, this gap in effective patent life persists even with the patent term extensions available to pharmaceutical companies under the 1984 Hatch-Waxman Act, which only partially compensate for patent life lost during clinical development.

clinical trial registry that has cataloged cancer clinical trials since the 1970s. The key feature of this data which makes it amenable to our analysis is that for each clinical trial, the registry lists each of the specific patient groups eligible to enroll in the trial - thus allowing a match between our measures of survival time and R&D activity (as measured by clinical trial investments) across cancer types and stages of disease.

Using this data, we document that - consistent with the fixed patent term distortion - patient groups with higher survival rates tend to have lower levels of R&D investment. Figure 1(a) gives a sense of this basic pattern using stage-level data. On average, metastatic cancer patients have a five-year survival rate of approximately 10 percent, and have nearly 12,000 clinical trials in our data. In contrast, localized cancer patients have a five-year survival rate of approximately 70 percent, and have just over 6,000 clinical trials in our data. A rough adjustment for market size - looking at the number of clinical trials per patient diagnosed - does little to change this basic pattern. When we investigate this relationship between survival time and R&D investments across cancer types and stages of disease, our baseline point estimates suggest that a 10 percentage point increase in the five-year survival rate is associated with a reduction in R&D investments on the order of 10 percent. The magnitude of this estimate in the raw data is similar when we condition on measures of market size and life-years lost from cancer.

This new fact - the negative correlation between survival time and R&D - is consistent with the fixed patent term distortion, but by itself is not evidence of inefficiency. We would expect that both higher demand for treatments and lower costs of clinical trials might also generate a negative correlation between survival time and R&D. We present evidence from two complementary analyses supporting the idea that the observed survival time-R&D correlation is more negative than is consistent with an efficiency argument based on demand and costs. First, to address the concern of unobserved heterogeneity in demand, we investigate surrogate endpoints (non-mortality based clinical trial endpoints) - which should break the link between patient survival and trial length. We document that there is *not* a negative relationship between survival time and R&D in the sample of cancers allowed to use surrogate endpoints - suggesting that heterogeneity in demand for treatments is not driving the observed survival time-R&D correlation in the full sample. Second, to address the concern of unobserved heterogeneity in the costs of R&D, we rely on the fact that higher costs also reduce social R&D incentives, so that any *additional* negative correlation between the survival rate and R&D investments among privately financed trials should be purged of heterogeneity in the costs of clinical trials that is correlated with the survival rate. We document that the survival time-R&D correlation is 30 percent larger for privately financed trials than for publicly financed trials. Finally, we document that all six FDA-approved cancer prevention drugs either relied on the use of surrogate endpoints or were approved on the basis of publicly financed clinical trials. Taken together,

this body of evidence provides support for the idea that fixed patent terms distort R&D.[3]

The idea that fixed patent terms may distort innovation has been discussed by several legal scholars (Abramowicz (2007); Eisenberg (2005); Roin (2010)), but to the best of our knowledge this paper is the first to formally investigate this distortion either theoretically or empirically. We also use our theoretical model to analyze the innovation and social welfare consequences of three policy interventions: a patent design change that would start the patent clock at commercialization, a policy change that would allow firms to rely on surrogate endpoints in clinical trials, and targeted R&D subsidies. Our empirical results also contribute more broadly to the literature on how patents affect the level and composition of R&D. Although theoretical models often assume a relationship between the strength of patent protection and the level of innovation, there is a remarkable dearth of empirical evidence on this link: for example, Lerner (2002) and Sakakibara and Branstetter (2001) find little evidence that stronger intellectual property rights induce more innovation. Moreover, with the exception of Moser (2005) - who documents that in the 19th century, inventors in countries without patent laws were relatively more concentrated in industries where secrecy was effective relative to patents - there is almost no evidence on how patents shape the *composition* of R&D. Our empirical evidence that fixed patent terms distort R&D can also be interpreted as evidence that patents affect both the level and composition of R&D: pharmaceutical innovation responds to plausibly exogenous variation in effective patent length (here, driven by patient survival times), and this variation in effective patent length has important effects on the composition of innovation investments.

Our empirical focus on cancer treatments is of substantive interest because of cancer's tremendous morbidity and mortality burden. In 2009, cancer was the second leading cause of death in the US (after heart disease), accounting for almost 25 percent of all deaths. Using an economic framework which values improvements in health based on individuals' willingness to pay, Murphy and Topel (2006) estimate that a permanent 1 percent reduction in cancer mortality has a present value to current and future generations of Americans of nearly $500 billion, and that a cure (if feasible) would be worth about $50 trillion.

The paper proceeds as follows. Section 2 presents the model. Section 3 describes our data. Section 4 documents the negative correlation between survival time and R&D investments, and Section 5 interprets this relationship. Section 6 concludes.

[3]Conversations with several venture capitalists also generated anecdotal evidence that fixed patent terms result in some scientifically promising projects failing to reach the market; for example, one venture capitalist we talked with noted: *It happens all the time that we pass on a drug, one we think would probably work, because there wouldn't be enough life left on its patent by the time it reached the market.*

2 Theory

In this section we present a simple theoretical model which is designed with three goals in mind. First, the model characterizes the conditions under which fixed patent terms will be expected to distort R&D investments. Second, we use the model to derive comparative statics that we then take to the data. Third, we provide a welfare analysis of patent design and other policy levers which could be used to address this distortion.

In the model, we conceptualize R&D as consisting of two stages: invention and commercialization. By invention we mean developing the basic idea for a product to the point where it is patentable: producing a new chemical compound, building a prototype, *etc.* By commercialization we mean all that is involved in bringing an invented product to market: getting FDA approval for the new chemical compound, producing the prototyped good at efficient scale, *etc.* The distinction between invention and commercialization matters because, in many industries, firms face strong incentives to file for patent protection at the time of invention rather than at the time of commercialization: delayed applications risk competitors filing conflicting patents, and risk independent researchers disclosing information (through scientific publications, for example) that would invalidate later patent applications. Our model clarifies that in cases where there is a long time lag between time of invention and time of commercialization - as induced by clinical trials for new drugs - this time lag can interact with fixed patent terms to generate distorted R&D incentives.

The model clarifies that the fixed patent term distortion is unlikely to arise in many industries, but as we discuss in Section 2.6, the conditions necessary for this distortion to exist naturally fit the pharmaceutical industry.[4] Reflecting this, we focus the model on the case of the pharmaceutical industry, although the analysis applies more generally.

2.1 Preliminaries

A representative firm conducts undirected R&D which stochastically yields inventions. Whenever the firm's undirected R&D yields an invention, it then must decide whether to invest directed R&D towards the goal of commercializing the specific invention. The firm proceeds with commercialization if the expected private benefits of commercialization exceed the expected private costs. An invention is characterized by the following parameters:

- Timing parameters: the year in which the invention is realized by the firm's undirected R&D is t_{invent}, which we normalize to 0. The number of years that the commercialization effort will take is

[4] Kitch (1977) argues based on case studies (see, *e.g.*, his Table 1) that long time lags between patenting and commercialization occur in many industries. As we will see in Proposition 2, the existence of such lags is necessary but not sufficient for the fixed patent term distortion to be relevant in other industries.

t_{comm}; t_{comm} can be interpreted as the number of years that it will take to conduct FDA clinical trials. We treat t_{comm} and several other parameters below as deterministic for simplicity; in practice many of the parameters would be stochastic. We focus the model on cases where firms are incentivized to file for patents at t_{invent} rather than t_{comm}; as motivated above, this is true for the pharmaceutical industry, and Proposition 2 clarifies that this is a necessary condition for the fixed patent term distortion to arise.

- Likelihood of successful commercialization: the commercialization effort yields a commercially viable product with probability p. The success parameter p can be interpreted as the likelihood that FDA clinical trials are successful.

- Obsolescence risk: if the product is successfully commercialized, then it is useful until superseded. We model obsolescence risk in a simple way, assuming that obsolescence occurs with probability $1-\gamma$ per year in each year following t_{invent}.[5] Obsolescence risk would more appropriately be modeled as an endogenous parameter (for example, a function of R&D investments); for simplicity we follow much of the previous literature in taking obsolescence risk as exogenous (*e.g.* Grossman and Lai (2004)).

- Monopoly profits and social value: if the product is successfully commercialized, non-obsolete, and protected by patent, it yields profits of π per year to the inventing firm. If the product were priced by a social planner instead of a monopolist, it would yield social value of v per year.

- Imitibility: if the product is successfully commercialized, non-obsolete, and not protected by patent, generic entrants may imitate the commercialized product. Imitation reduces the inventing firm's profits from π to $(1-\iota)\pi$, where $\iota \in [0,1]$ denotes the imitibility of the product (that is, vulnerability to generic competition). The case $\iota = 1$ corresponds to perfect imitibility, which drives the inventing firm's profits to zero.

- Discount factor: both the firm and society apply a discount factor of δ to the project.

- Cost of commercialization: the commercialization effort costs c, in net present value terms as of date t_{invent}. The cost c can be interpreted as the cost of conducting FDA trials.

Taken together, inventions can thus be characterized as a set of invention parameters $(p, \pi, v, \delta, \gamma, c)$, a degree of imitibility ι, and a commercialization lag t_{comm}.

[5] An alternative would be to incorporate obsolescence that occurs before t_{comm} into the probability of commercialization success p, and only use the term obsolescence to describe cases where the product is superseded after successful commercialization at t_{comm}. This is economically equivalent, but less convenient mathematically; see especially formula (1) below.

In a fixed-term patent system, patents for new inventions last t_{patent} years.[6] So long as an invention is protected by patent, *imitation is illegal*. Importantly, when firms patent at t_{invent} rather than t_{comm}, post-commercialization patent life is less than t_{patent} in cases where $t_{comm} > 0$.

2.2 Effective Monopoly Life and Effective Total Life

We define an invention's *Expected Monopoly Life* (*EML*) as the expected number of years, in present value terms, that the inventor can expect to earn monopoly profits from the commercialized product.

Initially, consider perfectly imitable inventions ($\iota = 1$). For such inventions, EML is the expected amount of time that the invention is commercially viable, protected by patent, and not yet superseded. If $t_{patent} \geq t_{comm}$ then this can be written as:

$$EML|_{\iota=1} = p \sum_{t_{comm}}^{t_{patent}} (\delta\gamma)^t = p\frac{(\delta\gamma)^{t_{comm}} - (\delta\gamma)^{t_{patent}}}{1-(\delta\gamma)}. \tag{1}$$

The key thing to notice about equation (1) is the role of the timing parameters: at best, the period of monopoly is from t_{comm} to t_{patent}. This best case occurs if the invention is successfully commercialized (which occurs with probability p) and not superseded as of time t_{patent} (which occurs with probability $\gamma^{t_{patent}}$). As soon as time reaches t_{patent}, the invention will be imitated and the monopoly position lost. Note as well that if $t_{patent} < t_{comm}$, then $EML = 0$: by the time the invention is commercialized, patent protection has expired.

Next, we allow for imperfectly imitable inventions ($\iota < 1$). For such inventions, we also need to include the time where the invention is off patent but not yet superseded, weighting by the degree of imitibility. The general formula, assuming $t_{patent} \geq t_{comm}$,[7] is:

$$EML = p\left(\sum_{t_{comm}}^{t_{patent}} (\delta\gamma)^t + (1-\iota)\sum_{t_{patent}}^{\infty} (\delta\gamma)^t\right) = p\frac{(\delta\gamma)^{t_{comm}} - \iota(\delta\gamma)^{t_{patent}}}{1-(\delta\gamma)}. \tag{2}$$

Last, we define *Expected Total Life (ETL)* as the expected number of years, in present value terms, that the invention will be commercialized and non-obsolete:

$$ETL = p\sum_{t_{comm}}^{\infty} (\delta\gamma)^t = p\frac{(\delta\gamma)^{t_{comm}}}{1-(\delta\gamma)}. \tag{3}$$

[6] We here abstract away from the provisions of the 1984 Hatch-Waxman Act, which awards some qualifying pharmaceutical firms extended patent terms; we discuss such policy levers in Section 2.5.1.

[7] If commercialization lag exceeds patent length, *i.e.*, $t_{comm} > t_{patent}$, then the formula is $EML = p\left((1-\iota)\sum_{t_{comm}}^{\infty}(\delta\gamma)^t\right) = p(1-\iota)\frac{(\delta\gamma)^{t_{comm}}}{1-(\delta\gamma)}$.

Observe that if the invention is perfectly non-imitable ($\iota = 0$) then EML and ETL coincide. If the invention is perfectly imitable ($\iota = 1$) then the key difference between ETL and EML is that monopoly life runs at best until t_{patent}, whereas total life runs indefinitely until the invention becomes obsolete.

2.3 Private and social incentives to invest

A profit-maximizing firm attempts to commercialize an invention if and only if the expected profits exceed the costs:

$$\text{Private Investment Occurs} \iff EML \cdot \pi \geq c. \tag{4}$$

In words, the firm can expect to enjoy monopoly profits of π for EML years. If $EML \cdot \pi$ exceeds the costs of commercialization c, it is optimal to commercialize.

Suppose instead that society owned the firm. If commercialization is successful, the social planner will price at marginal cost, and hence create social welfare of v per year. Hence the social planner attempts to commercialize the invention if and only if expected social welfare, if the good is priced at marginal cost, exceeds the costs of commercialization:

$$\text{Investment is Socially Optimal} \iff ETL \cdot v \geq c. \tag{5}$$

Notice that $ETL \geq EML$ and $v \geq \pi$ by definition. By construction, this ignores issues such as business stealing and R&D races which - although important - are not the focus of our analysis. Thus, in our framework, anytime a private firm would choose to commercialize an invention, so too would the social planner. The projects that the private firm does not pursue, but that society would pursue if it owned the firm, are those where:

$$\text{Private and Social Investment Differ} \iff \frac{EML \cdot \pi}{c} \leq 1 \leq \frac{ETL \cdot v}{c}. \tag{6}$$

In words, private and social investment decisions differ when the social return is positive but the private return is negative. The private market can under-provide R&D if either $\frac{EML}{ETL} < 1$ or $\frac{\pi}{v} < 1$.

2.4 Distortions in the level and composition of R&D

Our model yields distortions, relative to the social optimum, in both the level and composition of commercialization activity. By distortion in level, we mean simply that fewer inventions are commercialized by private firms than would be the case if the social planner made commercialization decisions. This is a standard result. By distortion in composition, we mean that the private market may choose to commer-

cialize A but not B, while a social planner would prefer to commercialize B over A. That is, the private sector not only pursues too little R&D relative to the social optimum, but also chooses the wrong projects relative to what the social planner would choose. We state this formally as follows:[8]

Proposition 1. *The private firm's commercialization activity differs from the social optimum in both the level and the composition:*

1. *(distortion in levels) Commercialization activity is lower than socially optimal, unless (a) exclusivity is infinite (so $EML = ETL$); and (b) monopolists capture full social surplus (i.e., $\pi = v$).*

2. *(distortion in composition) For two inventions, A and B, it is possible that the expected social return ($ETL \cdot v/c$) to pursuing invention A exceeds that of invention B, yet invention A is not pursued while invention B is. For this to be the case, at least one of the following must hold:*[9]

   (a) $\frac{\pi_B}{v_B} > \frac{\pi_A}{v_A}$, *i.e., monopolists capture more profit as a proportion of potential social value from invention B than from invention A.*

   (b) $\frac{EML_B}{ETL_B} > \frac{EML_A}{ETL_A}$, *i.e., a larger proportion of the total life of the invention is spent in the monopoly state for invention B than for invention A.*

As noted above, Part 1 of Proposition 1 is a standard result, which indicates that the private sector pursues too little inventive activity relative to the first best. Part 2 of Proposition 1 indicates that distortions in composition can arise from differences across inventions in either $\frac{\pi}{v}$ or $\frac{EML}{ETL}$.

An invention's profitability to social value ratio $\frac{\pi}{v}$ depends on the monopolist's ability to capture the value its invention creates.[10] One extreme case is if the monopolist can perfectly price discriminate, in which case $\frac{\pi}{v} = 1$. The other extreme case is inventions that are non-excludable, in which case $\frac{\pi}{v} = 0$. An example of the latter is a study on non-excludable forms of disease prevention: *e.g.*, a profit-maximizing firm would never conduct an expensive clinical trial to test whether a particular pattern of cardiovascular exercise reduces the risk of heart disease, because knowledge that cardiovascular exercise reduces the risk of heart disease would be non-excludable.

An invention's monopoly-life to total-life ratio, $\frac{EML}{ETL}$, corresponds to the proportion of an invention's expected total life that is protected under patent. It is straightforward to show that $\frac{EML}{ETL}$ is decreasing in

[8] Proofs are presented in Appendix A.

[9] We use subscripts A and B to denote the project specific parameters associated with these specific inventions (*e.g.*, π_A is the monopoly profits associated with successful commercialization of invention A).

[10] Past authors have estimated that on the whole, pharmaceutical firms appropriate only a small share of the social value of their innovations - generally between 2-20 percent (Philipson and Jena (2006); Lakdawalla et al. (2010); Lindgren and Jonsson (2012)).

commercialization lag, at an increasing rate. To see this, simply write out $\frac{EML}{ETL}$ for the case $t_{comm} \leq t_{patent}$:

$$\frac{EML}{ETL} = \frac{\frac{(\delta\gamma)^{t_{comm}} - \iota(\delta\gamma)^{t_{patent}}}{1-(\delta\gamma)}}{\frac{(\delta\gamma)^{t_{comm}}}{1-(\delta\gamma)}} = 1 - \iota(\delta\gamma)^{t_{patent}-t_{comm}}. \tag{7}$$

Comparative Static 1. *Comparative statics of an invention's proportion of monopoly life to total life,* $\frac{EML}{ETL}$*, on its commercialization lag,* t_{comm}*:*

1. $\frac{\partial \frac{EML}{ETL}}{\partial t_{comm}} < 0$ *if* $t_{comm} \leq t_{patent}$
2. $\frac{\partial^2 \frac{EML}{ETL}}{\partial t_{comm}^2} < 0$ *if* $t_{comm} \leq t_{patent}$
3. $\frac{EML}{ETL} = 1 - \iota$ *if* $t_{comm} \geq t_{patent}$

Comparative Static 1 is intuitive: as commercialization lag increases, the proportion of an invention's total life spent under monopoly protection decreases. The rate of decrease is increasing in the lag (part (2)), until, if the lag is so long that there is no patent protection remaining, $\frac{EML}{ETL}$ reaches its minimum of $1 - \iota$ (part (3)). If the commercialization lag is sufficiently long ($t_{comm} \geq t_{patent}$) and the invention is perfectly imitable by generic entrants ($\iota = 1$), then the patent system provides zero incentive for private commercialization activities.

