Working Paper Series
No. 14005

Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials

Eric Budish
University of Chicago

Benjamin N. Roin
Harvard Law School

Heidi Williams
MIT
National Bureau of Economic Research

September 2014
Also available at: http://economics.mit.edu/files/8651

Hoover Institution Working Group on
Intellectual Property, Innovation, and Prosperity
Stanford University

www.hooverip2.org

Do firms underinvest in long-term research? Evidence from cancer clinical trials*

Eric Budish
University of Chicago

Benjamin N. Roin
MIT

Heidi Williams
MIT and NBER

September 23, 2014

Abstract

This paper investigates whether private research investments are distorted away from long-term projects, by which we mean projects with long time lags between initial discovery ("invention") and the availability of a commercially viable product ("commercialization"). We present a simple theoretical model to formalize two potential sources of this distortion: first, excess impatience of private firms relative to the social planner; and second, the fact that patents – as currently designed – provide little effective incentive to develop technologies with long commercialization lags. We then explore this distortion empirically in the context of cancer research, where clinical trials - and hence, commercialization lags – are shorter 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 which together are consistent with private research investments being distorted away from long-term projects. Back-of-the-envelope calculations suggest that the value of life-years at stake is large. We discuss three specific policy responses - surrogate (non-mortality) clinical trial endpoints, targeted R&D subsidies, and patent design - and provide empirical evidence that surrogate endpoints can be effective in practice.

*This paper was previously circulated under the title "Do fixed patent terms distort innovation? Evidence from cancer clinical trials." We thank the editor, Pinelopi Goldberg, and three anonymous referees for detailed and thoughtful comments that greatly improved the paper. We thank Meru Bhanot, David Burk, Toby Chaiken, Brian Cordonnier, Max Cytrynbaum, Greg Howard, Niels Joaquin, Megan McGrath, and Ana Medrano-Fernandez for excellent research assistance. Daron Acemoglu, Chris Adams, David Autor, Nick Bloom, Tim Bresnahan, Raj Chetty, Joe Doyle, Dan Fetter, Amy Finkelstein, Ray Fisman, Alberto Galasso, Matt Gentzkow, Wes Hartmann, Amanda Kowalski, Anup Malani, Matt Notowidigdo, Felix Oberholzer-Gee, Ariel Pakes, David Ridley, Al Roth, Jon Skinner, Alan Sorensen, Scott Stern, Glen Weyl, and seminar participants at BEA, Carnegie Mellon, the CBO, Chicago Booth, the Cornell Empirical Patent Law Conference, ETH-Zurich, the FTC, Georgia Tech, Harvard Law School, LSE, MIT, the NBER (Health Care, Industrial Organization, Law and Economics, Productivity, and Public Economics), the Northwestern Kellogg Healthcare Markets conference, Princeton, Stanford, Toulouse, U-Arizona, UC-Berkeley, UC-Berkeley Haas, UCSD, UCLA Anderson, UIUC, U-Penn Wharton, UT-Austin, UVA, Wellesley, and Yale Law School provided very helpful comments. Research reported in this publication was supported by the National Institute on Aging and the NIH Common Fund, Office of the NIH Director, through Grant U01-AG046708 to the National Bureau of Economic Research (NBER); the content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NBER. 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 is also gratefully acknowledged. Contact: eric.budish@chicagobooth.edu, broin@mit.edu, heidiw@mit.edu.

1 Introduction

Over the last five years, eight new drugs have been approved to treat lung cancer, the leading cause of US cancer deaths.[1] All eight drugs targeted patients with the most advanced form of lung cancer, and were approved on the basis of evidence that the drugs generated incremental improvements in survival. A well-known example is Genentech's drug Avastin, which was estimated to extend the life of late-stage lung cancer patients from 10.3 months to 12.3 months.[2] In contrast, no drug has ever been approved to prevent lung cancer, and only six drugs have ever been approved to prevent any type of cancer. While this pattern could solely reflect market demand or scientific challenges, in this paper we investigate an alternative hypothesis: private firms may invest more in late-stage cancer drugs – and "too little" in early-stage cancer and cancer prevention drugs – because late-stage cancer drugs can be brought to market comparatively quickly, whereas drugs to treat early-stage cancer and to prevent cancer require a much longer time to bring to market. More broadly stated, we investigate whether private firms differentially underinvest in long-term research, by which we mean technologies with long time lags between the initial spark of an idea and the availability of a commercially viable product. We document evidence that such underinvestment is quantitatively significant in an important context – treatments for cancer – and analyze potential policy responses.

The idea that companies may be excessively focused on behaviors with short-run payoffs is an old one. A large policy- and practitioner-oriented literature has conjectured that managers may maximize short-term rather than long-term firm value (Porter, 1992; Council on Competitiveness, 1992; National Academy of Engineering, 1992). In the academic literature, Stein (1989) and others have argued that firms may be more impatient than neoclassical models would predict due to frictions such as agency problems within the firm. While such corporate short-termism has been widely discussed, little empirical evidence exists to either support or refute this view (see Stein (2003) for a survey and Asker, Farre-Mensa and Ljungqvist (2014) for a more recent contribution).

We propose an additional reason why private firms may be particularly likely to focus on the short term in the context of research and development (R&D): the structure of the patent system.[3] Patents

[1] See the lists of US Food and Drug Administration (FDA) approved hematology/oncology drugs by year: http://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm279174.htm.

[2] Specifically, Avastin was approved for "unresectable, locally advanced, recurrent or metastatic non-squamous NSCLC [non-small cell lung cancer]" patients and the clinical trial effectiveness estimate is posted on the Genentech website: http://www.gene.com/media/product-information/avastin-lung. As noted on the website, this is the first drug to extend median survival time for this patient population beyond one year.

[3] While the importance of patents has been debated in many industries, given our empirical focus on the pharmaceutical industry it is worth noting that a variety of evidence suggests that patents play a key role in motivating innovation in the pharmaceutical industry, including industry interviews (Mansfield, Schwartz and Wagner 1981; Mansfield 1986; Levin et al. 1987; Cohen, Nelson and Walsh 2000), the cost structure of new drug development relative to the generic production (DiMasi, Hansen and Grabowski 2003; Adams and Brantner 2006; Wroblewski et al. 2009), and the fact that standard investment

award innovators a fixed period of market exclusivity, e.g. 20 years in the United States. 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. This means that the patent system provides – perhaps inadvertently – very little incentive for private firms to engage in long-term research.[4] Our theoretical model clarifies that, in fact, there is a sense in which corporate short-termism and fixed patent terms reinforce each other in distorting private research dollars away from long-term investments. The fixed patent term reduces the number of calendar years for which private firms enjoy monopoly protection on investments, and excess discounting reduces the weight the private firm places on each of those years relative to the societal weight.

The idea that firms may underinvest in long-term research – while intuitive – is difficult to test empirically. The key prediction is that there is "missing" private R&D on scientifically feasible projects that would be developed but for their long commercialization lags. In practice, we do not observe the commercialization lags of projects that are never developed, and "missing" private R&D is hard to distinguish from alternative explanations such as a lack of market demand or a lack of scientific opportunities.

Two features of cancer markets allow us to make progress on quantifying this missing R&D. First, the treatment of cancer patients is organized around the organ (e.g. lung) and stage (e.g. metastatic) of disease, which provides a natural categorization of both observed and potential R&D activity. Second, for each such group of cancer patients we observe a good predictor of how long it would take to commercialize drugs for those patients: survival time. Survival time predicts commercialization lags because a firm commercializing a new cancer drug must complete FDA-required clinical trials showing evidence that the drug is safe and effective; and, for cancer clinical trials, "effective" is usually interpreted as improving survival.[5]

To illustrate, 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

models used by pharmaceutical firms pay close attention to effective patent length (Mayer Brown 2009). Informal interviews we conducted with venture capitalists for this paper also support this view, in the sense that these interviews highlighted the fixed term structure of the patent system as something that has important effects on research investments (see Appendix C).

[4] It has long been recognized that heterogeneity across inventions - such as variation in risk-adjusted costs of development - implies that any given fixed patent term will award "too much" market exclusivity to some inventions, and will be insufficient to motivate the development of others; on optimal patent length and optimal patent breadth see, e.g., Machlup (1958), Nordhaus (1969), Scherer (1972), Nordhaus (1972), Kaplow (1984), Klemperer (1990), Gilbert and Shapiro (1990), and Scotchmer (1991). Relative to this literature, the patent analysis in our model highlights a simple - and, we think, important - specific form of heterogeneity in patent-provided incentives arising from commercialization lags that has potentially important consequences for welfare.

[5] There are exceptions to this general statement, which our empirical work will take advantage of.

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). 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 a large difference in clinical trial length: 3 years for the metastatic patient trial versus 18 years for the localized patient trial. Consistent with the idea that commercialization lags differentially reduce private R&D incentives, 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.

We construct data on all clinical trials for cancer treatments over the period 1973-2011, which we match to data on patient survival times over the same period. Our survival data is drawn from patient-level cancer registry data, which we aggregate to cancer-stage level patient groups. Our measure of cancer treatment R&D is newly constructed from a clinical trial registry that has cataloged cancer clinical trials since the 1970s. The key feature of these R&D data which enables 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 link between our measures of expected commercialization lag (survival time) and R&D activity (clinical trial investments) across cancer types and stages of disease.

Using this data, we document that - consistent with our conjectured distortion - patient groups with longer commercialization lags (as proxied by 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. This pattern is even more stark if we contrast recurrent cancers (advanced cancers with very poor survival prospects) and cancer prevention: fewer than 500 trials in our data aim to prevent cancer, whereas recurrent cancers have more than 17,000 trials. A rough adjustment for market size - looking at the number of clinical trials per life-year lost from cancer - does little to change this basic pattern.

This new fact - a negative correlation between commercialization lags and R&D investments - is consistent with our conjectured distortion. However, by itself this fact is difficult to interpret for two reasons. First, this correlation need not reflect a causal relationship between commercialization lags and R&D investments. For example, if scientific opportunities are comparatively scarce for early-stage cancers

then a policy that shortened commercialization lags may have no effect on R&D investments. Second, even if this fact did reflect a causal effect of commercialization lags on R&D investments, on its own this fact need not be evidence of a distortion. As clarified by our theoretical model, the social planner is also more likely to pursue research projects that can be completed more quickly. To address these two concerns, we document evidence from two complementary empirical tests.

First, we document causal evidence that shortening commercialization lags increases R&D investments. The key idea behind this test is to take advantage of the fact that some types of cancers are allowed to use surrogate endpoints (that is, non-mortality based clinical trial endpoints), which break the link between patient survival rates and clinical 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. This suggests that our cross-sectional fact is unlikely to be explained by factors such as the pattern of available scientific opportunities. However, this test leaves open the possibility that the social planner and private firms symmetrically respond to commercialization lags, and thus does not provide direct evidence of a distortion.

Second, we contrast public and private R&D investments. Consistent with our model, we document that commercialization lags reduce both public and private R&D investments. But also consistent with our model – and consistent with the conjectured distortion – we document that the commercialization lag-R&D correlation is quantitatively and statistically significantly more negative for privately financed trials than for publicly financed trials.

As a complement to these empirical analyses, we also provide case study evidence documenting that all six FDA-approved cancer prevention technologies – technologies that should have long commercialization lags, and hence should be affected by the conjectured distortion – either relied on the use of surrogate endpoints or were approved on the basis of publicly financed clinical trials. That is, with the exception of a few instances where surrogate endpoints were able to be utilized, there have been zero privately developed chemoprevention drugs. Taken together, this body of evidence provides support for the idea that commercialization lags distort private R&D investments.

Our theoretical model describes two potential mechanisms for our empirical results – corporate short-termism and the patent distortion – but our results do not speak to which mechanism is quantitatively more important. The existing literature also provides little insight into the expected magnitudes of either mechanism. On one hand, the corporate finance literature has struggled to devise tests for the presence of short-termism bias, in part because the key theoretical implications often focus on behaviors that by construction are undertaken by managers but unobserved by the market. Perhaps most closely related is Bernstein (forthcoming), who documents that public firms pursue lower "quality" R&D than privately

held firms, but he lacks a direct measure of commercialization lags. On the other hand, the innovation literature has provided remarkably little evidence that stronger patent protection induces more R&D investments. For example, Lerner (2002) and Sakakibara and Branstetter (2001) find little evidence that stronger intellectual property rights induce more R&D.[6]

We use our theoretical model to analyze the innovation and social welfare consequences of three policy levers that could address this distortion: allowing firms to rely on surrogate endpoints in clinical trials, a patent design change that would start the patent clock at commercialization, and R&D subsidies targeting projects with long commercialization lags. Two aspects of this analysis are important to highlight. First, surrogate endpoints have benefits beyond just eliminating the distortion, because the social planner also values completing projects more quickly. Second, patent reforms would address only the distortion generated by patents, and would not address the distortion generated by corporate short-termism. Given that our empirical work does not quantify the relative importance of corporate short-termism and patents, our analysis of patent reforms as a policy lever should be considered suggestive rather than conclusive.

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. Taking advantage of our surrogate endpoint variation, we estimate counterfactual R&D allocations and induced improvements in cancer survival rates that would have been observed if commercialization lags were reduced. Based on these counterfactuals, we estimate that among one cohort of patients - US cancer patients diagnosed in 2003 - longer commercialization lags generated around 890,000 lost life-years; valued at $100,000 per life-year lost (Cutler, 2004), the estimated value of these lost life-years is on the order of $89 billion.

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 derives a back-of-the-envelope estimate of the life-years lost due to longer commercialization lags. Section 7 concludes.

[6]While the prior innovation literature has primarily focused on how patents affect the level of R&D, note that our model suggests a mechanism through which the structure of the patent system may also have important effects on the direction of R&D. This idea has been discussed by several legal scholars (Abramowicz (2007); Eisenberg (2005); Roin (2010)), but to the best of our knowledge has not previously been formally investigated either theoretically or empirically.

2 Theory

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 commercialization lag of an R&D project is the amount of time between invention and commercialization.

Our purposefully simple model shows why private-sector R&D may be distorted away from inventions with long commercialization lags. Note importantly that both private and social R&D incentives decline with commercialization lag – all else equal, both firms and society prefer inventions to come to market quickly. But, due to either excessive discounting or the fixed patent term, private incentives will decline more rapidly than social incentives, which is what gives rise to the distortion. Our model also analyzes three potential policy interventions which can be used to address this distortion.

