Archival Orange Book Patent Data

C. Scott Hemphill
Bhaven N. Sampat
Columbia University*

September 29, 2013

1 Background

The file `OBPATENTS8512` includes information on all patent listings from current and archival versions of the Food and Drug Administration's (FDA) "Orange Book." The Orange Book, officially named *Approved Drug Products with Therapeutic Equivalence Evaluations* lists FDA approved branded drugs and information about "therapeutically equivalent" generic drugs.[1] Patent information was added to the Orange Book starting with the 1985 edition, after the passage of the 1984 Hatch-Waxman Act.

The current version of the Orange Book is searchable on the FDA's website, and also downloadable in ASCII format. This is commonly called the "Electronic" Orange Book, or EOB. Until 2001, the FDA also published paper copies of the publication.

For longitudinal analyses of drugs and patenting, the EOB is limited since it does not list expired patents, or those for discontinued drugs that are no

*We thank the Robert Wood Johnson Foundation's Public Health Law Research program for helping support the development of this dataset, and Ravi Bhalla for research assistance.

[1]The "Orange Book" nickname refers to the initial choice of orange for the cover of the book's first edition, published in October 1980, shortly before Halloween.

longer marketed. We used data from older versions to develop a comprehensive listing of all patent listings from Orange Books published between 1985 and 2012, the drugs associated with these patents, and the patent expiration dates.

2 Codebook (Stata format)

The final file includes 1867 unique drugs and 3830 unique patents:

```
.
. codebook

-------------------------------------------------------------------------------
nda                                                                 NDA number
-------------------------------------------------------------------------------

                  type:  numeric (long)
                 range:  [4782,203388]                units:  1
         unique values:  1867                     missing .:  0/5799
                  mean:   27910.4
             std. dev:   33600.9
           percentiles:        10%       25%       50%       75%       90%
                             18936     20347     21086     21789     22352

-------------------------------------------------------------------------------
patent                                                           Patent number
-------------------------------------------------------------------------------

                  type:  string (str8), but longest is str7
         unique values:  3830                     missing "":  0/5799
              examples:  "4689320"
                         "5496545"
                         "6048901"
                         "6938796"

-------------------------------------------------------------------------------
expdate                                                     Expiration date
-------------------------------------------------------------------------------

                  type:  numeric daily date (float)
                 range:  [9258,25932]                 units:  1
       or equivalently:  [07may1985,31dec2030]        units:  days
         unique values:  2484                     missing .:  0/5799
                  mean:   19043.9 = 20feb2012 (+ 22 hours)
             std. dev:   3470.29
           percentiles:        10%       25%       50%       75%       90%
                             13487     17388     19789     21396     22849
                          04dec1996 10aug2007 07mar2014 31jul2018 23jul2022

.
```

3 Data sources

We began with a file obtained from the FDA through a Freedom of Information Act (FOIA) request (completed in 2004) which listed the new drug application (NDA) number, patent number, and expiration date for all Orange Book patents listed between 1985 and 2002 that had expired, and thus were not listed beginning in 2003. We supplemented this information with the PATENT.TXT file from annual EOB files downloaded each year between 2000 and 2012.[2] The 2000 to 2002 EOB files are largely redundant to the FOIA file. Of the NDA-patent dyads in these files expiring before 2003, 97 percent are also in the FOIA file. Nonetheless, we include the data from these files since they do add a few patents.

Together, the final file contain all expired and unexpired patents from FDA Orange Books published since these publications began listing patent information.[3]

4 Data processing

In the annual EOB patent files, each row includes a New Drug Application (NDA) number, product, patent number, and patent expiration date.[4]

An NDA is typically an active ingredient and dosage form. Each NDA can have multiple products (for multiple strengths). Under the 1984 Hatch-

[2]The EOB is updated monthly. We obtained versions of the EOB as late in a calendar year as possible. In practice the specific month does not matter much since information is typically repeated across multiple years.

[3]The paper versions of the Orange Book and the EOB files also contain information on therapeutic equivalence between drug products and non-patent exclusivity awarded to firms. Our focus here is on the patent information.

[4]The annual EOB patent files also list a drug substance flag (indicating that the sponsor submitted the patent as claiming the drug substance) and a drug product flag (indicating that the sponsor submitted the patent as claiming the drug product). Beginning in 2007, a patent delist flag is also included (indicating if the sponsor requested the patent be delisted), and beginning in 2009 a variable for application type is included (whether the application is an NDA or, on very rare occasions, an Abbreviated New Drug Application, or ANDA, filed by a generic drug maker). The patent file obtained by FOIA for pre-2003 listings contains information on each of these fields except for the patent delist flag and application type.

