Patents, Innovation, and Competition in Pharmaceuticals:

The Hatch-Waxman Act After Forty Years

Online Appendix

Appendix A: Data Construction and Analysis

Drugs

Stata do-file: make_fdalist.do

Our main dataset comprises new molecular entities (NMEs) approved by the Food and Drug Administration (FDA) between 1985 and 2020 that had at least one patent listed in the Orange Book over the period from 1985 to 2023.

The starting point for constructing this list is the FDA's Compilation of CDER New Molecular Entity (NME) and New Biologic Approvals (FDA 2023a). This list contains 1,236 drugs with an approval date (*approvaldate_b*) between 1985 and 2022. We drop 234 biologic drugs approved via biologic license applications (BLAs), which are not governed by Hatch-Waxman, leaving 1,002 drugs. We also drop 58 drugs approved outside of the 1985 to 2020 approval window, leaving 944 drugs.

Each drug has a unique active ingredient or combination of active ingredients[1] and is associated with a unique new drug application (NDA).[2] We use the unique number assigned by the FDA to each NDA (*applno*) to merge in patents and other data. In

[1] For three proprietary names (Akynzeo, Emend, and Trovan), there are two observations apiece. Each NDA in the pair has a unique set of active ingredients (one is a prodrug of the other) and a distinct New Drug Application.

[2] For 55 drugs, the FDA's list of NMEs reports more than one NDA. For these, we keep the first NDA listed. In nearly all cases, this is the numerically lowest NDA, and the NDA designated by FDA as an NME (as opposed to line extensions, such as new dosage forms or methods of administration).

addition to this variable, we keep the drug's trade or proprietary name (*proprietaryname*), its active ingredient(s) (*activeingredient*), FDA receipt date (*fdareceiptdate*), and FDA approval date (*approvaldate_b).* We use *approvaldate_b* to construct the FDA approval year for the brand drug (*approvalyear_b*) and the corresponding three-year FDA approval cohort (*cohort_b:* 1985-1987; 1988-1990; 1991-1993; 1994-1996; 1997-1999; 2000-2002; 2003-2005; 2006-2008; 2009-2011; 2012-2014; 2015-2017; 2018-2020).

Patents

Stata do-file: make_obpatents.do

For each drug, we merge in patents listed on the Orange Book using the drug's *applno*. For archival Orange Books published between 1985 and 2016, we use the NBER Orange Book dataset (NBER 2023) described in Durvasula et al. (2023). For Orange Books published between 2016 and 2023, we use annual electronic Orange Books obtained via the Internet Archive's Wayback Machine (FDA 2023b). The last Orange Book edition used is October 2023. From the Orange Books, we collect the number and expiration date of each patent. 118 drugs have no listed patents and are dropped.

The final dataset of 826 drugs, which we refer to below as the "NME List," is associated with 4,217 unique patents and 4,446 drug-patent pairs.

Patents Per Drug

Stata do-files: make_ppd.do

We calculate patents per drug (*ppd*) by counting the total number of unique patents ever listed for the drug, across all editions of the Orange Book. One complication for this calculation is reissued patents. A reissue is a correction to a predecessor patent,

requiring the surrender of the original patent. Counting both the reissue and its predecessor would result in double counting. Of the 4,217 unique patents in our data, 103 are reissues.

To address this issue, we match each reissue to its predecessor using the USPTO PEDS database (USPTO 2023a) searching for the text following "This application is a re-issue of" for each of the reissues in our data. In cases where the predecessor is also listed for the same drug, we only count the patent once for purposes of calculating *ppd*.

The variable *ppd* is potentially censored, since older drugs have more time to accumulate patents. Accordingly, we construct a second measure, *ppd_early*, which omits patents that first appear on the Orange Book four years or more after drug approval. For example, if a drug is approved in 1990, *ppd_early* counts patents listed in or before 1993.

Nominal Patent Term

Stata do-files: make_nominal.do

To calculate the expiration date for each patent, we take the latest expiration date listed across all Orange Book editions. We make this calculation at the NDA-patent pair level. For each drug, we calculate nominal patent term (*nominal*) as the time from drug approval to expiration of the last expiring patent. We also calculate *nominal_early* in a manner analogous to *ppd_early*, as nominal patent term omitting patents that first appear on the Orange Book four years or more after drug approval.

