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

Eric Budish, Ben Roin, and Heidi Williams

University of Chicago

MIT Sloan

MIT and NBER

Forthcoming, *American Economic Review*

- Over last five years, eight new drugs approved to treat lung cancer
- All eight were approved based on evidence of incremental survival improvements in patients with most advanced form of the disease
  - Well-known example: Genentech's Avastin (10.3 vs. 12.3 months)
- In contrast, no drug has ever been approved to prevent lung cancer, and only six drugs have ever been approved to prevent *any* cancer

- Over last five years, eight new drugs approved to treat lung cancer
- All eight were approved based on evidence of incremental survival improvements in patients with most advanced form of the disease
    - Well-known example: Genentech's Avastin (10.3 vs. 12.3 months)
- In contrast, no drug has ever been approved to prevent lung cancer, and only six drugs have ever been approved to prevent *any* cancer

- While this pattern could solely reflect market demand or scientific challenges, in this paper we investigate an alternative hypothesis: private firms may (differentially) underinvest in long-term research
    - Late-stage cancer drugs can be brought to market comparatively quickly, relative to early-stage treatments or preventatives
- We document that such underinvestment is quantitatively significant in markets for cancer drugs, and analyze potential policy responses

Why might private firms underinvest in long-term research?

We use a simple model to illustrate two potential sources of this distortion

1. Excess impatience of private firms relative to the social planner
    - Widely discussed, but little empirical evidence

Why might private firms underinvest in long-term research?

We use a simple model to illustrate two potential sources of this distortion

1. Excess impatience of private firms relative to the social planner
   - Widely discussed, but little empirical evidence

2. R&D markets, add'l potential mechanism: Structure of patent system
   - Patents award innovators a fixed (20-year) period of market exclusivity
   - Yet, many firms file patents at discovery ("invention") rather than first sale ("commercialization") $\Rightarrow$ inventions with long commercialization lags receive reduced - in extreme cases, zero - effective patent terms
   - Implies that in some markets, the patent system provides very little incentive for private firms to engage in long-term research

Testing for “missing” R&D

This idea - while intuitive - is difficult to test empirically

- Key prediction: “missing” R&D on long-term projects
- In practice, testing this prediction encounters two challenges:
  1. Measurement: don’t observe commercialization lags for missing projects
  2. Inference: “missing” R&D hard to distinguish from alternative explanations, e.g. lack of market demand or scientific opportunities

Testing for “missing” R&D

This idea - while intuitive - is difficult to test empirically

- Key prediction: “missing” R&D on long-term projects
- In practice, testing this prediction encounters two challenges:
  1. Measurement: don’t observe commercialization lags for missing projects
  2. Inference: “missing” R&D hard to distinguish from alternative explanations, e.g. lack of market demand or scientific opportunities

Two features of cancer markets allow us to make progress:

1. 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 observed and potential R&D activity
2. For each such group of cancer patients we observe a good predictor of how long it would take to commercialize a new drug: survival time

Two examples: Prostate cancer drugs

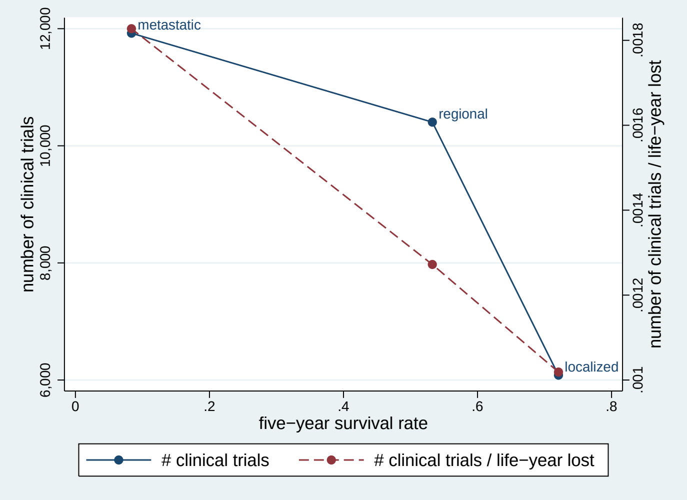
1. de Bono *et al.*: Metastatic patients (5-yr survival $\approx$ 20%)
    - Median follow-up time for measuring patient survival: 12.8 months
    - Trial length: 3 years
2. Jones *et al.*: Localized patients (5-yr survival $\approx$ 80%)
    - Median follow-up time for measuring patient survival: 9.1 years
    - Trial length: 18 years

Consistent with commercialization lags distorting private R&D incentives:

- Metastatic clinical trial funded by Cougar Biotechnology
- Localized clinical trial funded by US National Cancer Institute

We construct data on all such clinical trials over the last three decades, which we match to data on patient survival over the same period

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



Notes: See Figure 1(a) in paper.

