ISSN 1866-3494

July 11, 2022

ESMT Working Paper 22-05

Mapping Markush

Stefan Wagner, ESMT European School of Management and Technology

Christian Sternitzke, Sternitzke Ventures UG

Sascha Walter, University of Würzburg

Mapping Markush

Stefan Wagner[1], ESMT Berlin, stefan.wagner@esmt.org

Christian Sternitzke, Sternitzke Ventures UG

Sascha Walter, University of Würzburg

July, 2022

Forthcoming in Research Policy

Abstract

Markush structures are molecular skeletons containing not only specific atoms but also placeholders to represent broad sets of chemical (sub)structures. As genus claims, they allow a vast number of compounds to be claimed in a patent application without having to specify every single chemical entity. While Markush structures raise important questions regarding the functioning of the patent system, innovation researchers have been surprisingly silent on the topic. This paper summarizes the ongoing policy debate about Markush structures and provides first empirical insights into how Markush structures are used in patent documents in the pharmaceutical industry and how they affect important outcomes in the patent prosecution process. While not causing frictions in the patent prosecution process, patent documents containing Markush structures have an increased likelihood to restrict the patentability of follow-on inventions and to facilitate the construction of broad patent fences.

Acknowledgements: Financial support by the German Science Foundation (DFG) under the project WA 4148/2-1 is gratefully acknowledged. We thank three anonymous reviewers for valuable suggestions, Nate Liu Jingze for outstanding research assistance, the participants of the MPI Seminar for insightful comments, Josef Eiblmaier from Infochem for his patience and support, and Clarivate for data support.

[1] Corresponding author.

1. Introduction

At the heart of the patent system is a quid pro quo agreement between society and the patent holder to incentivize private investment into risky research activities. Society gets access to technical information about how to make the invention and the right to freely use it after patent expiry. The patentee gets a limited period of exclusivity conferred by the patent as set forth in the claim. However, even in the pharmaceutical industry, where the patent system is believed to function particularly well (Bessen and Meurer, 2008), there are doubts that pharmaceutical companies are not fulfilling their part of the quid pro quo trade.

Pharmaceutical companies often engage in strategic patenting practices to extend the duration of market exclusivity for original drugs beyond the initial 20-year exclusivity period by filing multiple related patents over the lifetime of a drug (European Commission 2009, Hemphill and Sampat, 2011, 2012; Howard, 2007; Kyle, 2016; Sternitzke, 2013). In addition, it has been increasingly questioned to what extent pharmaceutical patents disclose sufficient information to allow the making and use of the claimed invention without undue and costly experimentation. Disclosure requirements are of particular relevance as patentees in the pharmaceutical industry make use of *genus* claims – a broad patent claim covering a group of potential products (*species*) that incorporate the basic advance of the patented inventions.[2] Technically, this is achieved using Markush structures, which facilitate claiming a large class of chemical compounds without the necessity of specifying every single molecule. The use of genus claims allows pharmaceutical patentees to protect a potentially large number of related chemical compounds in a single patent filing.[3]

[2] In international patent systems, the legal criterion of 'unity of invention' requires that a patent application relates to a single invention or, at most, a group of inventions that are closely related to each other (see, for example, Rule 13 of the Patent Cooperation Treaty PCT, https://www.wipo.int/pct/en/texts/rules/r13.html, latest visit 25th of August, 2020).

[3] This is also common practice in the chemical industry as related unpublished work by the authors shows. In this paper, however, we focus on the pharmaceutical industry as the underlying economic questions differ due to the different nature of products and associated R&D activities.

Policy makers, courts, and practitioners have controversially discussed the use of Markush structures as genus claims in pharmaceutical patents. First, Markush structures have been criticized for burdening society by creating considerable challenges during patent prosecution. Since each alternative included in a Markush structure often raises different patentability questions and thus requires separate search and examination, Markush claims consume a disproportionate amount of examiner time and patent office resources compared to simpler claims containing a single molecule (species) (Bone and Kendall, 2008; Federal-Register, 2007). Second, there are concerns that Markush structures lead to patents of an excessive protective scope (Correa, 2008). Broad patents raise concerns about impeding follow-on R&D conducted by competitors as they restrict the patentability of follow-on inventions (Merges and Nelson, 1990). Finally, Markush structures facilitate constructing 'patent fences' around an original drug (Cohen et al., 2000; Sternitzke, 2013) to the extent that patentees are allowed to file subsequent patent applications on selected molecules contained in a Markush structure. Markush claims may thus increase the market power of patentees at the expense of social welfare if competing novel drugs are introduced at lower speed.

Although Markush structures are widely used in pharmaceutical patenting and hotly debated among stakeholders, there is, to date and to our best knowledge, no quantitative evidence on how Markush structures are employed and how they affect patent prosecution and innovation dynamics. Our study addresses these gaps in the context of pharmaceutical patenting in three ways. First, we provide a descriptive overview on the use of Markush structures in the pharmaceutical industry. We examine whether the Markush structures lead to frictions in the patent examination process such as increased workload for examiners and longer examination durations, as indicated by patent office communications (Federal-Register, 2007). Second, studying patent citations, we provide insights into the impact of Markush structures on

follow-on R&D in general, and more importantly, whether they limit the patentability of follow-on inventions by limiting the patentability of subsequent patent filings (Grimpe and Hussinger, 2014; Guellec et al., 2012; von Graevenitz et al., 2013). Finally, linking patent filings to drug development projects allows us to examine whether Markush structures are associated with broader patent fences (Hemphill and Sampat, 2011, 2012; Howard, 2007; Kyle, 2016; Sternitzke, 2013).

Our research requires assembling a novel and unique dataset. We combine Clarivate's proprietary 'Derwent Markush Resource' (DMR) database, which contains molecular skeletons extracted from Markush structures in patent documents (Barth et al., 2016), with the European Patent Office's (EPO) PATSTAT database and Clarivate's Cortellis database. The resulting dataset comprises Markush characteristics for the population of all pharmaceutical patents filed at the EPO between 1992 and 2008. Analyzing these data, we find that almost one quarter of all pharmaceutical patent filings in our sample claim Markush structures, i.e., rely on genus claims. Alleviating existing concerns on challenges during patent prosecution, our data provide little evidence that the use of Markush structures leads to extended delays during patent examinations. Patent applications claiming Markush structures are not associated with longer examination durations and are even processed faster. Citation patterns reveal differences between Markush and non-Markush patent documents. Markush filings receive a higher number of citations overall, and more frequently establish novelty-destroying prior art compared to other pharmaceutical patent documents. We interpret this as suggestive evidence that Markush claims hamper follow-on innovation. Finally, Markush patent filings are associated with significantly larger number of patents associated to drugs compared to drugs without Markush patents linked to them in the Cortellis database. This is a clear indicator that Markush structures allow originator companies to build large patent fences around drug development projects. This finding deserves attention, as patent fencing strategies have been a

concern for policy makers that often relate patent fences and secondary patents to a reduction in competition by generic entrants (European Commission 2009).

Taken together, this study provides first large-scale empirical evidence on how Markush structures are used in pharmaceutical patenting. It addresses existing concerns that broad (genus) claims overburden patent offices, hamper follow-on innovation and allow the creation of large patent fences. The results are relevant for practitioners and policy makers seeking to understand how the use of Markush structures impacts important policy parameters. This is of particular importance, as courts are increasingly questioning whether Markush claims meet necessary legal standards warranting the grant of exclusivity, whereas legal scholars are warning that abolishing Markush patenting "can have serious consequences for the effectiveness of chemical patents" (Karshtedt et al., forthcoming p. 4).[4]

2. Markush patents: Institutional background

2.1 Markush structures

The first use of the term Markush structures dates back to a 1923 decision by the Commissioner of Patents in the US. The decision allowed Eugene Markush to claim a chemical genus by enumerating species within the genus (see Ex parte Markush, 1925 CD 126 Com. Pat. 1924). This enabled Markush, in his original patent filing, to protect several related inventions in a single claim that recites a list of alternatively useable members. To date, reciting alternatives is not governed by a particular format, but often alternatives are set forth as "a material *selected from the group* consisting of A, B, and C" or "*wherein* the material is A, B, or C" (Valence 1961). The molecular groups summarized as A, B, or C are typically labelled

[4] Please note pharmaceutical patents are considered a subset of chemistry patents (Karshtedt et al., forthcoming).

with R<#>.[5] They constitute placeholders for a set of substituents. The total number of compounds encompassed by a Markush structure thus is the number of possible combinatorial combinations (enumerations) of all the listed substituents at the positions that permit variations. Depending on the formulation of a Markush structure, the number of compounds falling under its umbrella can vary widely from a small number of distinct molecules to a very large number. This facilitates claiming a large class of compounds without the necessity of writing out every single chemical entity of the class. Apart from pharmaceuticals, inventions in metallurgy, refractories, ceramics, chemistry, and biology most frequently rely on Markush structures (Simmons, 2003).

