What's the Value of Patent Citations? Evidence from Pharmaceuticals

David S. Abrams and Bhaven N. Sampat¹ Preliminary Draft: June 9, 2017

Abstract: The number of times a patent is cited by subsequent patents is commonly used in economic research as a proxy for both the private and social value of inventions. Patent citation measures are also increasingly used by policymakers assessing the impact of innovation policies. However, there is surprisingly little direct evidence that more highly cited patents are more valuable ones, and existing validation studies have not considered differences between private and social value. In this paper we examine the relationship between the number of citations to a drug patent and the private and social value of that drug. We find evidence that citations to drug patents are weakly related to both private and social value. We also provide new data on the share of social welfare gains appropriated by firms, an important issue in patent law and innovation policy on which there is also little evidence. Initial results indicate that a substantial share of drugs yield no added health benefits to patients. At the same time, on average, consumers gain a very large share of the aggregate surplus generated by new drugs.

¹Abrams: University of Pennsylvania (<u>dabrams@law.upenn.edu</u>) Sampat: Columbia University and NBER. The authors would like to thank Emma Boswell Dean for outstanding research assistance.

I. Introduction

The number of times a patent is cited by subsequent patents is commonly used in economic research as a proxy for both the private and social value of inventions (Gittleman 2012, Sampat and Ziedonis 2004). Patent citation measures are also increasingly used by policymakers assessing the impact of innovation policies. However, there is surprisingly little direct evidence that more highly cited patents are more valuable ones, and existing validation studies have not considered differences between private and social value. In this paper we make use of newly available data on drug efficacy, usage, and revenue to add to our understanding of patent value in the pharmaceutical industry.

The pharmaceutical industry provides a good context to examine relationships between patent citations and value. It is economically important: In the US alone, pharmaceutical sales were nearly \$400 billion last year, and drug costs are among the fastest growing components of U.S. healthcare expenditures.

There are also practical advantages to studying pharmaceuticals. Unlike most other industries, data on product-specific revenues are available, and marginal costs of production quite low, making it possible to measure the private value of new drugs using drug sales. Additionally, one may use quality adjusted life years (QALYs) generated by specific drugs, in order to compute the marginal social value. Unlike previous analyses which rely on econometric machinery to estimate social value of individual technologies (e.g. discrete choice models), in drugs we have a more direct measure – health improvements. Unlike many complex manufactured products that may involve hundreds or thousands of patents, drugs tend to depend on one or two key patents. Using information on the key patent on a drug generated by the regulatory process for patent extensions, we are able to directly link the underlying patents with the revenues and health benefits they generate.

The pharmaceutical industry is uniquely reliant on patents. There is a sixty-year empirical legacy in economics that patents are more effective at preventing imitation and appropriating returns from R&D in pharmaceuticals than in any other field (Levin et al 1987, Cohen et al 2000) and almost all new drugs are patented. The importance of patent protection in pharmaceuticals may also mean there is less strategic citation: firms have incentives to search for and cite relevant prior art, since the risks of patent invalidation are higher in this field than others.

Exploring the relationship between different measures of value is also particularly important given policy debates in pharmaceuticals regarding different dimensions of drug value. Some analyses suggest that new drugs deliver significant clinical and economic value on average (Lichtenberg 2001). However there is also considerable heterogeneity in the private and social value of new drugs (Garthwaite and Duggan 2012). In addition to examining the relationships between citations and private value, and between citations and social value, we also provide new data on the relationships between private and social value, and the share of social value captured by originating firms.

We proceed as follows. Section II provides background and literature review. Section III describes our data. Section IV presents basic descriptive statistics and section V shows main results, in the form of simple bivariate scatterplots. In Section VI we discuss the interpretation of the results and Section VI concludes.

II. Background

Citations and Value: History and Theory

The difficulty of measuring the value of innovation is a well-known problem in innovation policy (Griliches 1984, Griliches 1990). Though many analyses use patent data to do so, they provide an imperfect solution for several reasons. Most importantly, in many industries most innovations are not patented. The ways in which patents are used also varies by industry, with some sectors using patents mainly for defensive and other strategic purposes beyond appropriating returns from R&D (Cohen et al 2000). In addition, patents vary tremendously in their importance, making simple patent counts (e.g. at a firm, sector, or country level) very noisy measures of innovation (Griliches 1990).

The use of patent citation data began with efforts by bibliometricians during the 1970s. These studies drew on a long bibliometric tradition in evaluating and tracing scientific impact, and examined whether patent-patent citations similarly measured technological impact. These authors conducted various validation exercises, including whether expert rankings of the "significance" of new technical products correlated with the number of times they were cited (Narin, Carpenter and Woolf 1984; Albert, Avery, Narin, and MacAllister 1991). The theory here is that "if a previously issued U.S. patent is cited by the patent examiners in many issued U.S. patents, then the earlier, highly cited patent has been 'prior art' to many later patents and is likely to contain a significant advance that has

preceded these subsequent inventions" (Albert, Avery, Narin, and MacAllister 1991). The literature typically equated the "technical impact" and "quality" of inventions, drawing on parallels with the scientific citations and the idea that more highly cited papers are of higher scientific quality (Albert, Avery, Narin, and MacAllister 1991). The literature also argued that citations in patents are in ways better than those in papers since examiners functions as quality control agents:

"In patents the examiner's citations - citations from a front page of the U.S. patents which are placed there by the patent examiner - are somewhat more strict and controlled than citations in research papers. These examiner citations in patents play a specific legal role in delineating the prior art which limits the claims of the granted patent. These prior art citations are suggested by the patentee and his attorney, and selected and added to by the examiner. If a patent is cited many times in later patents, this indicates that many later inventions have claims that are adjacent to the claims of the invention granted in the cited patents. The U.S. patent examiner is under considerable pressure to make sure that he cites all the relevant prior art, and the quality of those patent citations is certainly more closely controlled than the overall quality of the citations within the scientific literature, which are placed there by the researcher for a myriad of reasons" (253).

Drawing on this bibliometric work, Trajtenberg (1990) was the first economist to use patent citation. Following the bibliometricians, Trajtenberg argued that the process of arriving at the final list of references "apparently does generate the right incentives to have all relevant patents cited, and only those". He concluded from this that the presumption that citations measure "technological importance" is well grounded. He also suggested that citations may reflect the value of inventions:

"The very existence of those later patents attests to the fact that the cited patents opened up the way to a technologically successful innovation. Moreover it presumably attests to the economic success ... since those subsequent patents are the results of costly innovational efforts undertaken mostly by profit-seeking agents ... If citations keep coming it must by that the innovation originating in the cited patent had indeed been valuable."

Economists use of citation data also imported the idea from bibliometrics that citations from one patent to another reflect "building" and thus knowledge flows (Jaffe and Trajtenberg, Ch 2) and also, as mentioned above, that the presence of an examiner as neutral arbiter makes them even better measures of the extent of building than scientific citations.

The early economic work, including validation exercises we will discuss below, assumed (and in some cases tested the hypothesis) that citations reflect private or social value of inventions. To our knowledge the literature offers limited theoretical justification for this linkage, or how the different concepts (citations as impact or quality of an invention; citations as social value; citation as private value) connect. On social value, it is possible that citations reflect knowledge or technological spillovers of inventions (if we accept the "building" metaphor) but unclear why they would correlate with market spillovers, or consumer surplus from an invention. Connecting this to private value requires an implicit assumption that more technologically impactful inventions generate higher profits. While this is not well developed in the literature, there are several potential channels through which this might be true (Sampat and Ziedonis 2009): (1) citations really measure social value, but some of this is appropriated by inventors of the cited patents; (2) citations represent entry into profitable areas; (3) citations indicate market interest in profitable fields (4) a "publicity effect" through which successful inventions are better known.²

Recent empirical research has also cast doubt on a strong version of the "building" assumption, that citations reflect knowledge flows from the cited to citing inventor. Abrams, et al (2013) introduce a model and empirical evidence showing that valuable patents may in fact receive few citations as the assignees take measures to impede further work in the field. Sampat (2010), Alcacer and Gittelman (2006), Alcacer, Gittelman and Sampat (2009) and others show that patent examiners account for a large share of citations. From this perspective at least, the presence of the examiner does not ameliorate noise, but creates it. A survey (Jaffe, Trajenberg, and Fogarty (2000) suggests that a significant share of citations were unknown to the inventor at the time the patent was filed.