Empirical evidence

To interpret this correlation between survival time and R&D investments, we document evidence from two complementary empirical tests:

1. Investigate "surrogate" (non-mortality) endpoints:
   Causal evidence that shorter commercialization lags increase R&D
2. Contrast public/private R&D investments:
   Direct evidence of a distortion in private R&D investments

Qualitative evidence: FDA-approved chemoprevention drugs

Policy responses

Analyze three policies: Surrogate endpoints, R&D subsidies, patent reform

- Surrogate endpoints have benefits beyond eliminating distortion
- Patent reform only affects component of distortion driven by patents

Taking advantage of our surrogate endpoint variation, we estimate counterfactual improvements in cancer survival rates that would have been observed if commercialization lags had been reduced

- Murphy and Topel (2006): Cure for cancer worth $\sim$ 50 trillion
- Estimated life lost among US cancer patients diagnosed in 2003:
  - Total estimated life-years lost: 890,000
  - Valued at \$100,000 per life-year (Cutler 2004): \$89 billion

Simple model

Conceptualize R&D as consisting of two stages:

1. "Invention": developing a basic idea to point where it is patentable
2. "Commercialization": bringing an invented product to market

Purposefully simple model show why private-section R&D may be distorted away from inventions with long commercialization lags

- Both private and social incentives decline with commercialization lags
- But either excessive discounting or a fixed patent term generates a distortion: private incentives decline faster than social incentives

Preliminaries

Monopoly profits π per period over
Expected Monopoly Life (EML):
$p \sum_{t_{comm}}^{t_{patent}-1} (\eta\delta)^t$

t_{invent}:
year of invention
(normalized to zero)

t_{comm}:
years required for
commercialization,
which costs c with
$\Pr(\text{success}) = p$

t_{patent}:
fixed patent term
(20 years)

T: expected end
of invention's
useful life, due to
obsolescence risk
$(1 - \gamma)$

Social value v per period over
Expected Total Life (ETL):
$p \sum_{t_{comm}}^{\infty} \delta^t$

Notes: The *EML* expression holds if $t_{\text{patent}} > t_{\text{comm}}$; if $t_{\text{patent}} \leq t_{\text{comm}}$ then $EML = 0$. We here focus on the case of perfect imitability if the product is successfully commercialized, non-obsolete, and not protected by patent ($\iota = 1$); in the paper we analyze imperfect imitability. We here focus on the case where firms file patents at the time of invention ($q = 0$); in the paper we analyze the choice of when to patent. The project's neoclassical risk-adjusted discount rate is r; society applies an obsolescence-risk weighted discount factor $\delta = \frac{\gamma}{1+r}$, and private firms apply the discount factor $\eta\delta$ where $\eta \leq 1$.

Private and social incentives to invest

- Firm expects to enjoy monopoly profits of π for EML years, so optimal to commercialize iff $EML \cdot \pi$ exceeds the cost of commercialization c

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

- Social planner commercializes iff expected social welfare, if the good is priced at marginal cost, exceeds the cost of commercialization c

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

- Anytime private firm would commercialize, so would social planner
  - By definition: $ETL \geq EML$, $v \geq \pi$ (ignores business stealing)
  - In words, private and social investment decisions differ when the social return is positive but the private return is negative

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

Distortions in the level and composition of R&D

Part 1 is a standard result. Part 2 indicates that distortions in composition can arise from differences across inventions in either $\frac{\pi}{v}$ or $\frac{EML}{ETL}$.

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; (b) firms are not excessively impatient; and (c) monopolists capture full social surplus.*
2. *(distortion in composition) For two inventions, A and B, it is possible that the expected social return 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:*
    1. $\frac{\pi_B}{v_B} > \frac{\pi_A}{v_A}$
    2. $\frac{EML_B}{ETL_B} > \frac{EML_A}{ETL_A}$

$\frac{EML}{ETL}$ ratio declines with commercialization lag t_{comm}

With either excessive impatience or finite patents that start at invention, private incentives decline more rapidly in commercialization lag than do 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}*.*
   1. *If* $t_{comm} < t_{patent}$ *the decline is strict:* $\frac{\partial \frac{EML}{ETL}}{\partial t_{comm}} < 0$
   2. *If* $t_{comm} \geq t_{patent}$ *then* $EML = 0$*. Hence* $\frac{EML}{ETL} = 0$.

Both mechanisms decline in commercialization lags

We can decompose $\frac{EML}{ETL}$ as follows (where EPL is EML using δ):

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

Both terms strictly decline with commercialization lag. ▸ Proposition 3 ▸ Examples

Will discuss three policy levers that could address this distortion.

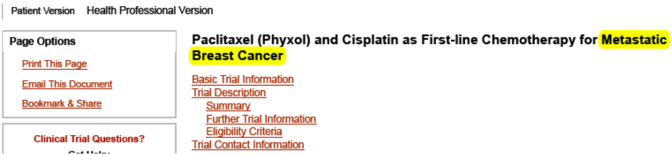
Why cancer R&D?

1. Substantive interest given cancer's morbidity, mortality burden
2. Unlike for many diseases, high-quality clinical data exists for cancer which accurately tracks patient survival times ▸ SEER data
3. Existence of a standardized classification system for cancer - organs (*e.g.* prostate) and stages (*e.g.* metastatic) - facilitates a relatively clean match between clinical data and R&D investments
   - *E.g.* Genentech's Bevacizumab FDA approved in 2004 for treatment of patients with metastatic carcinoma of the colon and rectum
4. Existence of patient-group specific R&D data ▸ NCI data ▸ FDA data

▸ Summary statistics

NCI data: Example

Web interface:



XML files:

```
<Diagnosis>
  <SpecificDiagnosis ref="CDR0000039108">stage IV breast cancer</SpecificDiagnosis>
  <DiagnosisParent ref="CDR0000038832">breast cancer</DiagnosisParent>
  <DiagnosisParent ref="CDR0000043666">body system/site cancer</DiagnosisParent>
  <DiagnosisParent ref="CDR0000041060">malignant neoplasm</DiagnosisParent>
  <DiagnosisParent ref="CDR0000040460">adult solid tumor</DiagnosisParent>
  <DiagnosisParent ref="CDR0000040461">solid tumor</DiagnosisParent>
</Diagnosis>
```

Extracted flat file:

Tagged_C~r	S0	S1	S2	S3	S4
ind_breast	0	0	0	0	1

1. Theory

2. Data

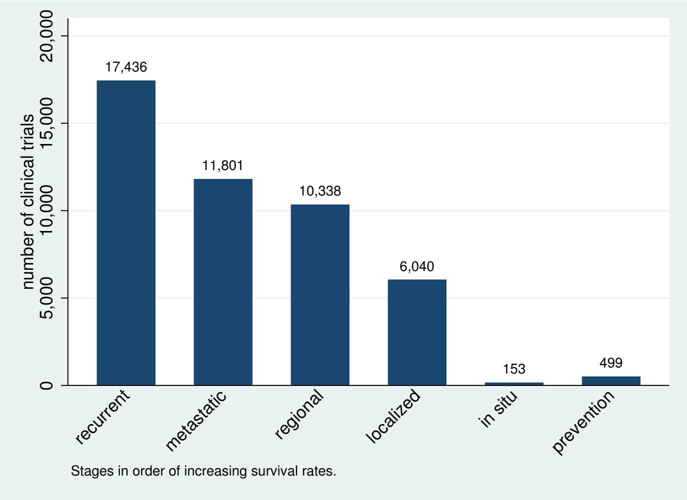
3. Empirical evidence
   - Descriptive analysis
   - Interpreting the correlation between survival time and R&D

4. Estimating the value of life lost due to commercialization lags

5. Policy analysis

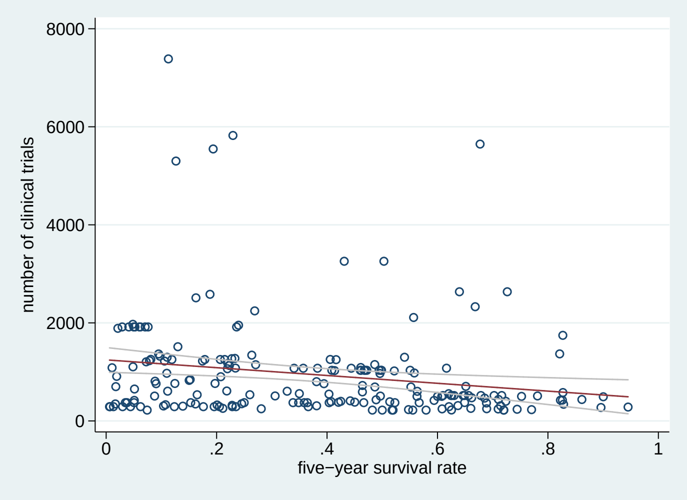
6. Discussion and conclusion

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



Notes: See Figure 1(b) in paper.

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



Notes: See Figure 2 in paper.

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

$$(\text{number of clinical trials})_{cs} = \alpha + \beta(\text{survival})_{cs} + \lambda'(\text{covariates})_{cs} + \varepsilon_{cs}$$

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: See Table 2 in paper. Cancer-stage observations. Estimates from quasi-maximum likelihood Poisson models. $N = 201$ in Columns (1) and (2), and $n = 192$ in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Robustness: Negative survival time-R&D correlation

1. Case study of "big four" cancers: breast, colon, lung, and prostate
   ▸ Scatterplot ▸ Market size-adjusted
2. Residualized scatter plots: Market size and life-years lost
   ▸ Residualized plot: Market size ▸ Residualized plot: Life-years lost
3. Cancer and stage fixed effects
   ▸ Table ▸ Residualized: Market size, cancer FE
   ▸ Residualized: Market size, stage FE ▸ Residualized: Market size, cancer FE, stage FE
4. Alternative survival time measures ▸ Table
5. Robustness across samples ▸ Table
6. FDA drug approvals ▸ Table

Empirical evidence

To test whether commercialization lags distort private R&D investments, we provide evidence from two complementary empirical tests:

1. Investigate "surrogate" (non-mortality) endpoints:
   Causal evidence that shorter commercialization lags increase R&D
2. Contrast public/private R&D investments:
   Direct evidence of a distortion in private R&D investments

Qualitative evidence: FDA-approved chemoprevention drugs

Investigating surrogate endpoints: Hematologic cancers

- Traditional FDA focus on survival / mortality-related endpoints
- Surrogate endpoints very controversial: Except for hematologic cancers (leukemias & lymphomas), used on a limited *ad hoc* basis
  - Example: "Complete response" for leukemia
  - Our data: 92% of drugs approved 1990-2002 for hematologic cancers relied on surrogate endpoints, vs. 53% for other cancers ($n = 39$)

- Surrogate endpoints shorten commercialization lag
- Model generates three testable predictions:
  - Prediction #1: Higher levels of R&D investment
  - Prediction #2: Less negative survival rate-R&D slope
  - Validation: Expect <u>no</u> change in commercialization at $t_{comm} = 0$