New Chemical Entity Exclusivity

Stata do-file: make_nces.do

We also determine which of the drugs on the NME List receive new chemical entity

(NCE) exclusivity. As discussed in the text, NCEs receive five years of exclusivity (even if the drug has no patent protection) under Hatch-Waxman, and patent challenges are not possible until four years after FDA approval. To identify them, we use the *Exclusivity* files from the FDA Orange Book. For archival Orange Books published between 1985 and 2016, we use the NBER Orange Book dataset (NBER 2023) described in Durvasula et al. (2023). For Orange Books published between 2016 and 2023, we use annual electronic Orange Books obtained via the Internet Archive's Wayback Machine (FDA 2023b).[3] The last Orange Book edition used is October 2023.

Nearly all NMEs are NCEs. The exceptions are drugs with a previously approved active moiety in a different form, and combination products with both a novel ingredient and a previously approved active ingredient. Out of 826 drugs on the NME List, 803 have NCE exclusivity. Of these, one (Kaletra, a combination of lopinavir and ritonavir) did not have full exclusivity, but rather inherited NCE exclusivity from a previous approval of one of its components (ritonavir). We create a dummy variable (*nce*) that flags NCE exclusivity, that takes on the value of 1 for the 802 drugs with full NCE exclusivity.

Generic Approval and Effective Market Life

Stata do-files: make_fdalist.do; make_eml.do; make_drugsatdfa.do.

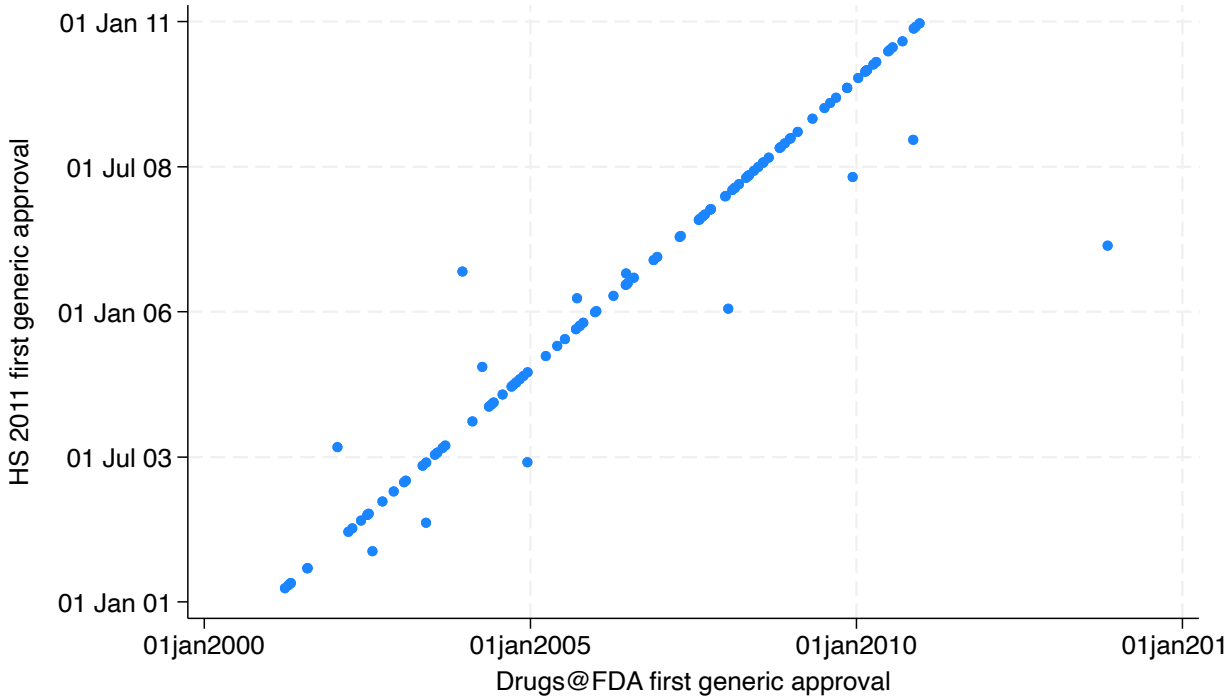
To assess generic entry for each drug, we use information from another FDA database, Drugs@FDA (accessed September 2023), to determine whether there is an approved generic and the timing of its approval (FDA 2023d). Using the *Applications*, *Products*, and *Submissions* files from this database, we collect all approvals of Abbreviated New Drug Applications (ANDAs), the associated active ingredients, and approval dates recorded

[3] As with patents, we include the electronic Orange Book for 2016 for redundancy with the last year of the NBER dataset. Any duplicate observations are dropped in subsequent processing steps.

by the FDA.

