A Public Database of European Drugs

Satyaki Chakravarty*†, Margaret Kyle‡, Fabio Montobbio†, and Gabriele Pellegrino†

Objective We describe the construction of a drug database comprising drugs authorized in France, Germany, and the United Kingdom using only public sources, in effect extending the Drugs@FDA database into the European Union. We compute exclusivity durations of the drugs, link their basic patents, and compute a window where generic entry could happen. We validate our database with popular commercial databases, such as IQVIA, Clarivate's Cortellis, and Patsnap's Synapse.

Big picture value of the data set Unlike Drugs@FDA—the centralized list of all the drugs approved in the United States, there is no comprehensive list of drugs available for Europe (when we refer to Europe, we also include the United Kingdom (UK) market in its definition). The United States spends the most, especially when branded drugs are considered. But spending by Europe is not trivial. Numerous questions need investigation, such as why drugs are launched later in the European Union, and what protects their exclusivities on the continent and the different member states—is it the intellectual property or are there stronger forms of exclusivities available (Kyle, 2025)?

Answering such questions requires comprehensive data on drug development—from market launch dates of branded drugs by originators, the various forms of exclusivities protecting the drug and their expiry dates, and the dates of generic entries. While there are commercial databases such as IQVIA and Clarivate Cortellis, among others, most researchers are unable to use them due to their enormous cost. Thus, it is not surprising that there are only a few, albeit seminal papers that focus on the European drug market (Kyle, 2007; Gaessler and Wagner, 2022; Dubois et al., 2022; Byrski and Wang, 2025).

Centrally approved medicines in the European Union through the European Medicines Agency (EMA) do have greater drug-level details but are inadequate in many aspects—most importantly, they miss the nationally approved medicines. Such information, however, is scattered through each EU member state's medicine agency websites. This project bridges this gap by creating a drug database along with their patents for three of the largest European drug markets, France, Germany, and the United Kingdom.

*Corresponding Author, satyaki.chakravarty@unicatt.it
‡Mines Paris - PSL, Paris, France
†Università Cattolica del Sacro Cuore, Milano, Italy

Linking European drugs to Drugs@FDA and Drugbank We start by collecting the full universe of drugs approved in France, Germany, and the United Kingdom from their respective national medicines agency webpages. After compiling the respective national sets, we create separate crosswalks between national molecule names to Drugbank.[1] The Drugbank database is free to use for academia and this acts as our bridge that consolidates molecule name variations. For French drugs, authorized by Agence nationale de sécurité du médicament et des produits de santé (ANSM), we convert the French molecule names to English using the Unified Medical Language System (UMLS) Metathesaurus and then connect to Drugbank. UK drugs are directly matched to Drugbank as the UK molecule names are in English. The German drugs authorized by their medicine agency Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) provide Chemical Abstracts Services (CAS) Registry Number for every molecule.[2] Drugbank also contains CAS Registry Numbers and this allows us to connect BfArM to Drugbank without matching names. For each of the three national sources, we rely on the following order of molecule name matching: direct name matching, fuzzy name matching with scores greater than or equal to 90, and matches using hand-corrections, in case the fuzzy matches have scores of less than 90. In this way, we match the three European drug sources to Drugbank. In the same way we connect the molecules in Drugs@FDA to Drugbank.[3] For simplicity, we identify each drug using the FDA application number, though they could be interchangeably used with pipe separated Drugbank accession numbers, as DrugBank IDs are assigned to molecules rather than drugs that can be combinations, while FDA application numbers are unique to each drug product.

Each drug's portfolio of patents can be identified in the United States using the Orange Book. Europe does not have anything similar. However, EU member states, just like the US, extend drug patents for up to five years if there were regulatory delays during a drug's approval, and it also provides six months for paediatric extensions. In the US it's called a Patent Term Extension (PTE) while the European states call it Supplementary Protection Certificates (SPC). Each state, and in our case, France, Germany, and the United Kingdom's intellectual property offices report the SPCs linked to patents, if they have one. We obtain such patents and their SPCs and link it back to the drugs they correspond to. For France, we rely on molecule or drug name matches, and this involves hand-corrections. But German and UK SPCs contain the corresponding marketing authorization (MA) numbers, and we use them to connect the SPCs to their drugs. We summarize this in Figure 1.

[1] Molecules are also called Active Pharmaceutical Ingredients (APIs).

[2] CAS Registry Numbers are assigned to chemical entities in the literature by the American Chemical Society.