Reflecting our empirical setting we focus the model on the pharmaceutical industry, though our analysis applies more broadly.

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. 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 t_{comm}, which we call *commercialization lag*. In the context of the pharmaceutical industry, commercialization lags arise both in research and in clinical development; to fix ideas, think of t_{comm} as the number of years that it will take to conduct US Food and Drug Administration (FDA)-required clinical trials. We treat t_{comm} and several other parameters below as deterministic for simplicity; in practice many of the parameters would be stochastic.

- Cost of commercialization: if the firm elects to commercialize the invention it incurs commercialization costs of c. For simplicity, we treat commercialization costs as a one-time cost incurred at time t_{invent}.[7] Conceptually, we think of the firm as deciding at time t_{invent} whether to allocate capital to

[7] An alternative approach would be to interpret c as the net present value of costs that are incurred over t_{comm} years, but this raises the issue of which discount factor to use for the purpose of computing this net present value – the neoclassical

the project; e.g., in pharmaceuticals, the firm decides at time t_{invent} whether to invest in conducting clinical trials.

- 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}.[8] 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 patent theory literature in taking obsolescence risk as exogenous (e.g. Grossman and Lai (2004)).[9]

- 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, and social value of v^{monop} per year.[10] If the product were priced by a social planner instead of a monopolist, it would yield social value of $v > v^{monop}$ per year.

- Imitability: if the product is successfully commercialized, non-obsolete, and not protected by patent, generic entrants may imitate the commercialized product.[11] Imitation reduces the inventing firm's profits from π to $(1-\iota)\pi$, where $\iota \in [0,1]$ denotes the imitability of the product (that is, vulnerability to generic competition). The case $\iota = 1$ corresponds to perfect imitability, which drives the inventing firm's profits to zero. We focus on $\iota = 1$ for most of the analysis, but note that even in

discount factor δ or the short-termism discount factor $\eta\delta$. Treating costs as incurred at time t_{invent} circumvents this issue, and captures the idea that clinical trials require similar financial resources whether they are funded by a private firm or the government. Our approach also abstracts from staged investment and the associated real-option considerations which, while important, are not directly related to the goals of our model; see, e.g., Gompers (1995) and Neher (1999) for analyses of staged financing.

[8] 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.

[9] Across industries, many inventions become obsolete long before their patents expire (Schankerman and Pakes 1986). However, this is generally not the case in the pharmaceutical industry, as many drugs are still in use long after their initial FDA approval date and generate significant sales revenues near the end of their patent term (Grabowski and Kyle 2007).

[10] A natural alternative assumption would be to model profits as endogenous to entry, since more competition could result in lower profits. We do not focus on this possibility here given that in our context, this would cut against our distortion: projects with short commercialization lags should have more entry, and be lower profit, which would in turn lower incentives for subsequent entry. Given that our data suggest that this dynamic is not sufficiently strong to offset our main finding - that projects with short commercialization lags have more entry - we focus on an exogenous profit parameter for simplicity.

[11] In the pharmaceutical industry, generic manufacturers are usually poised to enter the market as soon as patents expire (Grabowski and Kyle 2007; Hemphill and Sampat 2011). Such formal analyses are consistent with anecdotal evidence that industry analysts and e.g. the *Wall Street Journal* closely track patent expirations in the pharmaceutical industry, and these patent expirations tend to result in sharp changes in the profitability of branded drugs.

pharmaceuticals generic entry sometimes does not drive profits all the way to zero (see Bronnenberg et al. (2013)).

- Discounting and excess impatience: the project's neoclassical risk-adjusted discount rate is r. Following Stein (2003), corporate short-termism can be modeled as an excessive private-sector discount rate. For mathematical convenience we work with discount factors instead of discount rates, so corporate short-termism is reflected as a lower discount factor. Specifically, society applies the obsolescence-risk-weighted discount factor $\delta = \frac{\gamma}{1+r}$, whereas private firms apply the discount factor $\eta\delta$, with $\eta \leq 1$. The η term reflects excess impatience due to corporate short-termism.

- Patent term and timing of patent filing: in a fixed-term patent system, patents for new inventions last t_{patent} years from the filing date.[12] So long as an invention is protected by patent, imitation is illegal. Firms may choose whether to file for patent protection at the time of invention t_{invent} or at the time of commercialization t_{comm}. If they file at the time of invention they receive patent protection with probability one. If they wait until commercialization to file they receive patent protection with probability $q \leq 1$, reflecting the risk of disclosure, losing an R&D race, etc. Pharmaceutical firms face very strong incentives to file patents at the time of invention (Galli and Faller, 2003; Schreiner and Doody, 2006; Wegner and Maebius, 2001): delaying risks a competitor patenting first, or subsequent disclosures undermining the drug's novelty or non-obviousness for purposes of patentability (Thomas, 2007; Patrick, 2005; Zanders, 2011).[13] In practice, firms almost always have possession of the core patents over their drugs before entering clinical trials (Mossinghoff, 1999; Patrick, 2005; Thomas, 2007).[14] For this reason we focus on the case of $q = 0$ for most of the analysis.

2.2 Effective Monopoly Life and Effective Total Life

We define an invention's *Effective Monopoly Life* (*EML*) as the expected number of years, in present value terms as discounted by the private firm, that the firm can expect to earn monopoly profits from the

[12] 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 2.5.2.

[13] Zanders (2011), for example, argues: *"A question that is often raised during my courses is 'why don't companies wait as long as possible before patenting?' This is tempting, but given the fluid nature of employment in the industry and the general leakiness of information, this would be tantamount to commercial suicide."*

[14] Although the law is not settled, FDA clinical trials most likely constitute a public disclosure of the drug; see SmithKline Beecham Corp. v. Apotex Corp., 365 F.3d 1306, 1318 (Fed. Cir. 2004), *opinion vacated and superseded*, 403 F.3d 1331 (Fed. Cir. 2005). The SmithKline decision held that a drug's use in clinical trials puts it in the public domain, but since that opinion was vacated and the court decided the case on other grounds, the state of the law here is unclear. Once an invention is in the public domain, the inventing firm must file for patent protection within one year of public disclosure else they lose the right to patent (35 U.S.C. § 102).

commercialized product. This is the expected amount of time that the invention is commercially viable, protected by patent, and not yet superseded. We focus our analysis on the case of inventions that are imitable if not protected by patent ($\iota = 1$) and where firms must file for patent protection at invention in order to receive patent protection ($q = 0$). This is the most relevant case for the pharmaceutical industry; below we discuss other cases.

If $t_{patent} > t_{comm}$ then EML can be written as:

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

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} (obsolescence risk is incorporated into δ). As soon as time reaches t_{patent}, the invention will be imitated and the monopoly position lost. Note as well that if $t_{patent} \leq t_{comm}$, then $EML = 0$: by the time the invention is commercialized, patent protection has expired.

Next, we define an invention's *Effective Total Life (ETL)* as the expected number of years, in present value terms as discounted by society, that the invention will be commercialized and non-obsolete:

$$ETL = p \sum_{t_{comm}}^{\infty} \delta^t = p\frac{\delta^{t_{comm}}}{1-\delta}. \tag{2}$$

There are two differences between EML and ETL. First, monopoly life runs at best until t_{patent}, whereas total life runs indefinitely until the invention becomes obsolete. Second, monopoly life is measured according to the private-sector discount factor $\eta\delta$ whereas total life is measured according to the social discount factor δ.

If the invention is not perfectly imitable ($\iota < 1$) then the formula for EML would need to be modified to account for the fact that profits do not fall all the way to zero at t_{patent}.[15] In the extreme case of zero imitability ($\iota = 0$) and zero short-termism ($\eta = 1$), EML and ETL coincide. If the invention has q that is not only strictly positive but sufficiently large, then the formula for EML would need to be modified to account for the fact that firms may choose to file for patent protection at t_{comm} rather than t_{invent}.[16] In this case, the period of monopoly protection runs from t_{comm} to $t_{comm} + t_{patent}$, but the firm enjoys a

[15] The modified formula becomes $EML = p\left(\sum_{t_{comm}}^{t_{patent}-1}(\eta\delta)^t + (1-\iota)\sum_{t_{patent}}^{\infty}(\eta\delta)^t\right)$.

[16] The specific condition to check to see whether firms prefer to patent at t_{invent} or t_{comm} is which is larger of $p\sum_{t_{comm}}^{t_{patent}-1}(\eta\delta)^t$ or $pq\sum_{t_{comm}}^{t_{comm}+t_{patent}-1}(\eta\delta)^t$. Clearly, the former is larger for sufficiently small q (as is the case in pharmaceuticals) and the latter is larger for sufficiently large q.

successful, patent-protected invention with probability of just pq rather than p.

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:

$$\text{Private Investment Occurs} \iff EML \cdot \pi \geq c. \tag{3}$$

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:

$$\text{Investment is Socially Optimal} \iff ETL \cdot v \geq c. \tag{4}$$

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.[17] 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:

$$\text{Private and Social Investment Differ} \iff \frac{EML \cdot \pi}{c} \leq 1 \leq \frac{ETL \cdot v}{c}. \tag{5}$$

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 commercialize A but not B, while a social planner would prefer to commercialize B over A. That is, the private

[17] Bloom, Schankerman and Van Reenen (2013) provide a recent analysis estimating the magnitude of business stealing.

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:[18]

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 strictly lower than socially optimal, unless (a) patent terms are infinite (i.e., $t_{patent} = \infty$); (b) firms are not excessively impatient (i.e., $\eta = 1$); and (c) 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:*[19]

   (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., the ratio of monopoly life to total useful life is larger 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.[20] 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 a non-excludable form 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 a specific pattern of exercise reduces the risk of heart disease is non-excludable.

An invention's monopoly-life to total-life ratio, $\frac{EML}{ETL}$, describes the proportion of the invention's total useful life in which the private firm enjoys monopoly profits. Our central point is that an invention's $\frac{EML}{ETL}$

[18] Proofs are presented in Appendix A.

[19] 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).

[20] 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). Nordhaus (2004) estimates that this general conclusion holds outside of the pharmaceutical industry as well, arguing that only a minuscule fraction of the social returns from technological advances over the 1948-2001 period was captured by producers.

ratio declines with commercialization lag t_{comm}, due to both short-termism and the fixed patent term. To see this, write out the expression for $\frac{EML}{ETL}$ assuming that $t_{comm} \leq t_{patent}$:

$$\frac{EML}{ETL} = \frac{p\frac{(\eta\delta)^{t_{comm}} - (\eta\delta)^{t_{patent}}}{1-\eta\delta}}{p\frac{\delta^{t_{comm}}}{1-\delta}} = \frac{1-\delta}{1-\eta\delta}(\eta^{t_{comm}} - \eta^{t_{patent}}\delta^{t_{patent}-t_{comm}}) \tag{6}$$

Notice, first, that if $\eta = 1$ and $t_{patent} = \infty$ – there is no short-termism and patent length is infinite – then $\frac{EML}{ETL} = 1$ for all t_{comm}.[21] Commercialization lag reduces incentives to invest, but it reduces both private and social incentives to invest at exactly the same rate.

Notice, too, that if $\eta = 1$ and $t_{patent} = t_{comm} + k$ – that is, the patent term is finite but with the patent clock modified to start at commercialization, not invention (recall that we have normalized $t_{invent} = 0$) – then $\frac{EML}{ETL}$ again does not vary with t_{comm}. EML is strictly less than ETL under this patent design, but, just as with infinite patents, commercialization lag reduces private and social incentives at exactly the same rate.

However, if either $\eta < 1$ or the patent term is finite and starts at invention, then $\frac{EML}{ETL}$ declines with t_{comm}. The decline in private incentives is more rapid than the decline in social incentives.

Proposition 2. *Comparative statics of an invention's proportion of monopoly life to total life, $\frac{EML}{ETL}$, on its commercialization lag, t_{comm}:*

1. *If there is no short-termism ($\eta = 1$) and the patent term is either infinite ($t_{patent} = \infty$) or is finite but the clock starts at commercialization ($t_{patent} = t_{comm} + k$ for finite k), then the ratio of monopoly life to total life, $\frac{EML}{ETL}$, is constant in t_{comm}: $\frac{\partial \frac{EML}{ETL}}{\partial t_{comm}} = 0$.*

2. *If firms are excessively impatient ($\eta < 1$) or the patent term is finite and starts at invention, $\frac{EML}{ETL}$ is decreasing in t_{comm}.*

   (a) *If $t_{comm} < t_{patent}$ the decline is strict: $\frac{\partial \frac{EML}{ETL}}{\partial t_{comm}} < 0$*

   (b) *If $t_{comm} \geq t_{patent}$ then $EML = 0$. Hence $\frac{EML}{ETL} = 0$.*

This result, in combination with Proposition 1, shows that private-sector R&D is particularly distorted away from R&D projects with long commercialization lags, relative to projects with shorter commercialization lags. Moreover, there is a sense in which the effect of excess impatience on $\frac{\partial \frac{EML}{ETL}}{\partial t_{comm}}$ and the effect of the fixed patent term on $\frac{\partial \frac{EML}{ETL}}{\partial t_{comm}}$ reinforce each other. The fixed patent term means that increasing t_{comm} by 1 year reduces the number of calendar years of monopoly life by 1 year. Excess discounting

[21] Recall that while our analysis focuses on on the case of perfect imitability ($\iota = 1$), an economically equivalent condition to $t_{patent} = \infty$ is if $\iota = 0$. We discuss imperfect imitability in Section 2.5.2.

means that the private firm places too little weight on each of these years of monopoly life relative to their societal value. To see this decomposition formally, define an invention's *effective patent life* as $EPL = p\sum_{t_{comm}}^{t_{patent}-1}\delta^t = p\frac{\delta^{t_{comm}}-\delta^{t_{patent}}}{1-\delta}$; EPL is EML but using the social discount factor δ. We can decompose $\frac{EML}{ETL}$ into an excessive discounting term and a fixed patent term as:

$$\frac{EML}{ETL} = \underbrace{\frac{EML}{EPL}}_{\text{excess discounting}} \cdot \underbrace{\frac{EPL}{ETL}}_{\text{fixed patents}} \tag{7}$$

It is easy to see that both terms in this decomposition are strictly declining with commercialization lag:

Proposition 3. *Decomposition of* $\frac{\partial \frac{EML}{ETL}}{\partial t_{comm}}$ *into the effect of excess discounting and the effect of the fixed patent term:*

1. *If there is excess discounting – $\eta < 1$ – then* $\frac{\partial \frac{EML}{EPL}}{\partial t_{comm}} < 0$ *for* $t_{comm} < t_{patent}$.
2. *If there is a fixed patent term – a finite patent clock that starts at invention – then* $\frac{\partial \frac{EPL}{ETL}}{\partial t_{comm}} < 0$ *for* $t_{comm} < t_{patent}$.

