Orange Book Data Including Public-Sector Patent Information, 1985-2023 [update 2.0]

September 2024

Contents

1 Description **1**

2 Overview **2**

2.1 All Orange Book Patents, 1985-2023 2

2.1.1 Basic Orange Book Patent Data 2

2.1.2 Identifying Public-Sector Patents 2

2.1.3 Codebook 3

2.2 Patents for NMEs Approved from 1985 to 2022 4

2.2.1 The FDA NME List and Associated Patents 4

2.2.2 Codebook 4

3 Acknowledgements **5**

4 References **5**

1 Description

This repository provides data from the FDA's Orange Book on patents associated with marketed drugs, with additional information on whether each patent is a "public-sector" patent and "Bayh-Dole" patent. The dataset accompanies the following article:

- Lisa Larrimore Ouellette and Bhaven N. Sampat. 2024. "The Feasibility of Using Bayh-Dole March-In Rights to Lower U.S. Drug Prices."

Public-sector patents are defined as those with a government-interest statement, listed in NIH RePORTER or in the Government Patent Register, and/or with U.S. government assignees. Bayh-Dole patents are all public sector patents not assigned to a government agency.

2 Overview

2.1 All Orange Book Patents, 1985-2023

2.1.1 Basic Orange Book Patent Data

The first file in the repository (nda_govt_patents.dta) includes all Orange Book patent listings from 1985 to 2023. It starts with the NBER Orange Book dataset (`https://www.nber.org/research/data/orange-book-patent-and-exclusivity-data-1985-2016`), which includes patents from the 1985 to 2016 editions of the Orange Book, described by Durvasala et al. (2023). It includes all unique new drug application (NDA)-patent dyads from this data set, and appends all additional unique NDA-patent listings from subsequent editions of the Electronic Orange Book: the 11/2016 edition, 12/2017, 12/2018, 11/2019, 9/2020, 8/2021, 10/2022, 10/2023. We thank Maya Durvasula for discussions on updating the data; see Durvasula et al. (2023) Appendix A for more information on the process for doing so.

2.1.2 Identifying Public-Sector Patents

As described in Ouellette and Sampat (2024), we determined whether each of these patents was a public-sector patent based on information in the following sources:

1. USPTO PatentsView Government Interest file
    - Link: `https://patentsview.org/download/data-download-tables`
    - Filename: g_gov_interest.tsv
2. NIH RePORTER Patents File
    - Link: `https://reporter.nih.gov/exporter/patents`
    - Filename: Patents.csv
3. USPTO PatentsView Disambiguated Assignee File
    - Link: `https://patentsview.org/download/data-download-tables`
    - Filename: g_assignee_disambiguated.tsv

    Note: We manually reviewed all disambiguated assignees for Orange Book patents to determine whether they were U.S. government agencies.
4. The historical and modern Government Patent Register
    - Link: `https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BDCCN9`
    - Filename: Government_Patent_Register.dta

    Note: We focused here on Orange Book patents where either modern_register==1 or historical_register==1. See Gross and Sampat (2024).
5. USPTO Certificates of Correction

- Link: https://www.uspto.gov/patents/search/authority-files/certificates-correction
- Filename: EST-USPAT_AUTHORITY-COC-authorityFile_17000101-20231125.txt (last accessed December 2023; the filename changes with each update)

Note: We identified Orange Book patents with Certificates of Correction through the Correction Authority Listing above. The Authority file lists all patents with corrections, but not the substance of the correction. To do so, we needed to review the image of the patent files at (https://ppubs.uspto.gov/pubwebapp/static/pages/ppubsbasic.html. *We downloaded the PDFs of all Orange Book patents with Certificates of Correction, isolated the Certificates, converted the Certificates to searchable text using optical character recognition, and searched them for the term "government" "grant" "contract" "NIH" and others to identify additional patents with government interest statements. We reviewed all patents with these keywords manually to guard against false positives.*

Manual review of all "university" patents: We also manually reviewed the image PDFs of any patent where the disambiguated assignee name (https://patentsview.org/download/data-download-tables) corresponded to a university or academic institution, to capture any residual patents with government-interests not identified through the previous approaches. We also manually reviewed the image PDFs of a small number of pre-1976 patents on the Orange Book for government-interest statements and/or government assignment, since PatentsView dataset begins in 1976.