More importantly, the early bibliometric and economic literature appears to have been confused about why patents are cited and what citations represent in a legal sense. For example, not all inventions on which an application builds must be cited, only representative "prior art" that would be considered "material to patentability" (novelty or nonobviousness) to a reasonable examiner (Sampat 2010). While these references (together with references to other prior art, including scientific literature) are used by examiners to assess whether claims are patentable relative to prior art, they do not typically "delimit the scope of claims." (Though they can be used to reject claims during prosecution, and granted patents have a presumption of validity over cited references.) Moreover, applicants decide whether or not to search for and cite prior art for strategic

² Note that none of these channels, with the possible exception of (1) necessarily require one to assume that citations represent "building" of the cited invention on the citing.

reasons (Lampe 2010; Sampat 2010). There is concern both about underreporting of art (Sampat 2010) and "flooding the patent office" with too many references to bury important ones (Cotropia, Lemley, Sampat 2013). Moreover, the extent to which patent examiners cite previous patents varies by examiner (Lemley and Sampat 2012) and they may not even pay much attention to what applicants cite (Cotropia, Lemley, Sampat 2013). Examiners also have "pet" patents they consistently tend to cite. (Kortum et al 2003). In short, the idea that citations represent the impact of one invention on another is not quite right, nor is the assumption that applicant and examiner incentives are perfectly aligned to find and cite all relevant patents.

Citations and Value: Empirics

The idea that citations reflect different dimensions of value was typically based less on theory (or a model of legal practice) than posed as an empirical hypothesis to be tested. In the first use of citations by economists, Trajtenberg (1990) uses discrete choice demand models to estimate social value of CT scanner technology, and correlates temporal changes in social value with patent counts and citation weighted patent counts for CT scanners. He finds that citation-weighted patent counts perform better at predicting social value than simple patent counts, with correlation coefficients between .4 and .8 depending on assumptions about lag structure and the functional form of the relationship.

He concludes by emphasizing limitations of this approach and difficulties in validating: "All of my conclusions have been expressed in a qualified manner since they are based upon the findings from a single case study. It is important to emphasize, however, that the sort of validation of the citation-based patent index attempted here could hardly have been done in a wider context simply because the measure of the value of innovations that would be required for that purpose are nowhere to be found" (185).

Since then there has been widespread use of "forward" citation counts as proxy for not only the social value of inventions, but also private value (Gittleman 2012, Sampat and Ziedonis 2004). Previous validation studies have found mixed evidence for the assumption that citations correlate with private value (Sampat and Ziedonis 2004, Hall 2000, Abrams 2013, Moser et al 2015, Harhoff et al 1997). And, perhaps for reasons Trajtenberg anticipated in the quotation above, there are surprisingly few studies of citations and social value, or the related idea that citations reflect "spillovers" between citing and cited inventors (Jaffe et al 2000).

For reasons we alluded to in the introduction, the drug industry has several features that make examining citations and value of pharmaceuticals empirically tractable. Various databases resulting from FDA regulatory requirements (the Orange Book and PTO list of extended patents) allow for linking of patents to products. It is possible to track sales for individual drugs, and, since marginal costs of production are typically low, these are reasonable proxies for profits.

Profits are a measure of private value. Pharmaceuticals is one industry where it is more plausible to construct social value measures. Indeed, the feasibility of this idea motivates a number of recent policy proposals to replace or supplement patent based incentive systems with prize based systems, with the size of the prize based on the social value of drugs (Hollis and Pogge 2008).³ This is typically viewed as being easier to do in drugs than other fields because of the availability of standard pharmacoeconomic measures of drug value and standard (if not uncontroversial) measures of the economic value of additional life years. Specifically, we can use incremental quality adjusted life years (QALYs) associated with each drug. QALYs are a summary measure of the health gains (or losses) from an intervention as compared to an alternative, including the effects of interventions on mortality (life years) and morbidity (quality). An intervention that generates 1 QALY increases life expectancy by one year in perfect health.

QALYs were developed for use in decision analysis and the economic evaluation of new medical technologies. In some countries, public payors use QALY gains from an intervention as a basis for determining whether it provides good value for money, i.e. whether it is "worth it." In some health systems, most prominently the United Kingdom, comparison of the QALYs gained from a drug (relative to some alternative, typically the status quo treatment) to incremental costs--the so called incremental cost- effectiveness ratio--to a societally agreed upon willingness-to-pay per QALY is used to guide reimbursement decisions (Drummond 1992). In Australia, the Pharmaceutical Benefits Advisory Committee (PBAC) also use cost-effectiveness in making their reimbursement decisions, but also other factors (Drummond et al 2007).

In cost effectiveness analyses the threshold value is typically \$50,000/QALY, though this was locked in more by historical accident than any logic (Chambers et al 2014). Analysts have suggested it is too low and should instead be as high as \$150,000 to \$200,000 QALY (Chambers et al 2014). As a practical matter, determining QALY per drug is typically done through in depth analysis of individual treatments, often as component of clinical trials.

³ There are also proposals to uses QALYs as basis for value based pricing (Jayadev and Stiglitz 2009), on the idea that this could shift the direction of innovation away from those with only incremental benefits.

Our analyses below will use data from the Tufts Cost-Effectiveness Database ("Tufts") which reports standardized QALY information for a large number of drugs, based on reviews of the literature.

Using data on QALYs per drug, together with assumptions about willingness to pay per QALY, allow us to calculate social value of drugs in dollar terms which we then compare to private value. It is typically assumed that firms do not always capture full gain from inventive activity (Nelson 1958, Nordhaus 2004) since consumers also benefit from new goods (and improved quality or lower prices for existing goods). While "market failures" are the main focus of most economic analyses of innovation, scholars of pharmaceuticals also emphasize that innovations that generate significant private profits may have low social value, if, for example, consumers are shifted to those goods by advertising, or innovation is characterized by significant business stealing by "me-too" drugs (Jayadev and Stiglitz 2009).

There is limited direct evidence on the share of social returns appropriated. Analyses in Nordhaus (2004) suggest it is very small, through the author also conjectures that given high rates of profits in pharmaceuticals it may be higher there. Philipson and Jena (2006) calculate global life expectancy gains in HIV-AIDS since the 1990s due to new drugs, and suggest that innovators capture only 5 percent of benefits, concluding that there is considerable underinvestment in R&D.⁴ Sun et al (2009) use a similar approach, finding 5-19 percent of social returns from cancer drugs appropriated by firms. Goldman et al (2010) look at 5 specialty cancer drugs and claims from 71 health plans, and use data on income elasticity and assumptions about demand curve to calculate consumer surplus from the drugs, finding social benefits outweigh costs to consumers by a factor of four. Each of these analyses looks at a small number of select drugs in specific disease areas, and rely on assumptions about demand and modelling to calculate value. To our knowledge no previous analyses use data on QALYs to examine the gap between private and social value of a large set of drugs.

⁴ The authors go further, saying that if HIV/AIDS drugs are representative "patients and health plans are getting too good a deal in the short run, which hurts them in the long run by insufficient R&D."

III. Data

FDA List of New Molecular Entities Approved Between 1987 and 2011

Our first data source is list of all 642 new molecular entities approved by the U.S. Food and Drug Administration between 1987 and 2011, taken from the Appendix in Lanthier et al (2013). This dataset goes beyond what is available from standard FDA databases (e.g. Drugs@FDA) by coding each drug based on its "novelty compared to the existing base of pharmaceutical" (1433). Specifically, it distinguishes between first in class drugs (the first drugs approved in a pharmacologic therapy class, representing new pathways for treating a disease), advance in class (not first in class but representing advances in therapy), and addition to class drugs (those that function similarly to other drugs but do not represent major advances). We use these categories to construct measures of technological impact or innovativeness.