Surrogate endpoints: Level of R&D

$$(\text{number of clinical trials})_{cs} = \alpha + \beta(\text{survival})_{cs} + \gamma(0/1 : \text{hematologic})_c + \lambda'(\text{covariates})_{cs} + \varepsilon_{cs}$$

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)	

Notes: See Table 3 in paper. Cancer-stage observations. Estimates from quasi-maximum likelihood Poisson models. $N = 201$ in Columns (1) and (2), and $n = 192$ in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Drug approvals

Surrogate endpoints: Composition of R&D

$$(\text{number of clinical trials})_{cs} = \alpha + \beta(\text{survival})_{cs} \cdot (0/1 : \text{hematologic})_c + \delta(\text{survival})_{cs} + \gamma(0/1 : \text{hematologic})_c + \lambda'(\text{covariates})_{cs} + \varepsilon_{cs}$$

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: See Table 3 in paper. Cancer-stage observations. Estimates from quasi-maximum likelihood Poisson models. $N = 201$ in Columns (1) and (2), and $n = 192$ in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Drug approvals

Surrogate endpoints and R&D investments

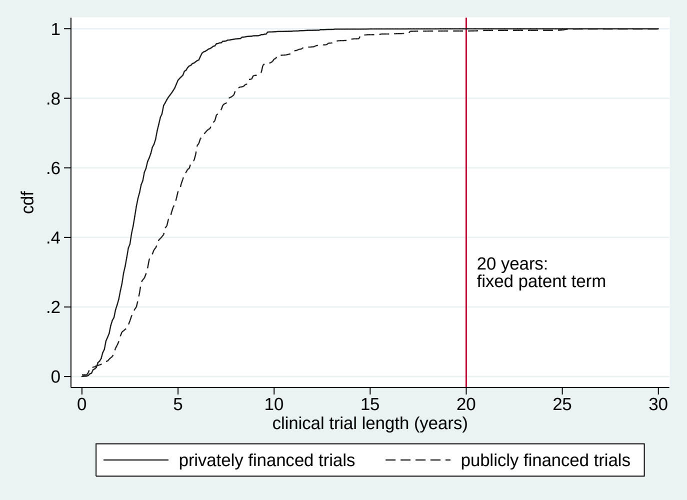
This suggests that there is a causal relationship: if commercialization lags were shortened, there are scientific opportunities available that would be pursued.

Notes: See Figure 4 in paper.

Interpretation

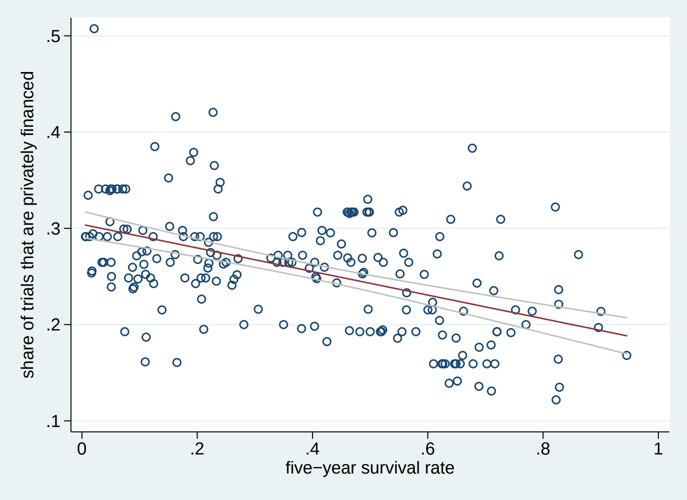
- Estimates suggest our cross-sectional fact is unlikely to be explained by factors such as the pattern of available scientific opportunities
- However, this test leaves open the possibility that the social planner and private firms symmetrically respond to commercialization lags, and thus does not provide direct evidence of a distortion

CDF of clinical trial lengths



Notes: See Figure 5(a) in paper. 95 percent of trials longer than 20 years are publicly financed; six exceptions appear to be typos.

Share of clinical trials that are privately financed



Notes: See Figure 5(b) in paper.

Share of clinical trials that are privately financed

(share of clinical trials that are privately financed)$_{cs}$ =
$\alpha + \beta$(survival)$_{cs}$ + λ'(covariates)$_{cs}$ + ε_{cs}

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)	

Notes: See Table 4 in paper. Cancer-stage observations. Estimates from ordinary-least-squares models. $N = 201$ in Columns (1) and (2), and $n = 192$ in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Contrasting public/private financing of clinical trials

$$(\text{number of clinical trials})_{cst} =$$
$$\alpha + \beta(\text{survival})_{cs} \cdot (\text{sponsor})_t + \delta(\text{survival})_{cs} + \gamma(\text{sponsor})_t + \lambda'(\text{covariates})_{cs} \cdot (\text{sponsor})_t + \varepsilon_{cst}$$

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)	