Waxman Act, branded firms are required to provide patent information for each NDA and product to the FDA.[5] Thus each product on an NDA can have one or more patents, and patent information can repeat within NDAs across products.

Each row also lists an expiration date for a patent. Patent terms are complex, and can depend on the interaction of issue dates, filing dates, and statutory patent term extensions[2, 3]. The FDA requires drug companies to list the final expiration date for each patent after taking account of these complexities.[6]

Some NDAs and products that meet certain testing requirements in children, also earn six months of pediatric exclusivity on their patents. In these cases, the row for the NDA/product is repeated with "*PED" following the patent number, reporting the expiration date as amended.

Here is a sample excerpt from the 1999 (paper) Orange Book:

APPL/PROD NUMBER		INGREDIENT NAME; TRADE NAME	PATENT NUMBER	PATENT EXPIRES
020977	001	ABACAVIR SULFATE;ZIAGEN	5089500	JUN 26, 2009
			5089500*PED	DEC 26, 2009
			5034394	JUN 26, 2009
			5034394*PED	DEC 26, 2009
020978	001	ABACAVIR SULFATE;ZIAGEN	5089500	JUN 26, 2009
			5089500*PED	DEC 26, 2009
			5034394	JUN 26, 2009
			5034394*PED	DEC 26, 2009
020482	001	ACARBOSE;PRECOSE	4904769	FEB 27, 2007
020482	002	ACARBOSE;PRECOSE	4904769	FEB 27, 2007

[5]For patents issued before branded drug approval, the drug maker is required to list any patent containing at least one claim that covers the drug's active ingredient, its formulation, or any method of use pertaining to an approved indication. For patents issued after drug approval, listing is not required, but there is a strong incentive to list. The incentive comes from the generic firm's obligation, when it files its ANDA, to challenge every listed patent or else wait until patent expiration before receiving FDA approval. All brand-name patents listed before ANDA filing are subject to this obligation. Moreover, listing provides an additional advantage in litigation. When the brand-name firm files a suit on a timely listed patent, the generic firm is subject to an automatic stay of FDA approval for up to 30 months, while the patent suit is considered by the district court [2, 3]. Not all patents are listed in the Orange Book, however. Methods of manufacturing the drug, formulations that do not cover the marketed drug product, and methods of use covering unapproved indications, are all barred. Moreover, for patents issued after NDA approval, listing is not required, and despite the incentives discussed above, not all eligible patents are actually listed[1].

[6]Note that the same patent can have different expiration dates for different NDAs (for example, if a patent were granted term extension or pediatric exclusivity for one NDA for which is it listed, but not another.)

5 Data Processing

After appending the Orange Book files, there were 86,887 rows of data. We dropped a small number of rows (10) where expiration dates were later than 2050, and hence apparent mistakes.

We aggregate to the NDA-patent level, listing all patents for each NDA, and the expiration dates for each patent. Each of the 5799 rows in the final dataset is thus a unique NDA-patent dyad. We report the expiration date inclusive of pediatric exclusivity for each NDA-patent (where this is granted), since this is generally the patent term relevant for potential generic entrants.[7] Where there are multiple expiration dates for a NDA-patent pair across Orange Books (e.g. because pediatric exclusivity is granted, or patent terms adjusted for other reasons) we report the one from the most recent version of the Orange Book available.

6 Validation against paper Orange Books

To verify completeness of the dataset compiled from the EOBs, we also compared this information to patent information from paper copies of the Orange Book between published between 1985 and 2000. The patent sections of the paper Orange Books were transcribed to ASCII files through double-key data entry. Since the paper Orange Book occassionally has printing errors (generally patent or NDA numbers off by a digit) that can be repeated across years, we focus on the NDA-patent dyads from the paper Orange Books that were included in at least two printed editions over this period. Of the 1377 NDA-patent dyads in the 1985-2000 paper Orange Books, only 60 (about 4 percent) were not found in the dataset constructed through the EOBs.[8]

[7] An exception to this expectation might arise if a pediatric extension is awarded unexpectedly after a generic firm decides to attempt market entry.

[8] These patents are available from the authors on request.

References

[1] Tahir Amin and Aaron S Kesselheim. Secondary patenting of branded pharmaceuticals: a case study of how patents on two HIV drugs could be extended for decades. *Health Affairs*, 31(10):2286--2294, 2012.

[2] C Scott Hemphill and Bhaven N Sampat. When do generics challenge drug patents? *Journal of Empirical Legal Studies*, 8(4):613--649, 2011.

[3] C Scott Hemphill and Bhaven N Sampat. Evergreening, patent challenges, and effective market life in pharmaceuticals. *Journal of Health Economics*, 31(2):327--339, 2012.