Notice that if the patent clock started at commercialization rather than invention, so $t_{patent} - t_{comm}$ was constant across inventions, then commercialization lag would no longer affect $\frac{EML}{ETL}$. (See Proposition 3 below for a formal statement). Commercialization lag reduces both EML and ETL: both the monopolist and the social planner prefer, other things equal, for an invention to reach market quickly. But, under this alternative patent design, commercialization lag reduces EML and ETL *at the same rate.* For this reason, we refer to the comparative static of $\frac{EML}{ETL}$ on commercialization lag as the *fixed patent term distortion.*

Equation (7) implies that there are several necessary conditions for the fixed patent term distortion to arise:

Proposition 2. *Under a fixed-term patent system, variation in commercialization lag* t_{comm} *across inventions distorts innovation as long as each of the following conditions is met:*

1. *Firms file for patent at* t_{invent} *rather than* t_{comm}
2. *Firms earn profits while their inventions are commercialized and protected by patent:* $\pi > 0$
3. *Patents are needed to protect against imitation:* $\iota > 0$
4. *Obsolescence risk is sufficiently low that inventions remain useful after patent expiration:* $\gamma > 0$

5. *Society is sufficiently patient that an invention's life after patent expiration is relevant:* $\delta > 0$

Proof. Follows immediately from equation (7).

□

2.5 Patent design and policy levers

In this subsection, we discuss the innovation and social welfare consequences of three policy interventions: a patent design change that would start the patent clock at commercialization, a policy change that would allow firms to rely on surrogate (non-mortality) endpoints in clinical trials, and targeted R&D subsidies.

2.5.1 Patent design policy lever

We have seen that commercialization lag distorts R&D under the current fixed-term patent system: in the extreme, if $t_{comm} \geq t_{patent}$, then the patent system provides zero incentive to private firms. Our next result shows that starting the patent clock at commercialization, rather than at invention, eliminates this distortion:

Proposition 3. *If the patent clock starts at commercialization,* $\frac{EML}{ETL}$ *is independent of commercialization lag,* t_{comm}*.*

Proof. Follows immediately from equation (7), since $t_{patent} - t_{comm}$ is now constant.

□

Thus starting the patent clock at commercialization, rather than at invention, eliminates one of the distortions in type of invention. Note of course that commercialization lag still reduces incentives to invest: other things equal, firms prefer inventions that come to market quickly over inventions that come to market slowly. The important thing is that, if the patent clock starts at commercialization, then commercialization lag reduces both private incentives and social incentives at the same rate. That is, both EML and ETL decrease, but the ratio $\frac{EML}{ETL}$, and hence the possibility of distortion, stays constant.

If we make an additional assumption on the distribution of invention possibilities, we can make a stronger claim, which is that starting the patent clock at commercialization strictly increases social welfare. In fact, the result says we should go further: social welfare is maximized by awarding *more* post-commercialization patent life to inventions with longer commercialization lag than inventions with shorter commercialization lag, in contrast to the current system which awards inventions with longer commercialization lag *less* post-commercialization patent life than inventions with shorter lag.

Proposition 4. *Suppose that the distribution of invention parameters* $(p, \pi, v, \delta, \gamma, c)$ *is orthogonal to an invention's commercialization lag* t_{comm}*, that profits and social value are positively correlated in the sense that* $\frac{\partial E(\pi|v)}{\partial v} > 0$ *and* $\frac{\partial E(v|\pi)}{\partial \pi} > 0$*, and that private firms make commercialization decisions according to equation (4). Suppose that the length of the patent award can be conditioned on* t_{comm} *but not on the other invention parameters. Then socially optimal patent policy requires that the number of years of post-commercialization patent protection increases monotonically with* t_{comm}*, whereas under the fixed-term patent system the number of years of post-commercialization patent protection decreases monotonically with* t_{comm}*.*

The intuition for this result, which was conjectured in Roin (2010), is as follows. Fix a level of t_{comm}, and consider an increase in post-commercialization patent life for inventions with this commercialization lag. This increase in patent protection has benefits and costs. The benefit is that more inventions with commercialization lag t_{comm} will be commercialized at the margin; technically, we have increased EML and hence made it more likely that equation (4) obtains. The cost is that, for inframarginal inventions that would have been pursued absent the increase in patent protection, there is more deadweight loss, for the standard reason that social value under monopoly is smaller than social value under perfect competition from generic entrants. The proof works by showing that the quality of the marginal invention is higher the higher is t_{comm}, where by higher quality we mean higher expected profits per unit cost $\frac{p\pi}{c}$ and higher expected social value per unit cost $\frac{pv}{c}$. This in turn implies that the benefits of increasing post-commercialization patent life are greater the larger is t_{comm}, and that the costs of doing so are lower: we get higher-quality inventions at the margin, and the invention parameters for which we suffer deadweight loss are a strict subset of the invention parameters for which we suffer deadweight loss under lower t_{comm}. Hence, the larger is t_{comm}, the larger should be post-commercialization patent life.

We wish to make four further remarks concerning this result. First, conditioning the length of patent award on t_{comm} should be feasible in practice, at least in the case of pharmaceuticals, since completion of FDA trials is intrinsically an observable event. Second, while the proposition assumes that the distribution of invention parameters is orthogonal to t_{comm}, in practice we might expect commercialization costs c to be increasing with commercialization lag; *e.g.*, longer FDA clinical trials are more expensive, other things equal. Having c be positively correlated to t_{comm} only strengthens the result. Third, the 1984 Hatch-Waxman Act (Public Law Number 98-417, 1984) contains a provision granting some qualifying firms a partial extension of patent life based on the time that the drug spent in clinical trials. Specifically, the act awards qualifying firms an additional half-year of patent life for every year spent in clinical trials, up to a maximum of 5 years not exceeding 14 total years. Our result says that the Hatch-Waxman extension

is directionally correct, but that optimal policy would go further. Finally, we are here abstracting away from strategic responses that could be "unintended consequences" from such a change in patent policy.[11]

2.5.2 Policy lever: Surrogate endpoints

A major factor determining the duration of a clinical trial is the amount of time needed to observe statistically significant differences in treatment outcomes among enrolled patients - known as the "follow-up period." The length of the follow-up period largely depends on two factors: the natural progression of the disease, and the clinical trial endpoints required by government regulators.

Prior to marketing a new drug, firms must submit clinical trial results to the US Food and Drug Administration (FDA) documenting that their product meets a set of safety and efficacy standards. Over time, the FDA's interpretation of which clinical trial endpoints can be used to support claims that a drug is effective have varied (see, *e.g.*, Johnson, Williams and Pazdur (2003)). Conventionally, clinical trials evaluate whether a candidate product provides a clinical benefit to mortality - be it overall survival or a closely related measure such as "disease free survival," which measures time until cancer recurrence. However, in recent years there has been increased interest in using surrogate endpoints as a substitute for the standard clinical endpoints in a drug trial. In the case of hypertension, for example, lower blood pressure is accepted as a surrogate for the clinical endpoint of preventing cardiovascular complications (Lee et al. (2006)). As we discuss in Section 5.1, blood cell counts and related measures have been accepted surrogate endpoints for hematologic malignancies (leukemias and lymphomas).

Surrogate endpoints have the potential to dramatically reduce the length of clinical trials necessary to test whether a drug is effective. However, surrogate endpoints have also been controversial. As described by Fleming (2005), although treatment effects on surrogate endpoints clearly establish some form of biological activity, changes in surrogate endpoints may not always correlate with changes in the clinical endpoint of interest. As an example, he discusses prostate specific antigen (PSA) levels: although PSA levels are correlated with the extent of prostate cancer, the PSA level itself is not a mechanism through which prostate cancer progresses, and thus it is unknown whether a treatment that reduced PSA levels in prostate cancer patients would necessarily generate improvements in survival. Reflecting this type of concern, most cancers use surrogate endpoints only on a limited, somewhat *ad hoc* basis.[12]

[11] More generally, we here restrict our attention to policy mechanisms that work within the existing patent system. More sophisticated policy mechanisms - for instance in conjunction with the ideas in Kremer (1998) and Weyl and Tirole (forthcoming) - could also be used.

[12] As discussed by US Food and Drug Administration (2007) and Johnson, Williams and Pazdur (2003), since 1992 the FDA's accelerated approval regulations have allowed for the following: for diseases that are serious or life-threatening, a drug can be FDA approved based on a surrogate endpoint that is reasonably likely to predict clinical benefit but is not established at a level that would support regular approval, under the condition that the applicant is required to perform a post-marketing study to demonstrate that treatment with the drug is indeed supported with clinical benefit. If the subsequent trials fail to

In the context of our model, surrogate endpoints can be conceptualized as strictly reducing commercialization lag t_{comm}: firms can always choose to use survival as an endpoint, and we assume that the surrogate endpoint can be observed strictly earlier than the survival outcome. For simplicity, we analyze the effect of an "ideal" surrogate endpoint - one that perfectly correlates with the true clinical outcome of interest. This simplifying assumption allows us to make the following simple point.

Proposition 5. *Allowing surrogate endpoints:*

1. *Strictly increases commercialization activity: some inventions that would not otherwise have been commercialized now are, and all inventions that would be commercialized even without surrogate endpoints still are.*

2. *Strictly increases firm profits and social welfare.*

3. *Let $\hat{t}_{comm}$ denote commercialization lag, in the absence of a surrogate endpoint, based on the time required to show an effect on patient mortality. Let $t_{comm} < \hat{t}_{comm}$ denote the commercialization lag if surrogate endpoints are allowed. If t_{comm} is orthogonal to $\hat{t}_{comm}$ - that is, if the time required to show impacts on the surrogate endpoint is orthogonal to the time required to show impacts on mortality - then allowing surrogate endpoints eliminates the fixed patent term distortion associated with $\hat{t}_{comm}$.*

Clearly this proposition is based on a strong assumption of the existence of an ideal surrogate endpoint. Our objective here is simply to show that there would be social welfare benefits from the discovery of valid surrogate endpoints.

2.5.3 Policy Lever: Targeted R&D subsidies

In principle, targeted R&D subsidies can improve social welfare. The basic logic is simple and standard. Take a particular invention that is not pursued by the private sector, but that would be pursued in the first-best world, *i.e.*,

$$EML \cdot \pi < c < ETL \cdot v. \tag{8}$$

Suppose that the deadweight loss of taxation is τ per dollar spent. Then, so long as the magnitude of the potential social gain is large enough relative to the magnitude of the private loss – that is, the magnitude of the first inequality in (8) is small relative to the magnitude of the second inequality in (8) – there is a potential for welfare-increasing intervention. To state this formally, suppose that an invention's social

demonstrate clinical benefit, or if the applicant does not conduct the required studies, the FDA can act quickly to remove the drug from the market.

value while the invention is still protected by monopoly is v^{monop}, with $\pi < v^{monop} < v$ for the usual reasons. The condition is:

$$\text{R\&D subsidy socially beneficial} \iff EML \cdot \pi < c \text{ and } (1+\tau)c < EML \cdot v^{monop} + (ETL - EML) \cdot v. \quad (9)$$

While condition (9) can obtain for inventions with any commercialization lag, it is especially likely to obtain for inventions with large commercialization lags. This is because such inventions spend more of their useful life off-patent than on-patent – the ratio $\frac{ETL-EML}{ETL}$ is increasing in t_{comm} – so it is more likely that on-patent life is not sufficient to incentivize private investment, while at the same time off-patent life is of sufficient importance that the value of public investment overcomes the deadweight loss of taxation. We can formalize this logic as follows.

Proposition 6. *Suppose that the distribution of invention parameters $(p, \pi, v, \delta, \gamma, c)$ is orthogonal to an invention's commercialization lag t_{comm}, and that private firms make commercialization decisions according to whether or not $EML \cdot \pi + s > c$, where s is an amount of government subsidy. Suppose that government R&D subsidies can be conditioned on t_{comm} but not on the other invention parameters. Then socially optimal subsidy policy requires that subsidies are strictly increasing in t_{comm}.*

The intuition for the proof of this result is similar to that for Proposition 4 on optimal patent length: the higher is t_{comm}, the higher is the quality of the marginally commercialized invention, and the smaller is the cost from needlessly subsidizing inframarginal inventions. As a policy matter, conditioning subsidies on t_{comm} might be achievable either through targeting diseases with long survival times, or through providing a subsidy based on the length of time an invention spends in clinical trials.

2.6 Empirical relevance of the model to the pharmaceutical industry

The model clarifies that the fixed patent term distortion is unlikely to arise in many industries, as a series of conditions are required in order for this distortion to arise. For example, in an industry with high rates of obsolescence risk where products become irrelevant long before patent expiration, we would expect this distortion to be irrelevant. In this section, we provide some context on why these conditions are likely satisfied in the case of the pharmaceutical industry.

First, firms must face a risk of losing the ability to patent if they wait until commercialization to file. A variety of sources confirm that pharmaceutical firms file patent applications as early as possible: see, *e.g.*, Maloney et al. (2003).

Second, the innovation process must involve a time lag between invention and commercialization.

Pharmaceutical firms face commercialization lags due to the FDA-required clinical trials process. The prostate cancer clinical trials discussed in the introduction illustrate why we would expect commercialization lags to be longer for clinical trials enrolling patients with longer expected survival times: because clinical trials must generally show evidence that treatments improve mortality-related outcomes, trials tend to be longer when enrolling patients with longer survival times. In Appendix A.5, we outline a power calculation of the type used to guide the design of clinical trials in order to fix ideas on this point.

Third, patents must be an effective means of generating monopoly profits. While the importance of patents has been debated in some industries, at least three sources of evidence suggest that patents play a key role in motivating innovation in the pharmaceutical industry. First, the pharmaceutical industry is one of the few industries in which executives report in interviews and surveys that patents are essential for recovering their R&D investments (Mansfield, Schwartz and Wagner (1981); Mansfield (1986); Levin et al. (1987); Cohen, Nelson and Walsh (2000)). Second, the costs of drug development are estimated to be very high relative to the costs of producing generic copies. Although estimates of the cost of drug development vary, studies have estimated that the average capitalized costs of bringing new drugs to market are on the order of $800 million (DiMasi, Hansen and Grabowski (2003), Adams and Brantner (2006)). While imitation costs vary, traditional (small molecule) drugs are relatively inexpensive to copy once the relevant patents expire - on the order of $1-5 million per drug (Wroblewski et al. (2009)). Third, the standard investment models used by pharmaceutical firms suggest they pay close attention to the effective patent length of their products (Mayer Brown (2009)).

Fourth, the technology must be imitable once patents expire. In the pharmaceutical industry, the structure of generic entry implies that the time window during which firms can generate revenues is largely defined by the duration of their patent-provided period of market exclusivity: once generic manufacturers enter the market, pharmaceutical firms must compete with products that are practically identical to their own. Generic manufacturers are usually poised to enter the market as soon as patents expire (Grabowski and Kyle (2007); Hemphill and Sampat (2011)), and pharmaceutical firms are estimated to lose about 70% of their sales within the first six months of generic competition (Datamonitor Group (2011)).

Finally, the technology must have a long "total useful life" relative to the length of the patent term. Across industries, many inventions become commercially irrelevant long before their patents expire (Schankerman and Pakes (1986)). However, this is generally not the case in the pharmaceutical industry, as most drugs generate significant sales revenues near the end of their patent term (Grabowski and Kyle (2007)) and are still in use long after their initial FDA approval date.

Given this evidence that the conditions necessary for the fixed patent term to distort innovation fit the structure of the pharmaceutical industry, we focus our empirical work on that context. As described

in Section 3, our empirical tests focus in particular on cancer drug development.

3 Data

Our empirical work focuses on cancer R&D for three reasons. First, unlike for many diseases, high-quality clinical data exists for cancer patients which accurately tracks patient-level characteristics such as survival time - a key variable needed for our analysis. Second, the existence of a standardized classification system for cancer - namely, standardized cancer organs of origin (such as breast and lung) and stages of cancers at the time of diagnosis (such as localized and metastatic) - facilitates a relatively clean match between aggregated patient-level clinical data and information on clinical trial investments relevant to different groups of patients. Cancer drug development tends to be specific to the organ and stage of the primary tumor: for example, Genentech's drug Bevacizumab was approved by the FDA in 2004 for the treatment of patients with metastatic carcinoma of the colon and rectum.[13] Cancer registry data records the organ and stage of the primary tumor at the time of diagnosis, allowing us to estimate the characteristics of patients (such as survival times) relevant to each cancer-stage. This mapping is of course imperfect: for example, the cancer registry data lacks the granularity required to precisely distinguish between hormone-receptor positive and hormone-receptor negative breast cancer patients. However, the level of clinical detail available in cancer registry data is remarkably complete relative to data available for other diseases. Finally, as discussed in the introduction, cancer is of interest from a substantive perspective given its high morbidity and mortality burden.

Section 3.1 and Section 3.2 describe our datasets, and Section 3.4 presents some basic summary statistics. Appendix B describes our data construction in more detail.

3.1 SEER cancer registry data

The clinical data we use is a standard patient-level research database called the Surveillance, Epidemiology, and End Results (SEER) data, compiled by the National Cancer Institute (NCI) and available from 1973-2009. SEER is considered the authoritative source of information on cancer incidence and survival in the US. The key variables we use for our analysis are the following:

- Cancer and stage of patients. Physicians diagnose cancer by the organ of origin and by stages that correspond to the extent of the disease's spread at the the time of initial diagnosis. We base our data construction on the standard SEER cancer classification system (including 80 cancer types)

[13] This overly-simplified description glosses over several important issues, including off-label use of cancer drugs, which we discuss more in Appendix B.

and the stage classification system that is most consistently available in the SEER data: localized, regional, and metastatic (listed in order of increasing extent of disease).[14] In addition to constructing cancer-stage-specific survival times, we also use information on the cancer and stage of diagnosis to construct a count of the number of patients diagnosed as a proxy for market size.

- Survival time. SEER is administratively linked to follow-up mortality data from the National Center for Health Statistics (NCHS) - in our data, as of 31 December 2009. Our primary measure of survival time is five-year survival, defined over all uncensored patient cohorts (1973-2004). We also use an early cohort of patients (1973-1983) with minimal censoring in our construction of the life lost measure described below.

- Basic patient demographics at the time of diagnosis. We use the year of diagnosis together with information on patient sex and age at diagnosis to merge on year-age-gender specific life expectancy data from the National Center for Health Statistics (NCHS). We combine this data on average life expectancy (in the absence of cancer) with our measure of observed survival time for the 1973-1983 cohort in order to make rough statements about the life lost due to cancer for each patient.

3.2 National Cancer Institute clinical trials registry

To measure R&D investments in cancer treatments, we construct a new clinical trials dataset drawing on data from the US National Cancer Institute (NCI)'s Physician Data Query Cancer Clinical Trials Registry. The NCI registry was established in 1971, and claims to be the most comprehensive cancer clinical trials registry. The intended purpose of the registry is to allow cancer patients and physicians to search for clinical trials currently accepting participants, and to allow them to access information and results from closed trials.