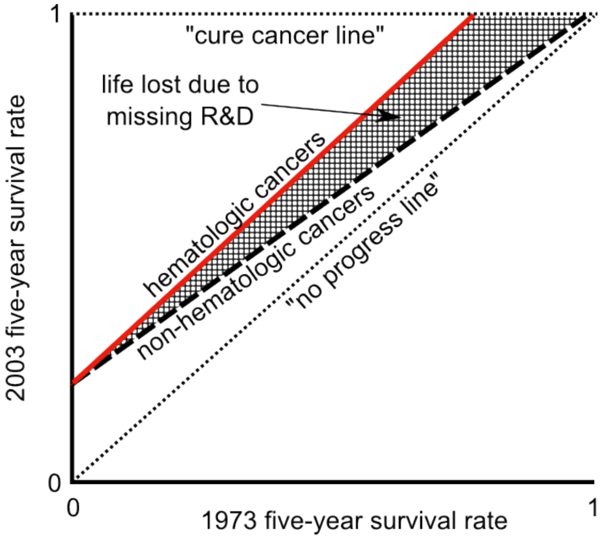
Notes: See Table 4 in paper. Cancer-stage-(0/1: *private*) observations, where (0/1: *private*) = 1 for privately sponsored observations and = 0 for publicly sponsored observations. Estimates from quasi-maximum likelihood Poisson models. $N = 201$ in Columns (1) and (2), and $n = 192$ in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Historical case studies of chemoprevention drugs

- Meyskens *et al.* (2011): six FDA approved chemoprevention drugs
- All six approvals either relied on the use of surrogate endpoints, or were approved on the basis of publicly financed clinical trials
  - Tamoxifen: prevention trials publicly financed
  - Cervical cancer vaccine: HPV incidence as endpoint

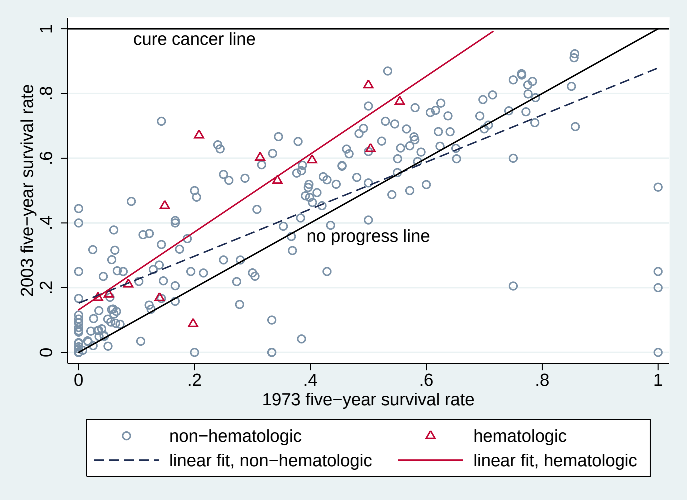
Taken together, this body of evidence - surrogate endpoints, public/private comparison, and case studies of chemoprevention drugs - provides support for the idea that commercialization lags distort private R&D investments.

Counterfactual: Survival gains, 1973-2003



Notes: See Figure 6 in paper.

Counterfactual: Survival gains, 1973-2003



Notes: See Figure 6 in paper.

Rough back-of-the-envelope: Value of lost life

Value of life lost among US cancer patients diagnosed in 2003:

1. Using the cancer registry data, we translate the gap between the hematologic and non-hematologic survival curves into an estimate of life-years lost per cancer patient: 1.07 life-years per patient
2. For each cancer-stage, multiply by the number of US patients$_{cs}$ diagnosed in 2003: 890,000 life-years lost for that cohort
3. Multiplying by a standard value of a statistical life-year (Cutler 2004: \$100,000) monetizes this lost life at a value of \$89 billion

$\Rightarrow$ Net present value over future cohorts of $\frac{\$89 \text{ billion}}{0.05-0.01} \sim \$2.2$ trillion

Mechanisms

Two potential mechanisms for our empirical results, but our results do not speak to which is quantitatively more important. Past literature also provides little insight into expected magnitudes of either mechanism:

1. Corporate finance literature has struggled to devise tests for the presence of short-termism bias
   - 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)
2. Innovation literature has provided remarkably little evidence that stronger patent protection induces more R&D investments
   - E.g. Lerner (2002) and Sakakibara and Branstetter (2001)

Policy analysis

Analyze innovation, social welfare consequences of three policy levers:

1. Policy design: Surrogate endpoints ▸ Proposition 4
   - Benefits beyond eliminating the distortion, because the social planner also values completing projects more quickly
2. Patent reforms ▸ Proposition 5 ▸ Proposition 6 ▸ Proposition 7 ▸ Interviews
   - Starting patent term at commercialization eliminates distortion
   - Currently provide patent protection that <u>decreases</u> in commercialization lag; our analysis suggests that if anything this should be <u>increasing</u>
   - Addresses patent distortion, but not short-termism distortion
3. Policy design: Targeted R&D subsidies ▸ Proposition 8
   - Direct public funds to R&D the private sector is unlikely to undertake

Conclusions

- Simple conceptual point: Commercialization lags may distort R&D away from inventions that take a long time to bring to market
- In the context of cancer R&D, this implies there may be too little R&D on cancer prevention and treatment of early-stage cancers
  - Empirical evidence is consistent with this distortion
- Analyze potential policy responses:
  surrogate endpoints, R&D subsidies, patent design
  - Empirical evidence suggests surrogate endpoints increased R&D investments and induced substantial improvements in survival outcomes

Closing example: Surrogate endpoints and heart disease

- Heart disease is the leading cause of death in the US, but the age-adjusted rate of death has dropped by 50% since 1968
- Decline largely attributed to beta-blockers, ACE-inhibitors, statins
- These drugs were approved based on blood pressure, LDL cholesterol
  - Surrogates first identified by decades-long Framingham Heart Study
  - Some have argued that w/o surrogate endpoints, these drugs may not have reached the market (Lathia *et al.* (2009); Meyskens *et al.* (2011))

Both our empirical evidence for cancer 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

Both mechanisms decline in commercialization lags

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* $t_{comm} < t_{patent}$.