For each new drug application (NDA) in the Orange Book dataset, we determined the active ingredient name and proprietary name for each drug using information from the 11/28/2023 version of the FDA's Drugs@FDA database:

- Link: https://www.fda.gov/drugs/drug-approvals-and-databases/drugsfda-data-files.
- Filename: Products.txt

Note: In cases where an NDA is associated with multiple names, we focus on the one linked to the first product number for that NDA. About 1.5 percent of the NDA numbers in the historical Orange Book dataset (and 1.1 percent of NDA-patent dyads) do not have drug names in the version of the Drugs@FDA database we used.

The final dataset is at the NDA-patent level (with variable names applno and patent_id, respectively). There are 13,992 observations, corresponding to 2832 distinct NDAs and 9860 distinct patents.

2.1.3 Codebook

```
. codebook, compact

Variable        Obs Unique      Mean   Min     Max  Label
-------------------------------------------------------------------------------
applno        13992   2832  113714.5  4782  217759  FDA NDA Number
proprietary   13832   2377         .     .       .  Proprietary Name
activeingr~t  13827   1633         .     .       .  Active Ingredient
patent_id     13992   9860         .     .       .  USPTO Patent Number
```

```
gi              13992      2   .0246569     0       1   Any GI Statement (1=yes)
coc             13992      2    .008076     0       1   Any GI Statement in C...
reporter        13992      2     .01508     0       1   In NIH RePORTER (1=yes)
govtass         13992      2   .0037164     0       1   Government Assigned (...
register        13992      2   .0126501     0       1   In Government Patent ...
public          13992      2   .0312321     0       1   Public-Sector Patent ...
bd              13992      2   .0274443     0       1   Bayh-Dole Patent (1=yes)
-------------------------------------------------------------------------------

. log close
```

2.2 Patents for NMEs Approved from 1985 to 2022

2.2.1 The FDA NME List and Associated Patents

The second data file, nme_govt_patents.dta, includes a subset of the file above, all Orange Book patent listings (1985-2023) for new molecular entities (NMEs) approved by the FDA between 1985 and 2022. We started with an FDA "Compilation of CDER New Molecular Entity (NME) Drug and New Biologic Approvals:

- Link: `https://www.fda.gov/drugs/drug-approvals-and-databases/compilation-cder-new-molecular-entity-nme-drug-and-new-biologic-approvals`
- Filename: Compilation-of-CDER-NME-and-New-Biologic-Approvals-1985-2022.xlsx

We focus on NMEs that are NDAs, dropping biologic license applications (BLAs) that do not have Orange Book listing requirements. We keep for each drug listing the NDA number (the first one, if there are multiple), the indicator for review designation, the active ingredient name, proprietary name, and FDA approval year. We merged in the dataset above to link these NMEs to patents, and determine which patents are public-sector patents. This dataset includes 4,748 observations, corresponding to 883 unique NME NDAs and 4,497 distinct patents.

2.2.2 Codebook

```
. codebook, compact

Variable       Obs Unique      Mean    Min     Max  Label
-------------------------------------------------------------------------------
applno        4748    883  120749.5  18217  216986  FDA NDA Number
approvalyear  4748     38  2008.888   1985    2022  FDA Approval Year
activeingr~y  4748    883         .      .       .  Active Ingredient/Moiety
proprietar~e  4748    880         .      .       .  Proprietary Name
prio          4748      2  .5037911      0       1  Priority Review (1=yes)
applicant     4748    437         .      .       .  Applicant
patent_id     4748   4497         .      .       .  USPTO Patent Number
govtass       4748      2  .0050548      0       1  Government Assigned (...
reporter      4748      2  .0227464      0       1  In NIH RePORTER (1=yes)
gi            4748      2   .033909      0       1  Any GI Statement (1=yes)
```

```
coc             4748      2   .0117944      0       1   Any GI Statement in C...
register        4748      2   .0208509      0       1   In Government Patent ...
public          4748      2   .0438079      0       1   Public-Sector Patent ...
bd              4748      2   .0385425      0       1   Bayh-Dole Patent (1=yes)
-------------------------------------------------------------------------------

. log close
```

3 Acknowledgements

Sampat acknowledges a grant from the National Institute of Healthcare Management (NIHCM) for a project "Can Improving Pharmaceutical Patent Quality Promote Competition and Reduce Drug Prices?" which supported the data collection and cleaning effort.

4 References

- Durvasula, M., et al. (2023). The NBER Orange Book dataset: A user's guide. *Research Policy*, 52(7), 104791.
- Gross, D. P., & Sampat, B. N. (2024). The Government Patent Register: A new resource for measuring US government-funded patenting. NBER Working Paper (w32136).
- Ouellette, L.L., & Sampat, B. N. (2024). The feasibility of using Bayh-Dole march-in rights to lower U.S. drug prices.