

Public and private investments in biomedical research: Online appendix

Maya Durvasula
Lisa Larrimore Ouellette
Heidi Williams

forthcoming, *American Economic Association Papers and Proceedings*

1 Drug approvals: drugs@FDA

Our data on drugs approved by the U.S. Food and Drug Administration (FDA) is drawn from the drugs@FDA database, available at <http://www.fda.gov/Drugs/InformationOnDrugs/ucm079750.htm>.

2 Drug patents: Approved Drug Products with Therapeutic Equivalence Evaluations (Orange Book)

Our data on patents linked to FDA-approved drugs is drawn from annual editions of the Approved Drug Products with Therapeutic Equivalence Evaluations (commonly known as the Orange Book). As noted in the text of our paper, patents are removed from these publications when they expire, so constructing a complete set of patents from the Orange Book requires reconstructing a list from each annual version of the publication. One of us (Williams) digitized the historical Orange Book patent and exclusivity tables for years 1985-2016 (no Orange Book was published in 1986), based on PDF versions obtained via a Freedom of Information Act (FOIA) request; that data is available here: <https://www.nber.org/research/data/orange-book-patent-and-exclusivity-data-1985-2016>. Since Orange Book patent listings began in 1985, we are unable to determine if drug patents were listed and removed before 1985.

3 Publicly funded patents: USPTO

The USPTO hosts bulk TXT and XML versions of US patent grants, which are available for download beginning 1 January 1976 here: <https://www.uspto.gov/learning-and-resources/xml-resources>. We parse the TXT files for patent grants issued before 1 January 2001 and XML files for all grants after this date to extract the key variables needed for our analysis.

3.1 Government interest statements

The 1980 Bayh-Dole Act – as well as earlier requirements imposed by some US government funding agencies – requires recipients of federally funded research grants who seek patent protection for their inventions to include a statement in the patent’s text referencing both the granting agency and the specific grant number. Such a statement appears in formats such as the example below, although in practice many government interest statements are formatted differently:

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT:

This invention was made with government support under [contract ID] awarded by [Federal Agency]. The government has certain rights in this invention.

In patent applications and grants, this section must appear either as the first section of a patent’s description or (if relevant) immediately following a section titled “Cross Reference to Related Application.”¹ In practice, this disclosure may appear elsewhere. Figure 1 includes an example of a government interest statement in a US patent grant (boxed in red).

Figure 1: Example of a government interest statement

US 2014/0065708 A1 1

ANTIBODIES AS T CELL RECEPTOR
MIMICS, METHODS OF PRODUCTION AND
USES THEREOF

CROSS REFERENCE TO RELATED
APPLICATIONS/INCORPORATION BY
REFERENCE STATEMENT

[0001] This application is a continuation of Ser. No. 12/380,605, filed Feb. 27, 2009; which claims benefit under 35 U.S.C. 119(e) of U.S. Ser. No. 61/067,328, filed Feb. 27, 2008, and U.S. Ser. No. 61/191,871, filed Sep. 12, 2008. The '605 application is also a continuation-in-part of U.S. Ser. No. 11/809,895, filed Jun. 1, 2007, now abandoned; which claims benefit under 35 U.S.C. 119(e) of U.S. Ser. No. 60/810,079, filed Jun. 1, 2006. The '895 application is also a continuation-in-part of U.S. Ser. No. 11/517,516, filed Sep. 7, 2006, now abandoned; which claims benefit under 35 U.S.C. 119(e) of provisional applications U.S. Ser. No. 60/714,621, filed Sep. 7, 2005; U.S. Ser. No. 60/751,542, filed Dec. 19, 2005; U.S. Ser. No. 60/752,737, filed Dec. 20, 2005; and U.S. Ser. No. 60/838,276, filed Aug. 17, 2006. The '516 application is also a continuation-in-part of U.S. Ser. No. 11/140,644, filed May 27, 2005, now abandoned; which claims benefit under 35 U.S.C. 119(e) of provisional applications U.S. Ser. No. 60/574,857, filed May 27, 2004; U.S. Ser. No. 60/640,020, filed Dec. 28, 2004; U.S. Ser. No. 60/646,338, filed Jan. 24, 2005; and U.S. Ser. No. 60/673,296, filed Apr. 20, 2005. The entire contents of each of the above-referenced applications are expressly incorporated herein by reference in their entirety.