Return

Two hypothetical examples

1. A vaccine administered to men at age 20 that prevents prostate cancer (which tends to affect men in their 50s or later)
    - Likely high social value v
    - Likely low (or zero) $\frac{EML}{ETL}$ ratio because of long required clinical trials
2. A drug administered to late-stage prostate cancer patients that extends life from, say, 6 months to 8 months
    - Likely lower social value v
    - Likely high $\frac{EML}{ETL}$ ratio because of short required clinical trials

In these examples, our distortion of interest - generated by the difference in $\frac{EML}{ETL}$ ratios - would be reinforced by differences in $\frac{\pi}{v}$.

Return

US National Cancer Institute SEER cancer registry

- Standard patient-level clinical dataset, available 1973-2009
- Considered authoritative source on cancer incidence, survival in US
- Key variables:
  - Cancer and stage of patient: used to construct incidence counts
    - SEER cancer sites (80 cancers)
    - Localized, regional, metastatic stages
  - Survival time:
    - Administrative link to NCHS mortality data as of 31 December 2009
    - Focus on 5-year survival for 1973-2004 (uncensored) cohorts
  - Gender / age and year of diagnosis:
    - Link to NCHS period year-age-gender specific life expectancy data
    - "Life lost": life expectancy without cancer, less observed survival

Return

US National Cancer Institute cancer clinical trials registry

- Key advantage: Large sample that directly codes relevant patients
  - Claims to be the most comprehensive cancer clinical trials registry
    - Established in 1971
    - Includes $> 30,000$ clinical trials
  - Explicit listing of which patient groups are eligible for each clinical trial
- Key disadvantage: Not intended as a research database
  - Designed for use by physicians and patients
  - Some missing data: Sponsorship observed for $\sim 50\%$ of sample
  - Data extraction not straightforward

Return

US Food and Drug Administration (FDA) drug approvals

- Approved oncology drugs from 1990-2002: 71 drugs
  - List published in Johnson-Williams-Pazdur (2003)
  - Paper specifies clinical trial endpoints used as basis for FDA approvals
- For 39 of 71 approvals, hand-collected cancer and stage for which drug was approved from the Drugs@FDA administrative database
  - FDA approval letters missing for other 32 drug approvals

Return

Summary statistics: Cancer-stage level data

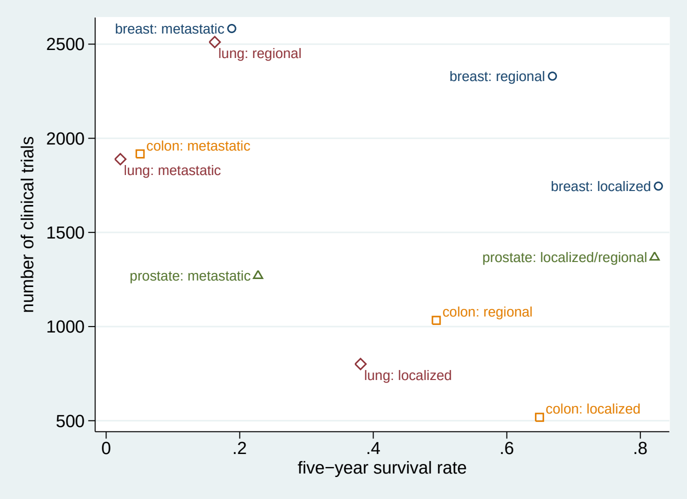
	mean	median	standard deviation	minimum	maximum
number of clinical trials, 1973-2011	945	556	1,015	221	7,385
number of drug approvals, 1990-2002	0.507	0	1.221	0	7
five-year survival rate, cases diagnosed 1973-2004	0.377	0.383	0.249	0.006	0.945
number of diagnoses (1000s), 1973-2009	12.423	3.159	29.429	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: See Table 1 in paper.

- Cancer-stage level data
- 201 observations:
  - 60 cancers appear in all stages (localized, regional, metastatic)
  - Prostate SEER-coded as two stages (localized/regional, metastatic)
  - 19 cancers are unstaged ⇒ appear as one observation

▸ Return

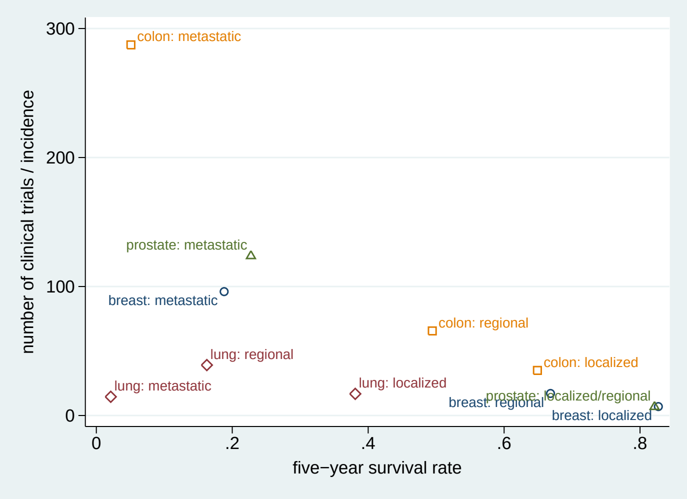
Survival time and R&D investments: Breast, colon, lung, and prostate cancer



Notes: See Appendix Figure D.1(a) in paper.