This is best illustrated using an example: Darunavir is the active ingredient of the HIV drug "Prezista" and is covered by the European patent EP 0810 209. The compound underlying Darunavir is part of a patented Markush structure (see Figure 1). Each placeholder represents a broad set of chemical (sub)structures which, in turn, can comprise many specific molecules.

[INSERT FIGURE 1 ABOUT HERE]

EP 0810 209 specifies that "P1 and P2 independently represent hydrogen, alkoxy-carbonyl, aralkoxycarbonyl, alkylcarbonyl […], and mono- and disubstituted aminocarbonyl and mono- and disubstituted aminoalkanoyl radicals wherein the substituents are selected from alkyl, aryl, aralkyl […], or where said aminoalkanoyl radical is disubstituted, said substituents along with the nitrogen atom to which they are attached form a heterocycloalkyl or heteroaryl radical." (Figure A.1 in the Appendix contains a complete representation of the first claim of EP 0810 209.). We have calculated that the Markush structures claimed in patent EP 0810 209

[5] Wegner (1992) dates this notation back to the German dyestuff laboratories that made use of it in the nineteenth century chemical literature, with R standing for "Rest" (implying "group" in English). This notation has been used in European and US patents at the end of the nineteenth century and thus before the Markush decision, which ultimately gave the name to it (Wegner, 1992).

contain 5.046*10^16 different compounds. The active ingredient of Prezista, Darunavir, is one of them, but is not exemplified in the patent.

2.2 Markush structures and pharmaceutical patenting

Genus claims based on Markush structures are often employed by originator companies at early stages of the drug development process (Dhulap and Kulkarni, 2019). Typically, the drug development process departs from a chosen therapeutic target molecule for which a large set of compounds is screened to identify promising drug candidates, so-called "leads" (Hughes et al., 2011).[6] After initial validation, the innovator protects promising leads with patents based on Markush claims. These initial patent filings may exemplify, without a legal obligation, some specific compounds in the patent filing. For example, the patent underlying Prezista (EP 0810 209) exemplifies about 100 specific compounds selected from the larger Markush structure. Note that the compound underlying the actual antiviral drug has not been exemplified in EP 0810 209 and that there is no legal requirement to do so.

The use of Markush patents is advantageous for originator firms as it saves them from filing multiple patent applications. If an originator were permitted to patent only a specific lead molecule, the principle of structure-activity relationship (SAR) would apply. SAR describes the relationship between the chemical structure of a molecule and its biological activity (Crum-Brown and Fraser, 1865). It would allow competitors to create a slightly different molecule that has a high chance of retaining the desired therapeutic properties with little effort. This would facilitate the creation of chemical similar molecules with comparable properties at the expense of ex ante incentives for innovation. However, under the unity of invention principle, Markush structures are permissible if all of those slightly different molecules within

[6] The screening is often done computationally to identify first "hits" or "leads" (i.e., compounds with a high likelihood to interact with the target in a desirable way). The leads are then validated in in vitro experiments.

the structure show similar properties. Then, patenting the entire core molecular structure of a compound with desired properties, that is, the broader *genus* (as opposed to *species* of exactly identified molecules), prevents third parties from using the disclosure in the patent as a blueprint for designing and synthesizing unpatented substitutes with similar pharmaceutical utility as the original molecule (Bone and Kendall, 2008; Holman, 2017). Markush claims have therefore been deemed important in maintaining incentives to innovate (Karshtedt et al., forthcoming).

While beneficial for originator companies, concerns have been raised in policy circles that Markush structures hamper the patent prosecution process, follow-on innovation, and facilitate the construction of large patent fences. First, determining the patentability of Markush claims given prior art has been suggested to inflate examination times (Federal-Register, 2007) because it requires the separate examination of all alternatives within a Markush claim. The search for prior art is further complicated by the fact that overlapping Markush structures in different patents are often nearly unrecognizable (Berks, 2001; Simmons, 2003).[7] It is time consuming, costly, and sometimes impossible to find out whether any member of a new Markush structure has been disclosed before, which would challenge the patentability of the entire claim. Moreover, it is difficult to understand whether patent applicants meet the disclosure standards in a given patent application as the patent examiner cannot anticipate whether a

[7] This can happen because the patentees use differing notation, but more often because one patent requires a substructure that is optional in another (Simmons, 2003).

selection of a single compound, and, if so, which selection from the broader genus of the Markush claim, will be used for a specific application (Braga et al., 2018; Rollins, 1985).[8]

Second, since Markush claims can host extremely large sets of different compounds, scholars have called for more research into their impact on follow-on innovation (Correa 2008). As claiming a large number of molecules via a Markush structure allows patent applicants to obtain patents protecting a genus of compounds which might share desirable pharmacological properties. It is only after a significant number of studies that the Markush patent holder decides which single compound should transit into the subsequent (and very costly) clinical trials. This might make follow-on innovation less attractive for competitors. Any attempt by competitors to scrutinize other alternatives than the one compound chosen by the Markush patent holder for desirable pharmacological effects is at risk of infringing the initial Markush patent. Additionally, the existence of a claimed Markush structure likely reduces the possibilities of obtaining subsequent patents for compounds with desirable pharmacological compounds (other than the one chosen by the originator company) as the Markush patent constitutes relevant prior art. Since patents with Markush structures establish broader protective scope than patents claiming specific molecules, they can be expected to impede follow-on R&D by restricting patentability of a large number of molecules.

Third, in addition to primary patents on the specific compounds that underly a drug (active ingredient), firms commonly attempt to acquire secondary patents on alternative forms

[8] The question of sufficient disclosure often gives rise to post-grant validity challenges when a molecule that is not exemplified in a Markush patent becomes the active ingredient of a drug. For instance, some of Pfizer's Markush patents protecting Sildenafil, the active ingredient of Viagra, have been invalidated in multiple jurisdictions for this reason. This is reflected in gradual changes to the examination guidelines of worldwide patent offices seeking to clarify criteria for prior art search and for rejection of Markush claims (Federal-Register, 2007; Prior, 2012). Changing US case law also makes it harder to defend genus claims during patent litigation. Exemplary cases are Wyeth and Cordis Corp. v. Abbott Labs., 720 F.3d 1380 (Fed. Cir. 2013); Enzo Life Scis., Inc. v. Roche Molecular Sys., Inc., 928 F.3d 1340 (Fed. Cir. 2019); Regents of the Univ. of Ca. v. Eli Lilly & Co., 119 F.3d 1559 (Fed. Cir. 1997); Idenix Pharm. LLC v. Gilead Sci. Inc., 941 F.3d 1149 (Fed. Cir. 2019). See Karshtedt et al. (forthcoming) for a discussion.

of molecules, different formulations, dosages, and compositions, and new uses as drug development progresses through the various stages of clinical trials. A strategy typically referred to as 'patent fencing' or 'evergreening' (Kapczynski et al., 2012; Sternitzke, 2013). Devising patenting strategies to extend periods of protection is an essential aspect of 'life cycle management' in the pharmaceutical industry (Burdon and Sloper, 2003; Ellery and Hansen, 2012; European Commission, 2009; Howard, 2007; Kapczynski et al., 2012; Sternitzke, 2013). In this context, Markush structures can help companies to extend the duration of patent protection. As described above, Markush patents are typically filed at the beginning of a drug development project and typically do not exemplify the compound investigated in clinical trials. Once companies select a specific compound from the larger genus they can file additional patents on these specific compounds. These patents are also known as *selection patents* as they protect a specific compound selected from a broader set of compounds. As selection patents are filed later than the underlying Markush patent, the joint use of Markush and selection patents has the potential to extend the duration of exclusivity for the compounds claimed in the Markush and the derived selection patent. Thus, Markush claims in patent filings might not only hamper follow-on R&D but also affect social welfare by extending the duration of market exclusivity of original drugs when selection patents are filed subsequently.