Two hypothetical examples can illustrate this distortion in the composition of R&D. A vaccine administered to men at age 20 which prevented prostate cancer (which tends to affect men in their 50s or later) would have a high social value v (given the high morbidity and mortality burden of prostate cancer), but would have a low (or zero) $\frac{EML}{ETL}$ ratio because of the long required clinical trials. In contrast, a drug administered to late-stage prostate cancer patients which extended life from, say, 6 months to 8 months would have a lower social value v, but a high $\frac{EML}{ETL}$ ratio because of the short required clinical trials. Note that in the case of these examples, our distortion of interest - generated by the difference in $\frac{EML}{ETL}$ ratios - would be reinforced by differences in $\frac{\pi}{v}$.

2.5 Policy Responses

Our empirical work will provide support for the idea that private-sector R&D activity is distorted away from projects with long commercialization lags. Given that evidence, in this subsection we discuss the innovation and social welfare consequences of three policy interventions that could be used to address this distortion: a policy change that would allow firms to rely on surrogate (non-mortality) endpoints in clinical trials; a patent design change that would start the patent clock at commercialization; and targeted R&D subsidies. Some readers may prefer to skip this section on a first reading, returning to our analysis of policy responses after reading the empirical analysis.

2.5.1 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 extremely 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 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 generate improvements in survival.[22] Reflecting this type of concern, most cancers use surrogate endpoints only on a limited, somewhat *ad hoc* basis.[23]

[22] A non-cancer example of the controversy around surrogate endpoints arose recently in the context of treatments for early-stage Alzheimer's disease. In a 2013 editorial in the *New England Journal of Medicine*, two FDA officials discussed the possibility of accepting new types of surrogate endpoints in clinical trials of treatments for early-stage Alzheimer's disease (Kozauer and Katz, 2013) - a proposal that was sharply criticized by the editorial board of the *New York Times* (Editorial Board, 2013), among others.

[23] 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 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. A recent President's Council of Advisors on Science and Technology (2012) report argued that the FDA should expand this accelerated approval program.

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 assumption allows us to make the following simple point.

Proposition 4. *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 independent of $\hat{t}_{comm}$ – that is, if the time required to show impacts on the surrogate endpoint is independent of the time required to show impacts on mortality – then allowing surrogate endpoints eliminates the distortion in composition associated with commercialization lag absent the surrogate endpoint: $\frac{\partial}{\partial x}\mathbb{E}\left(\frac{EML}{ETL}|\hat{t}_{comm} = x\right) = 0$.*

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 scientific discovery, validation, and allowance of valid surrogate endpoints.[24] Note that surrogate endpoints are valuable both because they eliminate the distortion in composition of R&D and because, even in the absence of a distortion, it is socially valuable to complete R&D projects sooner.

2.5.2 Patent design

In this section we discuss modifications to the fixed term patent design that address the distortion away from long-term R&D projects. Note, importantly, that the patent design policy response differs from our other policy responses in that it addresses only the fixed patent term as a source of distortion, and not excessive discounting. As we will discuss below, if patents are unimportant for motivating R&D (formally,

[24]The use of invalid surrogate endpoints could increase R&D investments but not generate any corresponding gains in survival. In the specific empirical context we analyze in Section 5.1, we will document evidence that surrogate endpoints for hematologic cancers appear to have increased R&D investments, and that this increase in R&D investments appears to have translated into real improvements in patient health.

imitability ι is zero), the patent design policy response will not be effective at addressing the distortion of interest, but in our simple framework this policy reform would also not be harmful.[25]

We begin with a simple result, analogous to Proposition 2 part (1), that shows that starting the patent clock at commercialization, rather than invention, eliminates the distortion in composition arising from the patent system.

Proposition 5. *If the patent clock starts at commercialization, i.e., $t_{patent} = t_{comm} + x$ for fixed and finite x, then $\frac{EPL}{ETL}$ is independent of commercialization lag, t_{comm}.*

If we make some admittedly stylized assumptions 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 6. *Make the following assumptions about the distribution of invention parameters: $\delta < 1$ and $\eta \leq 1$ are constant across inventions, so that EML varies only with commercialization lag t_{comm}, patent life t_{patent}, and success probability p; the social-to-private value ratios $\frac{v}{\pi}$ and $\frac{v^{monop}}{\pi}$ are constant across inventions; the density of inventions on the extensive margin, i.e., the expected number of new inventions elicited by a marginal increase in t_{patent}, is uniform; and, the expectation of costs, c, conditional on an invention being at the margin, is weakly increasing in t_{comm}. Suppose that private firms make commercialization decisions according to equation (3). 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 informally 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 (3) obtains. The cost is that, for

[25] Our model focuses on a Nordhaus (1969)-style trade-off between the incentives for developing a new innovation and the deadweight loss of higher prices during the life of the patent. By construction, this type of framework abstracts away from other ways in which patent reforms could impact social welfare, including business stealing, the effects of patents on follow-on innovation, litigation, or the benefits of the disclosure function of the patent system.

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 makes two key observations. First, the deadweight loss costs on the intensive margin are strictly decreasing with t_{comm} – both because the costs are pushed out further into the future and because the set of invention parameters for which private firms choose to commercialize is shrinking. Second, the benefits at the extensive margin are actually increasing with t_{comm}: for a private firm to be willing to commercialize an invention with higher t_{comm}, the invention must be higher quality in the sense of higher private value π – especially if the firm is excessively impatient – which in turn implies higher social value v^{monop} and v.[26] Intuitively, when t_{comm} is large, the inventions at the margin are especially worth encouraging, and the cost of doing so is comparatively low. 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 we acknowledge that our assumption of constant social-to-private value ratios $\frac{v^{monop}}{\pi}$ and $\frac{v}{\pi}$ is stylized, we note that the types of inventions that take longer to reach the market (e.g., treatments of early-stage disease and disease prevention) seem likely to have especially high such ratios. If these ratios increase with t_{comm}, then this increases the rate at which benefits at the extensive margin increase with t_{comm}, strengthening 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.[27] In practice, awarding FDA-granted exclusivity periods that run from the date of FDA approval would likely accomplish the same goal, be administratively simpler to implement, and avoid unintended problems that

[26] It is not necessary for the result that benefits at the extensive margin are weakly increasing with t_{comm}, only that they do not decrease too quickly (i.e., faster than do the deadweight loss costs on the intensive margin). For this reason, several of the assumptions in the proposition can be slightly relaxed. We have a numerical example, in which the density of the extensive margin is bimodal with a large decline between the two modes, which illustrates that the conclusion of the proposition is false if the density falls off too quickly. Intuitively, in the region in which the density on the extensive margins is very low, it is not sufficiently socially valuable to elicit inventions on the extensive margin to justify the deadweight loss costs for inventions on the intensive margin.

[27] 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 (2012) - could also be used.

could arise with revising the patent system.[28]

Our next result considers a more limited set of patent-design instruments than is allowed for by Proposition 6 and shows that there is still scope for improvement.

Proposition 7. *Suppose that the length of the patent term must be fixed, but that the patent clock can start either at invention or commercialization. Make the same assumptions regarding the distribution of invention parameters as in Proposition 6. Given any patent term that runs from the date of invention, there exists a patent term that runs from the date of commercialization that strictly increases social welfare. In particular, the optimal patent term that runs from the date of commercialization is superior to the optimal patent term running from the date of invention.*

Proposition 7 is useful for informing patent policy if it is possible to start the patent clock at commercialization, but difficult to condition the length of the patent award on the precise amount of time between invention and commercialization. As with the optimal policy considered above in Proposition 6, this more circumscribed policy proposal could be implemented via FDA-granted exclusivity periods as opposed to a restructuring of the patent system. A recent policy in the spirit of this result is a provision of the 2010 Patient Protection and Affordable Care Act (Public Law Number 111-148, 2010), which grants some qualifying drugs (specifically, biologic drugs) a 12-year exclusivity period running from the date of FDA approval, which runs concurrently with any remaining patent terms. Proposition 7 supports extending this type of post-approval exclusivity period to all drug approvals (but note that our analysis does not specify the optimal length of such an exclusivity period).

A caveat to the results in this section is that they presume that patents are an important way to incentivize research and development activity.[29] If patents do not increase research investments, the policy responses analyzed in this section would be ineffective. Formally, consider an industry in which imitability $\iota = 0$, so patents are not necessary to protect monopoly profits from projects. In such an industry, the modifications to patent design outlined in Propositions 5-7 will have no effect on R&D activity, although it is worth noting that in our simple framework these policy responses would not be harmful, only ineffective. By contrast, our other policy responses would be effective in such an industry provided that corporate short-termism is relevant ($\eta < 1$).

[28] FDA exclusivity precludes the approval of other drugs with the same active moiety, and is currently granted to new drug applications (three years for new indications; five years for new molecular entities); to orphan drugs (seven years); and to pediatric approvals (six months).

[29] As discussed in Footnote 3, while patents have been controversial in many industries, a variety of sources of evidence suggest that patents are likely to be important in the pharmaceutical industry.

2.5.3 Policy Lever: Targeted R&D subsidies

The logic that targeted R&D subsidies can improve social welfare 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.

Recall that we defined an invention's effective patent life as $EPL = p\sum_{t_{comm}}^{t_{patent}-1}\delta^t$, i.e. EPL is just like EML except that it uses the social discount factor δ rather than the private discount factor $\eta\delta$. The condition for the existence of a socially beneficial R&D subsidy is:

$$EML \cdot \pi < c \text{ and } c + \tau(c - EML \cdot \pi) < EPL \cdot v^{monop} + (ETL - EPL) \cdot v. \tag{9}$$

In words, the conditions are that, first, the private firm would not commercialize on its own, and, second, that the social value from commercialization exceeds the social costs – both the direct cost of commercializing, c, and the deadweight loss cost of the required subsidy. Notice that the private firm's commercialization decision (first without and then with the subsidy) depends on EML, whereas the value society gets from the commercialized invention during the period it is under patent protection depends on EPL.

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 a larger proportion of their useful life off-patent, 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 8. *Make the same assumptions regarding the distribution of invention parameters as in Proposition 6. Suppose that private firms make commercialization decisions according to whether or not $EML \cdot \pi + s \geq c$, 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, for any target level of total subsidy expenditures, 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 6 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, the most practical way to condition subsidies on t_{comm} might be to target subsidies at R&D that relates to treatment of early-stage disease and to disease prevention.

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.[30] 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. Such a match is possible in large part because 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.[31] Cancer registry data records the organ and stage of the primary tumor at the time of diagnosis, thus 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.

Sections 3.1, 3.2, and 3.3 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-

[30] 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.9, we outline a power calculation of the type used to guide the design of clinical trials in order to fix ideas on this point.

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

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) 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).[32] 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 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 estimate 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.[33] 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

[32] 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.

[33] Clinical trials are also used as a measure of R&D investments in Finkelstein (2004).

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 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) and have data on trials through 2011. 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.[34]

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

[34]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.

observations).

Table 1 presents some basic summary statistics on our cancer-stage level data. Between 1973-2011, an average 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 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 Descriptive analysis

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 life-years lost from cancer. The right-hand-side axis plots the number of clinical trials enrolling patients of each stage, divided by the number of life-years lost from that cancer-stage as a rough adjustment for

market size.[35] 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.[36] 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.[37] 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

[35] As described in Section 3, 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.

[36] To give a visual sense of the data for a few major cancers, Appendix Figure D.1 plots the relationship between the five-year survival rate and clinical trial activity for the "big four" cancers: breast, colon, lung, and prostate. Appendix Figure D.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). Appendix Figure D.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.

[37] Estimates from ordinary least squares models using the log of the number of clinical trials as the dependent variable are essentially identical (estimates not reported).

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 log of 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 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 on this correlation. First, we ask whether the survival time-R&D correlation is similar when estimated within cancers (cancer fixed effects) and within stages (stage fixed effects). Appendix Table D.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.[38] Second, we ask whether the survival time-R&D correlation is robust to alternative measures of patient survival time. Appendix Table D.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 D.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 D.4 confirms that the survival time-R&D correlation also holds in our sample of approved drugs.

[38] 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 D.2 shows residualized scatterplots corresponding to the regression specifications presented in Appendix Table D.1 on this same sample.

5 Interpreting the correlation between survival time and R&D investments

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 commercialization lags, as proxied by survival rates. This fact is consistent with the idea that private firms may underinvest in long-term research, because we observe lower levels of R&D investment on inventions that require longer commercialization lags. However, by itself this fact is difficult to interpret for two reasons. First, other factors - such as heterogeneity in demand and heterogeneity in the costs of R&D - could generate the same qualitative pattern. 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 long commercialization lags is more difficult, and that observed low R&D investments reflect a paucity of scientific opportunities. Second, even if this correlation does reflect a causal effect of commercialization lags on R&D investments, it need not be evidence of a distortion, because the social planner is also more likely to pursue research projects that can be completed more quickly.

To address these concerns, in this section we document estimates from two additional empirical tests. First, in Section 5.1 we document causal evidence that shortening commercialization lags increases R&D investments. This evidence suggests - for example - that heterogeneity in demand is unlikely to explain the cross-sectional relationship between survival time and R&D. However, this test leaves open the possibility that the social planner and private firms symmetrically respond to commercialization lags, and thus does not provide direct evidence of a distortion. Our second empirical test in Section 5.2 fills this gap by directly contrasting public and private R&D investments. Section 5.3 documents supporting qualitative evidence from historical case studies of FDA-approved chemoprevention drugs, which suggest that surrogate endpoints and public subsidies have played a key role in the development of chemoprevention drugs. Taken together, this body of evidence provides support for the idea that commercialization lags distort private R&D investments.

.