The NCI registry was not developed as a research database and - to the best of our knowledge - has not previously been used as a data source by other researchers. The key advantage of the NCI registry for our analysis - relative to other clinical trials databases such as the *NDA Pipeline* data or the Pharmaprojects data - is the fact that the NCI registry explicitly lists which groups of patients (as defined by cancer type

[14] For more details, see the SEER training website: http://training.seer.cancer.gov/ss2k/staging/review.html. We exclude in situ cancers from our analysis given that this category is relevant for only a few cancers (breast, cervical, and melanoma), but our results are similar if these cancers are included. Two other cancer categories are important but not monitored in the patient-level cancer registry data: remission and recurrence. A cancer is said to recur if it returns after being undetectable for a period of time, and the time during which the cancer is undetectable is referred to as remission. In general, recurrence is associated with poor survival prospects, but given that the cancer registry data do not monitor remission or recurrence, it is not possible to empirically assign a survival time to these groups of patients. Reflecting this data limitation, we do not examine trials enrolling only remission or recurrence cases in our analysis. As shown in Figure 1(b), in situ and recurrent cancers fit our model well - with excellent (poor) survival prospects corresponding to few (many) clinical trials, respectively.

and stage at diagnosis) are eligible to participate in each clinical trial. This feature enables us to construct a measure of the number of clinical trials in which different groups of patients (as defined by cancer type and stage) are eligible to enroll, providing a metric of firms' willingness to investigate candidate drugs on different groups of patients.

The NCI registry includes a handful of clinical trials with dates prior to 1973; we focus on trials from 1973 forward for consistency with the SEER registry data (which starts in 1973). For a subset of clinical trials in our data, we observe whether the clinical trial was publicly sponsored or privately sponsored.

3.3 FDA drug approvals data

While our main analysis focuses on the NCI clinical trials data, we also examine a data set of the 71 FDA approved oncology drugs from 1990-2002 from Johnson, Williams and Pazdur (2003). For 39 of these 71 drug approvals, we were able to hand-collect data on whether a surrogate endpoint was used, as well as the cancer and stage for which the drug was approved, from the Drugs@FDA database.[15]

3.4 Summary statistics for cancer-stage level data

We aggregate the patient-level cancer registry data and cancer clinical trials data into cancer-stage level observations. Our sample is constructed based on the 80 cancer types underlying the SEER site recodes, and the three non-in situ stages underlying the SEER historic stage A variable - localized, regional, and metastatic. After accounting for the details of how staging varies across cancers, our benchmark cancer-stage sample includes 201 observations: 60 cancers appear for all three stages (localized, regional, distant; 180 observations); prostate cancer is coded by SEER into two stages (localized/regional, distant; 2 observations); and 19 cancers are unstaged by SEER and hence only appear as one observation (19 observations).

Table 1 presents some basic summary statistics on our cancer-stage level data. Between 1973-2011, a given cancer-stage had roughly 1,000 clinical trials, but this average masks tremendous variation - ranging from a minimum of around 200 to a maximum of over 7,000. Between 1990-2002, the median cancer-stage had no drugs approved, ranging to a maximum of 7. Using the number of patients diagnosed with a given cancer-stage as a rough measure of market size, on average a cancer-stage has around 12,000 diagnoses in SEER catchment areas between 1973-2009, ranging from 100 to over 250,000. On average, the five-year

[15] Thirty-two of the approvals in the Johnson, Williams and Pazdur (2003) list had no information available in the Drugs@FDA database on the indication for which the drug was approved, and we are not aware of an alternative source for this data. Given the coarse stage information that is included in the indication descriptions, we code stage for the drug approval data as "early," "late," or "not specified" (rather than localized, regional, and distant). In our sample of 39 approvals, four are coded as early stage, 25 are coded as late stage, and 10 are coded as not specified.

survival rate (defined for cohorts diagnosed between 1973-2004, all uncensored cohorts) is 38 percent, but ranges from almost 0 to 94 percent. Finally, among trials reporting sponsorship data, around 75 percent report being publicly financed. Given that sponsorship data is missing for approximately half of our sample, it is difficult to know whether this is an accurate picture, or whether sponsorship is more likely to be reported for publicly funded trials relative to privately financed trials. While such systematic under-reporting of private sponsorship data could bias measurement of the level or share of trials that are privately financed, we do not expect such under-reporting to vary systematically with our survival time measure - in which case our empirical tests using sponsorship measures should still be valid.

4 Empirical estimates

4.1 Analysis by stage

Figure 1(a) plots two measures of clinical trial activity for each stage of cancer from 1973-2011 against the five-year survival rate of patients diagnosed with that cancer stage from 1973-2004. Whereas metastatic cancer patients have a five-year survival rate of around 10 percent, the five-year survival rate for regional patients is around 50 percent, and for localized patients is about 70 percent. The left-hand-side axis plots the corresponding number of clinical trials enrolling patients of each stage: metastatic cancer patients were the focus of nearly 12,000 clinical trials in our data, whereas regional cancer patients were the focus of around 10,000, and localized patients around 6,000.

Dating back at least to Schmookler (1966), economists have hypothesized that market size would be an important determinant of the level of R&D investments. Several recent papers have provided evidence for this idea in the context of the pharmaceutical industry (Acemoglu and Linn (2004); Finkelstein (2004); Dubois et al. (2012); Trusheim and Berndt (2012)). In our setting, a rough proxy for market size is the number of patients diagnosed. The right-hand-side axis plots the number of clinical trials enrolling patients of each stage, divided by the number of patients diagnosed with that cancer-stage as a rough adjustment for market size. This adjustment does little to change the basic pattern.

Figure 1(b) adds clinical trial counts for three other categories of disease for which the five-year survival rate is difficult to define: prevention trials, in situ cancers, and recurrent cancers. The bars are roughly positioned in order of increasing survival rates, for comparability with Figure 1(a). Very few clinical trials aim to prevent cancer (less than 500) or to treat in situ cancers (less than 200). In contrast, recurrent cancers have more trials than any other stage of disease (over 17,000).

4.2 Analysis by cancer-stage: Full sample

Figure 2 illustrates the relationship between our two key variables of interest in the full sample of cancer-stage observations: the five-year survival rate, and the number of clinical trials enrolling patients of that cancer-stage.[16] For cancer-stages with low survival rates, there is tremendous variation in the number of clinical trials, with some cancer-stages having a very high number of trials. In contrast, for cancer-stages with high survival rates, the distribution of clinical trial counts tends to be more compressed, and smaller in magnitude. The combination of these two patterns generates the downward-sloping relationship between the survival rate and R&D investments.

Table 2 formalizes this relationship between clinical trial activity and the five-year survival rate in a regression framework. For cancer-stage observation cs, we estimate the following:

$$Y_{cs} = \alpha + \beta S_{cs} + \lambda' X_{cs} + \varepsilon_{cs}. \tag{10}$$

The number of clinical trials Y for the cancer-stage is the outcome variable, and the coefficient on the survival rate variable S is the main estimate of interest. We investigate the robustness of this relationship by conditioning on various covariates X, described below. Reflecting the count nature of the clinical trials outcome, we show estimates from quasi-maximum likelihood Poisson models. We report heteroskedasticity-robust standard errors clustered at the cancer level.

Column (1) of Table 2 reports the raw correlation between the five-year survival rate and the number of clinical trials. The estimated coefficient implies that a ten percentage point increase in the five-year survival rate is associated with a 8.7 percent decrease in R&D investments. Column (2) adds a rough market size control (measuring the number of patients diagnosed with that cancer-stage), which does not substantively change the estimate of interest. This market size variable is clearly an imperfect measure of demand. As one attempt to refine this measure, we construct a measure of life lost at the individual level - measured as age-gender-year specific life expectancy (in the absence of cancer) in the year of diagnosis, less observed survival time in years. At the individual level, this measure attempts to proxy for willingness to pay, and summed across all individuals diagnosed with a given cancer-stage it may provide a more accurate measure of market size. Column (3) shows that the survival time-R&D correlation is similar if we condition on this alternative measure of market size. In Section 5.1, we investigate the concern of

[16] To give a visual sense of the data for a few major cancers, Figure C.1 plots the relationship between the five-year survival rate and clinical trial activity for the "big four" cancers: breast, colon, lung, and prostate. Figure C.1(a) plots the number of clinical trials enrolling patients of each cancer-stage, which decline with increases in the five-year survival rates. The points are labeled with the relevant cancer and stage, which enables a visual analysis of this relationship either within cancers (*e.g.* metastatic versus localized breast cancer) or within stages (*e.g.* localized lung cancer versus localized colon cancers). Figure C.1(b) adjusts the clinical trial count by the number of patients diagnosed as a rough adjustment for market size. Here, the downward-sloping relationship between the survival rate and R&D investments is much more clearly visible.

unobserved heterogeneity in demand more directly. Figure 3 presents the visual analog of these regression specifications, residualizing the survival rate using our two measures of market size.

In an appendix, we present a number of additional robustness checks. First, we ask whether the survival time-R&D correlation is similar when estimated within cancers (cancer fixed effects) and within stages (stages fixed effects). Appendix Table C.1 shows that the magnitude of the survival time-R&D correlation is quite similar after conditioning on cancer fixed effects, stage fixed effects, or both.[17] Second, we ask whether the survival time-R&D correlation is robust to alternative measures of patient survival time. Appendix Table C.2 shows that the estimated magnitude is quite similar using the one-year survival rate, as well as several parameterizations of a "pre-period" survival rate (1973 survival in years, the 1973 one-year survival rate, and the 1973 five-year survival rate). We focus on the five-year survival rate measured over a longer time period because we expect the survival rate to be more accurately measured on a larger sample, but the estimated magnitudes are not statistically distinguishable. Third, we investigate the robustness of the survival time-R&D correlation in various sub-samples of the data in Appendix Table C.3. The estimated correlation is quite similar, for example, if metastatic cancers are excluded - suggesting that the observed correlation does not only reflect a high level of research on end-of-life patients. Finally, Appendix Table C.4 confirms that the survival time-R&D correlation also holds in our sample of approved drugs.

5 Interpreting the relationship between survival time and R&D

Section 4 documents what is - to the best of our knowledge - a new fact: R&D investments on cancer treatments are strongly negatively correlated with expected survival time. This fact is consistent with the fixed patent term distortion, but by itself is not evidence of inefficiency. We characterize two broad classes of alternative explanations which we expect to generate the same qualitative pattern: heterogeneity in demand, and heterogeneity in the costs of R&D. On demand, while our regression analysis conditioned on indirect demand measures - such as market size and life-years lost - these variables may not capture the complex ways in which the survival rate may correlate with demand (see, *e.g.*, Hammitt and Haninger 2010; Philipson et al. 2010). On costs, it could be - for example - that the "science" of treating cancer-stages with high survival rates is more difficult, and that observed low R&D investments reflect a paucity of scientific opportunities. In this section, we empirically assess these two broad classes of alternative explanations,

[17] For comparability, we omit the 19 unstaged cancers from the sample in this table since these observations do not identify the relationship of interest once we include cancer fixed effects and by definition unstaged cancers do not correspond to localized, regional, or metastatic stage definitions. Appendix Figure C.2 shows residualized scatterplots corresponding to the regression specifications presented in Appendix Table C.1 on this same sample.

and argue that the observed survival time-R&D correlation is more negative than is consistent with an efficiency argument based on demand and costs.

5.1 Heterogeneity in demand: Investigating surrogate endpoints

To investigate the concern of unobserved heterogeneity in demand, we analyze surrogate (non-mortality) clinical trial endpoints. If heterogeneity in demand for treatments were driving the survival time-R&D correlation, the observed correlation should be independent of whether surrogate endpoints are used. In contrast, the fixed patent term distortion predicts (Proposition 5) that surrogate endpoints may eliminate the negative survival time-R&D correlation. In this section, we document that there is *not* a negative survival time-R&D correlation in the sample of cancers allowed to use surrogate endpoints - a fact consistent with the fixed patent term distortion, and inconsistent with the alternative explanation of heterogeneity in demand for treatments.[18]

As discussed by the US Food and Drug Administration (2007) and Johnson, Williams and Pazdur (2003), the most clearly established non-mortality related endpoint is "complete response" for leukemias. A historical example is helpful in illustrating why this surrogate endpoint has been useful. Mukherjee (2010) chronicles Sidney Farber's 1948 discovery of chemotherapy, which was made in the context of leukemia (Farber et al. (1948)). While investigating folic acid deficiencies, Farber hypothesized that folic acid antagonists could be of value in treating cancer patients - paving the way for the development of modern chemotherapy drugs. Mukherjee (2010)'s account of Farber's discovery argues that Farber was naturally inclined to test folic acid antagonists in the context of leukemia because white blood cell count monitoring offered an accepted method for testing whether the drug was effective in pushing the disease into remission. While monitoring technologies have clearly progressed since Farber's time, remission criteria in leukemias are still based on the same idea of blood cell counts and related bone marrow measures - outcomes which are generally agreed to closely correlate with improved survival. In addition to being used for monitoring, such measures have also been accepted by the FDA as the basis for approval of drug treatments for hematologic malignancies (leukemias and lymphomas; see Pazdur (2000) and Johnson, Williams and Pazdur (2003)).

To test the predictions of Proposition 5 on the effects of surrogate endpoints, we use both our clinical trials data and our drug approvals data.[19] In the sample of approved drugs, we can confirm that hema-

[18] Importantly, surrogate endpoints enable shorter trials, so this test does not address the possibility that clinical trial costs are correlated with the survival rate; we address this possibility in a separate test in Section 5.2.

[19] We use this drug approvals data in part to address a measurement error concern that could arise with our clinical trials data. Namely, the automated coding of our clinical trials data into cancer types could be less reliable for hematologic malignancies relative to other forms of cancer if text searches for organ names ("breast," "prostate," *etc.*) are more accurate than our text searches for different forms of leukemias and lymphomas (the names of which tend to be more complex). While

tological malignancies are more likely to be approved on the basis of surrogate endpoints: in our data, 92 percent of drugs approved by the FDA for hematological malignancies were approved on the basis of surrogate endpoints, relative to 53 percent of non-hematological malignancies.

We use these data to test the two predictions of Proposition 5 that relate to commercialization activity. Part 1 of Proposition 5 predicts that the use of surrogate endpoints should increase commercialization activity. To test this prediction, we ask whether - conditional on the five-year survival rate - hematological malignancies have a larger number of clinical trials. The estimated coefficient in Column (1) of Panel (A) in Table 3 suggests yes: interpreting the coefficient on this binary independent variable ($\beta = 0.339$) suggests a 40 percent increase in clinical trials for hematological malignancies relative to non-hematological malignancies ($(e^{\beta} - 1) \cdot 100 \approx 40\%$). This pattern is robust to the inclusion of controls for market size (Columns (2) and (3)). This result is consistent with the analysis of Trusheim and Berndt (2012), who observe that hematological malignancies have a larger number of clinical trials than would be expected based on their market size.

Part 3 of Proposition 5 predicts that - if survival time is orthogonal to the time required to show impacts on the surrogate endpoint - then the use of surrogate endpoints should eliminate the negative relationship between survival time and R&D investments. To test this prediction we estimate the following specification, where H_c is an indicator for hematological malignancies:

$$Y_{cs} = \alpha + \beta S_{cs} \cdot H_c + \gamma H_c + \delta S_{cs} + \lambda' X_{cs} + \varepsilon_{cs}. \tag{11}$$

Panel (B) in Table 3 presents these estimates. In contrast to the negative correlation between the five-year survival rate and the number of clinical trials for non-hematological malignancies (δ), we estimate a positive coefficient on the interaction term (β) - consistent with the idea that surrogate endpoints can eliminate the fixed patent distortion. The estimated interaction coefficient implies that in the sample of hematological malignancies we in fact observe a slight positive relationship between the five-year survival rate and the number of clinical trials.[20] This estimate is robust to the inclusion of controls for market size (Columns (2) and (3)). Appendix Table C.5 shows that this pattern of results also holds in the drug approvals data.

We draw two conclusions from the analysis in this section. First, the lack of a negative relationship between survival time and R&D investments among cancers allowed to use surrogate endpoints suggests

we aimed for the highest possible accuracy in cleaning the clinical trials data, because of the large sample size our cleaning of that data must be automated. In contrast, because there are a small number of drug approvals we can hand-code the cancer types relevant to each drug approval, reducing concerns about measurement error.

[20] Interpreting the interaction term in this non-linear model requires transforming the coefficient: blood cancers receive around 1,500 additional clinical trials, an increase of 160 percent over the sample mean.

that heterogeneity in demand for treatments is not driving the observed correlation between survival time and R&D investments in the full sample. Second, given the empirical support we find for the testable predictions of Proposition 5, these estimates support the idea that the establishment of valid surrogate endpoints may have a large social return.

5.2 Heterogeneity in costs: Investigating publicly financed clinical trials

To investigate the concern of unobserved heterogeneity in the costs of R&D, we contrast publicly and privately financed clinical trials. We rely on the fact that higher costs also reduce social R&D incentives, so that any *additional* negative correlation between the survival rate and R&D investments among privately financed trials should be purged of heterogeneity in the costs of clinical trials that is correlated with the survival rate. We document in this section that the relationship between survival time and R&D is 30 percent larger for privately financed trials relative to publicly financed trials.

As a first analysis of our trial sponsorship data, Panel (A) of Figure 5 presents the cumulative distribution functions (CDF) of clinical trial lengths in the trial-level data, separately for privately financed and publicly financed trials. The privately financed CDF lies above the publicly financed CDF at almost every clinical trial length. The vertical line at twenty years denotes the length of the fixed patent term: consistent with the idea that the patent system should offer zero incentive to develop drug compounds that take longer than twenty years to develop, very few trials in our data have a reported length of twenty years or longer. Of the approximately 120 clinical trials longer than 20 years that have non-missing data on sponsorship, 95 percent are publicly funded.[21]

Panel (B) of Figure 5 provides a second analysis of this sponsorship data, plotting the relationship between the five-year survival rate and the share of clinical trials enrolling patients of that cancer-stage which are privately financed.[22] The downward-sloping relationship is quantified in Panel (A) of Table 4: a ten percentage point increase in the five-year survival rate is associated with a 1.2 percent decrease in the share of clinical trials that are privately financed. The magnitude of this coefficient is quite similar conditional on our market size controls (Columns (2) and (3)).

Panel (B) of Table 4 presents estimates from a second test of how public and private R&D investments differ. Estimating Equation 10 separately on the sample of publicly financed trials and on the sample

[21] The longest privately financed trial in our data lasts 18.66 years, with the exception of six trials that are reported to last longer than sixty years. We suspect that these six trials have typographical errors in their start dates, but have not yet heard back from the sponsor (Bristol-Myers Squibb) in an inquiry on this point. If these six trials have typographical errors as we expect, then 100 percent of the trials with non-missing data on sponsorship that are longer than 20 years are publicly funded.

[22] In interpreting the scale of the graph, recall that as noted in Section 3.4 we suspect that sponsorship data is more likely to be reported for publicly funded trials relative to privately financed trials.

of privately financed trials, we would like to compare the estimated β coefficients to see whether the correlation between survival time and clinical trial activity is smaller in the sample of publicly financed trials relative to the sample of privately financed trials. Formally equivalent to estimating these two regressions separately is estimating a stacked regression where the unit of observation is a cancer-stage-type *cst* (where type is either privately financed or publicly financed):

$$Y_{cst} = \alpha + \beta S_{cs} \cdot T_t + \gamma T_t + \delta S_{cs} + \lambda' X_{cs} \cdot T_t + \varepsilon_{cst}. \tag{12}$$

Our T_t variable is defined as an indicator which equals one for observations counting privately financed trials, and equals zero for observations counting publicly financed trials. The coefficient of interest β measures the difference in the survival time-clinical trial activity correlation observed for privately financed trials relative to that observed for publicly financed trials.