In this context, it is worth pointing out that the legislation on the patentability of selection patents varies substantially across jurisdictions and it is beyond the scope of this paper to provide a comprehensive overview. The lowest common denominator of the existing case law on selection from a broader genus is that a novel selection (that is patentable) is possible only if the prior art document does not already contain an individualized disclosure. We refer to

Zeman and Zimmer (2008) and Fitt (2010) for a discussion of the case law at the EPO and in Germany. Rogers (2014) discusses the situation in the US, Wyld (2009) for the UK.[9]

3. Data and descriptive statistics

3.1 Empirical context and data construction

Since the most frequently used patent databases normally do not provide specific information on the existence of Markush structures in patent documents or on their chemical formulation, we had to access specialized chemistry databases.[10] Currently, only two databases offer machine-readable information on the complete chemical Markush structure. MARPAT is maintained by the American Chemical Society via its Chemical Abstract Service (CAS) and holds information on more than 1.3 million Markush structures.[11] Clarivate's DMR is more comprehensive and covers 1.8 million Markush structures.[12] DMR indexes Markush structures contained in patents' claims (rather than Markush structures in the description). The patent description might be consulted by the analyst who carries out the indexing to help clarify the claims if necessary. We rely on Clarivate's DMR because it allows us not only to identify patent documents with Markush structures, but also to enumerate all compounds they contain to derive a novel measure of their protective scope.

[9] We also refer to the examination guidelines of the different patent offices, for example, https://www.epo.org/law-practice/legal-texts/html/guidelines/e/f_v_3_2_5.htm, https://www.uspto.gov/web/offices/pac/mpep/s2117.html or http://www.ariga.co.jp/en/files/html/html-14/index.html (all accessed on October 27, 2021).

[10] This applies to databases like Google Patents, the NBER patent database, or the EPO's PATSTAT, but also to commercial databases such as Derwent World Patents Index, Lexis Nexis IP or PatSeer. The USPTO's Patent Claims Research Dataset is an exception. It contains a variable counting how often claims use the expression "*selected from*", which is sometimes (but not always) used to formulate Markush claims. It does not contain, however, information on the chemical structures of claimed molecular structures needed to compute the number of molecules contained in a Markush structure.

[11] See https://www.cas.org/support/documentation/markush, latest visit 22nd of March 2021.

[12] See https://support.clarivate.com/Patents/s/article/Derwent-Markush-Resource-Documentation, latest visit 22nd of March 2021.

To assemble our dataset, we start with identifying all EPO filings of pharmaceutical patent applications (International Patent Classification (IPC) code A61K, (Schmoch, 2008)) in the PATSTAT database. Our observational period is restricted to the years from 1992 to 2008 to allow for a time window large enough to observe post-filing outcomes (i.e., the outcome and duration of the patent prosecution process and patent citations). In total, we identify 93,915 patent applications in PATSTAT and collect bibliographic and procedural information on these filings (see below).[13] For each application, we identify whether it, or one of its DOCDB family members, are included in the DMR dataset. 32,258 EPO filings (or family members) have a corresponding entry in the DMR. We excluded DMR patents claiming single compounds which are not Markush structures.[14] Moreover, to account for the fact that a non-markush patent might list more than one specific compound we classify only patents claiming ten and more molecules as Markush. Robustness checks confirm that altering this cut-off point does not change the pattern of our findings. Overall, we identify a total of 22,259 patents claiming Markush structures, or about 23.7% of all pharmaceutical patent applications filed at the EPO between 1992 and 2008.[15]

For drug-level analysis, we then link the 93,915 pharmaceutical patent applications via their DOCDB families to drug development projects in Clarivate's Cortellis database.[16] Cortellis (as of 2018) covers 44,764 different compounds with associated drug development histories. While Cortellis links drugs to the underlying patents, this link is incomplete and comprises only 14,149 compounds (about 31.6% of all compounds listed in Cortellis). Of these

[13] Note that Clarivate's DMR contains Markush structures that have been extracted from priority patent filings at 33 international patent offices (Barth et al., 2016). According to Clarivate, this approximates the population of all Markush structures used in patent filings around the globe.

[14] Note that DMR, for the period before 1999, also contains patents claiming single compounds which are not Markush structures. Since the introduction of the Derwent Chemistry Resource (DCR) database in 1999, these patents enter DCR while Markush structures remain in DMR.

[15] DMR contains the Markush structures contained in priority filings. We verified that EP equivalent filings contain the corresponding Markush structure by manually comparing 100 randomly selected, non-EP priority filings contained in the DMR with the respective filings at the EPO. In all cases, the Markush structures were identical.

[16] Not all pharmaceutical patents will lead to subsequent (pre)clinical trials. Cortellis does not provide patent information on all drug development projects it contains.

14,149 compounds, 8,754 compounds are associated with at least one application in the population of all 93,915 pharmaceutical patent applications filed at the EPO during the 1992-2008 period. About half of the 8,754 compounds (i.e., 4,319 compounds) are protected by one or more Markush structures, which underscores the importance of Markush structures in the protection of drugs.

3.2 Variables

3.2.1 Dependent variables

Duration of the patent examination process

To investigate whether Markush structures are related to increased examination times of patent applications as indicated by the patent offices, we study the duration of patent prosecution: After a patent application has been filed, it is examined by a patent examiner. Harhoff and Wagner (2009) give a detailed discussion of patent examination at the EPO; if an application is not withdrawn by the applicant, the examination process can lead to a patent grant or a refusal. In our data, the *status of each application* is known up until October 2020. The examination was not terminated by this date only for a small fraction of patent applications, so these applications are coded as pending. PATSTAT includes the filing date of a patent application and the termination date of the subsequent examination procedure. We compute the *duration of the examination period* as the difference between the two dates measured in fraction of years.

Patent citations

Fine-grained patterns of citations derived from patent data allow us to provide first evidence on whether Markush structures impact follow-on R&D. Separate from a patent filing's overall impact (total number of citations), we identify to what extent Markush patents create a barrier for (or 'block' on) follow-on R&D by limiting the scope of patentability of subsequent

related inventions as follows: A significant part of patent examination is devoted to determining the novelty of a claimed invention in relation to the existing state-of-the-art. The state-of-the-art is documented in a so-called search report, issued during patent examination. The report contains a list of references to relevant prior patents or non-patent literature, such as scientific articles. This information allows the computation of how often a focal patent is cited by subsequent patent applications.[17] A first indicator describing patterns of citations is the total *number of citations* that a patent document receives over a five-year period by subsequent patent documents. We compute the number of citations a patent application receives as the sum of all citations to the DOCDB patent family it belongs to within five years of its date of filing. To avoid double counting citations originating from equivalent patent filings belonging to the same DOCDB patent family, we consider only citations from different patent families. Existing literature has firmly established that the number of citations not only indicates a patent's impact on subsequent R&D but is also (moderately) correlated with monetary value (Gambardella et al., 2008; Hall et al., 2005; Harhoff et al., 1999; Trajtenberg, 1990).

A very important feature of European search reports is the allocation of search codes to each reference, signifying relevancy to the patent application in question. Most importantly, Type X citations refer to prior art documents, which, taken by themselves, call novelty or the inventive step of a patent document's claim into question (see, e.g., Michel and Bettels (2001))[18]. The patent examiner classifies a citation as Type X if the cited document implies that (at least parts of) the invention claimed in the citing patent application is not novel and thus requires either changes to the citing application's claims or renders it completely unpatentable. Consequently, Type X citations have been used to identify the extent to which a

[17] We refer to 'references' when looking backward, and 'citations' when looking forward. A citation linkage between patent filing A and a subsequent filing B indicates that B refers to A, and A is cited by B.

