A Public Database of German Drugs

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

Objective We describe the construction of a database of German drugs along with their basic patents that were authorized between 1978 and 2024 using only public sources. After constructing the German drug database (henceforth BfArM) and drug patent database (Deutsches Patent- und Markenamt - Startseite, or DPMA), we describe the process of connecting this to the drugs available in the United States Food and Drug Administration's (FDA) Drugs@FDA, the NBER Orange Book, and Drugbank, all of which are publicly available sources of drugs. We validate our database with popular commercial databases, such as IQVIA.

Big picture value of the data set Unlike the drug dataset available through the FDA, called Drugs@FDA, which lists 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 (U.K.) market in its definition). While the United States spends the most, especially when branded drugs are considered, spending by Europe is not trivial. Numerous questions remain unanswered, on what really incentivizes drug development and their being launched in Europe. Why are drugs launched later there, and how much are pricing strategies and drug exclusivity legislations responsible for this situation?

We need more evidence on these, and answering such questions requires comprehensive data on drug development, from the start date of a drug project, its market launch dates by originators (first filer of an original drug), and the timing of generic entries. Finally, we also need information on the exclusivity provided by the European patent extensions, also known as the Supplementary Protection Certificates (SPC). While there are commercial databases curated by IQVIA and Clarivate Cortellis, most researchers are unable to use them due to the enormous cost they come with. Thus, it is not surprising that there are only a few, albeit seminal papers that focus on the European drug market.

While there are a few databases that consist of drugs that are approved centrally, they are inadequate in many aspects, including coverage and the details available per drug. Fortunately, all this information 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 the largest European drug market, Germany. We plan to expand this project to other geographies, but having the German drug database already addresses quite a few existing data bottlenecks.

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

Demonstrations to highlight the new components The first contribution of this project is the coverage of drugs. There are two main public sources of E.U. drugs that are comparable to the FDA. These are from the Head of Medicines Agency (HMA) and the European Medicines Agency (EMA). These two sources, even if combined, inadequately report the drug counts. Second, EMA and HMA are relatively new, established in 1995 and 2004, respectively.

A drug can enter an E.U. state from four sources: Central Approval, through the European Medicines Agency (EMA), National Authorization Procedure (NAP, directly filed at the national office), MRP (Mutual Recognition Procedure, already authorized in one EU state, found in HMA) and Decentralized Procedure (DCP, asking for authorization at selected EU states, also found in HMA). Throughout this document, we will use NAP when a drug is authorized in only one state, in our case Germany, and use "multi-state procedures" to broadly refer to simultaneous entry of a drug into multiple EU states, either through CP, DCP, or MRP.

In Figure 1, we plot the number of first authorizations for each Active Pharmaceutical Ingredient (API) by their authorization years and compare BfArM to the FDA and EMA. As we see, BfArM covers the years between 1978 and 2004 better than the multi-state procedures, and it also closely tracks FDA approvals. The increase in approvals between 1995 and 2004 is not an error. This is because of an increased filing and approval of old chemical entities already approved by the FDA in the 80s.

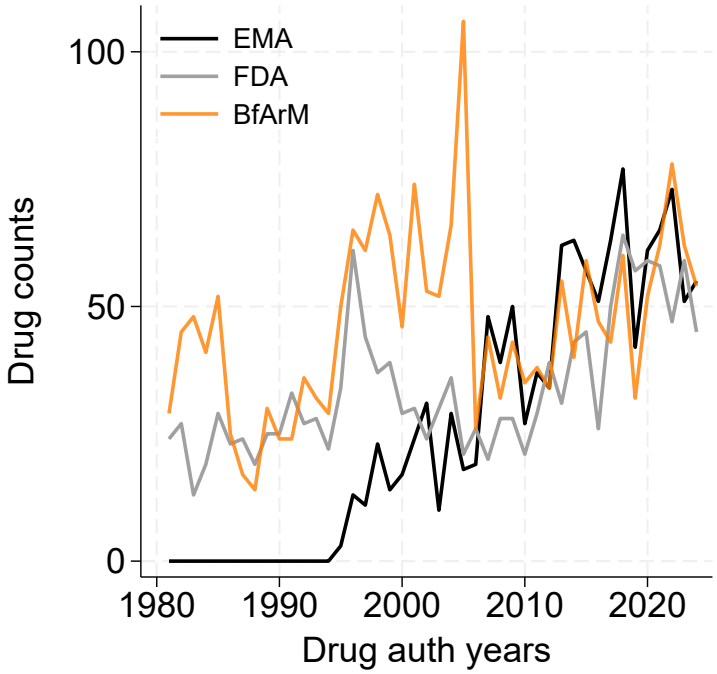


Figure 1: Authorized drugs comparisons

The Figure shows the number of new approvals authorized over the years from three sources, EMA, FDA, and BfArM. We take out veterinary medicines, vaccines, and ATC "V".

What we find is the national component — NAP accounts for 60 percent of drug approvals in Germany. This, of course, has fallen to 32 percent since 2004, after the Regulation (EC) No 726/2004 was enacted, mandating compulsory use of CP in case of biotechnology products and for products treating serious conditions. Thus, if we were to build a comprehensive drug database pertaining to the EU, considering NAP is important, for which the commercial databases are the only route.

The next component we highlight is the identification of generics through our constructed data. We report two artifacts in Figure 2, one, comparison of the number of first generic applicants for each API aggregated over generic application years from two data sources, and two, the average time taken for a generic to file their application for authorization in Germany (either through NAP or multi-state procedures) by API. Please note that we report generic filing and not the generic approval years, which we report in the full paper.