5.1 Investigating surrogate endpoints

If heterogeneity in demand for treatments or a paucity of scientific opportunities were driving the survival time-R&D correlation, the observed correlation should be independent of whether surrogate endpoints

are used. In contrast, our model predicts that surrogate endpoints should make the survival time-R&D correlation less negative (Proposition 4). 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.[39]

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)).[40]

To investigate the effects of surrogate endpoints on R&D activity, we use both our clinical trials data and our drug approvals data.[41] In the sample of approved drugs, we can confirm that hematological 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 three predictions of our model that relate to commercialization activity. First, Part 1 of Proposition 4 predicts that the use of surrogate endpoints should increase commercial-

[39] As highlighted above, surrogate endpoints enable shorter trials, so this test does not provide direct evidence of a distortion; we address this issue in a separate test in Section 5.2.

[40] Based on our reading of these FDA writings, our understanding is that both scientists and regulators have viewed the surrogate endpoints used for hematologic cancers as valid and uncontroversial. Although far from definitive, our empirical evidence in Section 6 is consistent with this view, suggesting that the additional R&D investments induced by the use of these surrogate endpoints have translated into improved survival gains.

[41] 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 (as detailed in Appendix B) 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 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.

ization 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 ($\beta = 0.753$) suggests a 112 percent increase in clinical trials for hematological malignancies relative to non-hematological malignancies ($(e^{\beta} - 1) \cdot 100 \approx 112\%$). 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.

Second, Part 3 of Proposition 4 predicts that - if survival time is independent of the time required to show impacts on the surrogate endpoint - then the use of surrogate endpoints should reduce the negative relationship between survival time and R&D investments. Third, in cases where surrogate endpoints do *not* decrease commercialization lag, our model implies that surrogate endpoints should *not* change R&D incentives. That is, for cancers that have a short commercialization lag even in the absence of using a surrogate endpoint, the option to use a surrogate endpoint should not change R&D incentives. Empirically, this means that we expect hematologic and non-hematologic cancers to have similar levels of R&D for the set of cancers that have short commercialization lags even in the absence of using surrogate endpoints (that is, for low survival time cancers).

To test these second and third predictions 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 second prediction of our model.[42] This estimate is robust to the inclusion of controls for market size (Columns (2) and (3)). This contrast in survival time-R&D correlations across hematologic and non-hematologic cancers is presented graphically in Figure 4.[43]

With respect to the third prediction of our model, we find that the estimated coefficient on the indicator

[42] Interpreting the interaction term in this non-linear model requires transforming the coefficient; the interaction coefficient of 2.266 in the first row of Panel (B) implies that an increase in the five-year survival rate of 10 percentage points predicts an increase in the number of trials for hematologic cancers that is greater than that of nonhematologic cancers by 300 trials (about 30 percent relative to the mean), and applying the delta method to obtain a standard error for this interaction term provides a t-statistic of 5.99. Figure 4 gives an alternative sense of the magnitude of the coefficients obtained from a linear model.

[43] Appendix Table D.5 shows that this pattern of results also holds in the drug approvals data.

variable for hematologic cancers is - statistically speaking - zero, and also relatively small in magnitude. In addition to being consistent with our model, this result is also important as a test of a key assumption underlying this counterfactual exercise - namely, that hematologic cancers and non-hematologic cancers would have similar R&D investments but for the more frequent use of surrogate endpoints for hematologic cancers. A priori, hematologic and non-hematologic cancers are very different for many reasons - for example, the science of treating hematologic cancers might be simpler for some reason. However, to the extent that such differences are common across all hematologic cancers, hematologic cancers with low five-year survival rates should have higher levels of R&D investments than do non-hematologic cancers with low five-year survival rates. But that is not what we see in the data: rather, hematologic and non-hematologic cancers have similar levels of R&D investments for the patient groups where surrogate endpoints should not change R&D incentives. This evidence is consistent with the "all else equal" assumption behind this hematologic/non-hematologic comparison.

What can we learn from this counterfactual exercise? We draw two conclusions. First, from the perspective of testing the model, our estimates are consistent with the idea that neither unobserved heterogeneity in demand nor a paucity of scientific opportunities is driving the observed negative survival time-R&D correlation in the full sample. Second, from a policy perspective our estimates support the idea (analyzed in Proposition 4) that valid surrogate endpoints may increase R&D investments, particularly on long-horizon R&D investments. The key caveat to interpreting this evidence as a test of our theoretical model is that because surrogate endpoints change the length of clinical trials, both the social planner and private firms should choose to increase research investments. Hence, this test does not provide direct evidence of a distortion; our second empirical test in Section 5.2 fills this gap by directly contrasting public and private R&D investments.

In Appendix A.10, we use this hematologic/non-hematologic comparison to provide a rough back-of-the-envelope estimate of the semi-elasticity of R&D investment with respect to a one-year change in commercialization lag: $\frac{\partial(\text{R\&D investment})}{\partial(\text{commercialization lag})}$.[44] Our main estimates of this semi-elasticity range between 7-23%.[45] It is worth noting that this elasticity is itself of policy relevance, as an input into how firms would be expected to respond to decreases in commercialization lags as provided by mechanisms such as FDA priority review vouchers (Ridley, Grabowski and Moe, 2006).

[44] As described in Appendix A.10, obtaining this semi-elasticity estimate requires scaling our estimate of how R&D investment changes in response to a change in the five-year survival rate ($\frac{\partial(\text{R\&D investment})}{\partial(\text{5-year survival rate})}$) by an estimate of how a change in the five-year survival rate translates into a change in commercialization lag ($\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$).

[45] We are not aware of any existing estimates against which this estimate can be compared.

5.2 Investigating publicly financed clinical trials

Our second empirical test directly contrasts public and private R&D investments. Consistent with our theoretical model, we document that commercialization lags reduce both public and private R&D investments. But also consistent with our model - and consistent the conjectured distortion - we will see that the commercialization lag-R&D correlation is quantitatively and statistically significantly more negative 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, essentially 100 percent are publicly funded.[46]

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.[47] 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 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-

[46] 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.

[47] 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.

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 35 percent larger for privately financed clinical trials relative to publicly financed clinical trials ($\frac{4.4}{4.4+8.6} \approx 35\%$). 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 models of public sector decision making: the public sector could have a different objective function than the private sector (as in our model), or the public sector could be compensating for underinvestment by the private sector. Both models are consistent with the existence of a 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.

5.3 Historical case studies of FDA-approved chemoprevention drugs

As a complement to our empirical analyses, we also document qualitative - case study - evidence on what motivated the development of existing chemoprevention drugs. Because cancer prevention trials typically examine cancer incidence as an outcome variable, we expect cancer prevention technologies to generally require long trials and thus to also be subject to our conjectured 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. Table 5 documents a summary of our work in Appendix E, which provides documentation for this assertion, and we here focus on briefly summarizing 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. Hence, the evidence from these case studies is quite consistent with the conjectured distortion: we expect cancer prevention trials to have long commercialization lags, and no cancer prevention technologies have been privately developed without relying on surrogate endpoints.

6 Estimating the value of life lost due to commercialization lags

In this section, we leverage our surrogate endpoint variation from Section 5.1 to estimate counterfactual improvements in cancer survival rates that would have been observed if commercialization lags were reduced.[48] Importantly, this exercise should not be interpreted as quantifying the size of our conjectured distortion, because as discussed surrogate endpoints generate social value beyond eliminating the distortion. As with our back-of-the envelope estimates of the semi-elasticity of R&D investment with respect to changes in the commercialization lag, this exercise is directly policy relevant as an input into how firms would be expected to respond to decreases in commercialization lags as provided by mechanisms such as the application of valid surrogate endpoints or FDA priority review vouchers (Ridley, Grabowski and Moe, 2006).

Figure 6 illustrates how we use variation in surrogate endpoints (across hematologic and non-hematologic cancers) to estimate counterfactual survival gains from 1973-2003. Panel (a) of Figure 6 illustrates our conceptual framework. If there had been no survival improvements between 1973 and 2003, all cancer-stage observations would locate along the 45 degree line ("no progress line"); in contrast, if all cancer-stages had been cured between 1973 and 2003, all cancer-stage observations would locate along the horizontal line where 2003 survival rates equal 1 ("cure cancer line"). As discussed in Section 5.1, we expect two patterns to emerge when contrasting survival improvements across for hematologic and non-hematologic cancers. First, survival improvements should be similar for hematologic and non-hematologic cancers

[48] While we would ideally quantify R&D-induced improvements in both morbidity and mortality, given data constraints we here focus on estimating the extent to which R&D increases patient survival.

in cases where surrogate endpoints do not shorten commercialization lags (that is, for cancers with low 1973 five-year survival rates). Second, the difference in survival improvements between hematologic and non-hematologic cancers should increase in commercialization lag (that is, increase in the 1973 five-year survival rate). Reflecting these predictions, the line marked "non-hematologic cancers" coincides with the line marked "hematologic cancers" at 0% survival, and the gap between the two lines increases as commercialization lag increases.

Panel (b) plots the observed 2003 five-year survival rates against the 1973 five-year survival rates. Strikingly, the data matches our illustrative figure in Panel (a) remarkably well. In particular, the linear fit lines for hematologic cancers and non-hematologic cancers meet for cancers with a very low 1973 five-year survival rate; the linear fit for hematologic cancers is close to a parallel shift of the 45 degree line (slightly steeper, as expected based on Figure 4); and the linear fit for non-hematologic cancers is much more shallow in slope. Note that given the dearth of quasi-experimental evidence documenting that increases in pharmaceutical R&D translate into improved survival (see, e.g., Lichtenberg (2012)), this evidence that the additional R&D investments induced by shorter commercialization lags (by relying on surrogate endpoints) translated into improved survival gains is itself of substantive interest.[49]

The area between the linear fit line for hematologic cancers and the linear fit line for non-hematologic cancers can be used to quantify the number of life-years that would have been gained if commercialization lags for non-hematologic cancers had been similar to commercialization lags for hematologic cancers. We formalize this estimation for the cohort of US cancer patients diagnosed in 2003 as follows. First, on the sample of hematologic cancers, we predict the 2003 five-year survival rate as a function of the 1973 five-year survival rate. Second, for the sample of non-hematologic cancers we use the estimated β from the hematologic cancers survival regression to predict a counterfactual 2003 five-year survival rate for non-hematologic cancers had commercialization lags for non-hematologic cancers been similar to commercialization lags for hematologic cancers. Third, we calculate δ_{cs}, the difference between the

[49] Welch et al. (2000) and others have argued that although five-year survival is a valid measure for comparing cancer therapies in a randomized trial, changes in five-year survival rates over time may be biased by changes in diagnosis patterns (known as 'lead-time bias'). For example, an expansion in mammography screening between 1973 and 2003 could have led to breast cancers being diagnosed at an earlier stage, which would have mechanically increased measured five-year survival rates even if there was no real change in patient health. In our context, changes in diagnosis would be expected to bias us away from finding that hematologic cancers saw larger gains in survival between 1973 and 2003 because the cancers that saw increases in screening over this period (such as breast and prostate cancer) are non-hematologic cancers. Empirically, if we construct an alternative version of Figure 6 Panel (b) that plots Welch et al.'s preferred outcome variable - the percent change in mortality from 1973 to 2003 - against the 1973 five-year survival rate, we observe a very similar pattern to that displayed in Figure 6 Panel (b): first, hematologic cancers on average had larger percent improvements in mortality from 1973 to 2003 than did non-hematologic cancers; second, as predicted by our model there is no gap between the hematologic and non-hematologic lines for patient groups with near-zero 1973 five-year survival rates; and third, the gap between the hematologic and non-hematologic lines increases in magnitude as the 1973 five-year survival rate increases. Taken together, these results suggest that changes in diagnosis patterns are not generating the differential patterns of survival changes across hematologic and non-hematologic cancers presented in Figure 6 Panel (b).

counterfactual and actual 2003 five-year survival rates, for each non-hematologic cancer-stage; on average, δ_{cs} is 13.2 percentage points. Fourth, we convert each δ_{cs} into a number of life-years lost per person based on the fact that, in our data, a change from 0 to 1 in the five-year survival rate corresponds to a gain of 8.1 additional years of life. Applying this conversion, the average δ_{cs} of 13.2 percentage points corresponds to $(8.1)(.132) = 1.07$ life-years per cancer patient. Fifth, we multiply each cancer-stage estimate of per-person life-years lost by the number of US cancer patients diagnosed in 2003 with that cancer-stage. We compute the number of patients in each cancer-stage using the SEER data, scaling up (dividing by 0.074) to account for the fact that SEER does not cover the entire US population. In total, this calculation suggests that among this cohort of patients - US cancer patients diagnosed in 2003 - the longer commercialization lags required for non-hematologic cancers generated around 890,000 lost life-years.

If we value each lost life-year at $100,000 (Cutler, 2004), the estimated value of these lost life-years is on the order of $89 billion per annual patient cohort. Applying a conservative social discount rate of 5% and assuming that patient cohorts grow with population growth of 1%, the net present value of the life-years at stake is $\frac{\$89\,\text{billion}}{.05-.01} = \$2.2\,\text{trillion}$.[50]

It is important to note that this life-lost estimate is rough at best. Our point estimate of the value of life lost per annual patient cohort is $89 billion, with a 95 percent confidence interval that ranges from $7 billion to $172 billion; the net present value point estimate of $2.2 trillion has a 95 percent confidence interval that ranges from $170 billion to $4.2 trillion.[51]

7 Discussion and conclusion

In this paper, we investigate whether private firms underinvest in long-term research projects. Our theoretical model clarifies how two factors - corporate short-termism and the structure of the patent system - may generate incentives that distort private research investments away from inventions that have both a long useful life and a long commercialization lag. We then investigate this distortion empirically in the context of the pharmaceutical industry, where drugs treating patients with short life expectancies can move through clinical trials more quickly than can drugs treating patients with longer life expectancies. Using a newly constructed data set on cancer clinical trial investments, we provide several sources of

[50] Note that other authors, such as Murphy and Topel (2006) and Weitzman (1998), have argued that a social discount rate of 2% or lower may be more appropriate; using such lower values would of course increase our estimate of the net present value of life-years at stake.

[51] To be conservative, we compute these confidence intervals using HC3 standard errors rather than robust standard errors, given the expected downward finite sample bias of robust standard errors in this small sample of hematologic cancers (see, e.g., the discussion in Angrist and Pischke (2009)). The analogous 95 percent confidence interval using robust standard errors is $15 billion to $164 billion (a net present value range from $365 billion to $4.1 trillion).

evidence which together are consistent with commercialization lags distorting private R&D investments away from drugs to prevent or treat early-stage cancers.