We merge this information with the NME List using *activeingredient*. For 341 drugs (out of 826), there is a matching first generic approval between 2000 and 2023 (*any_generic*). We also record the date of first generic approval (*first_g_approvaldate*) and its year (*first_g_approvalyear*). In addition, we construct a variable for first generic approval cohort (*cohort_g*: 2000-2002; 2003-2005; 2006-2008; 2009-2011; 2012-2014; 2015-2017; 2018-2020; 2021-2023).

We calculate effective market life (*effective*) as the time from drug approval to first generic approval. In a previous paper (Hemphill and Sampat 2011), we determined the timing of generic approval, using information from the FDA's first-time generic approval lists from 2001 to 2010. Of the 119 drugs in that data set, 112 appear in the new list constructed from Drugs@FDA. For these drugs, the two measures have a correlation of 0.97. Figure A.1 shows a scatter plot of entry dates, with each marker representing a drug, showing they match exactly for most drugs:



In this paper, as in the previous work, we use FDA records of first generic approval, rather than other evidence of actual generic launch, to measure effective market life. Hemphill and Sampat (2011) also compiled data (from Cardinal Health and other sources) on first generic launch date, showing the main results were robust to using this alternative measure of effective market life. While the Cardinal database is no longer publicly available, several recent papers estimate the launch date using the timing of first purchase by Medicaid.[4] In ongoing work we are extending our analyses to consider this measure as well.

[4] See, for example, Dave et al. (2020).

Patent Challenges

Stata do-files: make_p4list.do

To determine the existence and timing of a pre-expiration "Paragraph IV" challenge, we rely on a list of ANDAs containing Paragraph IV certifications that is maintained by the FDA (FDA 2023c). For each drug, we determine whether there has been a challenge (*p4*) by matching *applno* to the Paragraph IV list. A few Paragraph IV challenges on the FDA's list are missing NDA numbers; we correct these entries manually before merging.

Beginning in 2004, the FDA's list reports the date on which the first ANDA with a Paragraph IV certification was submitted. We record this date where present (*p4andadate*). We use this and *approvaldate_b* to construct the lag to first Paragraph IV submission (*p4lag*).

Patent Extensions

Stata do-files: make_extensions.do

A firm marketing a new drug is entitled to choose one patent for extension. The period of extension is, roughly speaking, the complete length of the FDA's NDA review period, plus 50 percent of the testing phase. Specifically, the extension is calculated as:

$$\text{Period of Extension} = (\text{RRP} - \text{PGRRP}) - \text{DD} - \tfrac{1}{2}(\text{TP} - \text{PGTP})$$

where:

- RRP: Total number of days in the regulatory review period.
- PGRRP: Number of days of the RRP on or before the patent issuance date.

- DD: Number of days during the RRP when the applicant did not act with due diligence.

- TP: Testing phase (time from IND submission to NDA submission).

- PGTP: Number of days of the TP on or before the patent issuance date

As the formula reflects, in cases where the patent issues after the regulatory review period begins, only the portion of that period after issue is counted, and similarly for the testing phase. The patent must be in force on the date of FDA approval, so post-approval patents added to the Orange Book cannot be extended. Any time lost to lack of due diligence by the applicant is also subtracted.

The extension is also subject to several caps. A patent can be extended for a maximum of five years, which is replaced by a two-year maximum for "pipeline" drugs that were in trials when Hatch-Waxman was enacted but not yet approved. There is also an upper bound, regardless of the length of the extension: the final term of the patent cannot exceed 14 years from approval.