These estimates are presented in Panel (B) of Table 4. The negative β estimate implies that the relationship between the five-year survival rate and R&D investments is more negative for privately financed trials relative to publicly financed trials - consistent with what we expected based on the analyses in Panel (B) of Figure 5. Interpreting the point estimate in Column (1) suggests that a ten percentage point increase in the five-year survival rate results in an additional 4.4 percent decrease in privately financed clinical trials, in addition to the 8.6 percent decrease observed for publicly financed clinical trials. These estimates imply that the relationship between survival time and clinical trial activity is on the order of 30 percent larger for privately financed clinical trials relative to publicly financed clinical trials ($\frac{4.4}{4.4+8.6} \approx 30\%$). The point estimates and their ratio are quite stable across specifications adding our market size controls (Columns (2) and (3)).

We wish to make two remarks concerning these estimates. First, this public-private contrast is consistent with two potential explanations: the public sector could - as in our model - have a different objective function than the private sector, or the public sector could be compensating for under-investment by the private sector. Both explanations are consistent with the existence of a fixed patent term distortion, and thus have the same qualitative interpretation, but the quantitative interpretation of the estimates would differ across the two models. Second, to the extent that a large share of publicly financed clinical trials investigate new uses of existing drugs, publicly financed trials may be constrained by "science" to mirror privately financed R&D investments.

Under the assumptions of our model, the direct (non-interacted) five-year survival rate should control for any differences in clinical trial costs that vary systematically with patient survival time. An important caveat to this interpretation is that in our model the firm and society apply a common discount factor to

each project. In a frictionless world, discount rates depend on the cash flows associated with a project and do not depend on who is financing the project. Of course, the world is not frictionless, and private firms are generally agreed to have higher discount rates than the government due to various frictions such as corporate short-termism (Stein (1989)). Given the lack of empirical evidence on corporate short-termism (Stein (2003)), it is difficult to assess the importance of this issue in our context. Importantly, the distinction between discounting and the fixed patent distortion is relevant to some but not all of the policies we analyze: starting the patent term at commercialization would address the fixed patent term distortion but not discounting, whereas surrogate endpoints and targeted R&D subsidies would address both.

5.3 Qualitative evidence

A first set of qualitative evidence supporting the idea that fixed patent terms distort innovation comes from the fact that academic clinicians, clinical researchers, and firms all report a reluctance to invest in drugs that require lengthy clinical trials because of their shorter effective patent life. For example, a medicinal chemistry textbook notes: "...patents normally run for 20 years from the date of application, ...some compounds are never developed because the patent protected production time available to recoup the cost of development is too short" (Thomas, 2003). We interviewed several venture capitalists for this paper, and while their confidentiality agreements with the companies they evaluate prevented them from naming any specific examples of drugs that failed to reach the market for this reason, they each claimed that it happens "all the time." Below are two excerpts from their comments:

- *The running clock on these patents is a huge deal. Companies absolutely choose not to go forward with drugs because their remaining patent life isn't sufficient. In the models we use to decide whether to in-license a drug, we measure the time and cost of doing the trials against the period of exclusivity and time until peak market share. You have a pretty good sense of how long it will take to get to approval, at least by the time you're in the Phase 2A trials, so these things happen pretty early in development. Companies de-prioritize those drugs. Quite often we've declined to take advantage of an opportunity because we thought there wouldn't be enough time under the patent term to earn a return on the investment.* (Venture Capitalist A)

- *The shorter the remaining patent term, the more certainty you need that the drug will work, and the more it needs to have a large market. Also, the ramp is important. You want at least a couple years of peak sales. It happens all the time that we pass on a drug, one we think would probably work, because there wouldn't be enough life left on its patent by the time it reached the market.* (Venture

Capitalist B)

A second set of qualitative evidence supporting the idea that fixed patent terms distort innovation comes from an investigation of what motivated the development of existing chemoprevention drugs. Given that cancer prevention trials typically examine cancer incidence as an outcome variable rather than cancer mortality, these trials do not mechanically fit into our conceptual framework, but we expect cancer prevention technologies to generally require long trials and thus to also be under-incentivized due to the fixed patent distortion. We start with the list of all - six - FDA approved chemoprevention drugs compiled by Meyskens et al. (2011). Our qualitative investigation of the history of these FDA drug approvals suggests that all six of these approvals either relied on the use of surrogate endpoints, or were approved on the basis of publicly financed clinical trials. In Appendix D, we provide documentation for this assertion, and here briefly summarize two of the case studies. First, the drug Tamoxifen was FDA approved for several cancer indications while on-patent; later, a publicly-funded clinical trial supported the 1998 FDA approval of Tamoxifen as a chemoprevention agent - preventing breast cancer incidence in high-risk groups. Second, the recent FDA approval of cervical cancer vaccines relied on the use of human papillomavirus (HPV) incidence as a surrogate endpoint for cervical cancer incidence.

6 Discussion and conclusion

In this paper, we ask whether fixed patent terms distort innovation. Our theoretical model clarifies a set of necessary conditions for this distortion to exist. Although unlikely to arise in many industries, these conditions naturally fit the pharmaceutical industry. Our empirical work focuses on cancer treatments, where drugs treating patients with short life expectancies can move through clinical trials more quickly - and thus receive longer effective patent terms - than drugs treating patients with long life expectancies. Using a newly constructed data set on cancer clinical trial investments, we provide several sources of evidence consistent with fixed patent terms distorting cancer R&D.

We also use our theoretical model to analyze the innovation and social welfare consequences of three policy interventions which could address the fixed patent term distortion: a patent design change that would start the patent clock at commercialization, a policy change that would allow firms to rely on surrogate endpoints in clinical trials, and targeted R&D subsidies. Our empirical analysis suggests that research establishing valid surrogate endpoints may have a large social return.[23]

The example of the Framingham Heart Study is helpful in illustrating the potential value of surrogate

[23]These results support informal arguments made by the cancer research community; see, for example, Korn and Stanski (2005), US Institute of Medicine (2008), Lathia et al. (2009), and American Society for Clinical Oncology (2011).

endpoints. Heart disease is the leading cause of death in the US, but since 1968 the age-adjusted rate of deaths from heart disease has dropped by 50 percent.[24] Although some of these gains are due to lifestyle changes, much of the decline in heart disease has been attributed to improved pharmacological treatments for cardiovascular disease, including the development of beta-blockers, ACE-inhibitors, and statins (Weisfelt and Zieman (2007)). Patients use these drugs to reduce the morbidity and mortality from heart disease, but very few of these drugs reached the market based on clinical trials using morbidity or mortality as the endpoint. Rather, almost all were approved based on evidence that these drugs lowered either blood pressure or LDL (low-density lipoprotein) cholesterol - outcomes that can be measured much more quickly than morbidity and mortality (Psaty et al. (1999)). These surrogate endpoints were first identified by the Framingham Heart Study, a large-scale, multi-decade, federally-funded observational study which found that high blood pressure and LDL cholesterol are critical risk factors in cardiovascular disease. Subsequent clinical trials helped to validate these prognostic factors, which led the FDA to accept them as surrogate endpoints in cardiovascular trials (Meyskens et al. (2011)). Researchers have argued that without these surrogate endpoints, it is unclear whether drugs such as beta-blockers, ACE-inhibitors, and statins would have reached the market as treatments for heart disease (Lathia et al. (2009); Meyskens et al. (2011)).

[24] See, for example, the discussion in Cutler and Kadiyala (2003).

References

Abramowicz, Michael, "The danger of underdeveloped patent prospects," *Cornell Law Review*, 2007, *92*, 1065–1121.

Acemoglu, Daron and Joshua Linn, "Market size in innovation: Theory and evidence from the pharmaceutical industry," *Quarterly Journal of Economics*, 2004, *119*, 1049–1090.

Adams, Christopher and Van Brantner, "Estimating the cost of new drug development: Is it really $802 million?," *Health Affairs*, 2006, *25* (2), 420–428.

Agha, Leila and David Molitor, "Technology adoption under uncertainty: Off-label prescribing in cancer care," mimeo, Boston University 2012.

Allison, Malorye, "Reinventing clinical trials," *Nature Biotechnology*, 2012, *30* (1), 41–49.

American Joint Committee on Cancer, "AJCC Cancer Staging Manual, Seventh Edition," Springer 2010.

American Society for Clinical Oncology, "Accelerating Progress Against Cancer: ASCO's Blueprint for Transforming Clinical and Translational Cancer Research," http://www.asco.org/ASCOv2/Department%20Content/Cancer%20Policy%20and%20Clinical%20Affairs/Downloads/Blueprint.pdf 2011.

Bartfai, Tamas and Graham Lees, *Drug Discovery: From Bedside to Wall Street*, Academic Press, 2006.

Chin, Richard and Menghis Bairu, eds, *Global Clinical Trials: Effective Implementation and Management*, Academic Press, 2011.

Cohen, Wesley, Richard Nelson, and John Walsh, "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)," National Bureau of Economic Research (NBER) working paper 7552 2000.

Collett, David, *Modelling Survival Data in Medical Research, Second Edition*, Chapman and Hall/CRC, 2003.

Cutler, David and Srikanth Kadiyala, "The Return to Biomedical Research: Treatment and Behavioral Effects," in Kevin Murphy and Robert Topel, eds., *Measuring the Gains from Medical Research: An Economic Approach*, London, UK: University of Chicago Press, 2003, pp. 110–162.

Datamonitor Group, "US most susceptible to brand erosion post patent expiry," Datamonitor Press Release http://about.datamonitor.com/media/archives/5293 2011.

de Bono, Johann, Christopher Logothetis, Arturo Molina, Karim Fizazi, Scott North, Luis Chu, Kim Chi, Robert Jones, Oscar Goodman, Fred Saad, John Staffurth, Paul Mainwaring, Stephen Harland, Thomas Flaig, Thomas Hutson, Tina Cheng, Helen Patterson, John Hainsworth, Charles Ryan, Cora Sternberg, Susan Ellard, Aude Fléchon, Mansoor Saleh, Mark Scholz, Eleni Efstathiou, Andrea Zivi, Diletta Bianchini, Yohann Loriot, Nicole Chieffo, Thian Kheoh, Christopher Haqq, and Howard I. Scher, "Abiraterone and increased survival in metastatic prostate cancer," *New England Journal of Medicine*, 2011, *364* (21), 1995–2005.

DiMasi, Joseph, Ronald Hansen, and Henry Grabowski, "The price of innovation: New estimates of drug development costs," *Journal of Health Economics*, 2003, *22*, 151–185.

Dubois, Pierre, Fiona Scott Morton, Paul Seabright, and Olivier de Mouzon, "Market size and pharmaceutical innovation," mimeo, Toulouse School of Economics 2012.

Eisenberg, Rebecca, "The problem of new uses," *Yale Journal of Health Policy, Law, & Ethics*, 2005, *717*.

Farber, Sidney, Louis Diamond, Robert Mercer, Robert Sylvester, and James Wolff, "Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid (aminopterin)," *New England Journal of Medicine*, 1948, *238* (23), 787–793.

Finkelstein, Amy, "Static and dynamic effects of health policy: Evidence from the vaccine industry," *Quarterly Journal of Economics*, 2004, *119*, 527–564.

Fisher, Bernard, Joseph Costantino, D. Lawrence Wickerham, Carol Redmond, Maureen Kavanah, Walter Cronin, Victor Vogel, Andre Robidoux, Nikolay Dimitrov, James Atkins, Mary Daly, Samuel Wieand, Elizabeth Tan-Chiu, Leslie Ford, Norman Wolmark, other National Surgical Adjuvant Breast, and Bowel Project Investigators, "Tamoxifen for prevention of breast cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study," *Journal of the National Cancer Institute*, 1998, *90* (18), 1371–1388.

Fleming, Thomas, "Surrogate endpoints and FDA's accelerated approval process," *Health Affairs*, 2005, *24*, 67–78.

Gilbert, Richard and Carl Shapiro, "Optimal patent length and breadth," *RAND Journal of Economics*, 1990, *21*, 106–112.

Goffin, John, "Introduction to clinical trials," in Shayne Cox Gad, ed., *Clinical Trials Handbook*, Wiley, 2009, pp. 1–22.

Grabowski, Henry and Margaret Kyle, "Generic competition and market exclusivity periods in pharmaceuticals," *Managerial and Decision Economics*, 2007, *28* (4-5), 491–502.

Grossman, Gene and Edwin Lai, "International protection of intellectual property," *American Economic Review*, 2004, *94*, 1635–1653.

Hammitt, James and Kevin Haninger, "Valuing fatal risks to children and adults: Effects of disease, latency, and risk aversion," *Journal of Risk and Uncertainty*, 2010, *40* (1), 57–83.

Hemphill, C Scott and Bhaven Sampat, "Evergreening, patent challenges, and effective market life in pharmaceuticals," *Journal of Health Economics*, 2011, *31* (2), 327–339.

Hovde, Mark, "Chapter 7: Management of clinical development costs," in Chi-Jen Lee, Lucia Lee, Christopher Wu, Benjamin Lee, and Mei-Ling Chen, eds., *Clinical Trials of Drugs and Biopharmaceuticals*, CRC Press, 2006, chapter 7, pp. 97–102.

Johnson, John, Grant Williams, and Richard Pazdur, "End points and United States Food and Drug Administration approval of oncology drugs," *Journal of Clinical Oncology*, 2003, *21*, 1404–1411.

Jones, Christopher, Daniel Hunt, David McGowan, Mahul Amin, Michael Chetner, Deborah Bruner, Mark Leibenhaut, Siraj Husain, Marvin Rotman, Luis Souhami, Howard Sandler, and William Shipley, "Radiotherapy and short-term androgen deprivation for localized prostate cancer," *New England Journal of Medicine*, 2011, *365* (2), 107–118.

Kaplow, Louis, "The patent-antitrust intersection: A reappraisal," *Harvard Law Review*, 1984, *97*, 1815–1892.

Kitch, Edmund, "The nature and function of the patent system," *Journal of Law and Economics*, 1977, *20*, 265–290.

Klemperer, Paul, "How broad should the scope of patent protection be?," *RAND Journal of Economics*, 1990, *21*, 113–130.

Kohan, Donald, John Cleland, Dan Theodorecsuc, Lewis Rubin, and Matthias Barton, "Clinical trials with endothelin receptor antagonists: What went wrong and where can we improve?," *Life Sciences*, 2012.

Korn, David and Donald Stanski, *Drug Development Science: Obstacles and Opportunities for Collaboration Among Academia, Industry and Government*, Association of American Medical Colleges, 2005.

Kremer, Michael, "Patent buyouts: A mechanism for encouraging innovation," *The Quarterly Journal of Economics*, 1998, *113* (4), 1137–1167.

Lakdawalla, Darius, Eric Sun, Anupam Jena, Carolina Reyes, Dana Goldman, and Tomas Philipson, "An economic evaluation of the war on cancer," *Journal of Health Economics*, 2010, *29* (3), 333–346.

Lamm, Donald, Brent Blumenstein, E. David Crawford, James Montie, Peter Scardino, H. Barton Grossman, Thomas Stanisic, Joseph Smith, Jerry Sullivan, Michael Sarosdy, John Crissman, and Charles Coltman, "A randomized trial of intravesical doxorubicin and immunotherapy with bacille calmette-guerin for transitional-cell carcinoma of the bladder," *New England Journal of Medicine*, 1991, *325* (17), 1205–1209.

Lathia, CD, D Amakye, W Dai, C Girman, S Madani, J Mayne, P MacCarthy, P Pertel, L Seman, A Stoch, P Tarantino, C Webster, S Williams, and JA Wagner, "The value, qualification, and regulatory use of surrogate end points in drug development," *Clinical Pharmacology & Therapeutics*, 2009, *86* (1), 32–43.

Leary, Warren, "F.D.A. backs a treatment for cancer of bladder," *New York Times*, 1990, *22 May.*

Lee, Benjamin, Chi-Jen Lee, Christpher Wu, and Lucia Lee, "Decision points in drug development," in Chi-Jen Lee, Lucia Lee, Christopher Wu, Benjamin Lee, and Mei-Ling Chen, eds., *Clinical Trials of Drugs and Biopharmaceuticals*, CRC Press, 2006, chapter 6, pp. 77–86.

Lerner, Josh, "Patent protection and innovation over 150 years," National Bureau of Economic Research (NBER) working paper 8977 2002.

Leveque, Dominique, "Off-label use of anticancer drugs," *Lancet Oncology*, 2008, *9*, 1102–1107.

Levin, Richard, Alvin Klevorick, Richard Nelson, Sidney Winter, Richard Gilbert, and Zvi Griliches, "Appropriating the returns from industrial research and development," *Brookings Papers on Economic Activity*, 1987, *1987* (3), 783–831.

Lindgren, Peter and Bengt Jonsson, "Cost-effectiveness of statins revisited: Lessons learned about the value of innovation," *European Journal of Health Economics*, 2012, *13* (4), 445–450.

Lippman, Scott and Ernest Hawk, "Cancer prevention: From 1727 to milestones of the past 100 years," *Cancer Research*, 2009, *69*, 5269.

_ **and Powel Brown**, "Tamoxifen prevention of breast cancer: An instance of the fingerpost," *Journal of the National Cancer Institute*, 1999, *91*, 1809–1819.

Machlup, Fritz, "An economic review of the patent system," Study of the Subcommittee on Patents, Trademarks, and Copyrights of the Committee on the Judiciary, United States Senate, 85th Congress, Second Session 1958.

Malani, Anup and Tomas Philipson, "Can medical progress be sustained? Implications of the link between development and output markets," National Bureau of Economic Research (NBER) working paper 17011 2011.

Maloney, Timothy, Richard Kaba, James Krueger, Rady Kratz, and Calista Mitchell, "Intellectual property in drug discovery and biotechnology," in Donald Abraham, ed., *Burger's Medicinal Chemistry & Drug Discovery*, 6 ed., Vol. 2 2003.

Mansfield, Edwin, "Patents and innovation: An empirical study," *Management Science*, 1986, *32*, 173–181.

_ , **Mark Schwartz, and Samuel Wagner**, "Imitation costs and patents: An empirical study," *The Economic Journal*, 1981, *91* (364), 907–918.

Mayer Brown, "Pharma & biotech - drug development: Valuing the pipeline: A UK study," http://www.mayerbrown.com/files/Publication/53189dc5-d8f8-45aa-9d4d-62646b17c2e8/Presentation/PublicationAttachment/c7199cda-f73d-4377-a09d-cc26afc93408/0728ten_Valuation_Report.pdf 2009.

Mcintyre, William, Michael Downs, and Sondra Bedwell, "Treatment options for actinic keratoses," *American Family Physician*, 2007, *76* (5), 667–671.