We use our theoretical model to analyze the innovation and social welfare consequences of three policy interventions which could address this distortion: a policy change that would allow firms to rely on surrogate endpoints in clinical trials, a patent design change that would start the patent clock at commercialization, and R&D subsidies targeting projects with long commercialization lags. While surrogate endpoints and targeted R&D subsidies would address the distortion regardless of the source, the patent design change only addresses the fixed patent term distortion.

Empirically, we document evidence - consistent with our theoretical model - that surrogate endpoints appear to increase R&D investments on innovations that would otherwise have long commercialization lags. We also use this surrogate endpoint variation to estimate counterfactual improvements in cancer survival rates that would have been observed if commercialization lags were reduced. We estimate that among one cohort of patients - US cancer patients diagnosed in 2003 - longer commercialization lags resulted in around 890,000 lost life-years. Valuing these lost life-years at $100,000 (Cutler, 2004) suggests that the estimated social value of the life-years lost in this one cohort of patients is on the order of $89 billion per year. This evidence suggests that - in the case of hematologic cancers - apparently-valid surrogate endpoints were effective in increasing R&D investments on innovations that would otherwise have had long commercialization lags, and that the resulting increases in R&D translated (in this case) into real gains in patient health. While much attention has been focused on the risks and costs of using surrogate endpoints that may imperfectly correlate with real improvements in patient health, our analysis is - to the best of our knowledge - the first attempt to use the historical record to quantify how the availability and use of a valid surrogate endpoint affected R&D allocations and patient health outcomes.

The example of the Framingham Heart Study is helpful in illustrating the potential value of surrogate 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.[52] Although some of these gains are due to lifestyle changes, much of the decline in heart disease has been attributed to improved pharmacological preventives and 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

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

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). Note that public subsidies - such as federal support for the Framingham study - were likely important in this context, because any individual firm's investment in discovering and validating surrogate endpoints would generate benefits that largely spill over to other firms. Both our empirical evidence on the effects of surrogate endpoints for hematologic cancers and this historical case study for heart disease suggest that research investments aimed at establishing and validating surrogate endpoints may have a large social return.[53]

[53]This type of argument has also been made informally 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). Collins (2012) provides an example of a technology - a chip that mimics how humans respond to a drug - that could serve the same role as a surrogate endpoint, by identifying promising candidate drugs more quickly.

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.

Angrist, Joshua and Jorn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press, 2009.

Asker, John, Joan Farre-Mensa, and Alexander Ljungqvist, "Corporate investment and stock market listing: A puzzle," mimeo, UCLA 2014.

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

Bernstein, Shai, "Does going public affect innovation?," Journal of Finance forthcoming.

Bloom, Nick, Mark Schankerman, and John Van Reenen, "Identifying technology spillovers and product market rivalry," *Econometrica*, 2013, *81*, 1347–1393.

Bronnenberg, Bart, Jean-Pierre Dube, Matthew Gentzkow, and Jesse Shapiro, "Do pharmacists buy Bayer? Sophisticated shoppers and the brand premium," mimeo, Chicago Booth 2013.

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.

Collins, Francis, "Testimony before the US House Subcommittee on Labor - HHS - Education Appropriations," Department of Health and Human Services, National Institutes of Health 2012.

Council on Competitiveness, "Capital Choices: Changing the Way America Invests in Industry," research report presented to the Council on Competitiveness and co-sponsored by the Harvard Business School 1992.

Cutler, David, *Your Money or Your Life: Strong Medicine for America's Health Care System*, Oxford University Press, 2004.

_ **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.

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.

Editorial Board, "Drugs for early-stage Alzheimer's," *New York Times*, 2013, *17 March.*

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.

Galli, Bruno and Bernard Faller, "Discover a drug substance, formulate and develop it to a product," in Camille Georges Wermuth, ed., *The Practice of Medicinal Chemistry*, 2 ed. 2003, pp. 857–866.

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.

Gompers, Paul, "Optimal investment, monitoring, and the staging of venture capital," *Journal of Finance*, 1995, *50* (5), 1461–1489.

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.

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

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

Kozauer, Nicholas and Russell Katz, "Regulatory innovation and drug development for early-stage Alzheimer's disease," *New England Journal of Medicine*, 2013, *e-publication ahead of print.*

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.

Lichtenberg, Frank, "The effect of pharmaceutical innovation on longevity: Patient-level evidence from the 1996-2002 Medical Expenditure Panel Survey and linked mortality public-use files," National Bureau of Economic Research (NBER) working paper 18552 2012.

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.

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.

Mossinghoff, Gerald, "Overview of the Hatch-Waxman Act and its impact on the drug development process," *Food & Drug Law Journal*, 1999, *54* (2), 187–194.

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.

National Academy of Engineering, *Time Horizons and Technology Investments*, National Academy Press, 1992.

Neher, Darwin, "Staged financing: An agency perspective," *Review of Economic Studies*, 1999, *66*, 255–274.

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.

_ , "Schumpeterian profits in the American economy: Theory and measurement," National Bureau of Economic Research (NBER) working paper 10433 2004.

Patrick, Graham, *An Introduction to Medicinal Chemistry*, 2 ed., Oxford University Press Oxford, 2005.

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.

Porter, Michael, "Capital choices: Changing the way America invests in industry," *Journal of Applied Corporate Finance*, 1992, *5* (2), 4–16.

President's Council of Advisors on Science and Technology, *Report to the President on Propelling Innovation in Drug Discovery, Development, and Evaluation*, http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-fda-final.pdf, 2012.

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 111-148, "Patient Protection and Affordable Care Act," 124 Stat. 119 2010.

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

Ridley, David, Henry Grabowski, and Jeffrey Moe, "Developing drugs for developing countries," *Health Affairs*, 2006, *25* (2), 313–324.

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.

Schreiner, Stephen and Patrick Doody, "Patent continuation applications: How the PTO's proposed new rules undermine an important part of the US patent system with hundreds of years of history," *Journal of the Patent and Trademark Office Society*, 2006, *88*, 556–569.

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.

_ , *Medicinal Chemistry: An Introduction*, Wiley-Interscience, 2007.

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.

Wegner, Harold and Stephen Maebius, "The global biotech patent application," in "Biotechnology Law: Biotechnology Patents& Business Strategies in the New Millennium" Intellectual property course handbook, Practising Law Institute, 2001, pp. 129–130.

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

Weitzman, Martin, "Why the far-distant future should be discounted at its lowest possible rate," *Journal of Environmental Economics and Management*, 1998, *36* (3), 201–208.

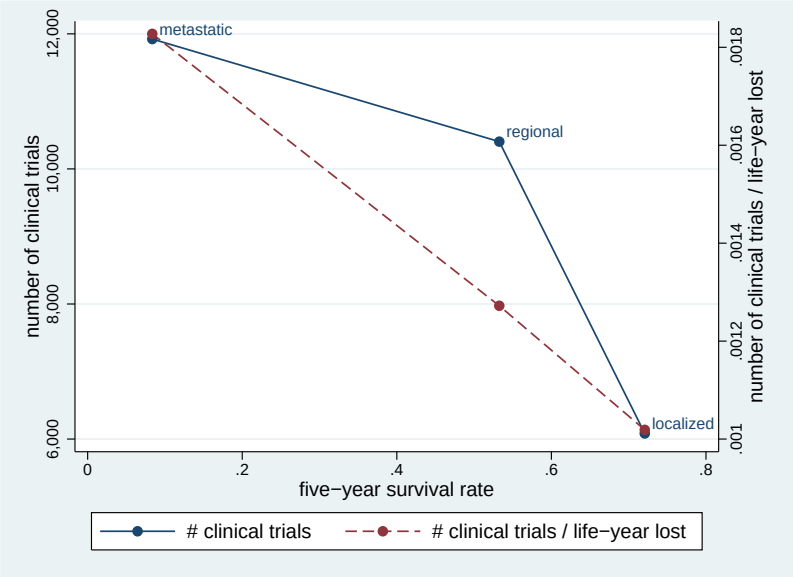
Welch, H. Gilbert, Lisa Schwartz, and Steven Woloshin, "Are increasing 5-year survival rates evidence of success against cancer?," *Journal of the American Medical Association*, 2000, *283* (22), 2975–2978.

Weyl, Glen and Jean Tirole, "Market power screens willingness-to-pay," *Quarterly Journal of Economics*, 2012, *127*, 1971–2003.

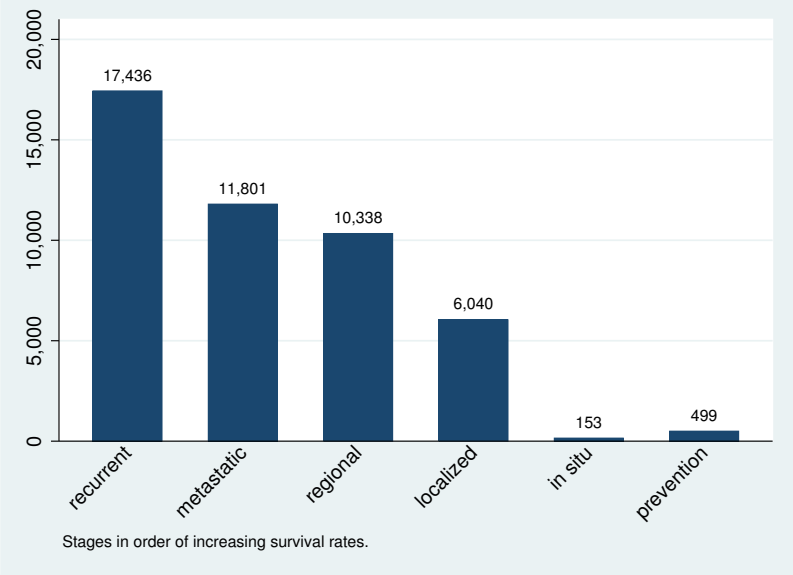
Wroblewski, Michael, Elizabeth Jex, Suzanne Drennon Munck, Christopher Garmon, Suzanne Michel, Alissa Wantman, and Ossie Neal, "Emerging health care issues: Follow-on biological drug competition," Federal Trade Commission Report http://www.ftc.gov/os/2009/06/P083901biologicsreport.pdf 2009.

Zanders, Edward, *The Science and Business of Drug Discovery: Demystifying the Jargon*, Springer Verlag, 2011.

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



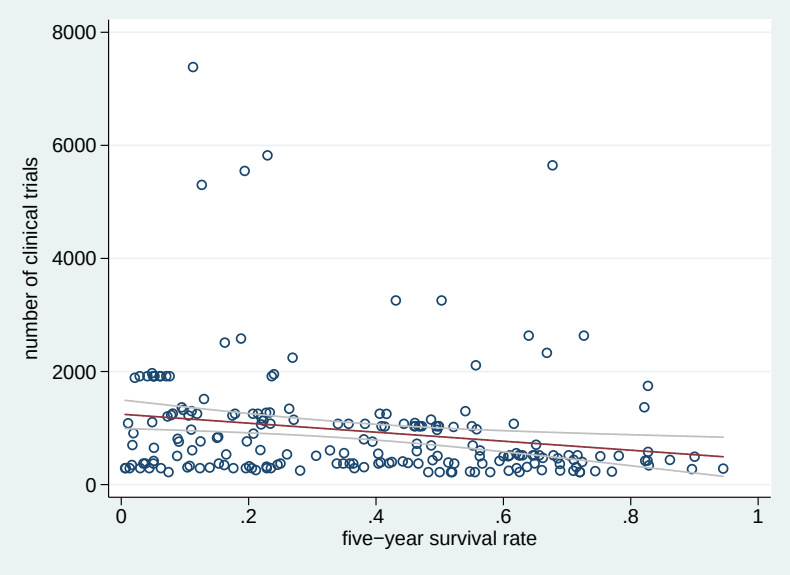
(a) R&D investments by five-year survival rates



(b) R&D investments by stage

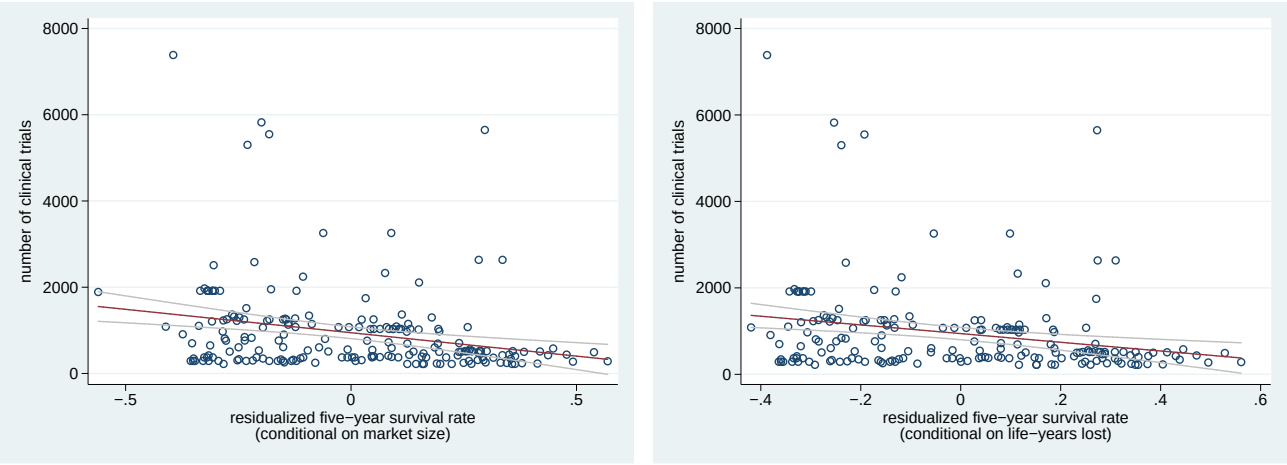
Notes: This figure plots measures of clinical trial activity for each stage of cancer from 1973-2011. Panel (a) plots two measures of clinical trial activity for each stage of cancer from 1973-2011 against the five-year survival rate among patients diagnosed with each stage between 1973-2004 (the cohorts 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 life-years lost - 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. 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. Note that because we here count the number of clinical trials patients of each cancer-stage are eligible to enroll in, a higher count of trials appears here than in Figure 1 because many trials enroll patients of more than one cancer-stage type. 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