For each patent we determine whether the patent was extended and for how long using a list of extensions maintained by the USPTO (USPTO 2023b) which we refer to as the "PTE List." The PTE List reports the tradename, patent number, original expiration date, and period of extension granted. After standardizing patent numbers (for example, removing spaces and commas) and correcting other typos, we merge each NME List patent to the PTE List patents. We manually review the matches and drop a small number of patent extensions where the basis for extension is a non-NME List drug. 71 percent (571/826) of NME List drugs have an extended patent (*any_extension=1*). For the these drugs we record the extended patent (*patent_id*), the

original expiration date from the PTE List (*expdate_orig*), the period of extension (*periodofextensiongranted*). We create dummy variables for whether the period of extension is 2 years or 5 years (*capped2* and *capped5*), and a dummy variable (*pipeline)* indicating the drug is a pipeline drug, which (as discussed above) makes it subject to the 2 year cap.[5]

We use the original expiration date and period of extension to calculate the final extension date (*term_extended*) and use this, together with *fdareceiptdate* (see above), to determine the extended patent term after extension (*term_extended*). We create a variable to infer whether the 14-year cap applies (*capped_14*) based on whether the calculated term as extended is between 13.98 and 14.02 years. As we explain in more detail in Appendix B, in a few cases this variable appears to understate the share of drugs subject to the 14-year cap.

To determine the length of clinical trial for each extended patent, we collect additional regulatory documents. (The PTE List does not report the duration of clinical trials or contain the information needed to calculate duration.) For each extended patent, we obtain the Notice of Final Determination (NFD) from the USPTO Patent Examination Data System (PEDS) database (USPTO 2023a) using the following procedure: [1] determine the patent application number from the USPTO PatentsView Database (USPTO 2023c); [2] search for responsive PEDS documents for this application number through the PEDS interface; [3] download documents with the correct identifier (for NFDs, those with code "TERM.PTO.ELC" or "TERM.PTO.NFD"). We extract from each NFD the date the regulatory review period began (the effective date of the IND, or when clinical trials began). For a small number of cases where NFDs are unavailable from PEDS, we instead determine the IND effective date from the historical Federal

[5] This variable is created from *fdareceiptdate* and *fdapprovaldate_b*, whose construction is discussed above.

Register (Federal Register 2023) notices pertaining to a drug's "Determination of Regulatory Review Period," searching by drug name and ensuring the patent number is identical to that on the PTE List. Since some Federal Register entries (especially older ones) have OCR errors, we review Register documents manually to ensure accuracy.

We calculate the clinical trial length (*indlength*) as the time between the start of the regulatory review period and the date of NDA filing, consistent with how the FDA determines the RRP component of patent extensions. For the NDA filing date (*fdareceiptdate*), we use information from the FDA's compilation of drug approvals, though this information is available from the NFDs and Federal Register entries as well.

Analyses of Public Data

Stata do-files: make_analysisfiles_public.do; an_main_public.do

The Stata do-file **make_analysisfiles_public.do** consolidates the data above into files used for our analyses. The final file for all analyses using public data is **analysisfile_public.dta**. The do file **an_main_public.do** runs analyses for Figures 1, 2, 4, and 5 and all performs all calculations reported in the text.

Restricted Data: Primary and Secondary Patents

Stata do-files: make_analysisfiles_restricted.do

To classify patents as primary or secondary, we use data acquired from Ark Patent Intelligence (IQVIA 2021), a commercial provider that markets its data to firms "to navigate the complex global patent landscape." Ark associates patents with seven categories and numerous subcategories, as set out below:

Ark Patent Categorization

Category	**Subcategory**
Molecule	Molecule patent Salts, hydrates and solvates Polymorphic forms Other molecule forms
Process and preparation	Intermediates and preparations thereof Final synthetic stages Complete synthesis Purification methods Fermentation methods Biotechnology
Formulation	General formulation and methods Route specific (injectable, oral, ophthalmic, otic, nasal, inhalation, topical, transdermal patch, rectal, vaginal, penile, urinary) Kits and packaging Excipients
Use	New use related to main indication Dosage regimen/administration conditions Drug with device
Combination	Novel combination; Use of combination
Assay	Assay methods Patient suitability
Device	Injection Respiratory Ophthalmic Diagnostic Energy dependent Oral admin

We obtain Ark patent coding for all patents listed on the Orange Book through 2019. For analyses using patent coding, we focus on NMEs approved through 2019 and Orange Book patents listed through 2019. There are 719 such NDAs, and 3,311 patents. Of these, Ark codes all but 80 patents. Ark codes patents to their strongest claims. We classify a coded patent as "primary" if it is (for any associated drug) in Ark's "Molecule" category and "Molecule Patent" subcategory. Otherwise, we classify the patent as "secondary."