Meyskens, Frank, Gregory Curt, Dean Brenner, Gary Gordon, Ronald Herberman, Olivera Finn, Gary Kelloff, Samir Khleif, Caroline Sigman, and Eva Szabo, "Regulatory approval of cancer risk-reducing (chemopreventive) drugs: Moving what we have learned into the clinic," *Cancer Prevention Research*, 2011, *4* (3), 311–323.

Moser, Petra, "How do patent laws influence innovation? Evidence from nineteenth-Century world fairs," *American Economic Review*, 2005, *95*, 1214–1236.

Mukherjee, Siddhartha, *The Emperor of All Maladies: A Biography of Cancer*, New York, NY: Scribner, 2010.

Murphy, Kevin and Robert Topel, "The value of health and longevity," *Journal of Political Economy*, 2006, *114*, 871–904.

Nordhaus, William, *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*, Cambridge, MA: MIT Press, 1969.

_ , "The optimum life of a patent: Reply," *American Economic Review*, 1972, *62* (3), 428–431.

Pazdur, Richard, "Response rates, survival, and chemotherapy trials," *Journal of the National Cancer Institute*, 2000, *92* (19), 1552–1553.

Philipson, Tomas and Anupam Jena, "Who benefits from new medical technologies? Estimates of consumer and producer surpluses for HIV/AIDS drugs," *Forum for Health Economics & Policy*, 2006, *9* (2).

_ , **Gary Becker, Dana Goldman, and Kevin Murphy**, "Terminal care and the value of life near its end," National Bureau of Economic Research (NBER) working paper 15649 2010.

Psaty, Bruce, Noel Weiss, Curt Furberg, Thomas Koepsell, David Siscovick, Frits Rosendaal, Nicholas Smith, Susan Heckbert, Robert Kaplan, Danyu Lin, Thomas Fleming, and Edward Wagner, "Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease," *Journal of the American Medical Association*, 1999, *282* (8), 786–790.

Public Law Number 98-417, "Drug Price Competition and Patent Term Restoration Act ("Hatch-Waxman") Act of 1984," 98 Stat. 1585 1984.

Roin, Benjamin, "Drug patent length," mimeo, Harvard Law School 2010.

Sakakibara, Mariko and Lee Branstetter, "Do stronger patents induce more innovation? Evidence from the 1998 Japanese patent law reforms," *RAND Journal of Economics*, 2001, *32* (1), 77–100.

Schankerman, Mark and Ariel Pakes, "Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period," *The Economic Journal*, 1986, *96* (384), 1052–1076.

Scherer, F.M., "Nordhaus' theory of optimal patent life: A geometric reinterpretation," *American Economic Review*, 1972, *62*, 422–427.

Schmookler, Jacob, *Invention and Economic Growth*, Cambridge, MA: Harvard University Press, 1966.

Scotchmer, Suzanne, "Standing on the shoulders of giants: Cumulative research and the patent law," *Journal of Economic Perspectives*, 1991, *5*, 29–41.

Stein, Jeremy, "Efficient capital markets, inefficient firms: A model of myopic corporate behavior," *Quarterly Journal of Economics*, 1989, *104*, 655–669.

_ , "Agency, information, and corporate investment," in "Handbook of the Economics of Finance," Elsevier, 2003, chapter 2.

Thomas, Gareth, *Fundamentals of Medicinal Chemistry*, John Wiley & Sons, Ltd, 2003.

Trusheim, Mark and Ernst Berndt, "The segmentation of therapeutic populations in oncology," mimeo, MIT 2012.

US Food and Drug Administration, "Guidance for industry: Clinical trial endpoints for the approval of cancer drugs and biologics," http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf 2007.

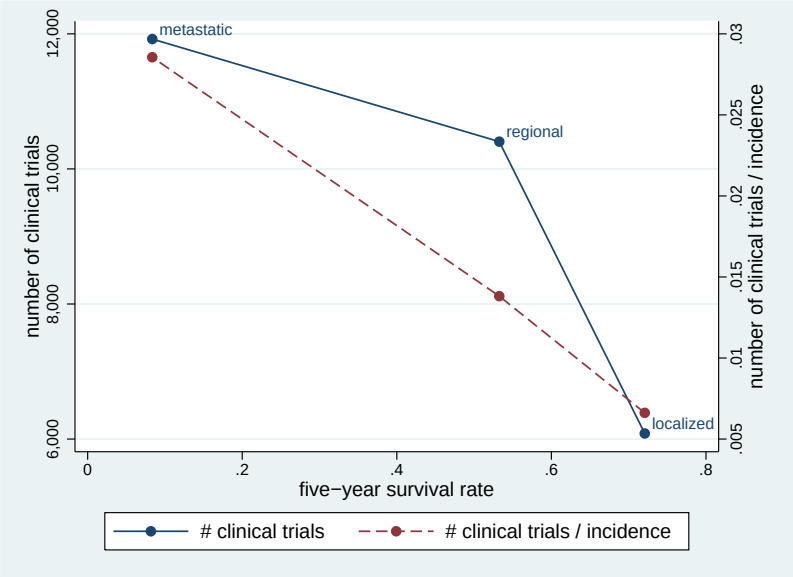
US Institute of Medicine, *Improving the quality of cancer clinical trials: Workshop summary*, The National Academies Press, 2008.

Weisfelt, Myron and Susan Zieman, "Advances in the prevention and treatment of cardiovascular disease," *Health Affairs*, 2007, *26* (1), 25–37.

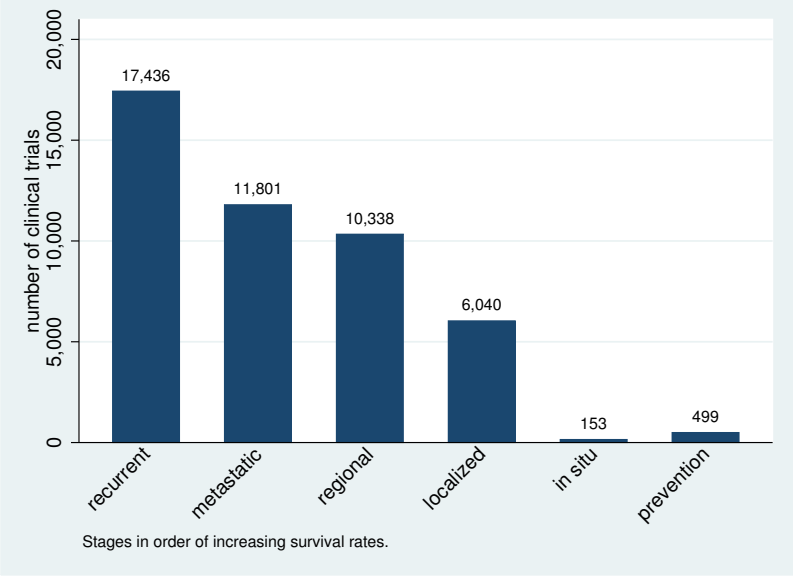
Weyl, Glen and Jean Tirole, "Market power screens willingness-to-pay," Quarterly Journal of Economics forthcoming.

Wroblewski, Michael, Elizabeth Jex, Suzanne Drennon Munck, Christopher Garmon, Suzanne Michel, Alissa Wantman, and Ossie Neal, "Emerging health care issues: Follow-on biological drug competition," Federal Trade Commission Report http://www.ftc.gov/os/2009/06/P083901biologicsreport.pdf 2009.

Figure 1: **Survival time and R&D investments: Stage-level data**



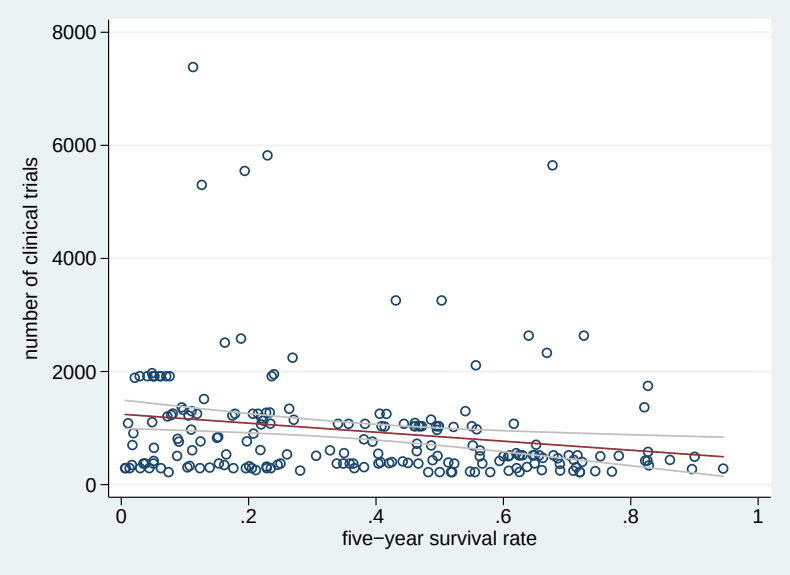
(a) R&D investments by five-year survival rates



(b) R&D investments by stage

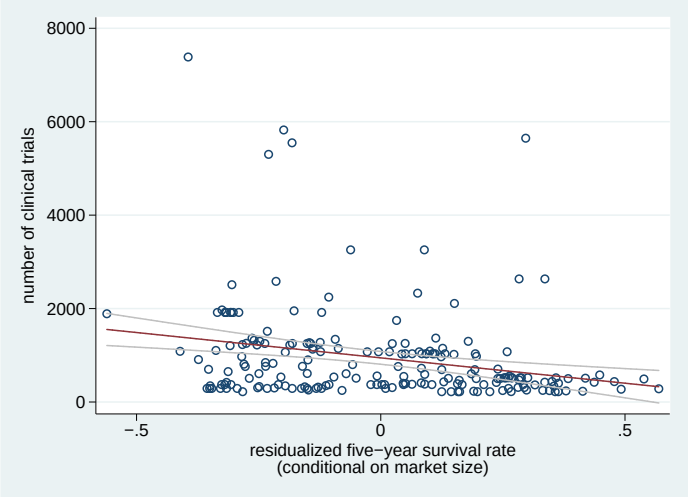
Notes: This figure plots measures of clinical trial activity for each stage of cancer from 1973-2011. Panel (a) plots two measures of clinical trial activity for each stage of cancer from 1973-2011 against the five-year survival rate among patients diagnosed with each stage between 1973-2004 (the cohorts for which five-year survival is uncensored). The left-hand-side axis plots the number of clinical trials enrolling patients of each stage from 1973-2011. The right-hand-side axis plots the number of clinical trials enrolling patients of each stage from 1973-2011 divided by the number of patients diagnosed with that stage from 1973-2009, as a rough adjustment for market size. Panel (b) is a bar chart plotting the same data for localized, regional, and metastatic cancers, but also including the number of trials for preventive technologies as well as in situ and recurrent cancers. For details on the sample, see the text and data appendix.

Figure 2: **Survival time and R&D investments: Cancer-stage data**

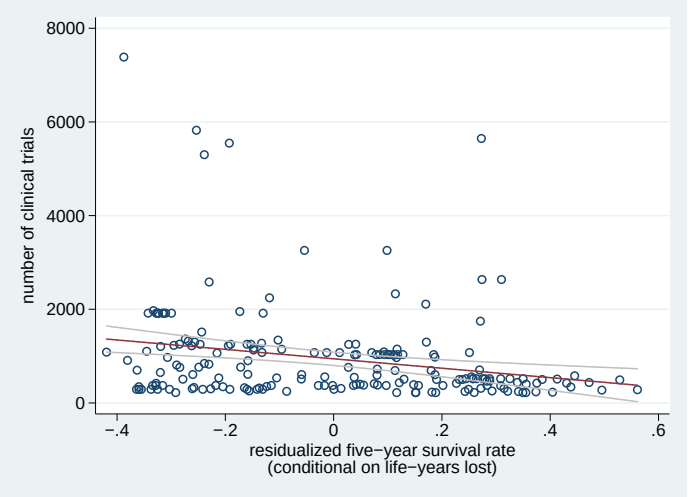


Notes: This figure shows the relationship between the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the number of clinical trials enrolling patients of each cancer-stage from 1973-2011. The level of observation is the cancer-stage. For details on the sample, see the text and data appendix.

Figure 3: **Survival time and R&D investments: Residualized cancer-stage data**



(a) Market size



(b) Life-years lost

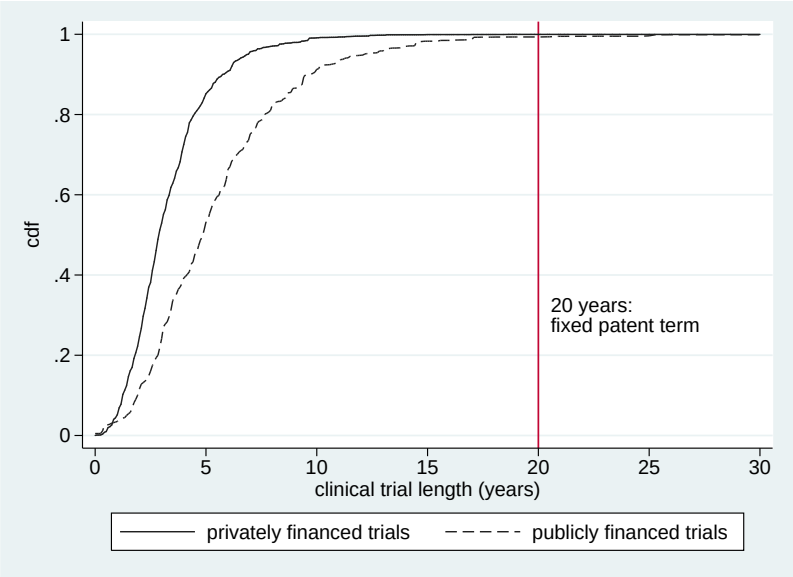
Notes: This figure shows the relationship between residualized versions of the five-year survival rate among patients diagnosed with that cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the number of clinical trials enrolling patients of each cancer-stage from 1973-2011. The level of observation is the cancer-stage. Panel (a) residualizes market size; Panel (b) residualizes life-years lost. *Market size* denotes the inclusion of a covariate measuring the number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* is measured as age-gender-year specific life expectancy (in the absence of cancer) in the year of diagnosis, less observed survival time in years, averaged over patients diagnosed with that cancer-stage between 1973-1983 (to minimize censoring) multiplied times market size. For details on the sample, see the text and data appendix.

Figure 4: **Surrogate endpoints, survival time, and R&D investments**

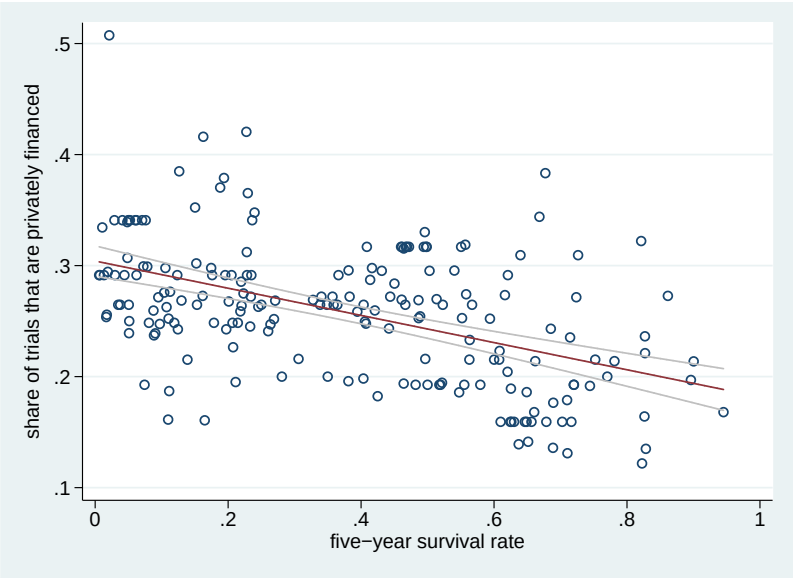


Notes: This figure shows the relationship between the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the number of clinical trials enrolling patients of each cancer-stage from 1973-2011, separately for hematologic and non-hematologic cancers. The level of observation is the cancer-stage. For details on the sample, see the text and data appendix.

Figure 5: **Survival time and financing of clinical trials**



(a) Cumulative distribution function of clinical trial length



(b) Survival time and private financing

Notes: This figure shows two analyses of how public and private financing of clinical trials differ. Panel (a) plots the cumulative distribution function of clinical trial length in years, omitting the handful of observations with length greater than 30 years for improved readability. The level of observation is the clinical trial. The vertical line at 20 years denotes the length of the fixed patent term. Panel (b) plots the relationship between the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the share of clinical trials enrolling patients of that cancer-stage from 1973-2011 that were privately financed. The level of observation is the cancer-stage. In interpreting the scale of the graph, recall that as noted in Section 3.4 we suspect that sponsorship data is more likely to be reported for publicly funded trials relative to privately financed trials. For details on the sample, see the text and data appendix.

Table 1: **Summary statistics: Cancer-stage data**

	mean	median	standard deviation	minimum	maximum
number of clinical trials, 1973-2011	945	556	1,015	221	7,385
number of drug approvals, 1990-2002	0.507	0	1.221	0	7
five-year survival rate, cases diagnosed 1973-2004	0.377	0.383	0.249	0.006	0.945
number of diagnoses (1000s), 1973-2009	12.423	3.159	29.429	0.010	252.593
estimated years of life lost (1000s), 1973-1983	114.433	35.663	233.576	0.583	1,658.804
share of trials privately financed	0.258	0.265	0.062	0.122	0.507

Notes: This table shows summary statistics for our cancer-stage level data. The level of observation is the cancer-stage. The clinical trials data is available from 1973-2011. The drug approvals data is available from 1990-2002. The SEER data starts in 1973 and ends in 2009, which is why the number of diagnoses variable is measured over that time period. The five-year survival rate is calculated over patients diagnosed between 1973-2004, the cohorts for which five-year survival is uncensored as of 2009. The life years lost measure is calculated on cohorts diagnosed from 1973-1983 to minimize censoring, as explained in the text. As explained in the text, we suspect that sponsorship data is more likely to be reported for publicly funded trials relative to privately financed trials. All variables have 201 observations except for the life lost measure which as 192, because 9 cancer-stages had no patients diagnosed between 1973-1983. For details on the sample, see the text and data appendix.

Table 2: **Survival time and R&D investments: Cancer-stage data**

Dependent variable: Number of clinical trials (mean = 945)						
	(1)		(2)		(3)	
five-year survival rate	-0.869	***	-1.167	***	-1.115	***
	(0.319)		(0.323)		(0.300)	
Market size	no		yes		no	
Life-years lost	no		no		yes	

Notes: This table shows the relationship between the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the number of clinical trials enrolling patients of that cancer-stage from 1973-2011. The level of observation is the cancer-stage. Estimates are from quasi-maximum likelihood Poisson models. Standard errors are clustered at the cancer level. *: $p<0.10$; **: $p<0.05$; ***: $p<0.01$. *Market size* denotes the inclusion of a covariate measuring the number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* is measured as age-gender-year specific life expectancy (in the absence of cancer) in the year of diagnosis, less observed survival time in years, averaged over patients diagnosed with that cancer-stage between 1973-1983 (to minimize censoring) multiplied times market size. The number of observations is 201 in Columns (1) and (2), and 192 in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. For details on the sample, see the text and data appendix.