[3] The fuzzy score is calculated using Levenshtein distance using `rapidfuzz`'s `fuzz.ratio`, defined as $\text{score}(a,b) = \big(1 - d(a,b)/(|a|+|b|)\big) \times 100$, where a is the FDA active-ingredient name, b the candidate DrugBank name, $|a|, |b|$ their lengths, and $d(a,b)$ the number of character insertions/deletions needed to turn a into b. The score ranges from 0 to 100 (100 = identical); we keep matches with score ≥ 90.

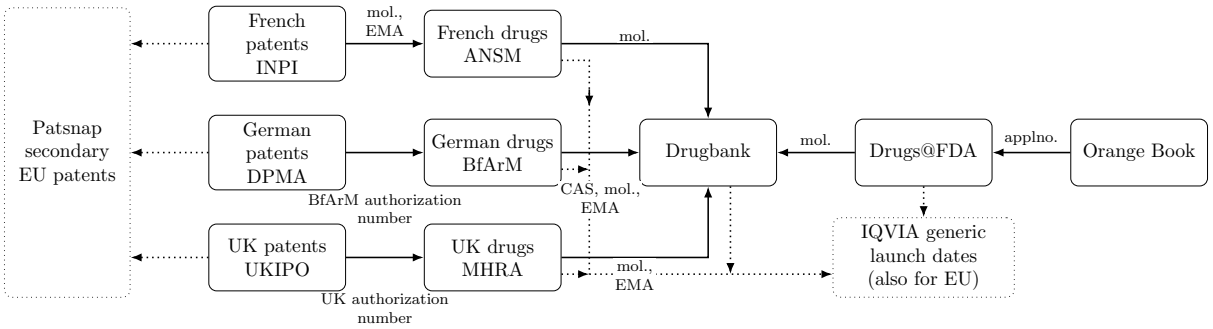


Figure 1: Linking European drugs to Drugbank and Drugs@FDA

Notes:

1. The BfArM authorization number is called the *Zulassungsnummer/Registrierungsnummer*, which translates to "Approval number/Registration number".
2. The UK authorization number is called the PL number.
3. Drugs@FDA and Orange Book merging is taken from the replication files of Hemphill and Sampat (2025). They have updated the publicly available clean Orange Book data called NBER Orange Book by Durvasula et al. (2023). Each drug is identified by their application number, or applno, assigned by the FDA.
4. Mol name refers to molecule name or API name matching. This involves fuzzy matching and hand corrections.
5. CAS refers to Chemical Abstract Services.
6. The solid boxes are publicly available data, and the dotted boxes: Patsnap and IQVIA are private data.

Highlight of our contribution We show our main contribution only for Germany in this extended abstract. Hemphill and Sampat (2025) start with 944 drug application numbers and use 826 in their analysis. We are able to find and connect 649 of these in Germany. The rest were not launched in Germany.

European countries provide a total market exclusivity of 10 years since 2004 (+1 if a new therapeutic area is added to the drug) regardless of patents. This is much more generous when compared with its US equivalent. The EU's 10 years is split into two parts: 8 years of data exclusivity, where generic companies cannot legally file Abbreviated New Drug Application (ANDA) equivalent applications in the EU states, and then 2 years of market exclusivity, where the generic companies can file but not launch their drug in the market.

For each drug available in Germany, using their marketing authorization date, we calculate their data exclusivity end date—where generics are allowed to file for authorization. Then, we get the actual SPC applications' expiration dates. Finally, for the same drugs, we get the first generic application dates from BfArM (the publicly available source) and the generic availability dates in hospitals and pharmacies from IQVIA's MIDAS. The latter is an accurate measure of generic availability, as their filing at BfArM for authorization does not reflect whether they were actually launched at a later date in Germany. Nevertheless, they act as an indicator that the drug will at some point be launched in the country.

We plot data exclusivity, SPC exclusivity, generic application filing date (as calculated from BfArM and IQVIA's MIDAS). We see a clear bunching of generic entry at the SPC expiration dates. This is significant because in the US, Paragraph IV patent invalidations are one of the significant sources of generic entry, and the invalidations are overwhelmingly obtained for secondary patents (Hemphill and Sampat, 2025). Instead, in Germany, SPCs assume the role. In the paper we show that this is largely true for the other jurisdictions. Kyle (2025, 2026) show that exclusivity lengths when compared across US and EU largely are the same. We add to this discussion by pointing to the specific instruments that are responsible for preservation of this duration. A significant number of secondary patents in each drug's portfolio is invalidated by the EPO due to validity concerns (Kyle, 2026). This is because the EPO is generally stricter in terms of patentability as compared to the USPTO (Frakes and Wasserman, 2023). Still, secondary patents do have a role in the EU, and the purple triangles appearing on the right side of the figure represent those drugs, and so does Reverse Payment Settlements (secondary patents and other forms of exclusivities are not marked in the figure).