STATEMENT REGARDING FEDERALLY
SPONSORED RESEARCH OR DEVELOPMENT

[0002] This inventive concept(s) was made with government support under Grant Number 70NANB4H3048 awarded by the Advanced Technology Program of the National Institute of Standards and Technology. The government has certain rights in the inventive concept(s).

BACKGROUND

Below, we describe the search process we employed to identify US patent grants that disclose federal funding in statements of this kind.

3.1.1 Search process

We apply three methodologies to identify government funding disclosures, and treat the union of the three as our definition of interest:

1. Rassenfosse et al. (2019) keyword search

- (a) We replicate the keyword search in Rassenfosse et al. (2019), who constructed a sample of government interest statements in US patent grants for the years 2005 to 2015. In brief, the Rassenfosse et al. method parses the *Description* section of the patent grant, searching for a set of keywords. First, it searches for the following keywords in the heading of a subsection of the Description section: government, federal, research, sponsored. If none of these words is found, the script searches for the following phrases in the free text of the description section: government has, government may have, with support under, government owns rights, pursuant to contract, government support, performance of work under.

2. Supplementary keyword search

- (a) We supplement the search process employed by Rassenfosse et al. with a broader keyword search, intended to flag blocks of text that may include a government interest statement. Specifically, we searched for “rights to inventions”, “sponsor”, and “federal”; we exported statements that followed any line containing these keywords. As detailed below, to address concerns about false positives, we extensively hand-checked output generated through this process.

3. XML tag search

¹See: <https://www.uspto.gov/web/offices/pac/mpep/s310.html>.

- (a) Between 1974 and 2004, the USPTO included, in some cases, separate “tags” in XML files that denote the beginning of a government interest statement. For this set of years, we extract all text that follows these XML tags (labeled as <GOVINT> in XML files) and employ the hand-checking strategies described in the following section.

3.1.2 Hand checks

For each of the three search processes, at least one research assistant hand-checked all output. Specifically, we employed the following steps in order to create flags for “publicly funded” and “no funding disclosure.” We construct both flags, but note that “no funding disclosure” is the complement of “public.”

1. Standardize formatting of extracted text (i.e. remove irregular characters, punctuation, capitalization).
2. Search for various forms of “not applicable”² in extracted disclosure statements, and designate these as “no funding disclosure.” We hand-check the set of patents that were flagged in this process, in order to ensure that we do not erroneously miscategorize patents that include this language but also contain a disclosure of public funding.
3. Flag and hand-check “negations” (e.g. “This invention was not made with federal funds.”)³ Negation statements that do not disclose public funding are tagged as “no funding disclosure.”
4. Flag and hand-check statements that obviously contain disclosures of government funding. These include any statements that include references to US federal agencies (e.g. “US Department of Energy”) or sentences like “funded with government support.”⁴ Any patents that included these statements were tagged as “publicly funded.”
5. Flag statements that explicitly state that no federal funding was used as “no funding disclosure.”⁵ As with previous text searches, we hand-checked output at this stage and made any necessary corrections to this flag.
6. Finally, we hand-checked all observations that were not flagged as either “publicly funded” or “no funding disclosure” in the previous steps.
7. We confirm that for the full sample of observations, every patent is assigned to one and only one of the categories (“publicly funded” or “no funding disclosure”) and resolve discrepancies by hand.

Extensive hand-checking of all output decreases the likelihood of false positives. However, hand-checking patent texts to identify false negatives – a sample of more than 7 million patent grants – was untenable. We instead elected to hand-check a random 10% sample (n=350) of full patent texts for patents appearing in the FDA Orange Book, in order to confirm that our three search strategies do, in fact, pick up the funding disclosures of interest. Additionally, we examined the full texts of sixteen patents reported as the output of NIH-funded research, which do not appear to have government interest statements (see Section 4.1.1). Neither of these validation exercises identified additional government interest statements in patent full texts.