[18] Starting with PATSTAT (April 2011), the EPO introduced Type I citations, which is Type X 'further explained'. In the new categorization, Type X prejudices novelty and Type I prejudices the inventive step (DOCDB User Documentation, 2020; Annex XIV; https://www.epo.org/searching-for-patents/data/bulk-data-sets/manuals.html; latest visit 4th of April 2021). Type I citations were counted as Type X citations.

cited document creates a barrier for (or 'blocks') follow-on R&D by limiting the scope of patentability of follow-on R&D (Grimpe and Hussinger, 2014; Guellec et al., 2012; Harhoff et al., 2016; Wagner and Wakeman, 2016). To analyze the extent to which patent documents containing Markush structures constitute barriers for subsequent R&D, we construct two different measures based on Type X citations. First, we compute the total *number of Type X citations* a patent family receives within five years of its application. Second, we construct a binary variable indicating whether a patent family received *at least one Type X citation* within five years or not.

Number of patent documents linked to a drug

For assessing the degree of fencing around drugs as influenced by Markush patents, we rely on the Cortellis database. Following Gaessler and Wagner (2022), we identify all EPO patent documents associated with a given drug from Cortellis.[19] Note, that drugs are often tested against different medical indications. Cortellis assigns patents to drugs on the drug-indication level, as some patents might be relevant only for selected indications. The total *number of patent documents* from unique patent families (for a drug-indication combination) measures the strength of the patent fence a company creates to protect a drug against generic entry.

3.2.2 Independent variables

Markush characteristics

Markush (0/1): Drawing on Clarivate's DMR database, we construct a variable indicating whether a patent application contains a Markush structure. *Markush* is coded 1 if a patent or a member of its DOCDB family has a corresponding entry in DMR which contains a

[19] Not all the patents assigned to a drug in Clarivate's Cortellis are in our sample. Some of them might have been filed before 1992 or after 2008, or were filed in IPC classes other than A61K.

Markush structure (genus claim) of ten or more molecules and coded 0 otherwise. In robustness checks, we tested alternative cut-off points, which largely confirmed our original findings.

Markush size: Clarivate's DMR provides Markush structures in a proprietary format. We process this information to compute the number of possible combinatorial variants (enumerations) of all the listed substituents at all positions of the molecular skeleton. The resulting count *Markush size* indicates the total number of molecules falling into a Markush structure. This was done by the proprietary software tool 'Smart Enumerator'[20] provided by InfoChem, a Germany-based software company for cheminformatics. As the number of molecules contained in a Markush structure is skewed, we use its logarithm in our regressions.

Patent characteristics

PCT application: We identify whether a patent was applied for under the rules of the Patent Cooperation Treaty (PCT) (Harhoff and Reitzig, 2004). PCT application gives the opportunity to extend patent protection globally in a streamlined process and allows the applicant more flexibility and institutional options.

Family size: The number of countries in which patent protection is sought is typically called family size (Putnam, 1996; van Pottelsberghe de la Potterie and van Zeebroeck, 2008). We compute family size as the number of jurisdictions in which patents within a DOCDB patent family from PATSTAT have been filed.

Number of claims: Each patent contains a set of claims that marks its boundaries. Existing research has associated the number of claims with patent scope (see Marco et al. (2019) for a discussion) and has established an association with various outcomes related to patent prosecution and enforcement. For example, there is a positive relation between the number of

[20] The smart enumerator does not count repeating structures like $(CH_2)_x$, with X given the repetitions of CH_2. Markush size thus is a lower bound for the total number of molecules contained in a Markush structure.

claims and the duration of patent examination (Harhoff and Wagner, 2009). We therefore included the *number of claims* included in a focal application in our regressions.

Number of 4-digit IPC classes: Patent documents are often assigned to more than one IPC class to capture all features relevant to the technical subject of the claimed invention. While all patent filings in our sample have been assigned to the A61K class, we include the total number of IPC4 classes they have been assigned to as a measure of technological breadth.

Number and composition of patent references: The search report published by the EPO yields information on prior art by referencing previous patent documents or non-patent literature. We consider the total number of patent references ('backward citations') as well as references to non-patent documents in our regressions. On the one hand, references are a measure of the examiner's search effort (Risch, 2007) and thus can be expected to increase the duration of examination. On the other hand, the composition of the references conveys important information on the quality of the application. If a search report contains many Type X or Type Y references, the claimed invention may not meet the requirements of novelty or inventive step.[21] A large share of Type X or Type Y references may affect the duration of patent examination (Harhoff and Wagner, 2009). For this reason, we not only include the absolute number of family-level references to patent and non-patent literature in our analyses, but also the family-level share of X and Y references contained therein.

Year of application: All our regressions include indicator variables for the year in which an application was filed at the EPO. These time-fixed effects allow us to control for unobservables which vary over time, such as the workload situation at the EPO (Harhoff and Wagner, 2009) or changes in the rules governing the patent examination process.

[21] Like Type X references, Type Y references limit the patentability of subsequent inventions but only in conjunction with other documents and not taken alone.

Applicant characteristics

Annual number of patent applications filed at the EPO: The ECOOM-EUROSTAT-EPO PATSTAT Person Augmented Table (EEE-PPAT) provides a harmonization of the applicant names listed on EP patents (for a full description, see (Magerman et al., 2010)). We use these harmonized names to compute the cumulative number of patent applications filed by a patent applicant on an annual basis (in '00s) as a rough proxy of size.

Applicant type: Based on the EEE-PPAT data, we classify applicants according to their organizational form as companies, government/non-profit, university/hospital, and individuals. This information allows us to control for organizational form by including indicator variables for each category in our regressions using 'company' as a reference category.

Country of origin: Finally, we characterize applicants by their country of origin and distinguish between applicants located in Europe, the US, Japan, and 'the rest of the world' (ROW). We include indicators in our regressions with Europe being the reference category.

Drug characteristics

Last stage of development: A major advantage of the Cortellis database is that it tracks drug development projects through different stages of clinical trials independently of whether they are successful or not (Gaessler and Wagner, 2022). It is important to account for a drug's last stage of development as companies often add additional patent filings to a development project as it progresses through the development funnel. We capture this by a set of indicator variables reflecting whether a drug's development was successful in completing preclinical trials, Phase I, II, or III clinical trials, and was eventually launched. These indicators are created on the drug-indication level as a drug is often tested separately against different indications.

Technology: We distinguish between small molecule drugs and biologicals, introducing a biologics indicator variable equaling 1 for biologics and 0 otherwise, based on Cortellis' classification.

ICD classification: Drugs are often tested against different indications. To control for structural differences in different medical indications, we map Cortellis' drug indications to their International Statistical Classification of Diseases and Related Health Problems (ICD-9) condition codes. Based on this mapping, we include an indicator equal to 1 for a drug's first indication it has been tested against and a set of indication fixed effects based on aggregate ICD-9 levels (Krieger et al., 2022) to our multivariate analyses.

3.3 Descriptive Statistics

Figure 2 presents the annual number of pharmaceutical patent filings at the EPO for the years 1992 to 2008. Patenting in the pharmaceutical area has increased strongly over the observation period with a small decline during the financial crisis starting in 2007. The dashed lines in Figure 2 represent the total number of EP patent applications containing Markush structures contained in Clarivate's DMR, further broken down into the number of EP filings with Markush structures and the number of EP filings with a DOCDB family member containing Markush structures.