In the paper, we go further explaining the role SPCs play in the European markets, as opposed to its US equivalent, the PTE, which plays a rather muted role in protecting exclusivities of drugs in the United States, and the final expiration is usually held by secondary patents. We also show that intellectual property is not the only reason for delayed originator launches as well as generics not applying for authorization at all.

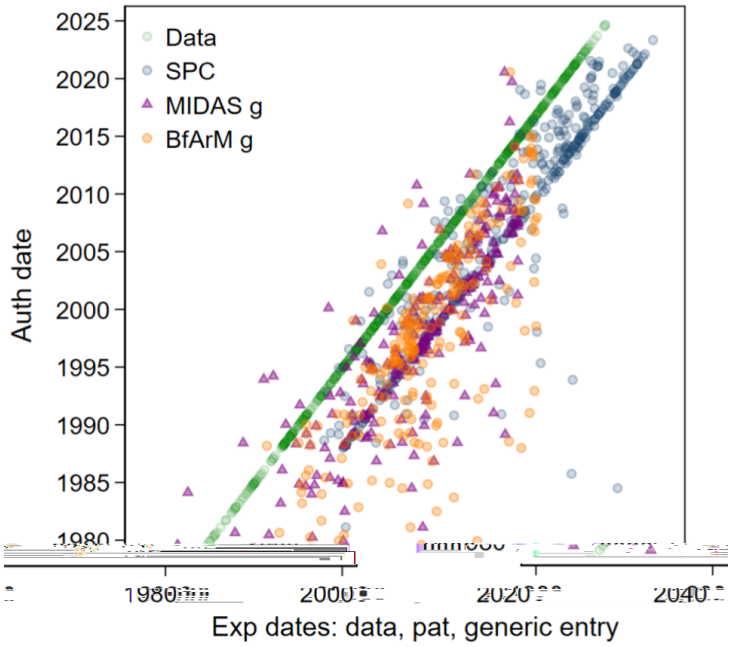


Figure 2: Decomposition of exclusivities in Germany

Public dissemination For the first phase of our work, we will make our code publicly available, and an accompanying data paper will describe the database construction and validate the data using the commercial sources as described. This will be presented at the I3 meeting. For the second phase, we plan to expand our work for the Dutch, Scandinavian, Swiss, and Italian markets.

References

Byrski, D. and L. X. Wang (2025). Marketing authorization and strategic patenting: Evidence from pharmaceuticals. *Journal of Public Economics 247*, 105415.

Dubois, P., P.-H. Moisson, and J. Tirole (2022). The economics of transferable patent extensions.

Durvasula, M., C. S. Hemphill, L. L. Ouellette, B. Sampat, and H. L. Williams (2023). The NBER Orange Book dataset: A user's guide. *Research Policy 52*(7), 104791.

Frakes, M. D. and M. F. Wasserman (2023). Investing in ex ante regulation: Evidence from pharmaceutical patent examination. *American Economic Journal: Economic Policy 15*(3), 151–183.

Gaessler, F. and S. Wagner (2022). Patents, data exclusivity, and the development of new drugs. *Review of Economics and Statistics 104*(3), 571–586.

Hemphill, C. S. and B. N. Sampat (2025). Patents, innovation, and competition in pharmaceuticals: The Hatch-Waxman Act after 40 years. *Journal of Economic Perspectives 39*(2), 27–52.

Kyle, M. K. (2007). Pharmaceutical price controls and entry strategies. *The Review of Economics and Statistics 89*(1), 88–99.

Kyle, M. K. (2025). Lessons for the united states from pharmaceutical regulation abroad. *Journal of Economic Perspectives 39*(2), 53–78.

Kyle, M. K. (2026). Potential and effective exclusivity of new drugs in Europe. In J. M. Barnett and B. Heiden (Eds.), *Bringing Medicines to Life: How Intellectual Property Enables Innovation in the Life Sciences*. Cambridge University Press. forthcoming.