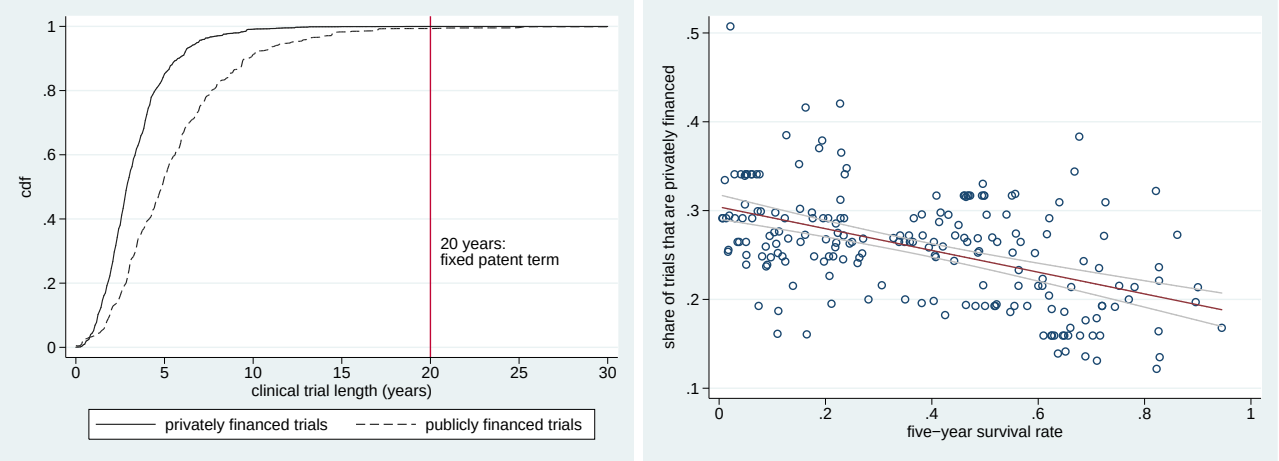
(b) Life-years lost

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**

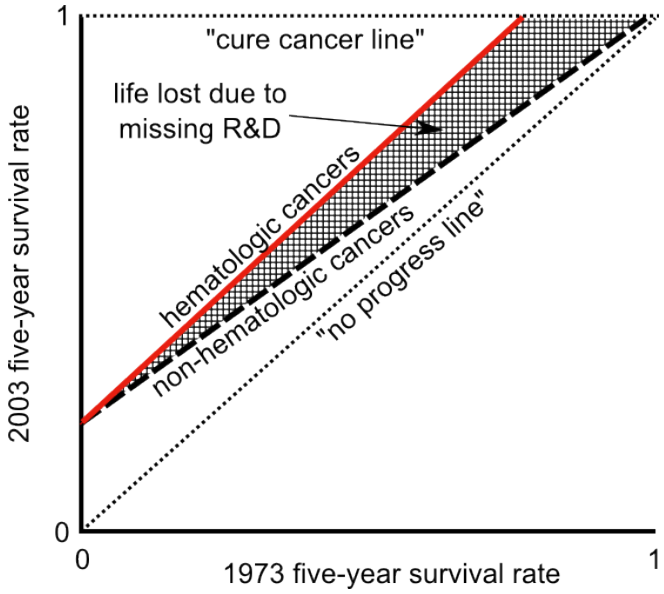


(a) Cumulative distribution function of clinical trial length

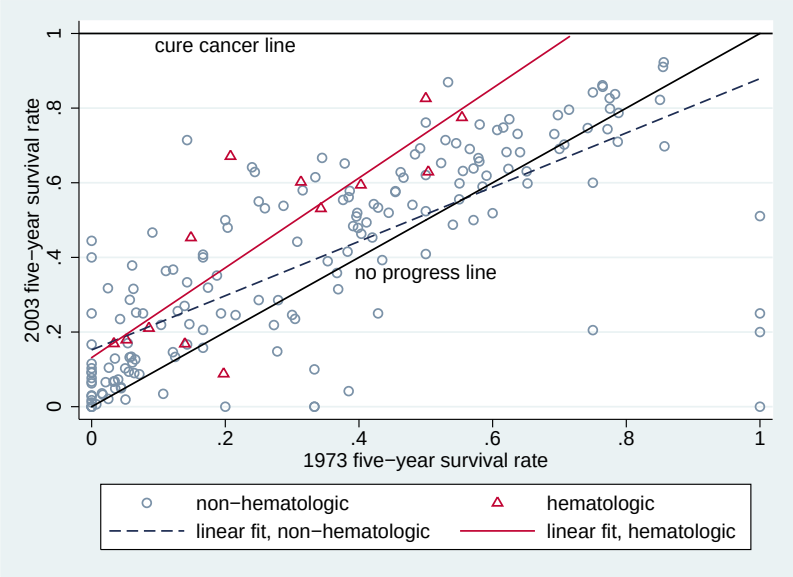
(b) Survival time and private financing

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. For details on the sample, see the text and data appendix.

Figure 6: **Survival gains, 1973-2003**



(a) Framework for analyzing survival gains, 1973-2003



(b) Observed survival gains, 1973-2003

Notes: This figure illustrates how we use variation in surrogate endpoints (across hematologic and non-hematologic cancers) to estimate counterfactual survival gains from 1973-2003 that would have been observed had commercialization lags for non-hematologic cancers mirrored the shorter commercialization lags realized for hematologic cancers. Panel (a) illustrates our conceptual framework. Panel (b) illustrates the empirical analog of Panel (a), plotting the 1973 five-year survival rate against the 2003 five-year survival rate. The level of observation is the cancer-stage. 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	-0.868	***	-1.113	***	-0.930	***
	(0.319)		(0.286)		(0.286)	
log(market size)	-		0.243	***	-	
			(0.055)			
log(life-years lost)	-		-		0.282	***
					(0.068)	

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 number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* denotes 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	-0.865	***	-1.108	***	-0.933	***
	(0.310)		(0.284)		(0.283)	
(0/1: *hematologic*)	0.753	***	0.578	***	0.466	**
	(0.185)		(0.176)		(0.201)	
log(market size)	-		0.231	***	-	
			(0.057)			
log(life-years lost)	-		-		0.261	***
					(0.073)	
Panel (B): Composition of R&D, Dependent variable: Number of clinical trials (mean = 945)						
	(1)		(2)		(3)	
(five-year survival rate)*(0/1: *hematologic*)	2.266	***	2.140	***	1.963	***
	(0.408)		(0.541)		(0.613)	
five-year survival rate	-1.122	***	-1.309	***	-1.133	***
	(0.343)		(0.297)		(0.303)	
(0/1: *hematologic*)	-0.077		-0.216		-0.261	
	(0.189)		(0.228)		(0.252)	
log(market size)	-		0.226	***	-	
			(0.056)			
log(life-years lost)	-		-		0.253	***
					(0.073)	

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 number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* denotes 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	***	-0.134	***	-0.119	***
	(0.016)		(0.017)		(0.014)	
log(market size)	-		0.009	***	-	
			(0.003)			
log(life-years lost)	-		-		0.008	***
					(0.003)	
Panel (B): Dependent variable: Number of clinical trials (mean = 244)						
	(1)		(2)		(3)	
(five-year survival rate)*(0/1: *private*)	-0.436	***	-0.500	***	-0.470	**
	(0.166)		(0.171)		(0.195)	
five-year survival rate	-0.866	***	-1.097	***	-0.932	***
	(0.314)		(0.287)		(0.285)	
(0/1: *private*)	-0.681	***	-0.723	***	-0.833	***
	(0.062)		(0.054)		(0.081)	
log(market size)	-		0.230	***	-	
			(0.063)			
log(market size)*(0/1: *private*)	-		0.003	***	-	
			(0.002)			
log(life-years lost)	-		-		0.257	***
					(0.076)	
log(life-years lost)*(0/1: *private*)	-		-		0.001	***
					(0.000)	

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 number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* denotes 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.

Table 5: **Historical case studies of FDA-approved chemoprevention drugs**

	approval indication	**surrogate endpoint used?**	**primarily publicly funded?**
BCG	bladder carcinoma in situ	no	yes
Diclofenac	squamous cell carcinomas	yes	no
Celecoxib	familial adenomatous polyposis (FAP)-related cancers	yes	no
Photofrin	esophageal carcinoma	yes	no
Tamoxifen	breast cancer	no	yes
Cervical cancer vaccines	cervical cancer	yes	no

Notes: This table summarizes our qualitative investigation of the history of all six FDA drugs approved as cancer prevention (chemoprevention) drugs. The key point of this table is to illustrate 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: no chemoprevention drugs have been privately developed in the absence of relying on surrogate endpoints. See the descriptions in Section 5.3 and Appendix E for more details.

A Appendix: Proofs (not for publication)

A.1 Proof of Proposition 1

Part 1 follows immediately from (5) since conditions (a) and (b) together imply $EML = ETL$ and condition (c) states $\pi = v$.

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.

A.2 Proof of Proposition 2

Follows immediately from equation (6) as described in the text.

A.3 Proof of Proposition 3

Proof of Part 1. For this proof we work with the integral forms of EML and EPL: specifically, $EML = \int_{t_{comm}}^{t_{patent}} (\delta\eta)^t \, dt$ and $EPL = \int_{t_{comm}}^{t_{patent}} \delta^t dt$. Observe that, for any $t_{comm} \leq t_{patent}$, the ratio $\frac{(\delta\eta)^t}{\delta^t} = \eta^t$ is positive and strictly decreasing in t over the interval $[t_{comm}, t_{patent}]$. This immediately implies that $\frac{\partial\left(\frac{EML}{EPL}\right)}{\partial t_{comm}} < 0$.

Proof of Part 2. $\frac{EPL}{ETL} = 1 - \delta^{t_{patent} - t_{comm}}$. Hence $\frac{\partial \frac{EPL}{ETL}}{\partial_{t_{comm}}} = \log(\delta)\delta^{t_{patent} - t_{comm}} < 0$.

A.4 Proof of Proposition 4

Proof of Part 1. This follows immediately from the private investment condition (3). Since surrogate endpoints decrease t_{comm}, they increase EML, which might cause additional investments to occur.

Formally, let t_{comm} and $\hat{t}_{comm}$ denote a drug's commercialization lag with and without surrogate endpoints, respectively. Consider a drug where surrogate endpoints strictly decrease commercialization lag, i.e. $t_{comm} < \hat{t}_{comm}$. Let $EML^{Surrogate}$ and $EML^{NoSurrogate}$ denote EML with and without the surrogate endpoint; $t_{comm} < \hat{t}_{comm}$ implies that $EML^{Surrogate} > EML^{NoSurrogate}$. 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 decrease commercialization lag, and hence always increase EML.

Proof of Part 2. The social welfare associated with a successfully commercialized invention is $EPL \cdot v^{monop} + (ETL - EPL) \cdot v$, where EPL, effective patent life, is defined according to $EPL = p \sum_{t_{comm}}^{t_{patent}-1} \delta^t$ (see discussion in Section 2.5.3). A reduction in t_{comm} strictly increases EPL, and has no effect on $ETL - EPL$, because both ETL and EPL increase by the expected number of additional years that the drug will be commercially available, in present value terms discounted at δ. 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 independent of $\hat{t}_{comm}$.

A.5 Proof of Proposition 5

The result follows immediately from the fact that $\frac{EPL}{ETL} = 1 - \delta^{t_{patent} - t_{comm}}$ since $t_{patent} - t_{comm}$ is now constant.

A.6 Proof of Proposition 6

Define social welfare as a function of commercialization lag, t_{comm}, and post-commercialization patent length, denoted x (i.e., $t_{patent} = t_{comm} + x$), by

$$W(t_{comm}, x) \quad = \quad \int_{EML \cdot \pi \geq c} (EPL \cdot v^{monop} + (ETL - EPL) \cdot v - c)\, dF_{t_{comm}}(\cdot) \tag{13}$$

where $dF_{t_{comm}}(\cdot)$ denotes the distribution of invention parameters conditional on t_{comm}; EML, EPL and ETL are defined as in the text of Section 2 but using the notation $t_{patent} = t_{comm} + x$; and the integral is taken over all inventions that satisfy the private investment condition $EML \cdot \pi \geq c$ as defined in equation (3). Using (13), we can define the optimal choice of post-commercialization patent length x as an implicit function of commercialization lag t_{comm}:

$$x^*(t_{comm}) \in \arg\max_x W(t_{comm}, x)$$

Our goal is to show that $x^*(t_{comm})$ is increasing in t_{comm}. By Topkis's theorem, it is sufficient for us to show that the cross-partial $\frac{\partial^2 W(t_{comm}, x)}{\partial x \partial t_{comm}}$ is strictly positive for all x and t_{comm}. (Topkis's theorem also tells us that a strictly positive cross-partial implies that the optimum $x^*(t_{comm})$ is unique for all t_{comm}).

To study the cross-partial $\frac{\partial^2 W(t_{comm}, x)}{\partial x \partial t_{comm}}$, we first decompose the partial $\frac{\partial W}{\partial x}$ into two components: the benefit from eliciting more inventions at the extensive margin, and the costs of additional deadweight loss from inventions on the intensive margin, i.e., inventions that would have been elicited even without the

increase in x. Write this as follows:

$$\begin{aligned}
\frac{\partial W}{\partial x} &= BenefitsExtensive(t_{comm}, x) - CostsIntensive(t_{comm}, x) \\
BenefitsExtensive(t_{comm}, x) &= k \cdot \mathbb{E}_{(EML \cdot \pi = c)} \left(EPL \cdot v^{monop} + (ETL - EPL) \cdot v - c \right) \\
CostsIntensive(t_{comm}, x) &= \int_{EML \cdot \pi \geq c} p\delta^{t_{comm}+x}(v - v^{monop}) dF_{t_{comm}}(\cdot)
\end{aligned}$$

with k denoting the rate at which new inventions are elicited on the margin, which we have assumed is uniform.[54] We will show that $\frac{\partial^2 W(t_{comm},x)}{\partial x \partial t_{comm}}$ is strictly positive by showing that $\frac{\partial CostsIntensive(t_{comm},x)}{\partial t_{comm}}$ is strictly negative and $\frac{\partial BenefitsExtensive(t_{comm},x)}{\partial t_{comm}}$ is weakly positive.[55]

First, consider $\frac{\partial CostsIntensive(t_{comm},x)}{\partial t_{comm}}$. There are two effects. First, increasing t_{comm} reduces the deadweight loss cost associated with invention parameter tuples on the intensive margin, because these costs are pushed out further in time. Second, increasing t_{comm} reduces the set of inventions for which deadweight loss is suffered. Both effects are negative. Formally,

$$\begin{aligned}
\frac{\partial CostsIntensive(t_{comm}, x)}{\partial t_{comm}} &= \int_{EML \cdot \pi \geq c} \frac{\partial}{\partial t_{comm}} p\delta^{t_{comm}+x}(v - v^{monop}) dF_{t_{comm}}(\cdot) \\
&\quad - \mathbb{E}_{(EML \cdot \pi = c)} \left(p\delta^{t_{comm}+x}(v - v^{monop}) \right) \cdot f_{t_{comm}}(EML \cdot \pi = c)
\end{aligned}$$

The first term simplifies to $\ln(\delta) CostsIntensive(t_{comm}, x)$, which is strictly negative since $\delta < 1$. The second term is weakly negative since $v \geq v^{monop}$, and strictly negative if $v > v^{monop}$. Hence, $\frac{\partial CostsIntensive(t_{comm},x)}{\partial t_{comm}} < 0$.