Patent Data

We also collected information on patents for each of the drugs from the USPTO list of patent extensions. Under the Hatch-Waxman Act, firms are given patent term extensions to compensate for patent time lost during the drug approval process. Firms are typically allowed to extend one patent per drug, which tends to be the main patent (the "active ingredient") patent on the drug (Hemphill and Sampat 2011). We collected this information from the USPTO list of extended patents (USPTO 2016) supplemented by information from the Federal Register archives. One downside to using this source is that some of the NMEs approved between 1987 and 2011 (196/642) do not have any patents extended since their approval time was very short or their patents were filed or issued late in the development process. The share of drugs without patents is 33 percent for first in class drugs, 30 percent for advance in class, and 37 percent for addition to class drugs.

The other source commonly used for analyses of drug patents is the FDA's Orange Book (Hemphill and Sampat 2011, 2012). Firms are required to list patents associated with their drugs on this document. On average, there are 3-4 Orange Book patents listed per drug, including a range of ancillary or secondary patents. Since we want to focus on the main patent, we use the extended patent list instead, even though it is not complete for all drugs. In the Appendix, we examine robustness of all results to using Orange Book patents instead. The extended patent list has another advantage over the Orange Book: it includes data on

biological drugs (BLAs) in addition to small-molecule drugs (NDAs). This is important since one of the datasets we use below, the Tufts Cost-Effectiveness Registry, includes a non-trivial number of BLAs as we will discuss below.

Tufts Cost-Effectiveness Database

We also collected data on the new molecular entities (NMEs) and biologics (BLAs) approved between 1999 and 2011 included in the Tufts Cost-Effectiveness Registry. As described in Chambers et al (2014) the Tufts CEA Registry is a comprehensive database of 5,655 cost-utility analyses, which compiles and standardizes information from published cost-effectiveness analyses. There were 102 NMEs with CEAs in that database, of the 279 NMEs approved over the same period. Studies were included only if they compared the drug of study to the current standard treatment, if there was one. For each drug, we use the aggregate QALYs and cost for a typical course of treatment, relative to the next best alternative.

For the BLA drugs that are in Tufts but not in the NME approval dataset we also collected information on the extended patent. Of the 102 drugs (NMEs or BLAs) in the Tufts database, 76 have extended patents.

Optum ClinFormatics database

We collected information on drug sales and utilization for drugs with extended patents in the FDA and/or Tufts dataset. The Optum ClinFormatics database consists of administrative health claims (including pharmacy claims) for 2000-2014 for a large managed care company, covering 47 million users over this period. The claims table data are submitted by pharmacies for drugs dispensed on an outpatient basis. Unlike other databases, Optum also includes data on inpatient drugs. Using the Optum data we construct data on total expenditures associated with each drug (regardless of who paid), and the total number of new users.

Citation Data

The main goal of the paper is to relate different drug value measures to patent citations. Accordingly for each patent in our dataset (extended patents for drugs on the FDA list and/or Tufts list) we collected USPTO data on all "forward" citations made by subsequently issued patents. In addition to overall citations, we collected citation data excluding "self"

⁵ http://healtheconomics.tuftsmedicalcenter.org/cear4/AboutUs/WhatistheCEARegistry.aspx

citations (those where citing and cited firm are the same) and examine robustness of main results to using this measure in the Appendix.⁶

Measures of Value

As discussed in Section II it is often unclear as to exactly what value forward citations may relate. Here we consider three. First, the technological impact of a drug, based on the Lanthier et al (2011) categorization. We assume that first in class drugs represent greater technological impact or novelty. Second, the marginal health benefit of a drug per user, which is QALY gain from a drug relative to the next best, based on the Tufts data on incremental QALYs for each drug. Third, the total marginal health benefit of a drug, which is the latter multiplied by the number of users (from Optum). While the previous two are both proxies for health benefits of a drug, we proxy for private value using total sales (from Optum).

IV. Citations and technological impact

Our first set of analyses relate forward citations to the technological impact of a drug. For these analyses we use the 449 drugs approved between 1987 ands 2011 with an extended patent.

In a first specification, we relate the number of citations to the innovation category (with "addition to class," the least innovative drugs, left out as the reference category) and drug approval year fixed effects. Model 2 adds patent issue year fixed effects. In each of these models first in class drugs and advance in class drugs receive more citations than simple additions to class. In Model 1 first in class drugs are the most highly cited, followed by advance in class drugs. The difference between the two is significant at the 1 percent level. After controlling for patent issue year effects, in Model 2, there is no significant difference between the two innovative categories, though each generate about 17 more citations than additions to class. In Model 3 we control for the actual drug class from Lanthier et al, to examine within class variation in innovativeness and importance. There, we see no significant difference in citations across the different categories. One concern, however, is

⁶ For patents issued after 2001 (about 60 percent of all the patents) we also distinguished between examiner and other citations (Sampat 2010) since previous analyses suggest these two sources provide different signals of value (Hegde and Sampat 2009). Though we do not use the data here we intend to explore in future analyses.

that these categories are too fine, limiting the amount of identifying variation. (Nearly half of the categories have only 2 drugs, and nearly 70 percent only 3.)

Accordingly, we also re-examined this using another technological categorization, the ATC codes for each drug, which we obtained from the World Health Organization ATC/DDD Index (https://www.whocc.no/atc ddd index/). The ATC system classifies drugs according to anatomy (A), therapeutic property (T), and chemical properties (C). We use "4 digit" ATC codes, the highest level of aggregation that captures each of these dimensions. Over 70 percent of the drugs in our sample are in 4-digit ATC groups with 5 or more other drugs. Model 4 shows results after including these fixed effects. Here too both first in class and advances in class are more heavily cited than the least innovative category, additions to class, but the magnitude of the effect is much greater for advance in class. Moreover, the difference between first in class and addition to class is insignificant.

In models 5 and 6, we combine the two categories, looking at innovative drugs vs. the additions to class, using the Lanthier et al. classes and ATC classes respectively. There is no significant difference between the number of citations for the innovative drugs and others.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Citations	Citations	Citations	Citations	Citations	Citations
First in class	18.03***	16.88**	-1.730	4.871		
	(6.112)	(6.678)	(10.74)	(7.734)		
Advance in class	15.56**	17.42***	4.229	14.70**		
	(6.022)	(6.689)	(12.41)	(7.398)		
First or advance in class					1.372	8.551
					(9.764)	(6.481)
Constant	61.14**	66.17*	127.2***	18.04	124.2**	20.17
	(29.22)	(34.78)	(48.44)	(44.93)	(47.98)	(44.47)
Observations	449	449	449	449	449	449
R-squared	0.117	0.188	0.837	0.522	0.836	0.519
Approval Year FE	Y	Y	Y	Y	Y	Y
Issue Year FE		Y	Y	Y	Y	Y
ATC FE				Y		Y
Drug Class FE			Y		Y	

Robust standard errors in parentheses

We see very similar patterns if we examine log-citations, after dropping the 4 drugs with no forward citations:

^{***} p<0.01, ** p<0.05, *<.1

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Log Citations	Log Citations	Log Citations	Log Citations	Log Citations	Log Citations
First in class	0.293**	0.260**	-0.208	0.0331		
	(0.122)	(0.129)	(0.251)	(0.179)		
Advance in class	0.215	0.246*	0.0615	0.211		
	(0.131)	(0.136)	(0.302)	(0.180)		
First or advance in class					-0.0696	0.101
					(0.249)	(0.160)
Constant	2.942***	3.028***	4.593***	2.324*	4.461***	2.362*
	(0.492)	(0.653)	(1.093)	(1.234)	(1.111)	(1.224)
Observations	445	445	445	445	445	445
R-squared	0.116	0.212	0.823	0.545	0.821	0.543
Approval Year FE	Y	Y	Y	Y	Y	Y
Issue Year FE		Y	Y	Y	Y	Y
ATC FE				Y		Y
Drug Class FE			Y		Y	

Robust standard errors in parentheses

^{***} p<0.01, ** p<0.05, * p<0.1

Overall, this evidence suggests that citations are associated with technological impact, but do not discriminate well between more and less important drugs within narrow or broad drug classes.