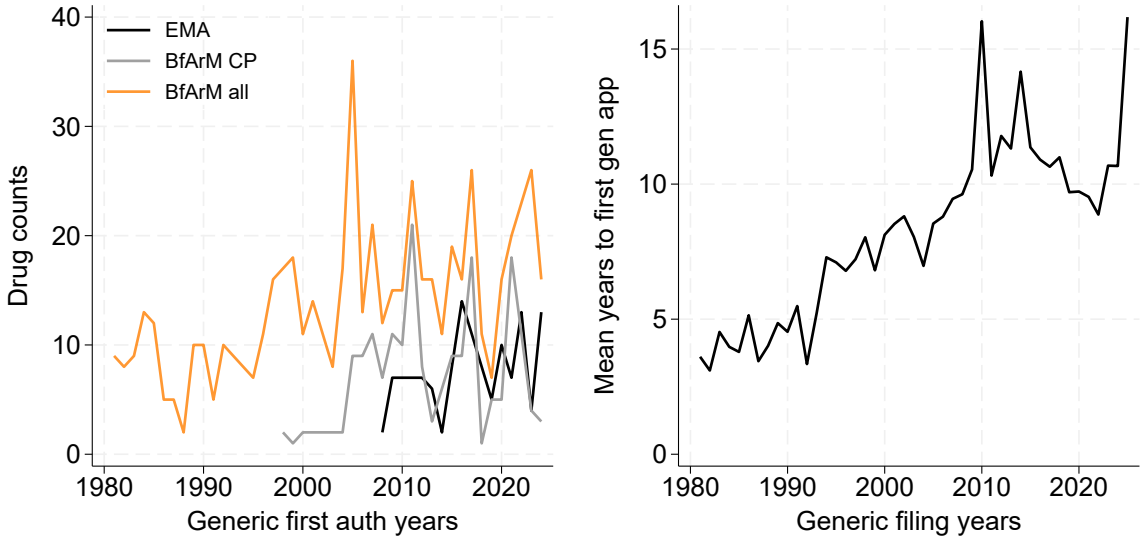


Figure 2: Generic drug applications

The Figure on the left plots the first generic approvals for each API from two sources, BfArM and EMA. BfArM CP plots CP drugs from the BfArM database, and BfArM all includes all drug approval types, NAP, MRP, DCP, and CP. The Figure on the right reports the average lag between originator drug approval and first generic filings over the years by each API.

From the first figure of Figure 2, we see that studying entry of generics solely through EMA data is grossly inadequate, one because not all drugs are covered through the multi-state procedures, and two, most generics prefer entering the EU states separately, as can be seen from the BfArM all line. In fact, there is considerable variation in the entry of generics by EU

states, and this remains understudied, partly because such information is only available through commercial databases.

In the second figure of Figure 2, we see that over time, the lag between the first filing of an API from an originator and the first generic application has steadily gone up, averaging at 8.2 years throughout the sample. This shows that the average legal exclusivity drugs enjoy has gone up over time. Why then are drug companies launching late in Europe?

Linking BfArM to Patents, Drugbank, and Drugs@FDA The third highlight of our project is linking the drug database BfArM to DPMA patents and other public drug databases. Germany, like other EU states and the United States, extends 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. We link each drug with DPMA patents to calculate the exclusivity each drug enjoys due to basic patents (called primary patents in the United States) and their extensions, and finally calculate generic entry dates to understand the actual exclusivity enjoyed by drugs in Germany. Thanks to the other public databases, we are also able to connect secondary patents listed on the Orange Book. We summarize the connections in Table 1 and describe the process in brief as follows.

The BfArM database controls each API using a number called "ASK Number" (ASK). In addition to this, the authority also releases the Chemical Abstracts Services (CAS) Registry Number attached to each ASK. CAS Registry Numbers are assigned to chemical entities in the literature by the American Chemical Society. The Drugbank database, which is free to use for academia, also contains CAS Registry Numbers. This allows us to connect BfArM to Drugbank without matching names. Next, Drugbank connects to the FDA New Drug Application (NDA) numbers. Thus, we can connect BfArM to the Drugs@FDA database. This also allows us to connect to the Orange Book to obtain US patents.

Finally, using BfArM drug authorization numbers, we attach the DPMA patents. This allows us to attach the BfArM drug database to the United States Patent and Trademark Office (USPTO) and European Patent Office (EPO) patents in two ways, through the Orange Book and through DPMA.

Table 1: Summary of connections to other databasees

Database	Connecting variable	Connection to
DPMA, drug patents	Drug Authorization Number	BfArM
BfArM, drugs	Chemical Abstract Services (CAS)	Drugbank
Drugbank	NDA, ANDA	NBER Orange Book Drugs@FDA
NBER Orange Book	Patent Number	USPTO EPO PATSTAT

The first two rows are our contribution, DPMA and BfArM.

In addition to all this, we are able to identify and harmonize company names in the BfArM database because the German authority tags each marketing authorization holder with a company number. Additionally, each originator drug is connected to its generic and parallel entrants through "drug relationships". A parallel entry or import is the process of launching drugs by importing the drug from other EU states, whether they have been launched in the EU state or not. We identify and harmonize drug company names and importers into originators, generics, parallel entrants, and API producers. Finally, we collate important administrative dates assigned to drugs by the German authority to understand differences in filing and authorization dates.

In summary, we contribute in three novel ways: one, we create a sample of all drugs launched in Europe only using public sources along with improved coverage of drugs approved before 2004, two, identification of drug filing companies, along with the generic producers, parallel importers, and the API producer, and three, connection to a few public drug databases, including patent documents.

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 a few commercial sources. This will be presented at the I3 meeting. For the second phase, we plan to replicate our work for the U.K., Dutch, and French markets, i.e., expand the database of drugs as well as their patents for those geographies.