²These include: Not applicable, no applicable, not-applicable, n/a, n.a., na, notapplicable, non-applicable, nonapplicable, and non applicable.

³Negations that we flagged and hand-checked include: was not, not the result of, were not, is not, have not, created without, were not made, and does not include.

⁴The full list includes: made in part with government support, made with government support, made with us government support, government has a paidup license, funded in part with government support, funded with government support, united states may have certain rights, supported in part by nih, government has rights, government may have certain rights, government may have rights, government may have certain rights, by or for the government, by or for the us government, by or for the united states government, entitle the government, with the united states, with the us, government may own rights, government may own certain rights, government therefore has certain rights, government therefore may have certain rights, government may therefore have certain rights, government may retain certain rights, army, airforce, air force, navy, department of the interior, government contract, national institute of health, department of energy, department of defense, department of health, atomic energy commission, granted to the government of the united states of america, licensed by the government, national institutes of health.

⁵We search for the following explicit statements of no federal funding: no federal, no rights, no rights are given, no government, no united states government.

3.1.3 Government interest statement sample

We construct a sample of patents that acknowledge government funding by combining “publicly funded” flags generated through each search process. Specifically, we set a field “public” equal to 1 if at least one of the three searches returned a government interest statement disclosing public funding. All other observations are set to zero.

Summary statistics comparing these three samples and the final “public” sample are provided in Table 1.

Table 1: Comparison across samples (total count)

sample	total records	any public	rassenfosse	supplementary keywords	xml tags
full universe of US grants (1976-present)	7,089,234	142,704	127,719	35,647	59,964
Orange Book patents (1981 - 2014)	5,185	90	77	6	65

Notes: Column (1), “total records,” reports total counts of patents in, respectively, the full universe of US patent grants and the FDA’s Orange Book. Column (2), “any public,” reports total counts of patents acknowledging public funding, as identified by at least one of the three search methods used. Columns (3)-(5), “rassenfosse,” “supplementary keywords,” and “xml tags,” list total patent counts for each of the three search processes. Note that these counts are not mutually exclusive.

3.2 Agency patents

We construct a sample of patents assigned to government agencies, following Sampat and Lichtenberg (2011). It seems likely that essentially all patents assigned to government agencies were constructed with public funding. We do note, however, as discussed in Kapczynski (2017), that many government laboratories and agencies elect not to hold patents. For those agencies that do hold patents, however, we treat all patents as publicly funded.

To identify this sample, we begin with this list of federal agencies: <https://www.usa.gov/federal-agencies/a#current-letter>. We restricted the universe of patent grants to assignees with addresses in the United States. We flagged all assignees whose names match the downloaded list of agencies and checked this sample for false positives. We also check for misspellings of “national institutes of health,” “national aeronautics and space administration,” and “united states of america” and repeat this search.

Next, we flagged all assignees whose names included: “united states of america represented by” or “united states.” Within this sample, we again searched for false positives.

Next, we used a list of government agency acronyms and flagged assignee strings that contain the acronym either at the beginning or the end of the string or that contain the acronym surrounded by spaces. We hand-checked this sample for false positives.

Additionally, we flagged assignee strings containing “national labs,” “us department of,” or “centers for disease control” and checked the resulting sample for false positives.

Finally, we flag all assignees who have “DC” as their listed state and hand-check this list for false negatives.

3.3 Patent certificates of correction

We reviewed all correction images published by the USPTO for patents in our Orange Book sample.

We downloaded the USPTO’s listing of all patent corrections, which includes patent number and date of correction, current through December 2020. 1,975 patents in our sample appear on this list of corrections.

We reviewed images of correction files published by the USPTO for all 1,975 patents. A random sample of 25% was separately checked by a second person on our team. We identify 19 Orange Book patents with certificates of correction that add government interest statements. We also identify 7 patents with correction statements that amend government interest statements, in order to provide more detail. This set of 19 corrections includes corrections that resolve discrepancies between NIH RePORTER and USPTO disclosures.

In Table 2, we list the 19 patent grants with corrections that add government interest statements, alongside the text of the new statement.