Next, we sign $\frac{\partial BenefitsExtensive(t_{comm},x)}{\partial t_{comm}}$. Using the relationships $v^{monop} = a \cdot \pi$ and $v = b \cdot \pi$, and the fact that $EML \cdot \pi = c$ on the extensive margin, we can rewrite $BenefitsExtensive(t_{comm}, x)$ as

$$BenefitsExtensive(t_{comm}, x) = k \cdot \mathbb{E}_{(EML \cdot \pi = c)} \left(\left(\frac{EPL}{EML} \cdot a - 1 \right) \cdot c + \left(\frac{ETL - EPL}{EML} \cdot b \right) \cdot c \right) \quad (14)$$

Hence we need to sign

$$\frac{\partial}{\partial t_{comm}} \mathbb{E}_{(EML \cdot \pi = c)} \left(\left(\frac{EPL}{EML} \cdot a - 1 \right) \cdot c + \left(\frac{ETL - EPL}{EML} \cdot b \right) \cdot c \right) \quad (15)$$

We will show that all of the terms in the main parenthetical of (15) are positive and weakly increasing in t_{comm} along the extensive margin $EML \cdot \pi = c$. First, the ratio $\frac{EPL}{EML}$ can be simplified to $\frac{1-\eta\delta}{1-\delta} \frac{(1-\delta^x)}{(1-(\eta\delta)^x)} \eta^{-t_{comm}}$. Since we are holding η and δ fixed they do not vary with t_{comm} at the extensive margin; hence the ratio is increasing in t_{comm} since $\eta^{-t_{comm}}$ is increasing in t_{comm} and all other terms stay constant. Since, in addition $\frac{EPL}{EML} \geq 1$ and $a \geq 1$, we have that the object $\left(\frac{EPL}{EML} \cdot a - 1 \right)$ is positive and weakly increasing. The ratio $\frac{ETL-EPL}{EML}$ simplifies to $\frac{1-\eta\delta}{1-\delta} \frac{\delta^x}{1-(\eta\delta)^x} \eta^{-t_{comm}}$. As with $\frac{EPL}{EML}$, the only

[54] Formally, what we call the density of inventions on the extensive margin – the rate at which additional inventions are elicited as x is increased – is $f(EML \cdot \pi = c) \frac{\partial EML}{\partial x}$ which we assume is constant in t_{comm} up to finite upper bounds on t_{comm} and x. The proof works equivalently if this density is weakly increasing in t_{comm}. The proof works with some modification if this density is decreasing in t_{comm} but not too rapidly; see the next footnote.

[55] As mentioned in the body of the text, it is not necessary for the result that $\frac{\partial BenefitsExtensive(x;t_{comm})}{\partial t_{comm}}$ is weakly positive; it can be negative, so long as it is less negative than $\frac{\partial CostsIntensive(x;t_{comm})}{\partial t_{comm}}$. For this reason, several of our assumptions can be slightly relaxed. For example, the density of inventions on the extensive margin, described in the previous footnote, could be decreasing in t_{comm} so long as the decline is not too rapid. Also, as mentioned in the body of the text, we have a numerical counterexample in which the density declines rapidly within a region; intuitively, in such a region, the benefit of eliciting additional inventions at the extensive margin does not justify the additional deadweight loss costs from the inventions on the intensive margin.

factor that varies with t_{comm} at the extensive margin is $\eta^{-t_{comm}}$, so the ratio is increasing in t_{comm}. Since $ETL \geq EPL$ and $b \geq 1$ we have that $\frac{ETL-EPL}{EML} \cdot b$ is positive as well. Last, expected costs c are weakly increasing in t_{comm} on the extensive margin by assumption. Hence, the parenthetical of (15) consists of positive terms that are all weakly increasing in t_{comm}, hence the sign of $\frac{\partial BenefitsExtensive(t_{comm},x)}{\partial t_{comm}}$ is positive.

For intuition, rewrite the rightmost parenthetical of (15) as $[(EPL \cdot a - 1) + (ETL - EPL) \cdot b] \cdot \pi$. Increasing t_{comm} (while holding fixed x) decreases both EPL and $ETL - EPL$, because it pushes both the period of patent protection and the period post-patent protection out into the future. However, in the other direction, increasing t_{comm} improves the quality of inventions at the extensive margin, as measured by π, which by assumption also improves quality as measured by social value v^{monop} and v. The sign of (15) thus tells us that the benefits from higher quality inventions exceed the costs from additional time discounting. Note as well the role of excess impatience. If $\eta = 1$ then the ratios $\frac{EPL}{EML}$ and $\frac{ETL-EPL}{EML}$ are constant in t_{comm}, whereas if $\eta < 1$ these ratios increase exponentially in t_{comm} at rate $\Omega(\eta^{-t_{comm}})$, which directly causes $\frac{\partial BenefitsExtensive(t_{comm},x)}{\partial t_{comm}}$ to grow exponentially in t_{comm}.

Putting this all together, we have

$$\begin{aligned} sign\left(\frac{dx^*(t_{comm})}{dt_{comm}}\right) &= sign\left(\frac{\partial^2 W(t_{comm},x)}{\partial x \partial t_{comm}}\right) \\ &= sign\left(\frac{\partial BenefitsExtensive(t_{comm},x)}{\partial t_{comm}} - \frac{\partial CostsIntensive(t_{comm},x)}{\partial t_{comm}}\right) \\ &= sign\left([\geq 0] - [< 0]\right) \end{aligned}$$

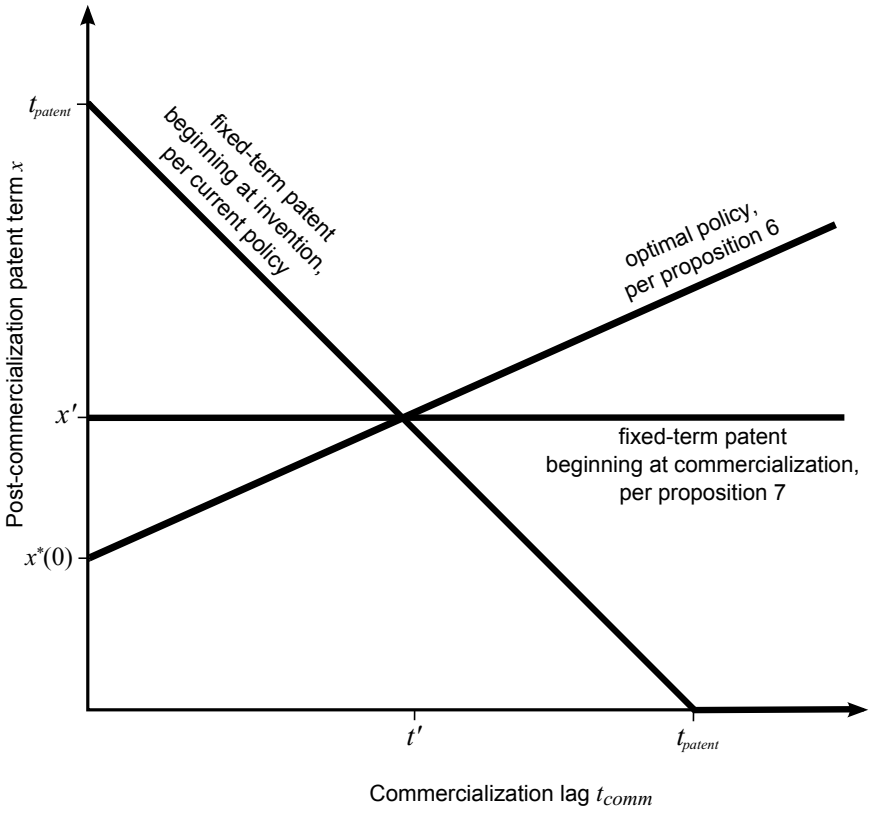
so $\frac{dx^*(t_{comm})}{dt_{comm}} > 0$, as required.

A.7 Proof of Proposition 7

Take as given a fixed patent term running from the date of invention, t_{patent}. This patent term induces a strictly negative relationship between commercialization lag and years of post-commercialization patent life. Letting x_{invent} denote the number of years of post-commercialization patent life when the clock starts at invention, we have $x_{invent}(t_{comm}) = \max(t_{patent} - t_{comm}, 0)$ which is strictly downward sloping in t_{comm} while $t_{comm} < t_{patent}$ and then flat at zero while $t_{comm} \geq t_{patent}$.

In contrast, Proposition 6 shows that the optimal post-commercialization patent life, x^*, is strictly *increasing* in t_{comm}. Suppose that the strictly increasing curve $x^*(t_{comm})$ and the strictly decreasing (then flat) curve $x_{invent}(t_{comm})$ intersect, which will occur iff $x^*(0) < x_{invent}(0)$; that is, if optimal policy involves awarding less than t_{patent} years of post-commercialization protection to inventions that have commercialization lag of 0. Let (t', x') denote the point of intersection. Construct a fixed patent term running from the date of commercialization using $x_{comm} = x'$. This policy, illustrated in Figure 7, strictly increases welfare relative to an t_{patent} term starting at invention. First, take an invention with $t_{comm} < t'$. For such an invention, we have $x^*(t_{comm}) < x_{comm} < x_{invent}(t_{comm})$, that is, optimal post-commercialization protection is smaller than our constructed policy x_{comm}, which itself is smaller than the given policy $x_{invent}(\cdot)$. By the same argument as in the proof of Proposition 6, reducing post-commercialization protection from $x_{invent}(t_{comm})$ to x_{comm} is welfare improving for these inventions (and we would be better off reducing further to $x^*(t_{comm})$): awarding more protection than $x^*(t_{comm})$ increases deadweight loss faster than it increases the gains from non-elicited inventions. Similarly, now take an invention with $t_{comm} > t'$. For such an invention we have $x^*(t_{comm}) > x_{comm} > x_{invent}(t_{comm})$, and our increase of patent protection from $x_{invent}(t_{comm})$ to x_{comm} increases the gains from eliciting more inventions faster than it increases deadweight loss.

Figure 7: **Illustration of the Proof of Proposition 7**



Notes: The post-commercialization patent term function under current policy, denoted $x_{invent}(\cdot)$ in the text, has a negative 45-degree slope until it reaches zero, then it is flat. The post-commercialization patent term under optimal policy, denoted $x^*(\cdot)$ in the text, is strictly increasing, but we do not know its shape or location.

Last, suppose that the curves $x^*(t_{comm})$ and $x_{invent}(t_{comm})$ do not intersect; this occurs iff $x^*(0) > x_{invent}(0)$. In this case, construct our post-commercialization patent term according to $x_{comm} = x^*(0)$. For all inventions, we have $x^*(t_{comm}) \geq x_{comm} > x_{invent}(t_{comm})$, and the argument in the preceding paragraph implies that this policy strictly increases welfare.

A.8 Proof of Proposition 8

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-commercialization-effort subsidy, of $s > 0$ dollars. We will argue that the marginal social welfare benefit of an increase in s for t'_{comm} inventions is strictly larger than that for t_{comm} inventions.

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 6, the marginal invention with lag t'_{comm} is weakly more socially valuable than the marginal invention with lag t_{comm}. Also by the same logic as in the proof of Proposition 6, 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.

A.9 Power calculation derivation

Our conceptual framework is based on the idea that inventions which require long commercialization lags may be under-incentivized. 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.[56] 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.[57] 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:

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

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

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

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

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

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. 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."[58] 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)). 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.

A.10 Elasticity calculation

In this section, we outline our rough estimate of the elasticity of R&D investment with respect to an additional year of commercialization lag.

From our empirical work in Section 4, we have an estimate of how R&D investment responds to the 5-year survival rate, $\frac{\partial(\text{R\&D investment})}{\partial(\text{5-year survival rate})}$. To translate this estimate into our elasticity of interest, we would like to scale $\frac{\partial(\text{R\&D investment})}{\partial(\text{5-year survival rate})}$ by an estimate of how commercialization lag varies with the 5-year survival rate. By combining these estimates, we could then estimate the elasticity of interest:

$$\frac{\frac{\partial(\text{R\&D investment})}{\partial(\text{5-year survival rate})}}{\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}} = \frac{\partial\,(\text{R\&D investment})}{\partial\,(\text{commercialization lag})}$$

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

The conceptual problem with estimating $\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$ is that - by construction - we only observe clinical trial length *conditional* on a drug compound being placed in clinical trials. Because - consistent with our model - we document that fewer drug compounds are placed in clinical trials for patients with longer survival times, we expect selection into clinical trials 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.[59] Given this selection, we cannot obtain an unbiased empirical estimate of $\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$. To overcome this selection problem, we instead calibrate the relationship between commercialization lag and the 5-year survival rate using the power calculation outlined in Appendix A.9.

We can approximate our estimate of $\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$ with an estimate of $\frac{\partial(\text{clinical trial length})}{\partial(\text{5-year survival rate})}$, given that we expect commercialization lag to scale one-for-one with clinical trial length. In the language of the power calculation outlined in Appendix A.9, we can re-write this elasticity as:

$$\frac{\partial\,(\text{commercialization lag})}{\partial\,(\text{5-year survival rate})} = \frac{\partial\,(\text{clinical trial length})}{\partial\,(\text{5-year survival rate})} = \frac{\partial k}{\partial \mu} = \frac{k\mu^{k-1} + Rk(1-R(1-\mu))^{k-1}}{-[\mu^k \ln \mu + (1-R(1-\mu))^k \ln(1-(R(1-\mu))]}$$

where μ is the per-period survival rate of untreated individuals, k is the number of periods of patient follow-up, and R is a constant per-period multiplicative treatment effect.