V. Citations, health benefits, and private value

Table 1 Summary Statistics

Variable	Mean	Median	Std. Dev.	Min	Max	Obs
Unique Users	135,564	9,346	328,200	0	1,762,840	101
Revenue (\$MM)	187	5.44	436	0	3,720	101
QALYs (Per User)	0.17	.03	0.88	-5.8	4.2	101
Total QALYs	5,004	283	24,308	-15,257	221,080	101
Revenue/ User	18,326	2,810	58,078	0	518,006	101
Marginal Revenue (\$MM)	541	34	2,527	-739	22,890	100
Aggregate Marginal Consumer Surplus (\$MM)	-37	-3.5	3,162	-23,521	16,673	100
Biologic	0.16	0	0.37	0	1.0	101
Total Cites to Extended Patent	52.8	38	57.39	0	276	75
Total Cites to Extended Patent (Excluding Self)	46.08	29	50.69	0	259	75
Total Cites to Orange Book Patents	90.07	56	112.91	2	613	75
Total Cites to Orange Book Patents (Excluding Self)	81.60	50	110.01	2	590	75

Table 1 reports summary statistics of the data. The average drug has about three-quarters of a million users, although the distribution is skewed with the median around 60,000. This means the distribution of many of the other aggregate variables will also be skewed. Total Expenditures is an example, with a mean of \$185 million and a median of \$5.4 million. For this reason, we report many of the analyses using both log and linear specifications. The mean drug generates 0.17 marginal QALYs per user, although a large share actually have a negative marginal

About 17 percent of the drugs are biologics. For the 75 drugs with a patent listing on the Orange Book, the average number of citations is 90 to all associated patents, and 81 excluding self citations. For the extended patents (whether or not on the Orange Book) there are fewer citations, but note that only one patent per drug is typically extended. (By comparison, drugs on the Orange Book have 3-4 patents on average, and each can generate citations.)

As Chambers et al (2014) report, the total number of QALYs associated with each drug averages .17 years, worth somewhere between \$8,500 and \$25,500. About a quarter of the drugs have negative QALYs, or provide fewer life year advantages than the previous standard of care. Turning back to the full set of 102 drugs, the total number of unique users recorded in the Optum data over the 2000-2014 period ranges from 0 to 10.2 million, and total sales from zero dollars to nearly \$4 billion.

Table 2 Unique Patient Counts

Unique Patients	Count
0-500	11
500-1000	4
1K-10K	36
10K-100K	24
100K-1MM	21
1MM+	4

Table 2 reports the distribution of unique patients over the full set of drugs. There is a large range of drug types. An appreciable share would be classified as orphan drugs, with 15 having under 1000 users. On the other end of the spectrum, there are 4 blockbuster drugs with over 1 million users each.



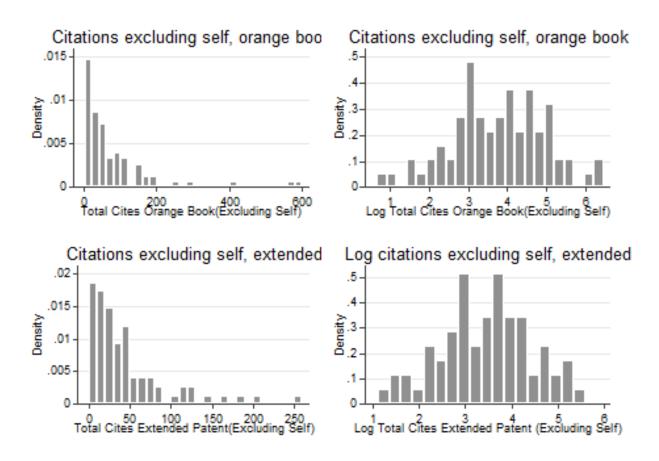


Figure 1 shows the distribution of forward citations to all patents in the data set. The upper two histograms compute citations to all Orange Book patents associated with a drug; the lower two use the citations to the Extended Patent (usually only one). It is clear from the histograms of the left that the distribution is non-normal. The figures on the right, using log citations, appear normal, thus we include log cites in much of the following analysis.

In order to validate the integrity of the utilization variables we examined how they evolve over time, and if they respond to known events (e.g. drug approval dates, withdrawals,

introduction of competitor drugs, and patent expiration). In Figure 2 we present time trends in total drug fills for 4 drugs⁷ where approval and generic entry dates are known.

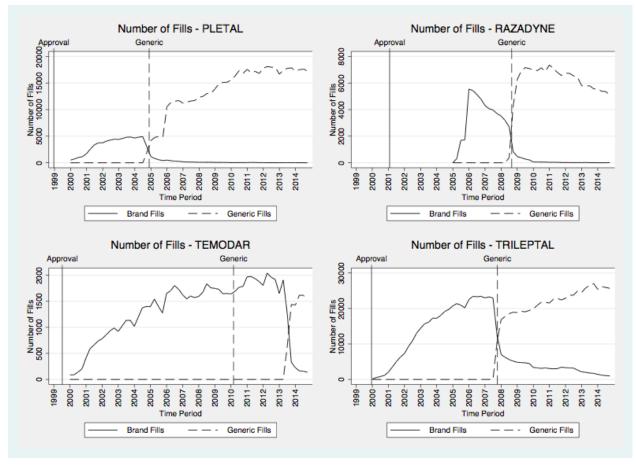


Figure 2

For all of the drugs, utilization does not begin until after drug approval, reassuringly. And for most we see sharp changes in utilization near or after generic approval data, with many more generic fills. Though not the focus of this analysis it is interesting that for nearly all of these drugs total utilization increases with genericization, contrary to the message of previous work (Lakdawalla and Philipson 2012) on this topic.

Citations vs. Private Value

In this section we present results from a set of cross-sectional regressions relating forward patent citations to different measures of value using data from 2000 - 2014. Forward

⁷ Dates derived from Hemphill and Sampat (2012)

⁸ For low sales drugs, and for drugs with various types of pay-for-delay settlements, generic launch may lag generic approval (Hemphill and Sampat 2012).

citations in Tables 3-5 are simply the aggregate number of citations received by the extended patent for a given drug. We control for differing ages by including a set of fixed effects for FDA approval year in all specifications. The first column of Table 3 shows that there is a positive and statistically significant relationship between citations and revenue. Revenue is derived from Optum data and is simply the producer surplus, assuming low marginal cost of production. Since producer surplus is highly skewed, the log specification in column 3 is also informative and shows a coefficient that is just insignificant at the 5% level. What we learn from these two regressions is that there does appear to be a weak positive relationship between citations and revenue, but one where the elasticity is relatively low. We see this relationship visually in Figure 3; Figure A3 in the appendix reports the analogous data using Orange Book patents, with similar results.

Table 3 Citations vs Private Value

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Citations	Citations	Log Citations	Log Citations	Citations	Log Citations
Revenue	0.0362**	0.0323**				
	(0.00774)	(0.00770)				
Biologic		-0.388		0.796		
		(20.54)		(0.505)		
Revenue*Biologic		0.0366*				
_		(0.0166)				
Log Revenue			0.119	0.159		
			(0.0664)	(0.0813)		
Log Revenue*Biologic				-0.140		
0				(0.134)		
Marginal Revenue				(5.25.)	0.00753	
marginal nevenue					(0.00513)	
Log Marginal Revenue					(0.00313)	0.201**
LOG IVIAI GIII AI NEVEITUE						(0.0572)
Observations	75	75	70	70	74	<u> </u>
Observations	75	75	70	70	74	57
R-squared	0.318	0.331	0.320	0.338	0.219	0.283

^{*}Robust standard errors in parentheses. Revenue (\$MM) is obtained from Optum, citations are extended patent forward citations. Incremental cost is aggregate incremental cost presented in millions. All regressions include year of FDA approval fixed effects.