To validate these measures, we compare the Ark categorizations to those from Hemphill and Sampat (2011), where we worked with a former pharmaceutical patent examiner to classify Orange Book listed patents listed through 2009. That analysis coded 315 patents associated with 119 drugs with first generic entry between 2001 and 2010. Of these 315 patents, 306 are coded in Ark. Among patents classified as primary in Hemphill and Sampat (2011), 94 percent are coded as primary under the Ark approach. Among patents classified as secondary patents in Hemphill and Sampat (2011), 95 percent are secondary under the Ark approach.[6]

Hemphill and Sampat (2011) Coding Versus Ark Coding

		Ark Patent Intelligence		
		Primary	Secondary	Total
Hemphill and Sampat (2011)	Primary	99	6	105
	Secondary	10	191	201
	Total	109	197	306

The coding is used to calculate in-text statistics on the share of secondary patents over

[6] Of the 9 patents missing from Ark, 7 are classified as secondary by Hemphill and Sampat (2011).

time, to assess which patents are "non-binding" for generic entry, to examine which types of patents are extended, and to determine the share of drugs and drug sales that have "live" primary patents, secondary patents, or both.

Restricted Data: SSR Sales

Stata do-files: make_analysisfiles_restricted.do

We obtain sales data for brand-name drugs from SSR Health Inc., covering the period from 2007 to 2022 (SSR Health 2024). SSR Health collects brand sales data based on disclosures by publicly traded drug firms. We obtain from SSR Health a panel of all drug level sales in their dataset from 2007 to 2022. After processing the SSR Health and FDA drug name data (standardizing case, removing punctuation, and normalizing abbreviations), we merge the NME List with the SSR Health data by brand name using the *fuzzywuzzy* module in Python 3.0. Following a manual review of the matches, we retain those with a score greater than 0.85, and manually review to verify that all remaining matches are true positives and that no valid matches are excluded. Finally, we adjust the sales data for each year to 2022 dollars using the CPI deflator.

In cases where we lack SSR Health sales data we do not know whether these are true zero sales drugs or simply not in the sampling frame of SSR Health. In analyses where we compare examine differences in patents, challenges, and effective market life by sales (for drugs with first generic entry 2008-2023), we observe no match for 26 percent of the drugs. We treat these drugs, and those with zero SSR Health reported sales in the year prior to generic entry, as "zero or missing" in the relevant figures. For analyses where we calculate the share of 2019 sales for drugs in primary versus secondary patent term, we are explicitly focused on share of SSR Health sales.

The Stata do-file **make_analysisfiles_restricted.do** creates the restricted data from raw

files and merges in with restricted data.

Analyses of restricted data

Stata do-files: an_main_restricted.do

The final file for all analyses using public data is **analysisfile_restricted.dt**a. The Stata do file **an_main_restricted.do** creates Figure 3 and performs all calculations using restricted data reported in the text.

Appendix B: PTE List Discussion and Robustness

Completeness

Of the 826 NMEs, 571 have an extended patent. Given the importance of patents in pharmaceuticals (see Section 1 of the paper for references) it may be surprising that nearly 31 percent do not. This raises a concern about potential missing patent term extensions from the PTE List. Older editions of the PTE List are known to have gaps (Leitzan and Acri 2020). The current edition of the list is thought to be a complete and accurate set of extensions (personal communication, Ali Salimi, USPTO).

To explore this, we determine patent application numbers for each Orange Book patent for the 826 drugs from PatentView, and use this to merge in "transaction histories" for each patent from the USPTO Patent Examination (PatEx) Research Dataset (USPTO 2023d). In this context, a transaction is a document filing, event, or decision. The transaction histories provide codes for all events in the prosecution history and publication history for patent applications. We use two distinct sources of transaction data in PatEx: USPTO Public PAIR and the Content Management System (CMS). PAIR transactions are recorded in the *transactions* file, and CMS transactions in

cms_documents. For each patent application, we identify all transactions linked to an application for Hatch-Waxman PTE, a Notice of Final Determination, and Patent Term Extension Certificates:

PAIR Transactions

Transaction	Code
Patent Term Extension Certificate	PTEC
Notice of Final Determination - Election Required	ELER
Notice of Final Determination - Eligible	NFDE
Patent Term Extension Application under 35 U.S.C. § 156 Filed	PTER

CMS Transactions

Transaction	Code
Patent Term Extension Certificate	TERM.PTO.C
Notice of Final Determination - Election Required	TERM.PTO.ELC

Notice of Final Determination - Eligible	TERM.PTO.NFD
Patent Term Extension Application Under 35 USC 156	TERM.REQ

Of the 826 drugs, 795 have one or more patent linked to a record in PAIR or CME (or both). For each of the 795 drugs, we determine whether the drug has one or more of the following recorded in PatEx (whether in PAIR or CMS): any patent with an application for PTE, Notice of Final Determination, or Patent Term Extension Certificate. We then compare this information against the PTE List. Of the drugs with Patent Term Extension certificates in PatEx, 95 percent also appear on the PTE List. Thus, nearly all drugs with patent term extensions are indeed on the PTE List.

We next investigate whether firms have tried but failed to obtain extensions. The PatEx data on PTE applications shows that for 12 percent of the drugs there is an application for term extension but no PTE grant, suggesting that the application was denied, abandoned, or is still pending. While failed or abandoned applications are part of the story, 22 percent of NMEs appear to have no PTE application. One potential explanation is that listed patents already have a term that expires more than 14 years after drug approval, the maximum available, in which case extension is unhelpful. A second possibility is a series of interim extensions obviating the need for formal extension. (We thank Professor Erika Lietzan for discussions on this point.) While we are exploring these issues in ongoing work, the analyses rule out missing entries from the current PTE List as a primary reason why nearly one-third of NMEs lack patent extensions.

Robustness to using earlier PTE List data

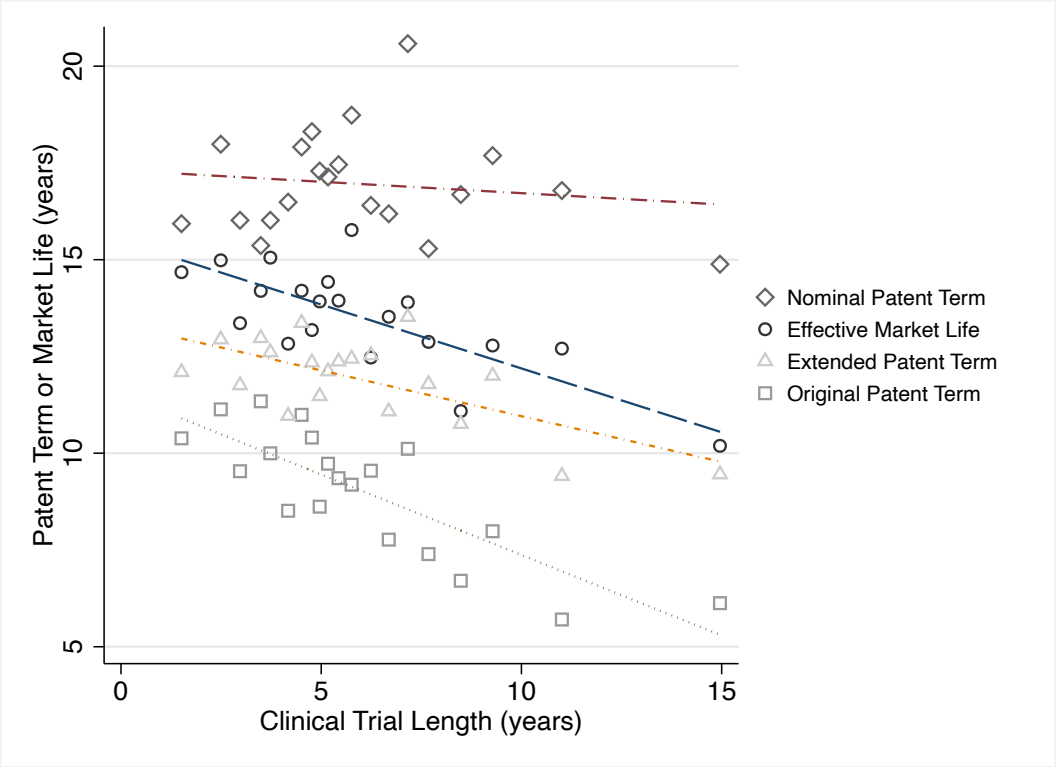
Stata do-files: appendix_pte_robustness.do

The PTE List does not incorporate certain changes to the pre-extension patent expiration date (including terminal disclaimers, where applicants give up patent term beyond a specific date, that are filed after grant of a certificate of extension). As a robustness check, we take advantage of the fact that earlier editions of the PTE List (March 2018 and earlier) attempt to make relevant adjustments.