Intuitively, μ and k come in pairs - not all μ and k are feasible conditional on a given technology (R). Here, we take the two (μ, k) pairs from the examples in the introduction given that by construction these are feasible pairs (given that the trials were completed), and that we know these trials looked at survival outcomes (rather than surrogate endpoints). We assume a technology of $R = 0.8$, which translates to a 20 percent improvement in the five-year survival rate; this choice of R is arbitrary but we explore robustness to alternative values of R below. Given the assumed value of R, our two introduction examples can be written as:

1. Metastatic prostate cancer: 5-year survival rate of 20 percent ($\mu = 0.2$)

    (a) Follow-up time of 12.8 months ($(12.8/12)/5 => k = 0.213$ units in 5-year increments)

    (b) Total trial length of 3 years ($3/5 => k = 0.6$ in 5-year increments)

2. Localized prostate cancer: 5-year survival rate of 80 percent ($\mu = 0.8$)

    (a) Follow-up time of 9.1 years ($9.1/5 => k = 1.82$ units in 5-year increments)

    (b) Total trial length of 18 years ($18/5 => k = 3.6$ units in 5-year increments)

[59] If we 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).

Plugging in these values for μ, k, and R into the above formula for $\frac{\partial k}{\partial \mu}$ gives estimates of 2.234 for metastatic prostate cancer, and 0.766 for localized prostate cancer. Those estimates are in units of 5-year increments, and multiplying them by 5 to translate them into a 1-year unit gives 11.170 and 3.827. In words, a change from 0 to 1 in the 5-year survival rate translates to between a 3.827-11.170 year increase in patient follow-up time.

Our estimate from Section 4 implies that a change from 0 to 1 in the 5-year survival rate translates into an 86.9% reduction in R&D investment. Scaling this estimate by our estimates of $\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$ implies an estimated semi-elasticity of R&D investment with respect to a one-year change in commercialization lag of between 7.779% (based on metastatic prostate cancer; $86.9/11.170 = 7.779$) and 22.707% (based on localized prostate cancer; $86.9/3.827 = 22.707$).

Alternatively, we can do the same calculation using total trial length (3 and 18 years) rather than follow-up times (12.8 months and 9.1 years). Reassuringly, we obtain nearly identical estimates: 7.993% (based on metastatic prostate cancer; $86.9/10.872 = 7.993$) and 23.416% (based on localized prostate cancer; $86.9/3.711 = 23.416$).

We can investigate sensitivity of our estimates to different assumed values of R, the quality of the technology. A 'reasonable' range of R might be between 0.15-0.95, in which case our estimated elasticities fall between 6-54%.[60]

[60] Our metastatic prostate cancer example from the introduction - where the treatment resulted in a gain of 3.9 months on average - corresponds to $R = 0.961$, which implies elasticity estimates between 6-20%. On the other extreme Gleevec, often referenced as a "miracle" drug, is estimated to have increased the five-year survival rate from 30% to 89% - implying $R = 0.157$, and elasticity estimates between 17-54%.

B Appendix: Data (not for publication)

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.[61]

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.[62] 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.[63] 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.[64] First, some cancers are not staged by SEER: for example, brain cancers are not staged.[65] 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.[66] Third, some individual observations that should be staged

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

[62] 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.

[63] 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.

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

[65] 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).

[66] 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.[67]

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).

[67]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.[68] 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.[69] 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

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

[69] 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.

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.[70] 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 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.
- **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,

[70] 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 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.[71] 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 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

[71] 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)."

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: Industry interviews (not for publication)

As discussed in the introduction, while the importance of patents has been debated in many industries, given our empirical focus on the pharmaceutical industry it is worth noting that a variety of evidence suggests that patents play a key role in motivating innovation in the pharmaceutical industry, including industry interviews (Mansfield, Schwartz and Wagner 1981; Mansfield 1986; Levin et al. 1987; Cohen, Nelson and Walsh 2000), the cost structure of new drug development relative to the generic production (DiMasi, Hansen and Grabowski 2003; Adams and Brantner 2006; Wroblewski et al. 2009), and the fact that standard investment models used by pharmaceutical firms pay close attention to effective patent length (Mayer Brown 2009). In this appendix, we document some additional evidence on this point including summarizing some informal interviews that we conducted for this paper.

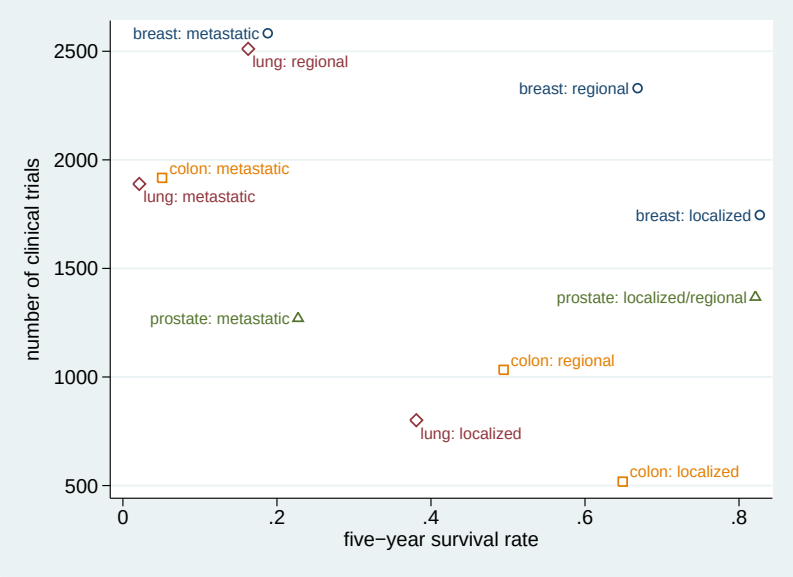
Academic clinicians, clinical researchers, and firms all report a reluctance to invest in drugs that - by nature of requiring lengthy clinical trials - receive short effective patent terms. For example, a medicinal chemistry textbook notes: "...patents normally run for 20 years from the date of application, ...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 due to short expected patent terms, 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)

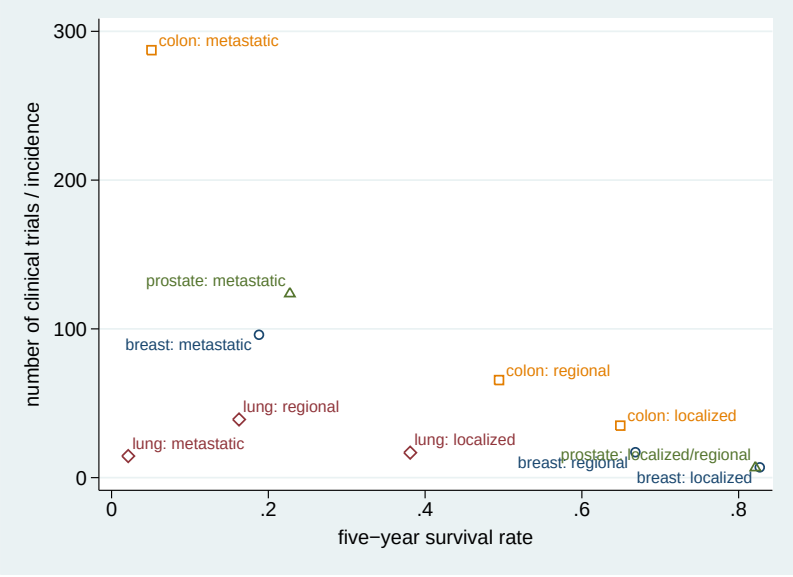
- *The shorter the remaining patent term, the more certainty you need that the drug will work, and the more it needs to have a large market. Also, the ramp is important. You want at least a couple years of peak sales. It happens all the time that we pass on a drug, one we think would probably work, because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

D Appendix: Additional figures and tables (not for publication)

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



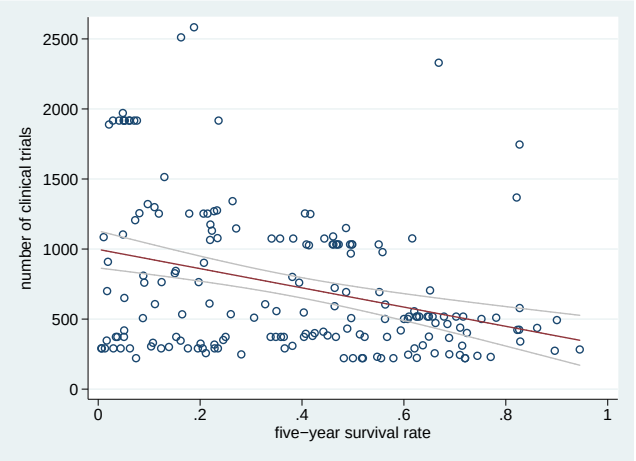
(a) Survival time and R&D investments



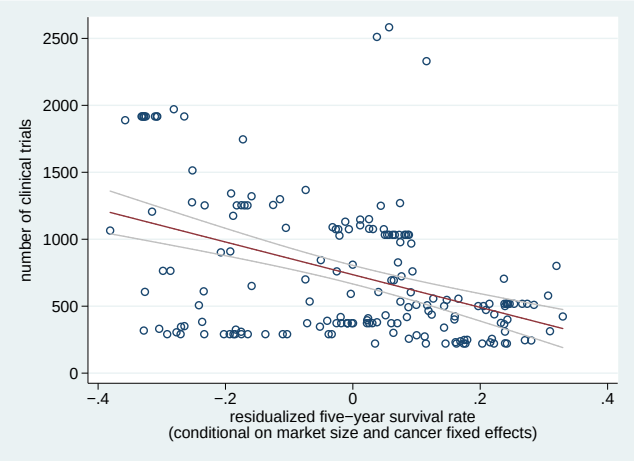
(b) Survival time and market-size adjusted R&D investments

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.

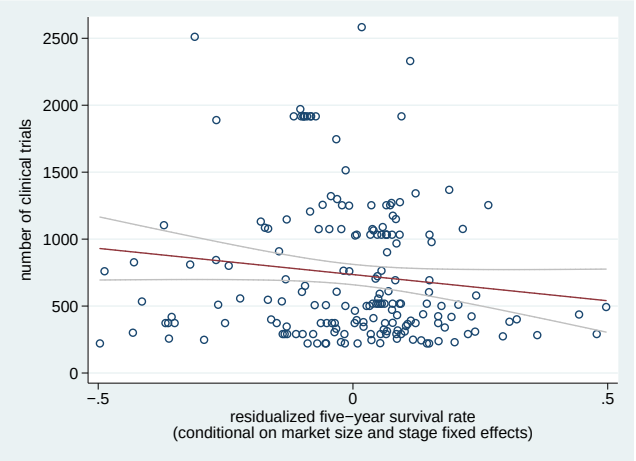
Figure D.2: **Survival time and R&D investments: Residualized cancer-stage data**



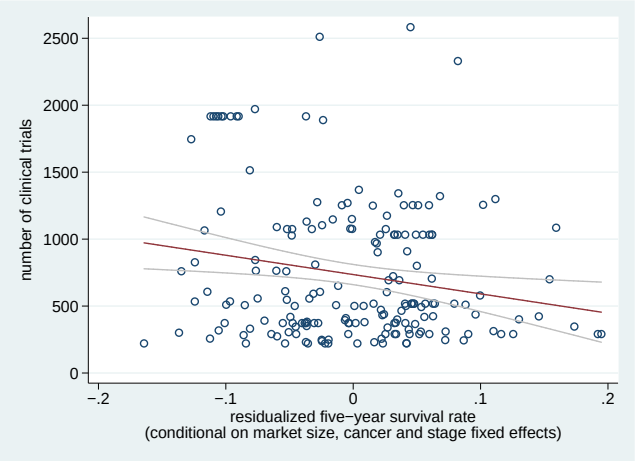
(a) Raw data



(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 D.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	***	-1.151	***	-1.588	***	-0.339		-1.360	***
	(0.236)		(0.188)		(0.132)		(0.305)		(0.315)	
log(market size)	-		0.189	***	0.098	**	0.193	***	0.059	
			(0.040)		(0.045)		(0.036)		(0.037)	
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 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 D.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			-0.868	***						
			(0.319)							
1973 survival (years)					-0.034	***				
					(0.013)					
1973 one-year survival rate							-0.597	**		
							(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 D.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	-0.868	***	-1.113	***	-1.241	**	-1.498	***	-0.963	***	-1.151	***
	(0.319)		(0.286)		(0.529)		(0.434)		(0.236)		(0.188)	
log(market size)	-		0.243	***	-		0.275	***	-		0.189	***
			(0.055)				(0.072)				(0.040)	
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 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 D.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	***	-2.341	***
	(0.912)		(0.798)		(0.823)	
log(market size)	-		0.393	***	-	
			(0.101)			
log(life-years lost)	-		-		0.438	***
					(0.133)	

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 number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* denotes 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 D.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	***	-2.815	***	-2.405	***
	(0.902)		(0.785)		(0.814)	
(0/1: *hematologic*)	1.250	***	1.178	***	1.032	**
	(0.458)		(0.393)		(0.432)	
log(market size)	-		0.398	***	-	
			(0.104)			
log(life-years lost)	-		-		0.413	***
					(0.141)	
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	***	6.543	***	6.075	***
	(1.668)		(1.622)		(1.622)	
five-year survival rate	-3.743	***	-3.925	***	-3.539	***
	(1.273)		(1.054)		(1.111)	
(0/1: *hematologic*)	-1.032		-1.190	*	-1.164	*
	(0.725)		(0.639)		(0.605)	
log(market size)	-		0.376	***	-	
			(0.109)			
log(life-years lost)	-		-		0.386	**
					(0.153)	

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 number of patients diagnosed with that cancer-stage between 1973-2009. *Life-years lost* denotes 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.

E Appendix: Development of chemoprevention drugs (not for publication)

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.

E.1 BCG

The 1990 FDA approval for bladder carcinoma in situ was supported by clinical trials funded by the National Cancer Institute (NCI).[72] 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)."*

E.2 Diclofenac

Diclofenac is a topical treatment for actinic keratoses, which is clinically recommended for treatment to prevent disease progression to squamous cell carcinomas.[73] 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.

E.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*

[72] 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.

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

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

E.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.

E.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."*

E.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."*