In columns 2 and 4 we add a dummy variable indicating whether the drug is a biologic as well as the interaction between it and the main regressor (revenue or log revenue). The results are inconsistent: Column 2 shows a coefficient on the interaction term of similar magnitude as the main coefficient, while in column 4 the interaction term has a negative

^{**} p<0.01, * p<0.05

sign and is statistically insignificant. Of course with only 17 biologics in the entire data set, the results are sensitive to a small number of outliers.

In the final two columns of Table 3 we examine the relationship between citations and a different possible measure of value, marginal revenue. The independent variable here is simply the aggregate producer surplus relative to the next-best technology. The linear specification reported in column shows a positive, but statistically insignificant relationship. However the coefficient in the log specification in column 6 is statistically significant and almost twice as large as the elasticity when examining aggregate producer surplus. One point to note is that this regression does not include the 17 observations where the marginal producer surplus is negative - i.e. the new drug is cheaper than the previous one.

Forward Citations vs. Revenue (2000-2014) Unique Citations to Extended Patents 6 Log Forward Citations 5 0 0 3 Ó 2 0 V Z, 'Ø, ďζ Log Total Revenue Log Total Cites (Excluding Self) Fitted values

Figure 3

Citations vs. Social Value

In Table 4 we examine the relationship between the citations and measures of social value and consumer surplus. Columns 1 and 2 report regressions of citations on incremental QALYs a drug confers, relative to the next best. This is the best measure of marginal social value of a drug, relative to the best alternative. QALYs here are used as the independent variable directly, without accounting for their value (which we do use to compute consumer surplus). Since QALYs can be negative, the log-log specification reported in columns 3 and 4 makes use of only 49 observations.

Table 4

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Citations	Citations	Log Citations	Log Citations	Citations	Log Citations
QALYs	-57.77	-43.81				
	(184.1)	(183.8)				
Biologic		7.382		1.003		
		(19.23)		(2.147)		
QALYs*Biologic		-382.9				
		(7,584)				
Log QALYs			0.169**	0.161*		
			(0.0599)	(0.0625)		
Log QALYs*Biologic				0.131		
				(0.304)		
Consumer Surplus					-0.00312	
					(0.00183)	
Log Consumer Surplus						0.108
						(0.0868)
Observations	75	75	49	49	74	27
R-squared	0.212	0.215	0.388	0.393	0.238	0.586

^{*}Robust standard errors in parentheses. QALYs are total incremental QALYs, presented in millions. Citations are extended patent forward citations. Consumer surplus is presented in millions. All regressions include year of FDA approval fixed effects.

While we don't observe any significant relationship between QALYs and citations in the first two columns, there is one in the log specification. The elasticity of .17 in column 3 is somewhat higher than that found in the analogous regression in Table 3 where log revenue was the variable of interest. Still, the finding should be interpreted with the fact that a third of the observations are missing due to their negative marginal QALYs. This may mean that

^{**} p<0.01, * p<0.05

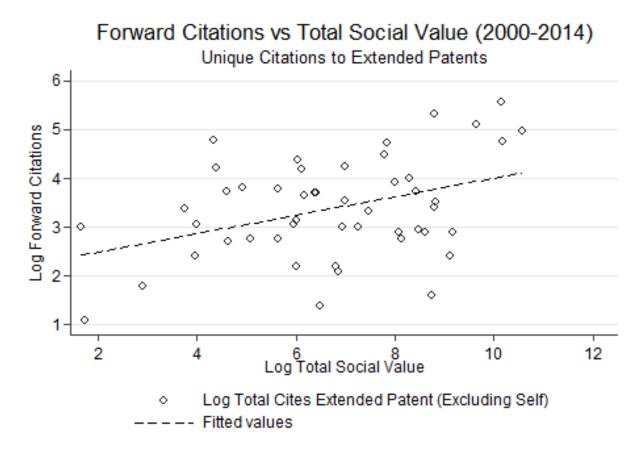
forward citations are a reasonable proxy for for marginal social value, but only when it is positive.

The final two columns of Table 4 examine the relationship between citations and marginal consumer surplus. Marginal consumer surplus is simply marginal health benefits minus marginal cost. Health benefits are computed in dollars by using a midrange value of \$100,000/QALY. There is no significant relationship between citations and marginal consumer surplus, and in fact the point estimate is negative. There are only 27 observations in the final log-log specification because so many of the drugs yield negative or zero marginal consumer surplus.

Figure 4 shows a related relationship, not explored in the regression table, that of log forward citations and marginal social value per drug. There does not appear to be a significant relationship, similar to what is found for Orange Book patents (Figure A4).

Figure 4

Total



As seen in Table 4 it seems likely that forward citations are a better proxy for total marginal social value, not marginal social value per user. After all, a high QALY drug for a small number of users may not attract as much entry (and thus generate later citing patents) as a medium QALY drug for a large population. Thus, Figures 5 and A5 display the results based on citations to extended and Orange Book patents, respectively, that are analogous to Column 3 from Table 4. Both show an increasing relationship between citations and marginal social value. Since these figures include patents of different ages, in the Appendix we reproduce Figure A5 but including only patents with at least 5 (Figure A6) or 10 years (Figure A7) of forward citations, and report the 5 and 10-year forward citation count as the dependent variable, respectively. With these adjustments for age, the relationship becomes stronger.

Forward Citations vs Total Social Value (2000-2014) Unique Citations to Extended Patents _og Forward Citations Log Total Social Value Log Total Cites (Excluding Self) Fitted values

Figure 5

Private vs. social value

So far we have attempted to determine whether there is a strong argument for using forward citations as a proxy for either private or social value of pharmaceuticals. The evidence has mostly shown that forward citations appear to be weakly correlated with both measures of patent value. Now we examine a different question - the relationship between private and social value. This relationship is crucial in several current policy debates including those surrounding incentives for pharmaceutical companies, drug pricing, and who benefits from new drug development.

Table 5

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable:	Revenue	Revenue	Log Revenue	Log Revenue	Revenue	Log Revenue
QALYs	4,067**	4,405**				
	(848.1)	(648.3)				
Biologic		71.81		0.394		
		(94.40)		(1.218)		
QALYs*Biologic		-77,558*				
		(34,651)				
Log QALYs			0.325**	0.338**		
			(0.0947)	(0.104)		
Log QALYs*Biologic				-0.0188		
				(0.157)		
Consumer Surplus					-0.00580	
					(0.0214)	
Log Consumer Surplus						0.408**
						(0.123)
Observations	101	101	68	68	100	46
R-squared	0.170	0.219	0.358	0.366	0.125	0.425

^{*}Robust standard errors in parentheses. Revenue presented in millions. QALYs are total incremental QALYs, presented in millions. Consumer surplus is presented in millions. All regressions include year of FDA approval fixed effects.

As in the analyses immediately above, we construct private value (revenue) based on total expenditures for the drug between 2000 and 2014 (from Optum) and marginal social value based on the average marginal QALY per drug (from Tufts) and number of unique users of

^{**} p<0.01, * p<0.05

the drug from 2000 to 2014 (from Optum). To facilitate comparison, we present the social value in dollars. Since there is no universally agreed upon yardstick, we constructed estimates based on a middle range estimate of \$100,000 per QALY.⁹

Table 5 reports results from regressions where the dependent variable is either revenue or log revenue and the independent variables are either measures of marginal social value or marginal consumer surplus. Columns 1 and 2 show that there is a highly significant relationship between producer surplus and marginal social value and that the relationship differs for biologics. Columns 3 and 4 show that the elasticity of producer surplus to marginal social value is approximately one-third. Column 6 indicates an even higher elasticity when the independent variable is consumer surplus. But this finding must be taken in conjunction with the facts that it excludes over half of the observations which have negative consumer surplus, and that the linear regression in column 5 shows no significant relationship.