Table 3: **Surrogate endpoints, survival time, and R&D investments**

Panel (A): Level of R&D, Dependent variable: Number of clinical trials (mean = 945)

	(1)		(2)		(3)	
five-year survival rate	-0.840	***	-1.142	***	-1.089	***
	(0.322)		(0.326)		(0.303)	
(0/1: *hematologic*)	0.339	**	0.368	**	0.308	*
	(0.165)		(0.163)		(0.185)	
Market size	no		yes		no	
Life-years lost	no		no		yes	

Panel (B): Composition of R&D, Dependent variable: Number of clinical trials (mean = 945)

	(1)		(2)		(3)	
(five-year survival rate)*(0/1: *hematologic*)	1.366	**	1.518	***	1.405	*
	(0.536)		(0.471)		(0.733)	
five-year survival rate	-0.900	***	-1.208	***	-1.151	***
	(0.336)		(0.339)		(0.313)	
(0/1: *hematologic*)	-0.028		-0.034		-0.069	
	(0.192)		(0.191)		(0.251)	
Market size	no		yes		no	
Life-years lost	no		no		yes	

Notes: This table shows two analyses of how cancer R&D differs on hematologic malignancies relative to other cancers, as a way of shedding light on how surrogate endpoints - which are more commonly used for hematologic malignancies - affect R&D investments. Panel (A) regresses the number of clinical trials enrolling patients of that cancer-stage from 1973-2011 on the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored) and an indicator for hematological malignancies. Panel (B) regresses the number of clinical trials enrolling patients of that cancer-stage from 1973-2011 on the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004, an indicator for hematological malignancies, and an interaction between these two variables. The level of observation is the cancer-stage. Estimates are from quasi-maximum likelihood Poisson models. Standard errors are clustered at the cancer level. *: $p<0.10$; **: $p<0.05$; ***: $p<0.01$. *Market size* denotes the inclusion of a covariate measuring the number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* is measured as age-gender-year specific life expectancy (in the absence of cancer) in the year of diagnosis, less observed survival time in years, averaged over patients diagnosed with that cancer-stage between 1973-1983 (to minimize censoring) multiplied times market size. The number of observations is 201 in Columns (1) and (2), and 192 in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. For details on the sample, see the text and data appendix.

Table 4: **Survival time and financing of clinical trials**

Panel (A): Dependent variable: Share of clinical trials that are privately financed (mean = 0.258)

	(1)		(2)		(3)	
five-year survival rate	-0.122	***	-0.136	***	-0.121	***
	(0.016)		(0.017)		(0.014)	
Market size	no		yes		no	
Life-years lost	no		no		yes	

Panel (B): Dependent variable: Number of clinical trials (mean = 244)

	(1)		(2)		(3)	
(five-year survival rate)*(0/1: *private*)	-0.436	***	-0.505	***	-0.407	**
	(0.166)		(0.166)		(0.164)	
five-year survival rate	-0.866	***	-1.137	***	-1.122	***
	(0.314)		(0.334)		(0.304)	
(0/1: *private*)	-0.681	***	-0.697	***	-0.752	***
	(0.062)		(0.052)		(0.051)	
Market size	no		yes		no	
Life-years lost	no		no		yes	

Notes: This table shows two analyses of how public and private financing of clinical trials varies with patient survival time. Panel (A) shows the relationship between the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the share of clinical trials enrolling patients of that cancer-stage from 1973-2011 that were privately financed; the level of observation is the cancer-stage, and estimates are from ordinary-least-squares (OLS) models. Panel (B) shows the relationship between the five-year survival rate and the number of publicly/privately financed clinical trials enrolling patients of that cancer-stage from 1973-2011; the level of observation is the cancer-stage-sponsor (where sponsor is either public or private), and estimates are from quasi-maximum likelihood Poisson models. Standard errors are clustered at the cancer level. *: $p<0.10$; **: $p<0.05$; ***: $p<0.01$. *Market size* denotes the inclusion of a covariate measuring the number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* is measured as age-gender-year specific life expectancy (in the absence of cancer) in the year of diagnosis, less observed survival time in years, averaged over patients diagnosed with that cancer-stage between 1973-1983 (to minimize censoring) multiplied times market size. The number of observations is 201 in Columns (1) and (2) of Panel (a), 402 (=201*two sponsor types) in Columns (1) and (2) of Panel (b), 192 in Column (3) of Panel (a), and 384 (=192*two sponsor types) in Column (3) of Panel (b), because 9 cancer-stages had no patients diagnosed between 1973-1983. For details on the sample, see the text and data appendix.

A Appendix: Proofs

A.1 Proof of Proposition 1

Part 1 follows immediately from (6).

For Part 2, the expected social return to A exceeding that to B can be written as:

$$\frac{ETL_A \cdot v_A}{c_A} \geq \frac{ETL_B \cdot v_B}{c_B}$$

Multiplying both sides by $\frac{EML_A}{ETL_A}\frac{\pi_A}{v_A}$ gives

$$\frac{EML_A \cdot \pi_A}{c_A} \geq \frac{ETL_B \cdot v_B}{c_B}\frac{EML_A}{ETL_A}\frac{\pi_A}{v_A}$$

Suppose that neither (a) nor (b) hold, *i.e.* $\frac{\pi_A}{v_A} \geq \frac{\pi_B}{v_B}$ and $\frac{EML_A}{ETL_A} \geq \frac{EML_B}{ETL_B}$. Then:

$$\frac{ETL_B \cdot v_B}{c_B}\frac{EML_A}{ETL_A}\frac{\pi_A}{v_A} \geq \frac{ETL_B \cdot v_B}{c_B}\frac{EML_B}{ETL_B}\frac{\pi_B}{v_B} = \frac{EML_B \cdot \pi_B}{c_B}$$

hence

$$\frac{EML_A \cdot \pi_A}{c_A} \geq \frac{EML_B \cdot \pi_B}{c_B}$$

Hence if invention B is pursued, so is invention A. A contradiction. QED

A.2 Proof of Proposition 4

Let $x^*(t_{comm})$ be the socially optimal post-commercialization patent length for inventions with commercialization lag t_{comm}. That is, having the patent clock expire at calendar time $t_{patent} = t_{comm} + x^*(t_{comm})$ maximizes expected social value for inventions of commercialization lag t_{comm} as given by

$$\mathbf{1}\{EML \cdot \pi \geq c\} \cdot \mathbb{E}(EML \cdot v^{monop} + (ETL - EML) \cdot v - c)$$

where the notation $\mathbf{1}\{\}$ denotes the indicator function for whether or not the invention is commercialized, the expectation operator is taken over invention parameters $(p, \pi, v, \delta, \gamma, c)$, and EML and ETL are as defined in equations (1) and (3). Note that EML and ETL vary with t_{patent}, and hence with $x^*(t_{comm})$.

We aim to show that $x^*(t_{comm})$ is increasing in t_{comm}. To do this, we will show that, if commercialization lag is $t'_{comm} > t_{comm}$, the marginal benefits of increasing x^* from $x^*(t_{comm})$ are higher than when commercialization lag is t_{comm}, and the marginal costs of increasing commercialization lag are lower. This, combined with $x^*(t_{comm})$ being optimal when commercialization lag is t_{comm}, implies that $x^*(t'_{comm}) > x^*(t_{comm})$.

The marginal benefit of increasing commercialization lag is that more inventions are commercialized. The marginal invention satisfies $EML = \frac{c}{\pi}$. Increasing t_{comm} while holding x^* fixed strictly decreases EML, for any realization of $(p, \pi, v, \delta, \gamma, c)$, due to discounting and obsolescence risk. (The number of calendar years of post-commercialization patent length stays the same, but these years are more heavily discounted.) Hence, the marginal invention at t'_{comm} has lower c and/or higher π, and hence is more socially valuable.

The marginal cost of increasing commercialization lag is that inframarginal inventions (*i.e.*, inventions

we would have received anyway) get more years of patent protection, and hence more years when they generate social value of $v^{monop} < v$. That is, the cost is the dead weight loss from additional years of monopoly. Since all tuples of invention parameters $(p, \pi, v, \delta, \gamma, c)$ that lead to commercialization when commercialization lag is t'_{comm} also lead to commercialization when commercialization lag is t_{comm}, but the reverse is not true, total deadweight loss from increasing x^* is larger at t_{comm} than at t'_{comm}.

The optimality of $x^*(t_{comm})$ for inventions with commercialization lag t_{comm} implies that the marginal benefits of increasing x^* exactly equal the marginal costs. Since, when commercialization lag is $t'_{comm} > t_{comm}$, the marginal benefits of increasing x^* are larger and the marginal costs are smaller, it follows that $x^*(t'_{comm}) > x^*(t_{comm})$, as required. QED

A.3 Proof of Proposition 5

Proof of Part 1. This follows immediately from the private investment condition (4). Since surrogate endpoints decrease t_{comm}, they increase EML, which might cause additional investments to occur. Formally, let $t_{comm}^{Surrogate}$ and $t_{comm}^{NoSurrogate}$ denote a drug's commercialization lag with and without surrogate endpoints, respectively. Consider a drug where surrogate endpoints strictly decrease commercialization lag, *i.e.* $t_{comm}^{Surrogate} < t_{comm}^{NoSurrogate}$. Hence $EML^{Surrogate} > EML^{NoSurrogate}$. Now choose π, c and p such that

$$EML^{Surrogate} \cdot \pi \geq \frac{c}{p} > EML^{NoSurrogate} \cdot \pi$$

Such an invention will get commercialized with surrogate endpoints but not without. The second part of the statement follows from our assumption that surrogate endpoints always at least weakly decrease commercialization lag, and hence always at least weakly increase EML.

Proof of Part 2. The social welfare associated with a successfully commercialized invention is $EML \cdot v^{monop} + (ETL - EML) \cdot v^{firstbest}$. A reduction in t_{comm} strictly increases EML, and has no effect on $ETL - EML$, because both ETL and EML increase by the expected number of additional years that the drug will be commercially available. Hence, the social welfare associated with any commercialized invention goes up, sometimes strictly. In combination with part (1) this yields that overall social welfare strictly increases.

Proof of Part 3. Follows immediately from the assumption that t_{comm} is orthogonal to $\hat{t}_{comm}$ and the definition of the fixed patent term distortion. QED

A.4 Proof of Proposition 6

Choose arbitrary commercialization lags t_{comm}, t'_{comm}, with $t_{comm} < t'_{comm}$. Initially, suppose that inventions with these commercialization lags receive the same per-commercialized-invention subsidy, of $k > 0$ dollars. We will argue that social welfare is improved by increasing the per-commercialized invention subsidy for t'_{comm} inventions while reducing the subsidy for t_{comm} inventions, in a cost-neutral manner.

Consider a marginal increase in the invention subsidy for commercialization lags t_{comm} and t'_{comm}. By the same logic as in the proof of Proposition 4, the marginal invention with lag t'_{comm} has lower c and/or higher π, and hence is more socially valuable, than the marginal invention with lag t_{comm}. Also by the same logic as in the proof of Proposition 4, the cost associated with providing a subsidy to inframarginal inventions (*i.e.*, inventions we would have received anyway) is smaller the longer is the commercialization lag, since any set of invention parameters that leads to commercialization with lag t'_{comm} also leads to commercialization when the lag is the shorter t_{comm}. Hence, at the margin, additional subsidy to t'_{comm} inventions is more valuable than to t_{comm} inventions. This implies that optimal subsidies are increasing in commercialization lag, as required. QED

A.5 Power calculation derivation

Our conceptual framework is based on the idea that inventions which require long commercialization lags may be under-incentivized by the patent system. Empirically, we focus on patient survival as a determinant of clinical trial length: because clinical trials must generally show evidence that treatments improve mortality-related outcomes, clinical trials tend to be longer when enrolling patients with longer survival times. In this section, we outline one example of a power calculation of the type used to guide the design of clinical trials in order to fix ideas on this point.

Approval of a drug compound by the US FDA requires evidence of efficacy and safety. Traditionally, "evidence of effectiveness" has been interpreted as evidence from controlled clinical trials. While most FDA approvals are based on placebo control groups, oncology trials instead compare the new drug compound to a non-placebo control of existing therapy. When testing the null hypothesis of no difference in mortality outcomes between the treatment and control groups, the traditional threshold for statistical evidence in oncology trials allows for a 1-in-20 chance of a false positive conclusion, or a p-value of 0.05.[25] This type of bar for statistical evidence motivates a calculation of what clinical trial design will be needed to achieve adequate statistical power to detect a statistically significant difference between the treatment and control groups.

An enormous literature exists on the design of clinical trials. Here, we simply focus on one type of calculation as an example. Collett (2003)'s *Modelling Survival Data in Medical Research* textbook includes a chapter on clinical trial design when survival is the outcome. Collett frames the design problem as a calculation of the required number of total deaths that must be observed.[26] Following this approach, at a given follow-up time k after treatment is administered, we can express the total number of deaths as $D = \frac{N}{2}(1-\mu^k)+\frac{N}{2}[1-(1-R(1-\mu))^k]$, where μ is the per-period survival rate of untreated individuals, k is the number of periods of patient follow-up, N is the sample size (equally divided between the treatment group and the control group), and R is a constant per-period multiplicative treatment effect. This expression can be derived as follows:

$$\begin{aligned} Pr(\text{die at time } t|\text{survival to time } t-1) &= \begin{cases} 1-\mu & \text{for Control} \\ R(1-\mu) & \text{for Treatment} \end{cases} \\ Pr(\text{survive at time } t|\text{survival to time } t-1) &= \begin{cases} \mu & \text{for Control} \\ 1-R(1-\mu) & \text{for Treatment} \end{cases} \end{aligned}$$

where μ is bounded by 0 and 1 and R is constrained such that $R(1-\mu)$ also is bounded by 0 and 1.

Consider first the control group. In the initial period there are $\frac{N}{2}$ individuals. In the subsequent period, there are $\frac{N}{2}\cdot\mu$, and in the kth period there are $\frac{N}{2}\cdot\mu^k$. Thus in the kth period there are $\frac{N}{2}-\frac{N}{2}\cdot\mu^k = \frac{N}{2}(1-\mu^k)$ deaths in the control group. Similarly, in the treatment group, at time k there are $\frac{N}{2}\cdot[1-R(1-\mu)]^k$ survivors and $\frac{N}{2}\cdot\{1-[1-R(1-\mu)]^k\}$ deaths. Thus the total number of deaths in the sample at time k is:

$$\begin{aligned} D &= \frac{N}{2}(1-\mu^k)+\frac{N}{2}\cdot[1-(1-R(1-\mu))^k] \\ &= \frac{N}{2}[2-\mu^k-(1-R(1-\mu))^k] \end{aligned}$$

Applying the implicit function theorem, we can derive the following two comparative statics:

[25] See, for example, http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDER/ucm103366.pdf.
[26] See the discussion in Chapter 10.

$$\frac{\partial N}{\partial \mu} = \frac{2D(k\mu^{k-1} + Rk(1-R(1-\mu))^{k-1})}{[2-\mu^k-(1-R(1-\mu))^k]^2}$$

$$\frac{\partial k}{\partial \mu} = \frac{k\mu^{k-1} + Rk(1-R(1-\mu))^{k-1}}{-[\mu^k \ln \mu + (1-R(1-\mu))^k \ln(1-(R(1-\mu))]}.$$

Since $0 < \mu < 1$ and $0 < R(1-\mu) < 1$, both of these partial derivatives are positive. Thus, we have two results. First, the required follow-up period k is increasing in the per-period survival rate μ: $\frac{\partial k}{\partial \mu} > 0$. Second, the required sample size N is increasing in μ: $\frac{\partial N}{\partial \mu} > 0$.

The first comparative static is the focus of our conceptual framework: clinical trials enrolling patients with longer expected survival times will - all else equal - require longer follow-up periods.[27] The second comparative static is related to our conceptual framework in a more nuanced way. In the absence of detailed data on clinical trial costs (which are confidential), it is difficult to know whether the financial cost of enrolling an additional patient is higher or lower than an equivalently effective lengthening of the trial. However, in addition to the financial cost of enrolling additional patients, there is also a *time cost* of an increase in sample size because of the time required to recruit patients.

A variety of sources have stressed the time required to recruit patients as a barrier to clinical development; for example, Bartfai and Lees (2006) argue: "[m]any trials take a long time because the rate of enrollment is low. It is not uncommon that a 90-day drug trial takes 18 months to complete for all enrolled patients; it might take 90 days for each patient, but by the time the selected centers reach the required numbers 1.5 years have flown by."[28] Importantly for our analysis, commentators often point to the running patent term as the reason why these delays due to patient recruiting are so costly. A book on clinical trial management notes, "access to patients remains critical for the success of clinical development programs" because "[s]low patient recruitment can delay product launch with revenue loss during the precious product patent life" (Chin and Bairu, eds (2011)). There are even reports of specific drugs being dropped from development because slow patient recruitment led to delays that consumed too much of their patent life (*e.g.* Kohan et al. (2012)). Thus, although at first blush clinical trial size might seem to be a mechanism for increasing the statistical power of clinical trials that is independent of trial length, this margin of adjustment also fits into our conceptual framework.

[27] The question of whether - empirically - clinical trials enrolling patients with longer expected survival times are longer in length is difficult to analyze given that we only observe clinical trial length *conditional* on a drug compound being placed in clinical trials. Because - consistent with our model - we will find that fewer drug compounds are placed in clinical trials for patients with longer survival times, we would expect selection into clinical trial investment to bias the relationship between patient survival and clinical trial length in the set of observed clinical trials. Perhaps the most natural selection story is that firms are only willing to place a drug compound in clinical trials for patients with long expected survival times if they receive permission to use a surrogate endpoint in place of survival as an endpoint; in this case, the relationship between patient survival and clinical trial length would be biased towards zero. If we nonetheless estimate this relationship in our data, we do estimate a statistically significant relationship; however, the magnitude is implausibly small, consistent with our prior that this relationship would be biased towards zero (a ten percentage point increase in the five-year survival rate is associated with a 1.5 percent increase in average clinical trial length - an increase on the order of one month)

[28] See also Hovde (2006), Goffin (2009), Malani and Philipson (2011), and Allison (2012).

B Appendix: Data

B.1 Description of SEER cancer registry data

The Surveillance, Epidemiology, and End Results (SEER) data is compiled by the National Cancer Institute (NCI), and is considered the authoritative source of information on cancer incidence and survival in the US.[29]

SEER collects data from population-based cancer registries covering approximately 28 percent of the US population. Specifically, the SEER data aims to be a comprehensive census of all cancer cases diagnosed among residents of geographic areas covered by SEER cancer registries. In order to focus on a geographically consistent sample over time, we analyze data from the seven original SEER registries that joined in 1973: the states of Connecticut, Iowa, New Mexico, Utah, and Hawaii and the metropolitan areas of Detroit and San Francisco-Oakland. Funding for the data collection varies by state, and is a mix of funding from the NCI, the Centers for Disease Control and Prevention (CDC), and state funding.