[INSERT FIGURE 2 ABOUT HERE]

Regarding the annual number of EP patent applications contained in Clarivate's DMR, two observations deserve attention. First, the number of filings contained in Clarivate's DMR is low in 1997 compared to the overall development of the time series for all pharmaceutical patent documents at the EPO. According to Clarivate, the drop-off in entries in Clarivate's DMR in 1997 is a one-time effect caused by a change in the data collection approach preced-

ing the release of the DCR database. Second, while the annual number of entries increases between 1998 and 2007, the increase is slower than the increase in overall patenting. Consequently, the relative share of EP patent filings that contain Markush structures steadily decreases in this period. We suspect that this is driven by an increasing importance of biological drugs in the pharmaceutical industry, which are made of complex antibodies, peptides, and so forth, for which Markush formulae are not as relevant (compared to small molecule drugs).[22] Further, drug lifecycle management and repurposing activities gained in importance relative to R&D dedicated to developing original drugs from around 2000 onwards (Dureja and Prajapati, 2012; Sleigh and Barton, 2010). This trend may be associated with an increase in so-called secondary patents protecting minor improvements or novel applications for existing compounds, which typically do not contain Markush structures. Amplifying these effects, in 1995, the EPO's Board of Appeals restricted the permissible scope of Markush claims in its AgrEvo vs. Triazole (T 0939/92) decision, which reduced the attractiveness of filing Markush claims at the EPO.[23]

[INSERT FIGURE 3 ABOUT HERE]

Markush structures contain an average of 7.62*10^39 different molecules (about 91.831 on the logarithmic scale), but the number of molecules in a Markush structure is highly skewed with a median number of molecules equaling 288,540 different molecules (about 12.572 on the logarithmic scale). Figure A.3 in the Appendix presents a histogram of the size distribution of Markush structures in our data. The median size of Markush structures

[22] We collected data from the US Food and Drug Administration (FDA) on drug approvals, distinguishing NMEs (new molecular entities such as small molecule drugs) and BLAs (biologics license applications). Figure A.2 shows the results for FDA approval years 1985-2019. NMEs based on biological drugs gain importance over time.

[23] In this decision, the board concluded that the technical effect on which an invention relies needs to exist over the whole chemical area claimed in Markush structure, see https://www.epo.org/law-practice/case-law-appeals/recent/t920939ex1.html, latest visit 7th of April 2021.

is increasing over time (see Figure 3). In the early 1990s, the median number of molecules (logarithm) in a Markush structure was about 9 but sharply increased from 1998 onwards. This increase can be explained by two major technological advancements around this time. First, breakthroughs in combinatorial chemistry and high-throughput sequencing during the 1990s, combined with improved software tools (Galande, 2000), considerably extended the boundary of what type of molecules can be synthesized (Martin et al., 1995). Second, software support for drafting Markush structures became available in the mid-1990s and facilitates the design of Markush structures of vast sizes (Newall, 1996).

In Table 1, we present average durations of the patent examination process (our first dependent variable) for Markush and non-Markush patents, distinguishing patent grant, rejections of the application, and withdrawals. Table 1 reveals that – contrary to expectations – patent applications with Markush structures are examined faster than non-Markush filings across all outcomes. For instance, examination procedures of Markush applications leading to a patent grant take 6.02 years, while the duration is 6.66 years for non-Markush applications. The difference is similar for refusals with 6.21 years for Markush applications vs. 6.98 years for non-Markush patents, but less pronounced for applicant-driven withdrawals (5.43 years vs. 5.85 years). Moreover, Table 1 reveals that applications containing Markush structures have a higher average grant rate than non-Markush filings (49.25% vs. 47.76%) excluding pending cases, but a lower average withdrawal rate (47.25% vs. 47.81%).

[INSERT TABLE 1 ABOUT HERE]

[INSERT TABLE 2 ABOUT HERE]

Table 2 reports patterns of citations for Markush and non-Markush patent applications. Markush patent filings receive a higher number of total citations from subsequent patent documents within five years of application than non-Markush patent documents (3.300 citations

vs. 2.730 citations). This could be interpreted as an indication of a higher value of Markush patent filings (Hall et al., 2005; Harhoff et al., 1999; Trajtenberg, 1990). To scrutinize to what extent Markush structures affect follow-on R&D, we compute the average number of Type X citations received within five years as well as the share of patent documents that have received at least one Type X citation. Markush patent applications receive, on average, a significantly higher number of Type X citations than non-Markush patent documents. However, the difference is small (0.712 Type X citations vs. 0.627 Type X citations). We observe a similar pattern when focusing on the likelihood of receiving at least one Type X citation with a higher share of Markush patent filings compared to non-Markush filings (39.40% vs. 35.91%). While the difference is not large, it is a sign that Markush structures potentially negatively affect follow-on R&D.

[INSERT TABLE 3 ABOUT HERE]

Turning to the question of protecting drugs against product market competition, and ultimately generic entry, we report the number of patents assigned to drugs in the Cortellis database (on the drug-indication level) as a measure of patent fences in Table 3. We have been able to identify the number of patents for a total of 26,066 drug-indication combinations. We report how these 26,066 development projects advance through the stages of clinical trials. With all projects in Cortellis having entered preclinical trials (100%), we observe the typical attrition with only 70.9% of projects entering Phase I trials, 57.4% Phase II trials, 27.9% Phase III trials, and only 18.1% of all projects leading to market approval.[24] Table 3 shows stark differences between drug projects that are associated with at least one Markush patent filing, and those which are not. Reflecting companies' strategies to build patent fences, the

[24] This relatively high approval rate is a consequence of Cortellis oversampling of successful drugs when linking patents to drug development projects (Gaessler and Wagner, 2022).

number of patents associated with a given development project is increasing as a drug advances in the development funnel. Table 3 shows that development projects associated with Markush patents are not only characterized by higher chances of success at each stage of the development funnel, but also have a significantly higher number of patents associated with them. For instance, while drugs that are not linked to a Markush filing are protected by an average of about 9.39 patents at market approval, Cortellis links 68.26 patents to projects with at least one Markush patent. This can be interpreted as a sign that companies use Markush patents strategically to establish larger patent fences surrounding drugs. Note, however, that Cortellis is not a regulatory document like the list of Approved Drug Products with Therapeutic Equivalence Evaluations ('Orange Book') published by the U.S. Federal Drug Administration under the Hatch-Waxman Act. Cortellis has been curated to facilitate freedom-to-operate searches and competitive landscaping and might therefore include additional patents not associated to drugs in specific regulatory filings.

4. Multivariate analyses

4.1 Overview

We employ multivariate regressions to estimate the effect of Markush structures on important outcomes such as duration of the examination procedure, citation patterns, and the construction of patent fences. For the different outcomes *y*, we estimate regressions of the form

$$E(y|X) = \beta_0 + \beta_1 * (\text{Markush}{=}1) + \beta_2 * \text{Markush\_Size} + \zeta X + \varepsilon,$$

where β_1 identifies the difference between Markush and non-Markush patent documents, while β_2 captures the effect of the number of molecules comprised in a Markush structure. ζ is a vector of coefficients corresponding to X the vector of controls. For patent-level outcomes (duration of examination and citation patterns), these include patent-specific procedural and bibliographic information, and applicant information and yearly indicator variables.

For drug-indication-level outcomes (patent fence), these include drug-specific and applicant information, as well as yearly indicator variables.

4.2 Duration of patent examination

We follow Harhoff and Wagner (2009) and model the duration of patent examination using survival models, as our data contains pending cases for which the examination process is still ongoing (see Table 1) which needed to be excluded in common regression frameworks despite conveying information on process durations. Employing accelerated failure time (AFT) models allows us to address this problem. AFT models express the natural logarithm of survival time as a linear function of the covariates X with $\ln T = X\beta + \varepsilon$. If ε follows a logistic distribution, the log-logistic regression model is obtained (Kalbfleisch and Prentice, 2011). We need to further consider that the process of patent examination can be terminated by three different mutually exclusive outcomes. Therefore, we apply competing risks models which are based on different random variables T, describing the duration until examination ends via one of the three exits: grant, withdrawal, or refusal (Cox and Oakes, 1984).