Table 2: Patent grants with corrections that add government interest statements

Patent number	Corrected statement	Correction text
6455499	new disclosure	This invention was made with government support under grant HD33172 awarded by the National Institutes of Health. The government has certain rights in the invention.
4619939	new disclosure	Research leading to the conception and reduction to practice of the invention claimed herein was supported in part by Grant No. 5-R01-02227-03 issued by the National Institutes of Health. The United States Government has certain rights in and to the claimed invention.
4673563	new disclosure	This invention was made with United States Government support under Grant No. HL 10384, awarded by the National Institutes of Health. The United States Government has certain rights in the invention.
4971998	new disclosure	This invention was made with Government support under Grant Number NIH-5M01-RR00088 awarded by the National Institutes of Health. The Government has certain rights in the invention.
5134127	new disclosure	This invention was made with government support under N01- CM-67912 and N01-CM97546 awarded by the National Institutes of Health. The government has certain rights in the invention.
5376645	new disclosure	This invention was made with government support under N01- CM-67912 and N01-CM97546 awarded by the National Institutes of Health. The government has certain rights in the invention.
5407914	new disclosure	This invention was made with government support under Grant Nos. HL 23584 and GM 37696 from the National Institutes of Health and Grant No. N00014-89-K-0029 from the Office of Naval Research. The U.S. government may have certain rights in the invention.
5563175	new disclosure	Research in this application was supported in part by a grant from the National Institute of Health (NIH Grant No. NS15703). The Government has certain rights in the invention.
5614560	new disclosure	This invention was made with government support under Grant No. R01 EY09024 by the NIH. The government has certain rights in the invention.

6028071	new disclosure	This invention was supported by NIH grant number CA56517. The US government has certain rights in this invention.
6197819	new disclosure	Research in this application was supported in part by a grant from the National Institute of Health (NIH Grant No. NS15703). The Government has certain rights in the invention.
6365127	new disclosure	The invention herein was made in the course of work under Grant No. R01-CA-23185 from the National Institute of Health. The United States Government has certain rights in this invention.
6423686	new disclosure	This invention was made with government support under Grant Number HD33172 awarded by the National Institutes of Health. The government has certain rights in the invention.
6692763	new disclosure	This invention was made with government support under Grant No. AG05407 awarded by the National Institutes of Health. The government has certain rights in the invention.
6958335	new disclosure	This invention was made with government support under CA065823 awarded by the National Institutes of Health. The government has certain rights in the invention.
7270800	new disclosure	This invention was made with United States government support under Grant Nos. AG001039 and AG018402, awarded by the National Institutes of Health. The United States government has certain rights in the invention.
8236282	new disclosure	This invention was made with United States government support under Grant Nos. AG001039 and AG018402, awarded by the National Institutes of Health. The United States government has certain rights in the invention.
8299078	new disclosure	This invention was made with government support under grant number CA092074 awarded by National Institutes of Health. The government has certain rights in the invention.

For seven additional Orange Book patents, we identified certificates of correction that amended existing government interest statements. These include patents numbered: 6890898, 7078381, 7459428, 7807135, 7851509, 8101663, 8506929.

3.4 Patent continuations

Legal and regulatory sources suggest that disclosures of public funding in “parent” patents apply to derived continuation (“child”) patents, which are separate patents based on the same disclosure and priority date as their parent. In order to incorporate these patents into our measure of public funding, discussed in detail in the primary text, we used patent continuation histories reported in the USPTO Public PAIR database: <https://www.uspto.gov/learning-and-resources/electronic-data-products/patent-examination-research-dataset-public-pair>.

We identify 22 patents in our sample of Orange Book patents that are linked to a government interest statement through disclosure in a parent application.

4 Publicly funded patents: NIH RePORTER

The National Institutes of Health ExPORTER provides bulk data from NIH RePORTER records and includes project details, project abstracts (separated due to size considerations), publications citing support from NIH projects, patents citing support from NIH grants, and clinical trials citing support from NIH grants: <https://exporter.nih.gov/>. This dataset includes projects funded by the NIH, Administration for Children and Families, Agency for Healthcare Research and Quality, Centers for Disease Control and Prevention, Health Resources and Services Administration, the US Food and Drug Administration, and Veterans Affairs. This data does not include output that is the result of work conducted at government labs. Coverage for NIH grants begins in 1985, and most files are updated weekly.⁶ Coverage for other federal agencies varies.