It is useful to get a sense of the distribution of private and social value per user. These results may be seen in Figures 6 and 7, respectively.

⁹ We report results based on an upper bound of \$150,000 per QALY in the appendix.

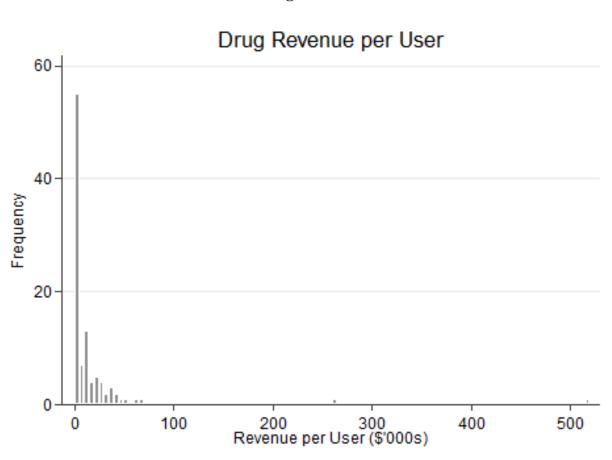


Figure 6

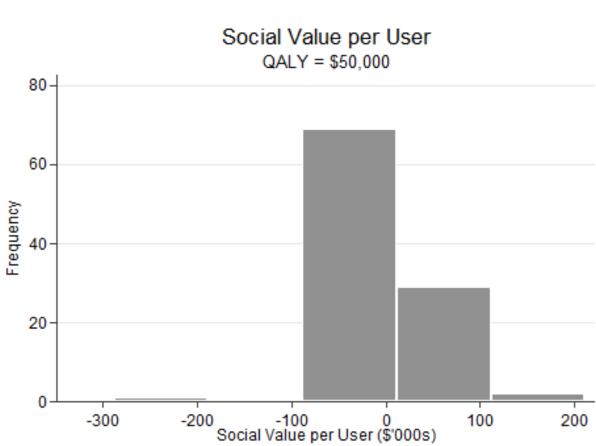


Figure 7

Several interesting things may be observed. When examining the per-user figures, it is clear that there is far more variation in the social value per user gained from a drug. The social value per user ranges from around -\$2 million to + \$3.6 million. Recall that social value may be negative since a new drug may be worse than the current treatment. In fact, 32 of the 101 drugs in this analysis are worse than the current treatment in terms of having negative marginal QALYs. Of those, 19 both have negative marginal QALY and higher cost relative to the current treatment. Private value (revenue per drug) ranges from 0 to about \$500,000 per user.

 $^{^{10}}$ This analysis uses \$50,000 for 1 QALY, the lower bound of what is typically used.

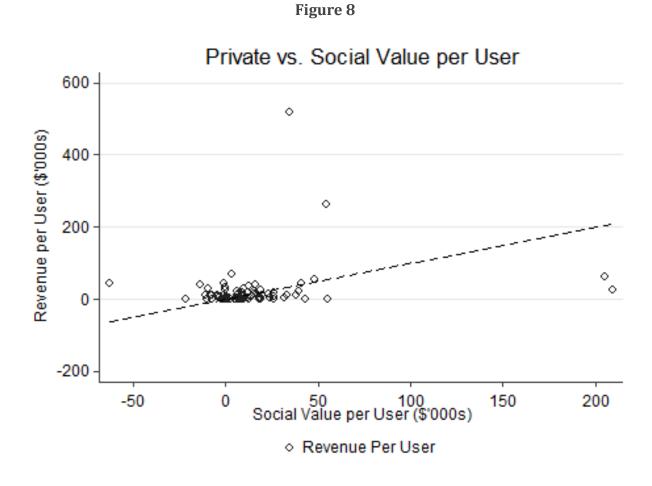
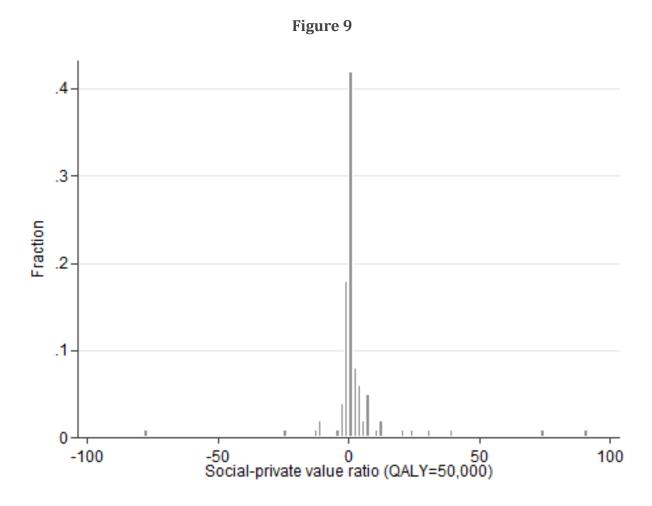


Figure 8 shows private vs. social value per user using the \$50,000/QALY value (removing a handful of outliers for readability). The dashed line is a 45 degree line to indicate the point at which private and social value are equal. Points below and to the left of the line represent drugs with greater social than private value; those above and to the right have greater private value than social value. While it is hard to see it on this figure, of the 101 drugs, 43 have lower social than private value. This of course includes the 32 that have negative social value, along with 11 other drugs where the expenditures per user exceed the health gains. The distribution of the ratio of social to private value across drugs is reported in Figure 9. (The ratio is negative for all drugs with negative QALYs.)



The mean ratio is 10 and the median 2, consistent with the idea that most of the time the benefits of new drugs outweigh their costs. (The aggregate value of the QALYs across all drugs and users in our dataset is \$255 billion, compared to \$19 billion in sales.) But Figures 8 and 9 also shows for about a third of the drugs (accounting for \$6 billion in aggregate sales) incremental QALYs are zero or negative, and for many other drugs quite low.

On one hand, the overall high average ratio of social-private value in our data is consistent with previous work suggesting firms under-appropriate returns from R&D (Philipson and Jena 2006). But by looking at a large cross-section of drugs we also see significant heterogeneity that the averages conceal. The results so far are similar to that from broader work on the economics of medical technology emphasizing that while new innovations generate a lot of value on average, and thus technological change in health care is "worth it" (in the sense of covering its costs), there is also a lot of wasted expenditure (Cutler 2004). This is consistent with the results of the final two regressions reported in Table 3 which show a positive and statistically significant relationship between private and social value.

They also reveal a low coefficient, indicating consumers capture are large share of the surplus on average.

VI. Discussion

We have presented a number of new findings about the relationship between patent citations and drug value, as well as how different measures of value are related. The nature of available information on pharmaceuticals requires a number of assumptions to reach these conclusions. Here we discuss the assumptions implicit or explicit in the analysis and how they impact what we may learn.

First one must note the small number of observations (102) on which there is any data available for health benefits. When combined with other data sources, such as revenue, the number drops for most of the analyses. While small in number, this does represent the universe of NME's for which this data is available from 1999-2011. As such the data may not be representative of all drug patents, and is likely biased towards drugs where there were incentives to perform cost-benefit studies.

We do not consider patents for drugs that were not commercialized, as those would have no revenues. If anything, this likely drops a number of patents that would have few citations and no revenues, and likely would simply add noise to our estimates. For our main analyses we focus on the primary patent associated with a drug, as identified by the Extended patent list. We do so for simplicity and because most studies that make use of patent citations seem to implicitly begin with the assumption of a single patent per product. This may understate the number of total citations associated with a drug, and so we also perform the analysis using Orange Book patents and find similar results.