Return

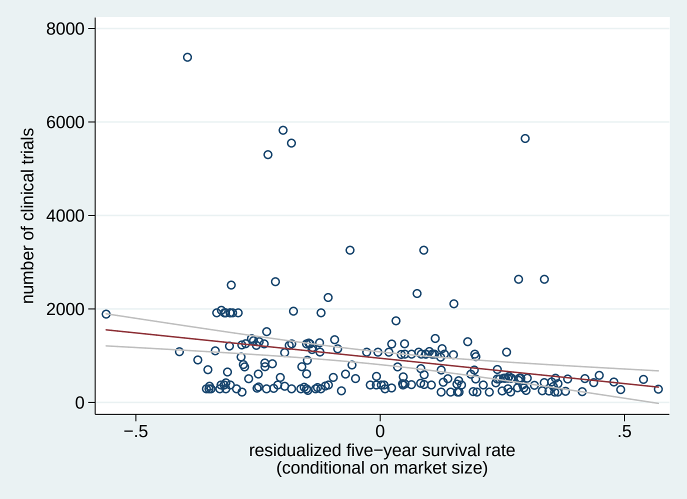
Survival time and market-size adjusted R&D investments: Breast, colon, lung, and prostate cancer



Notes: See Appendix Figure D.1(b) in paper.

Return

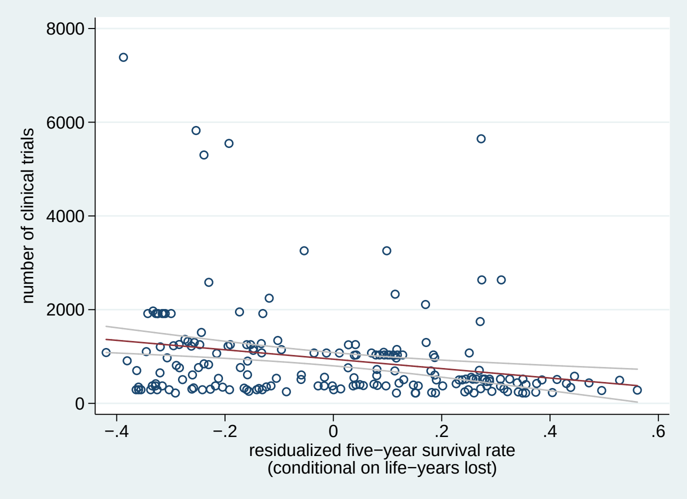
Survival time and R&D investments: Market size residualized



Notes: See Figure 3(a) in paper.

Return

Survival time and R&D investments: Life-years lost residualized



Notes: See Figure 3(b) in paper.

Return

Survival time and R&D investments: Robustness to cancer and stage fixed effects

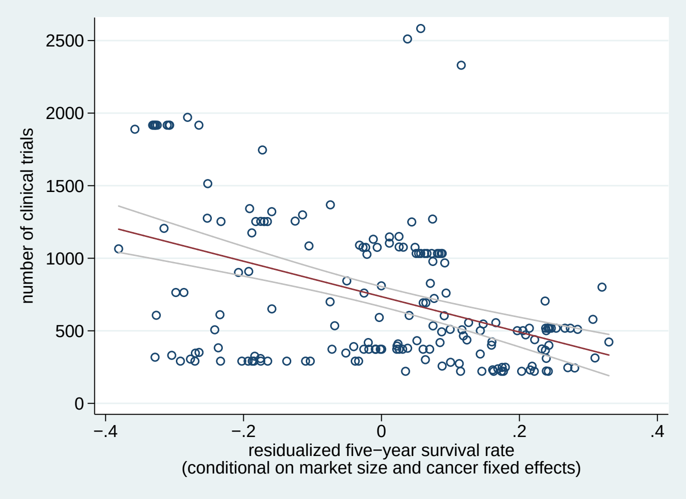
$$(\text{number of clinical trials})_{cs} = \alpha + \beta(\text{survival})_{cs} + \lambda'(\text{covariates})_{cs} + \varepsilon_{cs}$$

	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: See Table D.1 in paper. Cancer-stage observations. Estimates from quasi-maximum likelihood Poisson models. Standard errors clustered at the cancer level. $N = 182$. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Return

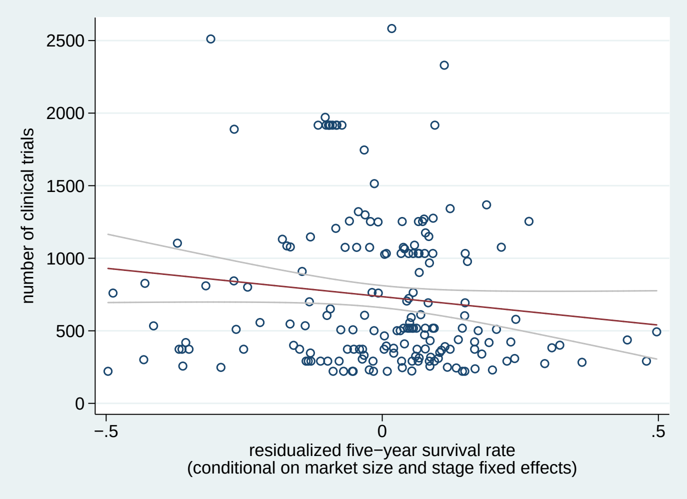
Survival time and R&D investments: Residualized cancer-stage level data



Notes: See Appendix Figure D.2(b) in paper.