The SEER registries collect detailed information on cancer patients near the time of diagnosis, including data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, and first course of treatment.[30] This data is administratively linked to follow-up mortality data from the National Center for Health Statistics (NCHS).

We use the 1973-2009 SEER Research Data (ASCII text format) as downloaded on 25 June 2012, which includes patients diagnosed from 1973-2009.[31] The follow-up mortality data cutoff date is 31 December 2009. The key variables that we obtain from the SEER data are the following:

- **Cancer information**. We use the SEER site recode with kaposi sarcoma and mesothelioma variable to identify the cancer type for each individual in our sample. For example, a value of 20010 for this variable corresponds to a diagnosis of lip cancer. There are 80 unique cancer categories, as listed here: http://seer.cancer.gov/siterecode/icdo3_d01272003/. This variable is non-missing for all observations.

- **Stage information**. We use the SEER historic stage A variable to identify the stage of cancer for each individual in our sample: in situ, localized, regional, metastatic, or unknown. As described on the SEER website, stage information is not available for all observations for three reasons.[32] First, some cancers are not staged by SEER: for example, brain cancers are not staged.[33] Second, some cancers are not staged in a subset of years: for example, between 1973-1982 nose, nasal cavity, and middle ear cancers were not staged.[34] Third, some individual observations that should be staged

[29] For more details, see http://www.seer.cancer.gov.

[30] Importantly for our purposes, the SEER website notes that the SEER data is the only comprehensive source of population-based information in the US that includes stage of cancer at the time of diagnosis and patient survival data.

[31] This data is available via a research data agreement; see https://www.seer.cancer.gov/seertrack/data/request/ for details. The citation for this data is: Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2009), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2012, based on the November 2011 submission.

[32] For more information, see http://seer.cancer.gov/seerstat/variables/seer/yr1973_2009/lrd_stage/index.html.

[33] The SEER site recodes that are not staged by SEER are: brain (31010); cranial nerves and other nervous system (31040); pleura (22050); hodgkin and non-hodgkin lymphoma (both nodal and extranodal; 33011, 33012, 33041, and 33042); myeloma (34000); acute lymphocytic leukemia (35011); chronic lymphocytic leukemia (35012); other lymphocytic leukemia (35013); acute myeloid leukemia (35021); acute monocytic leukemia (35031); chronic myeloid leukemia (35022); other myeloid/monocytic leukemia (35023); other acute leukemia (35041); aleukemic, subleukemic, and not otherwise specified leukemias (35043); kaposi sarcoma (36020); and miscellaneous (37000).

[34] The SEER site recodes that are staged by SEER for a subset of years are: nasopharynx (not staged 2004 and later; 20060); peritoneum, omentum, and mesentery (not staged 1973-1987; 21120); nose, nasal cavity, and middle ear (not staged

have missing stage data. The first two categories - for which missing stage data are "expected" - result in stage data missing for 19 percent of the SEER sample; the third category - for which missing stage data is "unexpected" - results in stage data missing for an additional 5 percent of the SEER sample. The exceptions to the standard staging categories are as follows:

- Prostate cancer. Prostate cancer is staged by SEER starting in 1995, but uses a combined localized/regional category rather than separate localized and regional stages. We code the localized/regional prostate cases as regional cancers.
- Bladder cancer. All in situ cases of bladder cancer (29010) in the SEER data were re-coded by SEER to appear as localized cancers.

For consistency, we code all unstaged cancers and cancers that utilize only one stage into an "unstaged" stage classification in our analysis.

- **Survival time**. Because the SEER data are linked to follow-up mortality data from the National Center for Health Statistics, for each individual in our sample we know survival time in months as calculated using the date of diagnosis and one of the following: date of death, date last known to be alive, or follow-up cutoff date of 31 December 2009. This variable is non-missing for all observations.
- **Year of diagnosis**. The SEER data record the year of diagnosis for each patient, defined as the year the tumor was first diagnosed by a recognized medical practitioner, whether clinically or microscopically confirmed. The year of diagnosis varies from 1973 to 2009, and is non-missing for all observations. We use the year of diagnosis together with information on patient sex and age at diagnosis to calculate life expectancy at the time of diagnosis (in the absence of cancer) for each individual in the SEER sample.
- **Age at diagnosis**. The SEER data record the patient's age in years at diagnosis. This variable is missing for 692 of 3,245,656 individuals (0.02 percent of the sample). Because we need information on age at diagnosis in order to calculate life expectancy at the time of diagnosis, we drop these 692 individuals from the sample.
- **Sex**. The SEER data record the sex of the patient at diagnosis. This variable is non-missing for all observations. We use this variable together with information on year of diagnosis and patient age at diagnosis to calculate life expectancy at the time of diagnosis for each individual in the SEER sample.

Between 1973 and 2009, 3,245,656 individuals were diagnosed in catchment areas of the seven original SEER registries. Our only sample restriction is to exclude the 692 individuals missing data on age at diagnosis (0.02 percent of the sample), leaving us with a final SEER sample of 3,244,964 individuals.

SEER also produced population data which can be used to normalize the cancer incidence data into rates per population. We use the 1969-2009 SEER population data (ASCII text format) for the catchment areas of the seven original SEER registries as downloaded on 28 June 2012.[35]

1973-1982; 22010); larynx (not staged 2004 and later; 22020); lung and bronchus (not staged 1973-1987; 22030); trachea (not staged 2004 forward; 22060); vagina (not staged 2004 forward; 27050); prostate (not staged 1973-1994; 28010); other endocrine including thymus (not staged 2004 and later; 32020); and mesothelioma (not staged 2004 and later; 36010).

[35]The citation for this data is: Surveillance, Epidemiology, and End Results (SEER) Program Populations (1969-2009) (www.seer.cancer.gov/popdata), National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released January 2011.

B.2 Life expectancy data

We use year-age-gender-specific period life expectancy data for 1973-2006 from the National Center for Health Statistics (NCHS) files posted at http://www.cdc.gov/nchs/products/life_tables.htm. For 2000-2006, digitized files are available from NCHS. For 1973-1999, the data was entered by the firm Digital Divide Data (http://www.digitaldividedata.org/) and was funded by NIA Grant Number T32-AG000186 to the NBER. No data is available for 1979 nor 1981.

Based on the year of diagnosis, age at diagnosis, and gender of each patient, we use this NCHS data to construct year-age-gender-specific life expectancy for each patient - in the absence of cancer - at the time of diagnosis. Because in most years the life tables end at age 85, we apply the life expectancy numbers for 85-year-old individuals to all individuals age 85 and older. Because data is not available for years 1979, 1981, 2007, 2008, and 2009, we fill in the data as follows: apply the 1978 life expectancy data in 1979; apply the 1980 life expectancy data in 1981; and apply the 2006 life expectancy data in 2007, 2008, and 2009. Using this life expectancy data, we calculate the life lost for each individual as their life expectancy at the time of diagnosis minus their survival time in years.

We focus on measuring life lost among patients diagnosed between 1973-1983 to minimize censoring. In this sample, median survival time by cancer-stage is almost never censored; in the handful of cases where censoring is an issue, we top code survival time at 25 years.

B.3 Description of classification method for research investment datasets

Unlike the SEER data, our data measuring research investments in cancer treatments were not originally developed as research data, and hence required a large amount of restructuring to be converted to a format useable for our analysis. We detail this restructuring below, but first describe the method we use to classify the cancer and stages to which a given clinical trial is relevant.

We start by compiling a classification system which can consistently code observations that vary in the aggregation level at which the cancer type was identified in the original data. For example, while some cancer clinical trials enroll stage III breast cancer patients, others are open to all patients with "solid tumors" and we need a way of classifying which cancer types are solid tumors. We base our classification around the SEER site recode ICD-O-3 (1/27/2003) definition.[36] These SEER site records define the major cancer sites (*e.g.* breast, stomach, prostate) and are the standard set of cancer classifications used by cancer researchers. Importantly, the SEER cancer registry data include SEER site recodes, so using the SEER site recodes as the basis of our classification of the research investment datasets is what enables a cross-walk between the SEER cancer registry data and the research investment data.

For each research investment observation, we search the textual description of the cancer type for which the observation is relevant in order to match the observation to one or more of the SEER site recodes.[37] Most of the search words are drawn directly from the SEER site recode title (*e.g.* "lip" for lip cancer, SEER site recode 20010), but we also search for variations on cancer names that are frequently observed in the data (*e.g.* searching both "pancreas" and "pancreatic" for cancer of the pancreas, SEER site recode 21100). We allow a given observation to be labeled as relevant to multiple cancer types (*e.g.* an observation labeled as being relevant for hematologic/blood cancers is classified as relevant to all hematologic/blood cancers, such as both acute myeloid leukemia and chronic myeloid leukemia). While there are surely imperfections in this classification system, it allows for a consistent coding of our data.

One additional issue that deserves discussion is off-label use of drugs. Off-label prescription of a drug refers to use of the drug outside of what is prescribed on its FDA-approved label (Leveque (2008)). Off-label use of drugs is generally thought to be widespread, particularly in cancer, although very few studies have actually measured the extent of off-label drug use in representative populations. An exception is

[36] Available at http://seer.cancer.gov/siterecode/icdo3_d01272003/.

[37] Detailed documentation on the precise search words used in this classification system are available on request.

a recent study by Agha and Molitor (2012), who estimate that 22 percent of cancer drug prescriptions were off-label in the Medicare population between 1998-2008. Why could off-label drug use be important in this context? If off-label use were completely unrestricted, firms would face an incentive to always approve drugs for the group of patients for which trials were the least expensive, and then *ex post* have the drugs be prescribed for all patients. In practice, although in the US physicians are free to prescribe drugs for off-label uses, it is illegal for pharmaceutical firms to actively advertise/promote those uses, and additional constraints are often imposed by insurers for reimbursement. Perhaps the clearest evidence that restrictions on off-label use appear to be binding comes from the fact that we very frequently observe firms making large R&D investments to re-approve a given drug compound for an additional indication (see, *e.g.*, Eisenberg (2005)), which they would have no need to do if off-label use restrictions were not binding. For all of these reasons, although off-label use is important in this market, it seems unlikely that the potential for off-label use is generating the aggregate relationship between survival time and clinical trial activity observed in our data.

B.4 Description of NCI cancer clinical trial registry data

The Physician Data Query (PDQ) Cancer Clinical Trials Registry is the National Cancer Institute (NCI)'s cancer clinical trials database. The NCI registry was created via the National Cancer Act of 1971, and claims to be the world's most comprehensive cancer clinical trial registry. The intended purpose of this registry is to allow cancer patients and physicians to search for clinical trials now accepting participants, and to allow them to access information and results from closed trials.

We use the 12 July 2011 version of the NCI registry data (XML format), which includes all clinical trials entered into the registry prior to that date.[38] The registration of clinical trials in the NCI registry is strictly voluntary but strongly encouraged. The NCI registry is thought to include most clinical trials sponsored by the NCI, as well as a substantial share of clinical trials sponsored by pharmaceutical companies, medical centers, and other groups. For example, the NCI registry includes all cancer clinical trials registered under the requirements specified by Section 113 of the Food and Drug Administration Modernization Act of 1997 (phase II and higher drug treatment trials), all cancer clinical trials registered under the requirements of the International Committee of Medical Journal Editors (phase II and higher trials that have a comparison or control group), and all cancer clinical trials that are included in the US National Institutes of Health (NIH) `Clinicaltrials.gov` database.

Many trials in our sample enroll multiple patient "types" as measured by the cancer-stages eligible for participation in the trial, and when we expand the data to unique trial-cancer-stage observations. We make three sample restrictions. First, cancer stages are sometimes reported in the NCI data at a finer level of granularity than we observe in the SEER data: for example, a given trial may list breast cancer stage IIA and breast cancer IIB patients as eligible for enrollment, but we do not consistently observe cancer stage at that level of detail in the SEER data. To avoid double-counting, we remove duplicate observations at the trial-cancer-stage level. Second, because we do not observe remission in the SEER data, we are unable to construct measures of the patient population eligible for these trials and thus drop trials enrolling only remission cases from the sample. Third, because we do not observe recurrent cases in the SEER data, we are again unable to construct measures of the patient population eligible for these trials and thus drop trials enrolling only recurrent patients from the sample.

The key variables that we obtain from the NCI registry data are the following:

- **Cancer information**. We identify the types of cancers eligible for enrollment in each clinical trial, coded by the SEER site recodes. By construction, this variable is non-missing for all observations.

[38] This data is available via a research licensing agreement; see `http://www.cancer.gov/licensing` for details. The scripts used to extract the XML files are available on request.

- **Stage information**. In the NCI registry data, the cancer stages eligible for enrollment in each clinical trial are most frequently identified in the following categories: stage 0, stage 1, stage 2, stage 3, stage 4, recurrent cancers, cancers in remission, and localized cancers. As discussed above, we drop trials enrolling only remission or recurrent patients from the sample. We then need a crosswalk which maps the remaining stage categories in the NCI registry data - stage 0, stage 1, stage 2, stage 3, stage 4, and localized cancers - to the SEER historic stage A categories in the SEER data (in situ, localized, regional, and metastatic). We follow the *AJCC Cancer Staging Manual* (American Joint Committee on Cancer (2010)) and use the following mapping: stage 0 maps into in situ; stages I, II, and localized map into localized; stage III maps into regional; and stage IV maps into metastatic.[39] In addition, to harmonize the NCI registry stage coding with the SEER stage coding we make the following revisions:

  - Prostate cancer. In the SEER data, the localized and regional prostate cases (28010) are coded into a joint localized/regional category which as described above we code as regional cancers. Analogously, in the NCI registry data we code trials for either localized or regional prostate cancer as being for regional prostate cancer.
  - Bladder cancer. In the SEER data, all in situ cases of bladder cancer (29010) are coded as localized cancers. Analogously, in the NCI registry data we code trials for in situ bladder cancers as being for localized bladder cancers.

  For consistency, we code all cancers which SEER codes as either unstaged or utilizing only one stage into an "unstaged" stage classification in our analysis.

- **Clinical trial sponsorship.** Approximately 50 percent of clinical trials in the NCI registry data are listed as being either publicly sponsored or privately sponsored. We define publicly-sponsored trials as trials that are solely publicly sponsored, and define privately-sponsored trials as trials that are solely privately sponsored; in our sponsorship analysis, we treat the approximately 1 percent of trials that are listed as being both publicly sponsored and privately sponsored as missing sponsorship data.

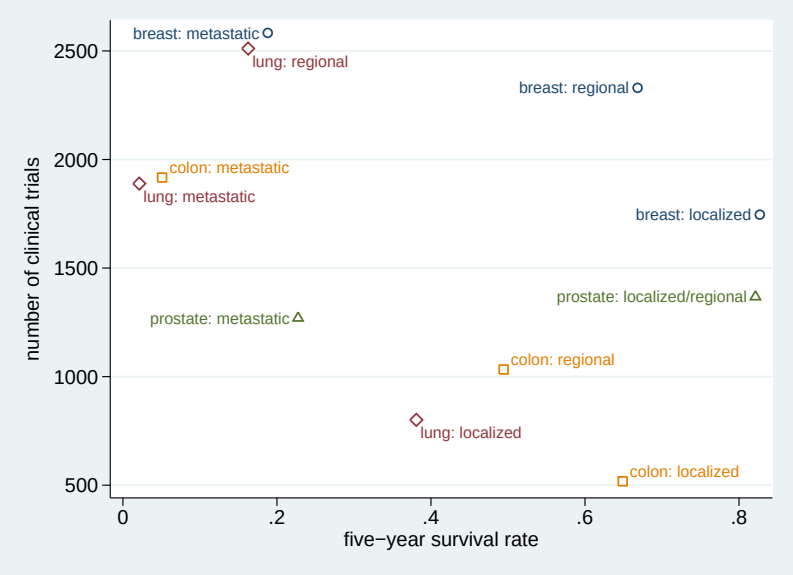
- **Clinical trial length.** Length of clinical trials is very rarely reported in the NCI registry data. Our understanding is that this is because the NCI registry is primarily oriented towards the recruitment of patients into clinical trials, whereas trial lengths are typically reported at the time of trial completion. In order to obtain data on clinical trial length, we take advantage of the fact that the NCI registry includes - where available - a `Clinicaltrials.gov` trial ID number. `Clinicaltrials.gov` is a registry *and results* database of clinical trials: likely because `Clinicaltrials.gov` includes clinical trial results, trial length is much better reported relative to the NCI registry data. The NCI registry claims to include all cancer clinical trials listed in the `Clinicaltrials.gov` registry, and approximately 70 percent of the NCI trials are included in the `Clinicaltrials.gov` registry. While we rely on the more complete NCI registry for the main analysis, we use the `Clinicaltrials.gov` subsample in order to examine data on trial length. Approximately 60 percent of trials in the NCI registry which appear in the `Clinicaltrials.gov` registry have non-missing data on trial length. Much of the missing data appears to be explained by trial length being more frequently reported in

[39] The exact language on page 12 is as follows: "Stage I is usually assigned to tumors confined to the primary site with a better prognosis, stages II and III for tumors with increasing local and regional nodal involvement, and stage IV to cases with distant metastatic disease. In addition, a group termed stage 0 is assigned to cases of carcinoma in situ (CIS)."

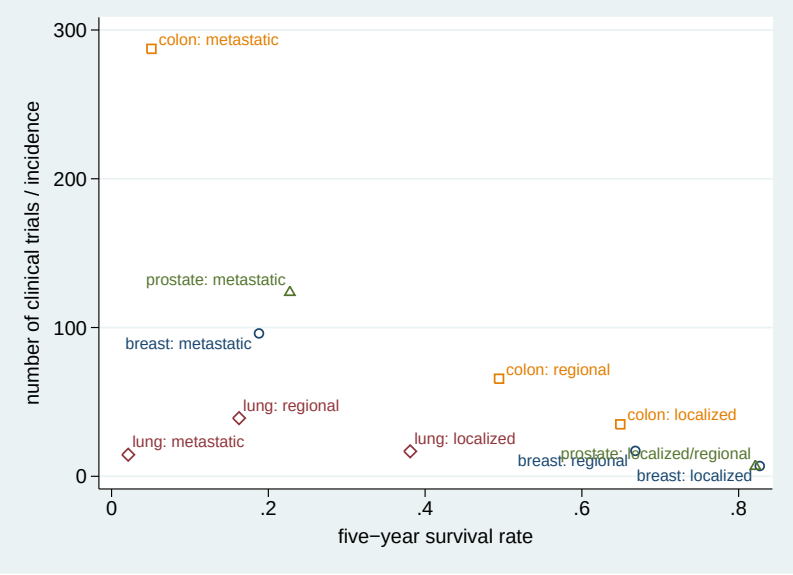
more recent years (even given that we would expect missing data for ongoing trials to bias upwards the share of trials with missing data in more recent years): on the order of 80 percent of trials starting in 1997 have missing data on trial length, compared to 50 percent in 2003 and 25 percent in 2011. This increased reporting over time likely in part reflects increased incentives for reporting: for example, there was a tightening of reporting requirements affecting clinical trials initiated in or ongoing as of September 2007; see http://clinicaltrials.gov/ct2/info/results for details.