All of the analyses incorporating revenue information draw on the Optum data set. Ideally we would have access to complete national pricing and usage information for all drugs, both inpatient and outpatient. While this dataset does not exist, the Optum data is one of the closest approximations. It is one of the most comprehensive commercially available datasets on patient claims, covering over 47 million unique patients from 2000-2014. The population is geographically diverse, spanning all 50 states. It almost certainly underweights the elderly population as Medicare patients are only included through 2006. As such, one should not interpret these results as a random sample from the universe of insured American patients, but as a large, diverse subset.

Pharmaceutical patents are certainly not representative of all patents, and therefore it is difficult to extrapolate from the results here to patents more broadly. However, pharmaceuticals are the industry that is often pointed to as the best argument for the patent system, where billion dollar investments are often protected only by IP rights. It is the sector where patent applicants take searching for prior art most seriously. This field is perhaps the only one where it is possible to determine private and social value for an individual patent.

VII. Conclusions

Measuring the value of innovation is important for evaluating a range of science and technology policies. Patent-citation based proxies for private and social value are commonly used, but with limited direct validation. In this paper we leveraged unique administrative data from pharmaceuticals (many of which are available as an unintended result of regulations: FDA patent listing rules, Hatch-Waxman patent extension procedures, the need for CEA for reimbursement in some countries) to link patents to products and products to measures of private and social value. While our analyses are still preliminary, they provide very little support for the hypothesis that citations correlate with private or social value. Though evidence from earlier research on citations and private value is mixed (Jaffe and Rassenfosse 2016), our results are in stark contrast to previous work suggesting citations are good proxies for social value (Trajtenberg 1990). Importantly, our social value measures are based not on discrete choice demand models but QALYs from different cost-effectiveness analyses; each approach has its strengths and weaknesses.

Though our analyses were all cross-sectional, in future work we hope to more explicitly incorporate a time dimension (modeling the timing of citations and evaluation of private and social value, and robustness of results to various approaches to discounting the value measures), to look at differential informational value of different types of citations (examiner vs. applicant), examine the extent to which citations respond to shocks in value (FDA approval and other clinical milestones, generic entry, withdrawal), and conduct complementary qualitative studies of specific drugs to better understand the citations-value relationship (or lack thereof).

Our paper also provides new data on the private and social value of pharmaceuticals. The assumption that private and social value may not be aligned motivates a number of policy proposals (e.g. innovation prizes). Our analysis provides support for this assumption: there

are many drugs where private value exceeds social value, and vice versa. In future work it would be interesting to examine what factors (patent strength, extent of competition, advertising) affect the share of social returns appropriated. While our results so far are consistent with previous work suggesting that social returns for drugs are on average high, they also suggest there is considerable heterogeneity, with a large number of drugs where social value is less than private. While the assumption that social returns are on average high motivates classic and contemporary discourse on science and technology policy, including patent policy, the significant heterogeneity across drugs in the private-social welfare gap has been neglected by policymakers and academics alike.

Our social value measures are based on QALYs. The exact value of a QALY is a matter of some debate, as we have discussed. More importantly, adding QALYS is based on utilitarian idea of value, and some have raised objections to adding up life years regardless of a patient's context or situation. Therapies that represent significant improvements for some patients but not the majority would not generate many incremental QALYs, but from some perspectives way still be considered valuable. There are also technical issues with these measures. Estimating QALYs for different health states is difficult, involving, for example choices about whose preferences count in determining utility (Unlike QALYs another health related quality of life measure, the disability adjusted life year or DALY, uses expert judgement of the disutility of conditions, rather than patient judgement.) There is also a question about when in the lifecycle of a therapy to measure a QALY (before or after post-marketing learning) and the generalizability of QALY estimates (often collected as part of clinical trials) to the population. While we acknowledge this as an important limitation, such issues would complicate any attempt to measure "value" of an innovation to patients, of course.

The paper also illustrates the difficulties involved in moving from cost-effectiveness measures to social welfare measures, and assumptions necessary to make along the way, which any reward system based on social welfare (including prize-based mechanisms) would have to confront. We had the benefit of a database of published cost-effectiveness analyses on a select set of drugs. But any institutionalized reward system based on QALYs or similar measures would also need to be accompanied by a significant technology assessment infrastructure to work well, which could be the most costly aspect of such a system (Hollis and Pogge 2008).

This points to the main limitation of our analysis, the selected nature of our sample. We examined drugs on which cost-effectiveness analyses of sufficient quality to merit inclusion in the Tufts database were performed. While there is not much we can do about the lack of

high quality CEAs, in future work we also plan to assess the extent of selection and its implications for our findings.

Works Cited

- Abrams, David, Ufuk Akcigit, and Jillian Popadak. "Patent Value and Citations: Creative Destruction or Strategic Disruption?" (2013). Print.
- Chambers, J. D., T. Thorat, J. Pyo, M. Chenoweth, and P. J. Neumann. "Despite High Costs, Specialty Drugs May Offer Value For Money Comparable To That Of Traditional Drugs." *Health Affairs* 33.10 (2014): 1751-760. Print.
- Cohen, Wesley, Richard Nelson, and John Walsh. "Protecting Their Intellectual Assets:

 Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)."

 (2000). Print.
- Cutler, David M. *Your Money or Your Life: Strong Medicine for America's Health Care System.*Oxford: Oxford UP, 2004. Print.
- Drummond, Michael F. "Economic Evaluation of Pharmaceuticals." *PharmacoEconomics* 1.1 (1992): 8-13. Print.
- Garthwaite, Craig, and Mark Duggan. "Empirical Evidence on the Value of Pharmaceuticals."

 Oxford Handbooks Online (2012). Print.
- Gittelman, Michelle. "Patent Citations." *The Palgrave Encyclopedia of Strategic Management*.

 Print.
- Goldman, Dana P., Anupam B. Jena, Darius N. Lakdawalla, Jennifer L. Malin, Jesse D. Malkin, and Eric Sun. "The Value of Specialty Oncology Drugs." *Health Services Research*.

 Blackwell Science Inc, Feb. 2010. Web. 03 Jan. 2017.
- Griliches, Zvi. "Patent Statistics as Economic Indicators: A Survey." (1990). Print.

- Griliches, Zvi. "R&D, Patents and Productivity." (1984). Print.
- Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg. "Market Value and Patent Citations: A First Look." (2000). Print.
- Harhoff, Dietmar, Francis Narin, Frederic M. Scherer, and Katrin Vopel. *Citation Frequency and the Value of Patented Innovation: Research Area, Market Processes and Corporate Management.* Berlin, 1997. Print.
- Hegde, Deepak, and Bhaven Sampat. "Examiner Citations, Applicant Citations, and the Private Value of Patents." *Economics Letters* 105.3 (2009): 287-89. Print.
- Hemphill, C. Scott, and Bhaven N. Sampat. "Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals." *Journal of Health Economics* 31.2 (2012): 327-39.

 Print.
- Hemphill, C. Scott, and Bhaven N. Sampat. "When Do Generics Challenge Drug Patents?" *Journal of Empirical Legal Studies* 8.4 (2011): 613-49. Print.
- Hollis, Aidan, and Thomas Pogge. *The Health Impact Fund: Making New Medicines Accessible*for All: A Report of Incentives for Global Health. New Haven, CT: Incentives for Global

 Health, 2008. Print.
- Jaffe, Adam, Manuel Trajtenberg, and Michael Fogarty. "The Meaning of Patent Citations:

 Report on the NBER/Case-Western Reserve Survey of Patentees." (2000). Print.
- Jaffe, Adam, and Gaétan De Rassenfosse. "Patent Citation Data in Social Science Research:

 Overview and Best Practices." (2016). Print.
- Kapczynski, Amy, Chan Park, and Bhaven Sampat. "Polymorphs and Prodrugs and Salts (Oh My!): An Empirical Analysis of "Secondary" Pharmaceutical Patents." *PLoS ONE* 7.12

(2012). Print.