[INSERT TABLE 4 ABOUT HERE]

Table 4 reports the coefficient estimates from the AFT competing risks models. The coefficients can be interpreted approximately as percentage change of process durations, with negative coefficients reducing process durations and positive coefficients indicating longer delays. We report two specifications for each possible outcome of the patent examination procedure. In Columns 1, 3, and 5 of Table 4, we present results relating important patent and applicant characteristics to observed process durations, without including our Markush variables. Similar to Harhoff and Wagner (2009), we find that patent characteristics indicating higher private value (PCT application, family size) correlate with faster patent grants, but delay withdrawals. For valuable applications, patentees have an incentive to speed up processes

that can be expected to lead to a patent grant. On the other hand, if a patent application is of higher value, an applicant will be less willing to withdraw an application early on.

Variables correlated with the complexity of the examination task, such as the number of claims to be examined, the number of references (to previous patent filings and to non-patent literature), or technological complexity (indicated by the number of different 4-digit IPC classes assigned to an application), are associated uniformly with longer examinations. Our findings also indicate that the composition of references contained in a search report affect process durations; a larger share of novelty-destroying Type X references (both to patent literature as well as non-patent documents) increases process durations leading to patent grants while having a much less pronounced shortening effect on the withdrawals.

In Columns 2, 4, and 6 of Table 4, we report the results from specifications that include our Markush indicator, and additionally, the log of the number of molecules contained in the Markush structure. Markush characteristics are significantly associated with process durations across all outcomes, but do not change the effect of the other variables. In line with the descriptive findings reported in Table 2, the results from our multivariate duration analyses suggest that applications with Markush structures are processed faster by the EPO than pharmaceutical applications without. For instance, grant decisions are issued about 6.73% (-0.0745 + 0.000576*12.572 = -0.0673, see column 2 of Table 4) or about 5.2 months faster if an application contains a Markush structure. The size of the Markush structure further reduces the duration until a grant decisions and withdrawals, but it significantly delays refusals of applications. A patent filing containing a Markush structure of median size (12.572 on the logarithmic scale) is withdrawn about 9.00% faster (-0.0900 = -0.0686 - 0.0017*12.572), see column 4 of Table 4), and refused about 4.17% later (0.0417 = 0 + 0.00332*12.572, see column 6 of Table 4) compared to non-Markush patent applications. (Simmons, 1991, 2003) argues that, compared to text-based chemical claims, Markush structures are less ambiguous in their meaning and thus prior art search aided by software-based searches in chemical databases

(possible from the early 1990s onwards, Tokuno (1993)) is easier. This might explain shorter examination durations.

4.3 Type X citations

The number of Type X-citations received by a patent document within five years of its filing is a count variable and we choose count data regression models accordingly. The distribution of citations is characterized by overdispersion, which makes the more flexible negative binomial regression model a preferable choice over the more rigid Poisson model (Hilbe, 2011). We apply it to model the count of Type X citations received within five years. Note that in negative binomial regressions, the difference in the logs of expected counts changes by the respective regression coefficient, given the other predictor variables in the model are held constant. For small values, coefficients thus approximate the percentage change in the expected count of citations received. To rule out that our findings regarding Type X citations are simply be driven by the overall number of citations, we additionally model the likelihood that a patent document receives at least one Type X citation, an indicator that it was used as to restrict the patentability of follow-on R&D at least once, in probit regressions.

[INSERT TABLE 5 ABOUT HERE]

Columns 1 and 2 of Table 5 report the estimates obtained from negative binomial specifications. In line with existing literature, we find that indicators such as the DOCDB family size, measures of patent breadth (e.g., claims and the number of different IPC 4 classes a patent application has been assigned to), and the number of references contained in a patent document are positively associated with the number of Type X citations received. This holds for the likelihood of receiving at least one Type X citation (see Columns 3 to 6 of Table 5).

Regarding Markush characteristics, we find a significant effect on both the number and occurrence of Type X citations. Markush patent documents receive significantly more

Type X citations than non-Markush patent documents (see Column 2 of Table 5). In addition to the highly significant Markush dummy, the overall number of molecules contained in a Markush structure further increases the number of Type X citations received. A median-sized Markush patent document receives 0.0996 more Type X citations and has a 2.66% (0.0192 + 0.000587*12.572 = 0.0266) higher likelihood of receiving at least one Type X citation than non-Markush patent documents, see Column 6 of Table 5. Compared to an unconditional mean of 39.40% (Table 2), this corresponds to a relative increase of 6.75%. As highlighted above, Type X citations indicate that the cited document restricts the patentability of follow-on R&D. As Markush patent documents are thus more likely to be used as novelty-destroying prior art for follow-on innovation, they might make follow-on R&D less attractive. This, aligned with existing literature, suggests that broad patents might slow down follow-on innovation (Merges and Nelson, 1990; Scotchmer, 1991).

4.4 Patent fences

We model the number of patents associated to a given drug development project ('patent fence') subject to drug development and company characteristics using negative binomial regression models that account for overdispersion (Table 6). The dependent variable is constructed on the drug-indication level as a drug is often tested against different indications. As a result, the same drug can be contained multiple times in the regression. We therefore cluster standard errors on the drug level.

Reflecting the descriptive findings of Table 3, we find that the number of patents associated with a development project increases as a project progresses through the development funnel. The coefficients on the indicators of a project's last stage of development are increasing monotonously from Phase I to Approval (with preclinical trials being the reference group) in all specifications of Table 6. For approved drugs, patent fences contain between 149% and 106% more patents compared to drugs that fail in preclinical trials. This reflects the increasing

effort of companies to delay generic entry by increasing the number of patents surrounding as approval becomes more likely. Patent fences are significantly smaller when a drug is tested for the first time, as indicated by the negative coefficient of the first indication variable. This finding confirms existing evidence on lifecycle management strategies of companies seeking to find additional indications for drugs that have been approved for a first indication. Biologicals are protected by a larger number of patents, though the effect is significant only once we include the Markush variable in the regressions.

In Columns 2, 3, and 4 of Table 6, we gradually add characteristics of Markush patent filings associated with drug development projects. Column 2 of Table 6 highlights that drugs associated with at least one Markush patent filing are characterized by a patent fence that comprises significantly more patents (about 117.7%) compared to drugs without Markush patents. Column 4 of Table 6 provides a more nuanced view of how Markush patent documents are related to patent fences, as it additionally contains the number of different Markush patents associated with a drug development project, as well as the average number of molecules they comprise (on a logarithmic scale). Column 3 shows that – in addition to the baseline effect of having at least one Markush patent associated to a drug – each additional Markush patent filing increases the total number of patents by another 7.07%. The size of the patent fence surrounding a drug development project is also positively and significantly related to the size of a Markush structure, though the effect size is small. An increase in the average number of molecules in the Markush patent documents linked to a drug by one unit on the logarithmic scale further increases the number of patents surrounding a drug by about 0.357%. While not directly testing underlying patent filing strategies, these results suggest that Markush patent filings inherently provide sufficient opportunities for creating multiple subsequent patent filings surrounding a given drug.

[INSERT TABLE 6 ABOUT HERE]

5. Discussion and conclusion

As genus claims, Markush structures allow patentees to claim a large class of chemical compounds without writing out every single molecule. While beneficial from the viewpoint of research active pharmaceutical companies, the use of Markush patents has been discussed controversially by policy makers and legal scholars. Concerns have been raised regarding the complexity of examining Markush claims in patent applications and their effect on follow-on innovation and generic entry. Despite the relevance, the topic has received little attention by economists interested in patenting strategies specifically, and innovation more broadly. The dearth of empirical evidence on the use of Markush structures in pharmaceutical patenting can be linked to sparsity of available data sources and the difficulty in processing chemical information on Markush structures.