Return

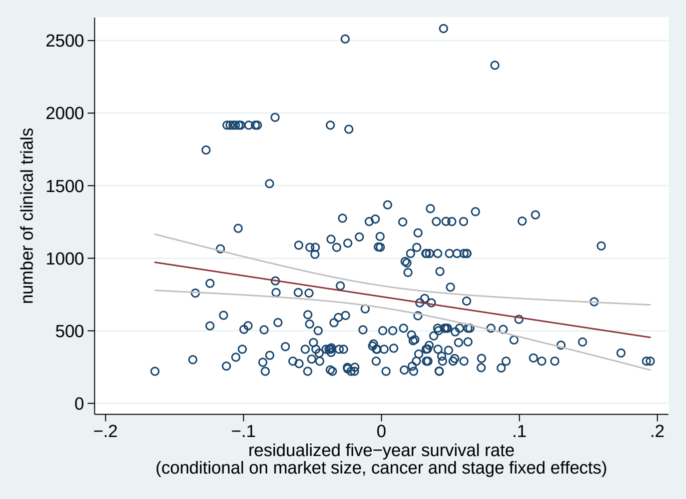
Survival time and R&D investments: Residualized cancer-stage level data



Notes: See Appendix Figure D.2(c) in paper.

▸ Return

Survival time and R&D investments: Residualized cancer-stage level data



Notes: See Appendix Figure D.2(d) in paper.

Return

Survival time and R&D investments: Robustness to alternative survival measures

$$(\text{number of clinical trials})_{cs} = \alpha + \beta(\text{survival})_{cs} + \lambda'(\text{covariates})_{cs} + \varepsilon_{cs}$$

	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: See Table D.2 in paper. Cancer-stage observations. Estimates from quasi-maximum likelihood Poisson models. $N = 201$ in Columns (1) and (2), and 187 in Column (3), because 14 cancer-stages had no patients diagnosed in 1973. Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Return

Survival time and R&D investments: Robustness across samples

$$(\text{number of clinical trials})_{cs} = \alpha + \beta(\text{survival})_{cs} + \lambda'(\text{covariates})_{cs} + \varepsilon_{cs}$$

	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: See Table D.3 in paper. Cancer-stage observations. Estimates from quasi-maximum likelihood Poisson models. $N = 201$ in Columns (1) and (2), $n = 140$ in Columns (3) and (4), and $n = 182$ in Columns (5) and (6). Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Return

Survival time and R&D investments: FDA drug approvals

$$(\text{number of FDA approvals})_{cs} = \alpha + \beta(\text{survival})_{cs} + \lambda'(\text{covariates})_{cs} + \varepsilon_{cs}$$

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: See Table D.4 in paper. Cancer-stage observations. Estimates from quasi-maximum likelihood Poisson models. $N = 201$ in Columns (1) and (2), and $n = 192$ in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Return

Surrogate endpoints: Number of FDA drug approvals

$$(\text{number of FDA approvals})_{cs} = \alpha + \beta(\text{survival})_{cs} + \gamma(0/1 : \text{hematologic})_c + \lambda'(\text{covariates})_{cs} + \varepsilon_{cs}$$

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)	

Notes: See Appendix Table D.5 in paper. Cancer-stage observations. Estimates from quasi-maximum likelihood Poisson models. $N = 201$ in Columns (1) and (2), and $n = 192$ in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Return

Surrogate endpoints: Composition of FDA drug approvals

$$(\text{number of FDA approvals})_{cs} = \alpha + \beta(\text{survival})_{cs} \cdot (0/1 : \text{hematologic})_c + \\ + \delta(\text{survival})_{cs} + \gamma(0/1 : \text{hematologic})_c + \lambda'(\text{covariates})_{cs} + \varepsilon_{cs}$$

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: See Appendix Table D.5 in paper. Cancer-stage observations. Estimates from quasi-maximum likelihood Poisson models. $N = 201$ in Columns (1) and (2), and $n = 192$ in Column (3), because 9 cancer-stages had no patients diagnosed between 1973-1983. Standard errors clustered at the cancer level. *: $p < 0.10$; **: $p < 0.05$; ***: $p < 0.01$.

Return

Policy design: Surrogate endpoints

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:* $\frac{\partial}{\partial x}\mathbb{E}\left(\frac{EML}{ETL}|\hat{t}_{comm} = x\right) = 0$.

Note: Expect <u>no</u> change in commercialization for inventions at $t_{comm} = 0$ Return

Policy design: Patent reform

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}.

Return

Policy design: Patent reform

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 (1). 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}.

Return

Policy design: Patent reform

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.

Return

Anecdotal evidence from industry interviews

- *...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.*
- *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.*

Return

Policy design: Targeted R&D subsidies

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}*.*

Return