- Lakdawalla, Darius, and Tomas Philipson. "Does Intellectual Property Restrict Output? An Analysis of Pharmaceutical Markets." *The Journal of Law and Economics* 55.1 (2012): 151-87. Print.
- Lanthier, M., K. L. Miller, C. Nardinelli, and J. Woodcock. "An Improved Approach To

 Measuring Drug Innovation Finds Steady Rates Of First-In-Class Pharmaceuticals,

 1987-2011." *Health Affairs* 32.8 (2013): 1433-439.
- Lichtenberg, F. R. "Are The Benefits Of Newer Drugs Worth Their Cost? Evidence From The 1996 MEPS." *Health Affairs* 20.5 (2001): 241-51. Print.
- Moser, Petra, Joerg Ohmstedt, and Paul Rhode. "Patent Citations and the Size of the Inventive Step Evidence from Hybrid Corn." (2015). Print.
- Nelson, Richard R. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy* 67.3 (1959): 297-306. Print.
- Nordhaus, William. "Schumpeterian Profits in the American Economy: Theory and Measurement." (2004). Print.
- Philipson, Tomas J., and Anupam B. Jena. "Who Benefits from New Medical Technologies?

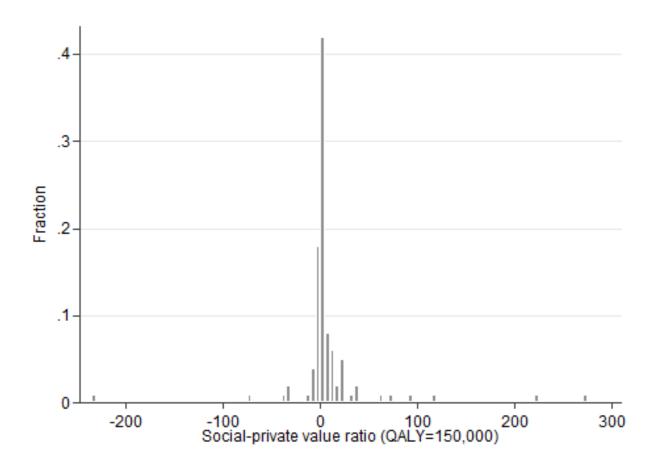
 Estimates of Consumer and Producer Surpluses for HIV/AIDS Drugs." *Forum for Health Economics & Policy* 9.2 (2006). Print.
- Sampat, Bhaven N. "When Do Applicants Search for Prior Art?" *The Journal of Law and Economics* 53.2 (2010): 399-416. Print.
- Stahl, Jessica C. "Mergers and Sequential Innovation: Evidence from Patent Citations." *SSRN Electronic Journal*. Print.

- Sun, Eric C., Anupam B. Jena, Darius N. Lakdawalla, Carolina M. Reyes, Tomas J. Philipson, and Dana P. Goldman. *An Economic Evaluation of the War on Cancer*. Cambridge, MA:

 National Bureau of Economic Research, 2009. Print.
- Sun, Eric, Anupam Jena, Darius Lakdawalla, Carolina Reyes, Tomas Philipson, and Dana Goldman. "An Economic Evaluation of the War on Cancer." (2009). Print.
- Trajtenberg, Manuel. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *The RAND Journal of Economics* 21.1 (1990): 172. Print.

Appendix

Figure A1 Social vs. private value using \$150,000/QALY threshold:



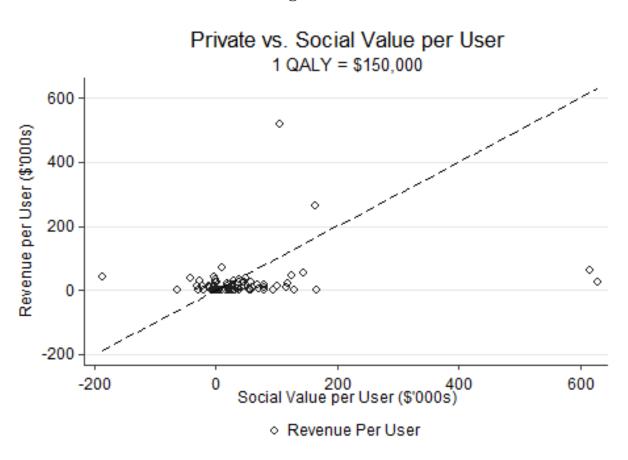
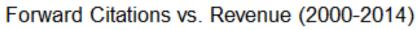


Figure A2

Figure A3



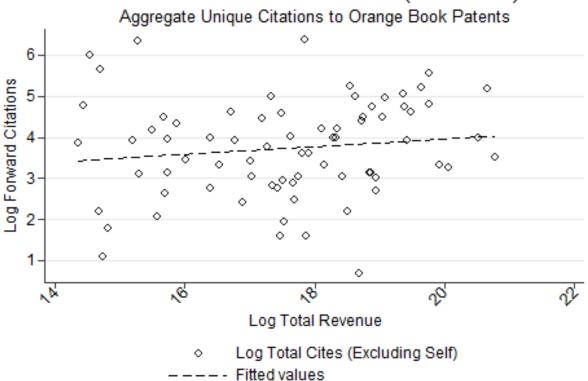
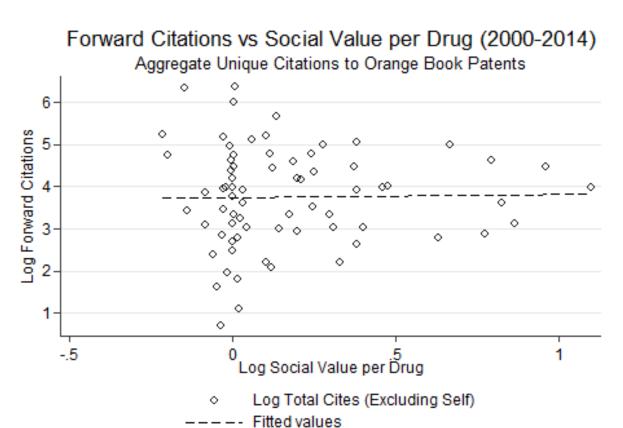


Figure A4





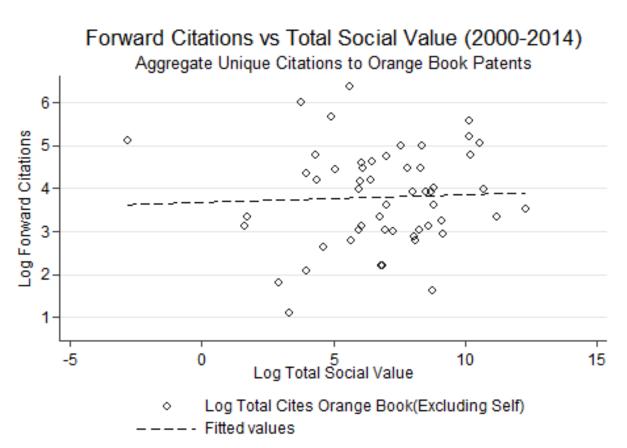


Figure A6

5 Year Citations vs. Revenue (Drugs Approved Until 2009
Unique Citations to Extended Patents

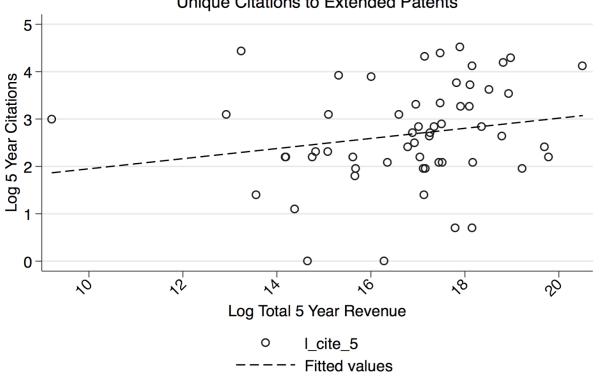


Figure A7

10 Year Citations vs. Revenue (Drugs Approved Until 2004 Unique Citations to Extended Patents