In this paper, we present a first study using chemical information on Markush structures contained in patent filings on a large scale. We have obtained chemical information on the population of Markush structures contained in EP patent filings between 1992 and 2008 and have used cheminformatics tools to compute the total number of molecules contained in a Markush structure as their key characteristic. Linking Markush structures to patent information allows us to derive important insights for the ongoing policy discussion. First, we find little evidence that Markush patent applications cause significant frictions in the patent examination process despite concerns expressed by patent offices. Applications containing Markush structures are processed faster by patent offices. These findings not only contribute to the existing literature on the functioning of the patent system in general and the examination process in particular (e.g., (Frakes and Wasserman, 2019; Guerrini, 2013; Hall et al., 2004; Mann and Underweiser, 2012; Wagner, 2009), but they also alleviate concerns that Markush structures represent a substantial burden for patent examiners. Second, citation patterns are informative

on Markush patent documents' impact on subsequent R&D. Focusing on Type X citations exclusively, which indicate a document's potential to restrict the patentability of follow-on R&D, we find similar results. Patent filings claiming Markush structures are more frequently cited as Type X than other patent applications. This indicates their increased potential to hamper follow-on R&D compared to non-Markush patents. Third, we rely on information on how many patent documents are associated with drug development projects in the Cortellis database to study how Markush patent documents are related to specific IP strategies of pharmaceutical companies. We find a strong association between how many patents surround a development project ('patent fence') and the presence of Markush filings related to a drug. While existing literature suggests that the accumulation of patent rights surrounding a drug product is a strategy that is often employed to delay generic entry (Burdon and Sloper, 2003, Howard, 2007, European Commission, 2009, Sternitzke, 2013, Ellery and Hansen, 2012, Kapczynski et al., 2012, Gupta 2020), our data does not allow us to directly measure the actual delay caused by patenting strategies relying on Markush claims. This would require additional data collection efforts which are beyond the scope of this paper. Moreover, the correlation between the number of patents surrounding drug development projects reported in Cortellis and the presence of Markush patent(s) might be driven by unobserved variables, such as market size, that drive the filing of Markush claims as well as the creation of comprehensive patent fences.

Our study is not without further limitations. First, while not trivial to compute, measuring the size of a Markush structure simply by the number of molecules it contains is a simplistic approach. Two Markush structures containing a comparable number of molecules might in fact differ significantly depending on their chemical structures. If all molecules in a Markush structure are highly similar, the Markush structure claims only a 'narrow' chemical space. If the molecules it contains, however, are highly dissimilar, the chemical space claimed is much larger. Future research must address this shortcoming by exploiting existing chemical

information to a greater extent. A measure of chemical scope can be developed, for example, by computing the maximum dissimilarity within a Markush structure obtained from all pairwise comparisons of its molecules (Krieger et al., 2022). Further, our study is confined to associations between the inclusion of Markush structures in patent filings and our outcomes of interest. The identification of causal relations requires a more refined identification strategy than the cross-sectional regressions presented here. In future research, we seek to further expand our research on identifying the causal effect of Markush structures on follow-on innovation. Linking patents to drug development and associated target compounds will allow us to devise more complex identification strategies and also to study R&D competition between firms more closely. An important question in this regard is to what extent the existence of broad Markush patents within a certain mechanism of action will shift competitors' R&D activities to different, possibly novel, drug classes trying to exploit different mechanisms of actions. We are unable to address this type of question with the data assembled for this paper.

References

Barth, A., Stengel, T., Litterst, E., Kraut, H., Matuszczyk, H., Ailer, F., Hajkowski, S., 2016. A novel concept for the search and retrieval of the derwent markush resource database. Journal of chemical information and modeling 56, 821-829.

Berks, A.H., 2001. Current state of the art of markush topological search systems. World Patent Information 23, 5-13.

Bessen, J.E., Meurer, M.J., 2008. Patent failure: How judges, bureaucrats, and lawyers put innovators at risk. Princeton University Press.

Bone, R.G.A., Kendall, J.T., 2008. Markush under threat: Us pto considers alternatives. Industrial Biotechnology 4, 246-251.

Braga, E.J., de Souza, A.R., de Lima Soares, P.L., Rodrigues, R.C., 2018. The role of specification in patent applications: A comparative study on sufficiency of disclosure. World Patent Information 53, 58-65.

Burdon, M., Sloper, K., 2003. The art of using secondary patents to improve protection. Journal of Medical Marketing 3, 226-238.

Cohen, W.M., Nelson, R.R., Walsh, J., 2000. Protecting their intellectual assets: Appropriability conditions and why us manufacturing firms patent (or not). NBER working paper.

Correa, C., 2008. Guidelines for the examination of pharmaceutical patents: Developing a public health per-spective. WHO – ICTSD – UNCTAD Working paper.

Cox, D.R., Oakes, D., 1984. Analysis of survival data. CRC press.

Crum-Brown, A., Fraser, T., 1865. The connection of chemical constitution and physiological action. Transactions (Royal Society of Edinburgh) 25, 257.

Dhulap, S., Kulkarni, M., 2019. Nonobviousness of pharmaceutical inventions: Implications for patent prosecution and litigation. Pharmaceutial Patent Analyst 8, 91-107.

Dureja, H., Prajapati, V., 2012. Product lifecycle management in pharmaceuticals. Journal of Medical Marketing: Device, Diagnostic and Pharmaceutical Marketing.

Ellery, T., Hansen, N., 2012. Pharmaceutical lifecycle management: Making the most of each and every brand. John Wiley & Sons.

European, C., 2003. Third european report on science & technology indicators, Towards a knowledge-based economy, Brussels.

EuropeanCommission, 2009. Pharmaceutical sector inquiry - final report. https://ec.europa.eu/competition/sectors/pharmaceuticals/inquiry/staff_working_paper_part1.pdf

Federal-Register, 2007. 72, 44992-45001.

Fitt, R., 2010. Selection patents and markush claims in europe. Biotechnology Law Report 29, 17-20.

Frakes, M.D., Wasserman, M.F., 2019. Irrational ignorance at the patent office. Vanderbilt Law Review 72, 975-1030.

Gaessler, F., Wagner, S., 2022. Patents, data exclusivity, and the development of new drugs. The Review of Economics and Statistics, (in print).

Galande, A., 2000. Combinatorial technology: An overview. Indian Journal of Pharmaceutical Sciences 62, 84-91.

Gambardella, A., Harhoff, D., Verspagen, B., 2008. The value of european patents. European Management Review 5, 69-84.

Grimpe, C., Hussinger, K., 2014. Resource complementarity and value capture in firm acquisitions: The role of intellectual property rights. Strategic Management Journal 35, 1762-1780.

Guellec, D., Martinez, C., Zuniga, P., 2012. Pre-emptive patenting: Securing market exclusion and freedom of operation. Economics of Innovation and New Technology 21, 1-29.

Guerrini, C.J., 2013. Defining patent quality. Fordham Law Review 82, 3091-3141.

Hall, B.H., Graham, S., Harhoff, D., Mowery, D.C., 2004. Prospects for improving us patent quality via postgrant opposition. Innovation Policy and the Economy 4, 115-143.

Hall, B.H., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. RAND Journal of Economics 36, 16-38.

Harhoff, D., Narin, F., Scherer, F.M., Vopel, K., 1999. Citation frequency and the value of patented inventions. Review of Economics and Statistics 81, 511-515.

Harhoff, D., Reitzig, M., 2004. Determinants of opposition against epo patent grants-the case of biotechnology and pharmaceuticals. International Journal of Industrial Organization 22, 443-480.

Harhoff, D., von Graevenitz, G., Wagner, S., 2016. Conflict resolution, public goods, and patent thickets. Management Science 62, 704-721.

Harhoff, D., Wagner, S., 2009. The duration of patent examination at the european patent office. Management Science 55, 1969-1984.

Hemphill, C.S., Sampat, B.N., 2011. When do generics challenge drug patents? Journal of Empirical Legal Studies 8, 613-649.

Hemphill, C.S., Sampat, B.N., 2012. Evergreening, patent challenges, and effective market life in pharmaceuticals. Journal of Health Economics 31, 327-339.

Hilbe, J.M., 2011. Negative binomial regression. Cambridge University Press.

Holman, C.M., 2017. In defense of secondary pharmaceutical patents: A response to the un's guidelines for pharmaceutical patent examination. Indiana Law Review 50, 759.

Howard, L., 2007. Use of patents in drug lifecycle management. Journal of Generic Medicines 4, 230-236.