Of the 571 drugs in the 2024 edition, 362 are also in the 2018 edition. For these 362 drugs, the extension period matches 94 percent of the time, and the original expiration date matches 79 percent of the time.[7] In the 2018 edition, 70 percent of drugs are subject to the 5-year or 2-year extension cap, or the 14-year extended term cap, compared to 65 percent for the same set of drugs in the 2024 edition.

Figure A.1 replicates Figure 5 from the main text, using data from the 2018 edition instead of the 2024 edition. Notwithstanding some differences between the two editions, the overall results remain consistent.

[7] The 2018 edition also reports the calculated extended expiration date. For 77 percent of drugs, this date matches what we calculate (using original expiration date and extension period) from the 2024 edition.



Appendix References

Dave, Chintan V., Michael S. Sinha, Reed F. Beall, and Aaron S. Kesselheim. 2020. "Estimating the Cost of Delayed Generic Drug Entry to Medicaid." *Health Affairs*, 39 (6): 1011–17.

Durvasula, Maya, C. Scott Hemphill, Lisa Larrimore Ouellette, Bhaven Sampat, and Heidi L. Williams. 2023. "The NBER Orange Book Dataset: A User's Guide." *Research Policy* 52 (7): 104791.

Hemphill, C. Scott, and Bhaven N. Sampat. 2012. "Evergreening, Patent Challenges, and Effective Market Life in Pharmaceuticals." *Journal of Health Economics* 31 (2): 327–39.

IQVIA. n.d. "ARK Patent Intelligence." https://www.iqvia.com/solutions/industry-

segments/generics/ark-patent-intelligence (accessed November 2021).

NBER. 2023. *Orange Book Patent and Exclusivity Data—1985-2016.* https://www.nber.org/research/data/orange-book-patent-and-exclusivity-data-1985-2016 (accessed September 1, 2023).

Office of the Federal Register, National Archives and Records Administration. 2023. *Federal Register Documents Search*. https://www.govinfo.gov/app/collection/FR (accessed September 1, 2023).

SSR Health. 2024. *SSR Health Drug Sales Database* (accessed June 25, 2024).

US Food and Drug Administration. 2023a. *Compilation of CDER New Molecular Entity (NME) Drug and New Biologic Approvals.* https://www.fda.gov/drugs/drug-approvals-and-databases/compilation-cder-new-molecular-entity-nme-drug-and-new-biologic-approvals (accessed September 1, 2023).

US Food and Drug Administration. 2023b. *Approved Drug Products with Therapeutic Equivalence Evaluations (Orange Book)*. https://www.fda.gov/drugs/drug-approvals-and-databases/approved-drug-products-therapeutic-equivalence-evaluations-orange-book (accessed September 1, 2023).

US Food and Drug Administration. 2023c. *Patent Certifications and Suitability Petitions.* https://www.fda.gov/drugs/abbreviated-new-drug-application-anda/patent-certifications-and-suitability-petitions (accessed September 1, 2023).

US Food and Drug Administration. 2023d. *Drugs@FDA: FDA-Approved Drugs.* https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm (accessed September 1, 2023).

US Patent and Trademark Office. 2023a. *Patent Examination Data System (PEDS).*

https://ped.uspto.gov/peds/# (accessed September 1, 2023).

US Patent and Trademark Office. 2023b. *Patent Terms Extended Under 35 U.S.C. § 156.* https://www.uspto.gov/patents/laws/patent-terms-extended (accessed September 1, 2023).

U.S. Patent and Trademark Office (USPTO). 2023c. *PatentView.* https://www.uspto.gov/ip-policy/economic-research/patentsview (accessed September 1, 2023).

U.S. Patent and Trademark Office (USPTO). 2023d. *Patent Examination Research Dataset:.* https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-examination-research-dataset-public-pair (accessed September 1, 2023).