4.1 NIH patents with government interest statements

There are 44 Orange Book patents that have both disclosures of public funding in parsed government interest statements and grant numbers listed in NIH RePORTER. Parsed USPTO grant numbers are a strict subset of the grant numbers reported in NIH data in all 44 cases, based on a review of both sets of records. To illustrate, Table 3 compares grant numbers reported in NIH RePORTER and grants reported in USPTO full text files, within government interest statements, for 10 observations.

⁶Project publications and links are updated annually. All other files are updated weekly.

Table 3: NIH grant information: NIH RePORTER versus parsed text

Patent number	NIH grant number(s)	Parsed grant information from USPTO files
05149794	R01GM031459 / R01CA024872	NIH: CA 24872 and GM 31459
08841422	R01CA106504	NIH: CA106504
05977322	U01CA051880	Army Grant No. DAMD17-94-J-4433; NIH Grant No. U01 CA51880.
06529752	R44HL062077; R43HL062077	National Heart, Lung and Blood Institute 2R43HL62077-02
07335679	U19CA052995; U01CA052995	NIH/NIG Grant No. U01 CA 052995.
05691460	R01AR040520; R29AR036546	NIH Grants AR-36546 and AR-40520.
05869335	R01AR037909	Minnesota Agricultural Experiment Station (project 377-3457), the Oklahoma Agricultural Experiment Station (projects 1669 and 1433), and grants from the USDA/NRI Competitive Grants Program, No. 94-37204-0449, 91-37204-6360, and Grant No. AR37909 from the National Institutes of Health
05674281	R01HL013426	N01-HV-88105 with the National Heart, Lung and Blood Institute and Grant No. R01-HL-13426 with the National Heart, Lung and Blood Institute
10279011	R01AI101406 R01AI080455 R01CA107096 R01HL095075 R01HL069929	HL69929, AI101406, CA107096, AI080455 and HL095075 awarded by the National Institutes of Health and W81XWH-09-1-0294 awarded by the Army Medical Research and Materiel Command
08753815	R01AA012964 R01AA010522	Grant Nos. AA010522-12, AA0032903, AA001016 and AA012964 awarded by the National Institutes of Health
05866325	P01CA040046	CA 40046 from The National Cancer Institute (e.g., the National Institutes of Health)

4.1.1 NIH patents without government interest statements

There are 5,650 patents in the full universe of USPTO granted patents that are flagged as NIH funded by the RePORTER data, but not as publicly funded (i.e. we found no government disclosure statement).

We identify sixteen patents in our Orange Book sample which are reported in the NIH RePORTER data, but which do not directly include a government-interest statement. All sixteen report government funding through either a certificate of correction, a parent patent, or both. Specifically, twelve had certificates of correction that alter the text of the original patent to include a disclosure of NIH funding, and six are continuations of parent patents that include government-interest statements. Two of the sixteen patents fall in both categories. We reviewed full texts of each of these sixteen patents to confirm that our government interest statement search process did not miss disclosures of funding; full text reviews confirmed that none of these patents included funding statements. Patent numbers are listed below:

Table 4: NIH patents without government interest statements

Patent number	Public parent	Certificate of correction	Resolved?
4619939	no	yes	yes
4673563	no	yes	yes
4971998	no	yes	yes
5110806	yes	no	yes
5407914	yes	no	yes
5563175	no	yes	yes
5614560	no	yes	yes
6028071	no	yes	yes
6365127	yes	yes	yes
6423868	yes	no	yes
6455499	no	yes	yes
6692763	yes	no	yes
6958335	yes	yes	yes
7270800	no	yes	yes
8236282	no	yes	yes
8691185	no	yes	yes

Notes: Full texts and USPTO image files were examined for each of the sixteen patents listed above.