C Appendix: Additional figures and tables

Figure C.1: **Survival time and R&D investments: Breast, colon, lung, and prostate cancers**



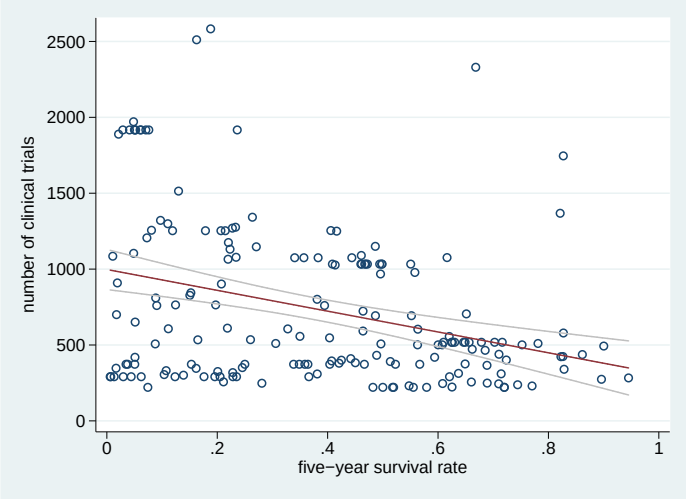
(a) Survival time and R&D investments



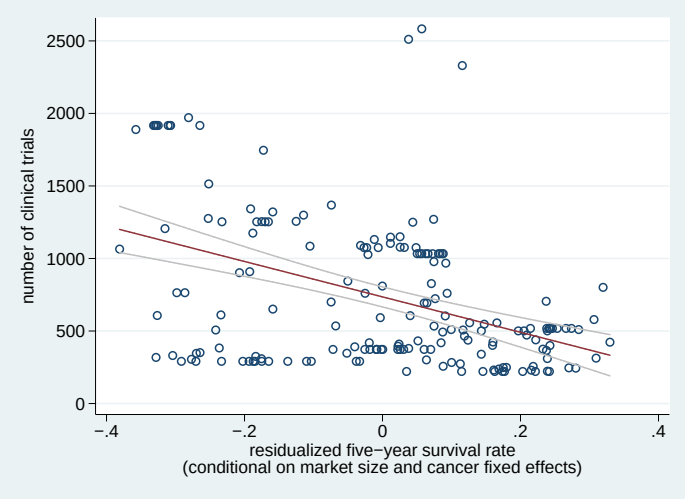
(b) Survival time and market-size adjusted R&D investments

Notes: This figure shows the relationship between the five-year survival rate among patients diagnosed with a given cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and two measures of clinical trial activity for that cancer-stage from 1973-2011, for the "big four" cancers: breast (26000), colon (specifically, ascending colon; 21043), lung (22030), and prostate (28010). The level of observation is the cancer-stage. Panel (a) plots the number of clinical trials enrolling patients of each cancer-stage from 1973-2011; Panel (b) plots the number of clinical trials enrolling patients of each cancer-stage from 1973-2011 divided by the number of patients diagnosed with that cancer-stage from 1973-2009, as a rough adjustment for market size. For details on the sample, see the text and data appendix.

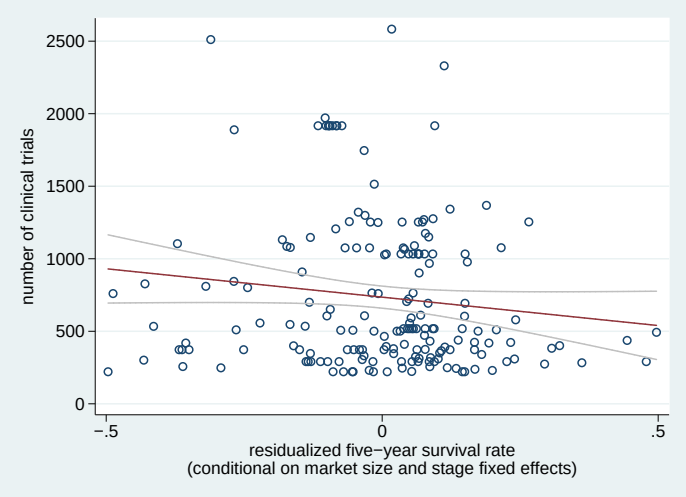
Figure C.2: **Survival time and R&D investments: Residualized cancer-stage data**



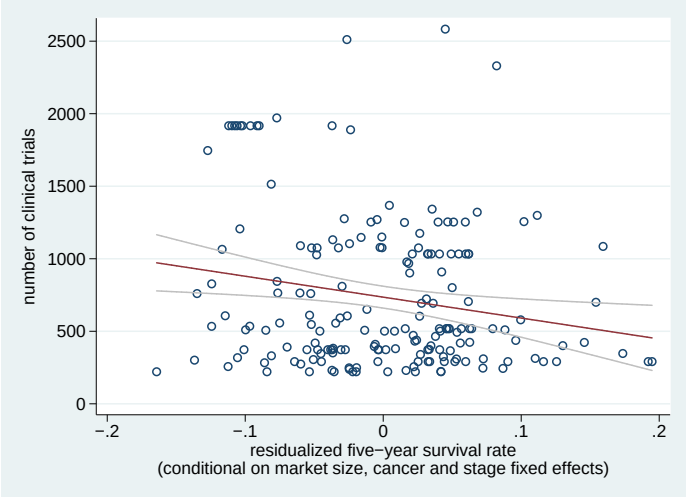
(a) Raw data



(b) Market size and cancer fixed effects



(c) Market size and stage fixed effects



(d) Market size, cancer and stage fixed effects

Notes: This figure shows the relationship between the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the number of clinical trials enrolling patients of each cancer-stage from 1973-2011. The level of observation is the cancer-stage. Panel (a) shows the raw data; Panel (b) residualizes market size and cancer fixed effects; Panel (c) residualizes market size and stage fixed effects; and Panel (d) residualizes market size, cancer fixed effects, and stage fixed effects. *Market size* denotes the inclusion of a covariate measuring the number of patients diagnosed with that cancer-stage between 1973-2009. As explained in the text, unstaged cancers are omitted from these figures since these observations do not identify the relationship of interest once we include cancer fixed effects and by definition unstaged cancers do not correspond to localized, regional, or metastatic stage definitions; Figure 2 shows an analogous scatterplot which includes unstaged cancers. For details on the sample, see the text and data appendix.

Table C.1: **Survival time and R&D investments: Robustness to cancer and stage fixed effects**

Dependent variable: Number of clinical trials (mean = 945)										
	(1)		(2)		(3)		(4)		(5)	
five-year survival rate	-0.963	***	-1.214	***	-1.551	***	-0.576	**	-1.709	***
	(0.236)		(0.161)		(0.095)		(0.289)		(0.269)	
Market size	no		yes		yes		yes		yes	
Cancer fixed effects	no		no		yes		no		yes	
Stage fixed effects	no		no		no		yes		yes	

Notes: This table shows the relationship between the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the number of clinical trials enrolling patients of that cancer-stage from 1973-2011. The level of observation is the cancer-stage. Estimates are from quasi-maximum likelihood Poisson models. Standard errors are clustered at the cancer level. *: $p<0.10$; **: $p<0.05$; ***: $p<0.01$. *Market size* denotes the inclusion of a covariate measuring the number of patients diagnosed with that cancer-stage between 1973-2009. As explained in the text, unstaged cancers are omitted from these regressions since these observations do not identify the relationship of interest once we include cancer fixed effects and by definition unstaged cancers do not correspond to localized, regional, or metastatic stage definitions; n=182. For details on the sample, see the text and data appendix.

Table C.2: **Survival time and R&D investments: Robustness to alternative survival measures**

Dependent variable: Number of clinical trials (mean = 945)					
	(1)	(2)	(3)	(4)	(5)
one-year survival rate	-0.781 **				
	(0.325)				
five-year survival rate		-0.869 ***			
		(0.319)			
1973 survival (years)			-0.034 ***		
			(0.013)		
1973 one-year survival rate				-0.598 **	
				(0.297)	
1973 five-year survival rate					-0.731 **
					(0.309)

Notes: This table shows the relationship between various measures of the survival rate among patients diagnosed with each cancer-stage and the number of clinical trials enrolling patients of that cancer-stage from 1973-2011. The level of observation is the cancer-stage. Estimates are from quasi-maximum likelihood Poisson models. Standard errors are clustered at the cancer level. *: $p<0.10$; **: $p<0.05$; ***: $p<0.01$. The number of observations is 201 in Columns (1) and (2), and 187 in Column (3), because 14 cancer-stages had no patients diagnosed in 1973. For details on the sample, see the text and data appendix.

Table C.3: **Survival time and R&D investments: Robustness across samples**

Dependent variable: Number of clinical trials (mean in Columns (1), (2) = 945)						
	(1)	(2)	(3)	(4)	(5)	(6)
five-year survival rate	-0.869 ***	-1.167 ***	-1.241 **	-1.662 ***	-0.963 ***	-1.214 ***
	(0.319)	(0.323)	(0.529)	(0.515)	(0.236)	(0.161)
Market size	no	yes	no	yes	no	yes
Excluding metastatic cancers	no	no	yes	yes	no	no
Excluding unstaged cancers	no	no	no	no	yes	yes

Notes: This table shows the relationship between the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the number of clinical trials enrolling patients of that cancer-stage from 1973-2011. The level of observation is the cancer-stage. Estimates are from quasi-maximum likelihood Poisson models. Standard errors are clustered at the cancer level. *: $p<0.10$; **: $p<0.05$; ***: $p<0.01$. *Market size* denotes the inclusion of a covariate measuring the number of patients diagnosed with that cancer-stage between 1973-2009. $N = 201$ in Columns (1) and (2); given the sample restrictions noted in the table, n=140 in Columns (3) and (4), and n=182 in Columns (5) and (6). For details on the sample, see the text and data appendix.

Table C.4: **Survival time and FDA drug approvals: Cancer-stage data**

Dependent variable: Number of approved drugs (mean = 0.507)						
	(1)		(2)		(3)	
five-year survival rate	-2.306	**	-2.719	***	-2.559	***
	(0.912)		(0.778)		(0.872)	
Market size	no		yes		no	
Life-years lost	no		no		yes	

Notes: This table shows the relationship between the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored), and the number of drugs approved by the US FDA for that cancer-stage from 1990-2002. The level of observation is the cancer-stage. Estimates are from quasi-maximum likelihood Poisson models. Standard errors are clustered at the cancer level. *: $p<0.10$; **: $p<0.05$; ***: $p<0.01$. *Market size* denotes the inclusion of a covariate measuring the number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* is measured as age-gender-year specific life expectancy (in the absence of cancer) in the year of diagnosis, less observed survival time in years, averaged over patients diagnosed with that cancer-stage between 1973-1983 (to minimize censoring) multiplied times market size. The number of observations is 201 in Columns (1) and (2), and 192 in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. For details on the sample, see the text and data appendix.

Table C.5: **Surrogate endpoints, survival time, and drug approvals**

Panel (A): Level of R&D, Dependent variable: Number of approved drugs (mean = 0.507)

	(1)		(2)		(3)	
five-year survival rate	-2.327	***	-2.798	***	-2.604	***
	(0.902)		(0.742)		(0.847)	
(0/1: *hematologic*)	1.250	***	1.348	***	1.269	***
	(0.458)		(0.434)		(0.423)	
Market size	no		yes		no	
Life-years lost	no		no		yes	

Panel (B): Composition of R&D, Dependent variable: Number of approved drugs (mean = 0.507)

	(1)		(2)		(3)	
(five-year survival rate)*(0/1: *hematologic*)	6.632	***	7.075	***	6.896	***
	(1.668)		(1.389)		(1.412)	
five-year survival rate	-3.743	***	-4.198	***	-4.125	***
	(1.273)		(0.804)		(0.943)	
(0/1: *hematologic*)	-1.032		-1.073		-1.133	*
	(0.725)		(0.702)		(0.682)	
Market size	no		yes		no	
Life-years lost	no		no		yes	

Notes: This table shows two analyses of how cancer R&D differs on hematologic malignancies relative to other cancers, as a way of shedding light on how surrogate endpoints - which are more commonly used for hematologic malignancies - affect R&D investments. Panel (A) regresses the number of drugs approved by the US FDA for that cancer-stage from 1990-2002 on the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004 (the cohorts for which five-year survival is uncensored) and an indicator for hematological malignancies. Panel (B) regresses the number of drugs approved by the US FDA for that cancer-stage from 1990-2002 on the five-year survival rate among patients diagnosed with each cancer-stage between 1973-2004, an indicator for hematological malignancies, and an interaction between these two variables. The level of observation is the cancer-stage. Estimates are from quasi-maximum likelihood Poisson models. Standard errors are clustered at the cancer level. *: $p<0.10$; **: $p<0.05$; ***: $p<0.01$. *Market size* denotes the inclusion of a covariate measuring the number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* is measured as age-gender-year specific life expectancy (in the absence of cancer) in the year of diagnosis, less observed survival time in years, averaged over patients diagnosed with that cancer-stage between 1973-1983 (to minimize censoring) multiplied times market size. The number of observations is 201 in Columns (1) and (2), and 192 in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. For details on the sample, see the text and data appendix.

D Appendix: Development of chemoprevention drugs

In a review article on chemoprevention drugs in the journal *Cancer Prevention Research*, Meyskens et al. (2011) compile a list of the FDA approved drugs which prevent human cancers: BCG for bladder carcinoma in situ, Diclofenac for actinic keratoses, Celecoxib for familial adenomatous polyposis (FAP)-polyps, Photofrin for Barrett's esophagus, Tamoxifen/Raloxifene for breast cancer, and vaccines (Gardasil and Cervarix) to prevent cervical cancer. As summarized in Section 5.3, our qualitative investigation of the history of these FDA drug approvals suggests that each of these six approvals was either financed by the public sector (Tamoxifen and BCG) or relied on the use of surrogate endpoints (Diclofenac, Celecoxib, Photofrin, and cervical cancer vaccines). In this appendix, we provide documentation for this assertion.

D.1 BCG

The 1990 FDA approval for bladder carcinoma in situ was supported by clinical trials funded by the National Cancer Institute (NCI).[40] A popular press citation in the *New York Times* (Leary (1990)) noted the approval was supported by *"controlled, multi-center trials sponsored by the National Cancer Institute."* Lippman and Hawk (2009) cite the importance of one particular trial by Lamm et al. (1991) as supporting this approval, which was NCI-funded. Lippman and Hawk (2009) note: *"The FDA approved BCG for preventing recurrence of superficial bladder cancer in 1990 based on several clinical trials including one by the Southwest Oncology Group (Lamm et al. (1991))."* The acknowledgements in the Lamm et al. (1991) paper note: *"Conducted by the Southwest Oncology Group and supported in part by Public Health Service Cooperative Agreement grants from the National Cancer Institute (CA-04915, CA-37429, CA-42777, CA-04919, CA-27057, CA-13512, CA-1238, CA-36020, CA-22433, CA-16385, CA-20319, CA-37918, CA-13238, CA-35109, CA-12213, CA-12644, CA-35090, CA-461433, CA-35996, CA-35261, CA-14028, CA-03096, CA-35274, CA-22411, CA-35178, CA-35117, CA-35176, CA-35281, CA-28862, CA-03389, and CA-32102)."*

D.2 Diclofenac

Diclofenac is a topical treatment for actinic keratoses, which is clinically recommended for treatment to prevent disease progression to squamous cell carcinomas.[41] See, for example, the FDA approval letter and medical review for Diclofenac: http://www.accessdata.fda.gov/drugsatfda_docs/appletter/2000/21005ltr.pdf and http://www.accessdata.fda.gov/drugsatfda_docs/nda/2000/21005_Solaraze_medr_P1.pdf.

D.3 Celecoxib

The clinical trial endpoint for Celecoxib was a reduction in the number of adenomatous colorectal polyps, as a surrogate endpoint for gastrointestinal and other familial adenomatous polyposis (FAP)-related cancers. The medical review for Celecoxib's FDA approval notes: *"The sponsor has submitted clinical efficacy and safety data in support of the following new indication for Celebrex [celecoxib]: reduction in the number of adenomatous colorectal polyps in familial adenomatous polyposis patients...based on improvement*

[40] Note that there seems to be a typo in the FDA approval date (1978) listed by Meyskens et al. (2011), because the FDA approval seems to have been in 1990.

[41] See, for example, Mcintyre et al. (2007), who note: *"Actinic keratoses should be treated because of their potential to progress to squamous cell carcinomas."*

in a surrogate endpoint" (http://www.accessdata.fda.gov/drugsatfda_docs/nda/99/21156-S007_Celebrex_medr.pdf).

D.4 Photofrin

The clinical trial endpoint for Photofrin was "complete ablation of high-grade dysplasia in patients with Barrett's esophagus," as a surrogate endpoint for the incidence of esophageal carcinoma. See, for example, the label for Photofrin: http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/020451s020lbl.pdf.

D.5 Tamoxifen

Lippman and Brown (1999) note that the 1998 FDA approval of Tamoxifen as a chemoprevention agent was supported by the National Surgical Adjuvant Breast and Bowel Project's Breast Cancer Prevention Trial (Fisher et al. (1998)), which was funded by the National Cancer Institute and the National Institutes of Health. Lippman and Brown (1999) note: *"Tamoxifen as a chemopreventive agent has produced a fundamental change in the outlook for controlling breast cancer. Tamoxifen in the National Surgical Adjuvant Breast and Bowel Project (NSABP) P-1 Breast Cancer Prevention Trial (BCPT) achieved a striking 49% reduction in the incidence of invasive breast disease in women at increased risk of breast cancer (Fisher et al. (1998)). With this finding, the Food and Drug Administration (FDA) approved tamoxifen for risk reduction in this setting, marking the historic first FDA approval of any agent for cancer risk reduction."* The acknowledgements in the Fisher et al. (1998) paper note: *"This investigation was supported by Public Health Service grants U10-CA- 37377 and U10-CA-69974 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services."*

D.6 Cervical cancer vaccines

The clinical trial endpoints for cervical cancer vaccines (Gardasil and Cervarix) were the incidence of cervical CIN 2/3 (cervical intraepithelial neoplasia grade 2/3) and cervical AIS (cervical adenocarcinoma in situ), as surrogate endpoints for the incidence of cervical cancer. See, for example, the label for the Gardasil vaccine: http://www.fda.gov/downloads/biologicsbloodvaccines/vaccines/approvedproducts/ucm111263.pdf, which states: *"CIN 2/3 and AIS are the immediate and necessary precursors of squamous cell carcinoma and adenocarcinoma of the cervix, respectively. Their detection and removal has been shown to prevent cancer; thus, they serve as surrogate markers for prevention of cervical cancer."*