Hughes, J.P., Rees, S., Kalindjian, S.B., Philpott, K.L., 2011. Principles of early drug discovery. British journal of pharmacology 162, 1239-1249.

Kalbfleisch, J.D., Prentice, R.L., 2011. The statistical analysis of failure time data, 2nd ed. John Wiley & Sons.

Kapczynski, A., Park, C., Sampat, B., 2012. Polymorphs and prodrugs and salts (oh my!): An empirical analysis of "secondary" pharmaceutical patents. PLOS ONE 7.

Karshtedt, D., Lemley, M.A., Seymore, S.B., forthcoming. The death of the genus claim. Harvard Journal of Law and Technology.

Krieger, J.L., Li, D., Papanikolaou, D., 2022. Missing novelty in drug development. The Review of Financial Studies 35, 636–679.

Kyle, M.K., 2016. Competition law, intellectual property, and the pharmaceutical sector. Antitrust Law Journal 81, 1-36.

Magerman, T., Van Looy, B., Song, X., 2010. Exploring the feasibility and accuracy of latent semantic analysis based text mining techniques to detect similarity between patent documents and scientific publications. Scientometrics.

Mann, R.J., Underweiser, M., 2012. A new look at patent quality: Relating patent prosecution to validity. Journal of Empirical Legal Studies 9, 1-32.

Marco, A.C., Sarnoff, J.D., Charles, A., 2019. Patent claims and patent scope. Research Policy.

Martin, E.J., Blaney, J.M., Siani, M.A., Spellmeyer, D.C., Wong, A.K., Moos, W.H., 1995. Measuring diversity: Experimental design of combinatorial libraries for drug discovery. Journal of Medicinal Chemistry 38, 1431-1436.

Merges, R.P., Nelson, R.R., 1990. On the complex economics of patent scope. Columbia Law Review 90, 839-916.

Michel, J., Bettels, B., 2001. Patent citation analysis. A closer look at the basic input data from patent search reports. Scientometrics 51, 185-202.

Newall, C.E., 1996. Software: Putting structures in a database. Analytical Chemistry 68, 554A-555A.

Prior, K.J., 2012. The uspto's historic struggle with markush claims: Will the 2011 guidelines provide relief? Law School Student Scholarship Paper 114.

Putnam, J.D., 1996. The value of international patent rights, PhD thesis, Yale University.

Risch, M., 2007. The failure of public notice in patent prosecution. Harvard Journal of Technology and Law 21, 179-232.

Rogers, D.L., 2014. Federal circuit's obviousness test for new pharmaceutical compounds: Gobbledygook. Chicago-Kent Journal of Intellectual Property 14, 49-105.

Rollins, A.D., 1985. 35 usc 119-description and enablement requirements. Journal of the Patent & Trademark Office Society 67, 386-394.

Schmoch, U., 2008. Concept of a technology classification for country comparisons. Final report to the World Intellectual Property Organisation (WIPO), WIPO.

Scotchmer, S., 1991. Standing on the shoulders of giants: Cumulative research and the patent law. Journal of Economic Perspectives 5, 29-41.

Simmons, E.S., 1991. The grammar of markush structure searching: Vocabulary vs synthax. Journal of Chemical Information and Computer Sciences 31, 45-53.

Simmons, E.S., 2003. Markush structure searching over the years. World Patent Information 25, 195-202.

Sleigh, S.H., Barton, C.L., 2010. Repurposing strategies for therapeutics. Pharmaceutical Medicine 24, 151-159.

Sternitzke, C., 2013. An exploratory analysis of patent fencing in pharmaceuticals: The case of pde5 inhibitors. Research Policy 42, 542-551.

Tokuno, H., 1993. Comparison of markush structure databases. Journal of chemical information and computer sciences 33, 799-804.

Trajtenberg, M., 1990. A penny for your quotes: Patent citations and the value of innovations. RAND Journal of Economics 21, 172-187.

van Pottelsberghe de la Potterie, B., van Zeebroeck, N., 2008. A brief history of space and time: The scope-year index as a patent value indicator based on families and renewals. Scientometrics 75.

von Graevenitz, G., Wagner, S., Harhoff, D., 2013. Incidence and growth of patent thickets - the impact of technological opportunities and complexity. Journal of Industrial Economics 61, 521-563.

Wagner, R.P., 2009. Understanding patent-quality mechanisms. University of Pennsylvania Law Review 157, 2135-2173.

Wagner, S., Wakeman, S., 2016. What do patent-based measures tell us about product commercialization? Evidence from the pharmaceutical industry. Research Policy 45, 1091-1102.

Wegner, H.C., 1992. Patent law in biotechnology, chemicals & pharmaceuticals. Stockton Press.

Wyld, O., 2009. High court considers markush claims and the doctrine of uk selection patents. Journal of Intellectual Property Law & Practice 4, 147-149.

Zeman, S.M., Zimmer, F.-J., 2008. Chemical selection inventions in the epo and germany: Continued divergence in legal interpretation. Biotechnology Law Report 27, 293-303.

Figures

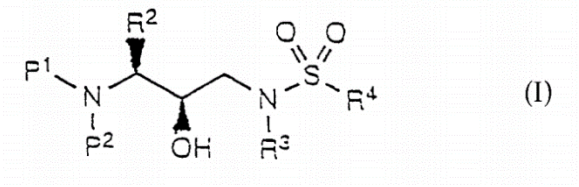


Figure 1: Molecular structure of the compound claimed in Claim 1 of European Patent EP 0 810 209 (B1), which protects Darunavir.

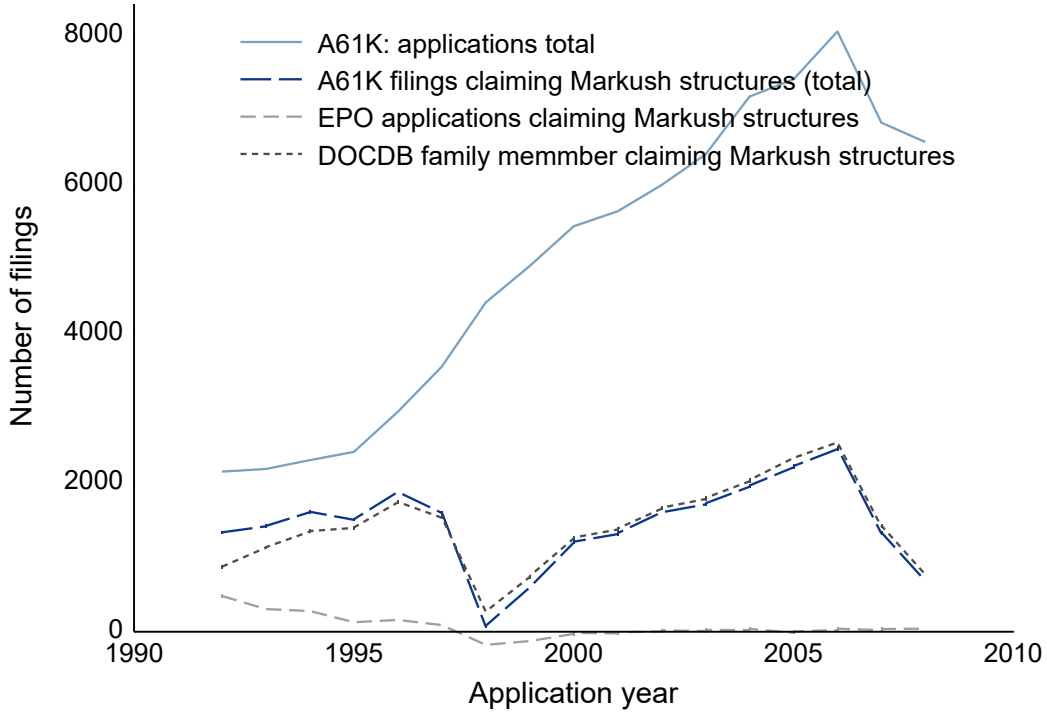


Figure 2: Pharmaceutical patent applications (main IPC A61K) filed at the EPO between 1992 and 2008. The dashed lines represent patent filings containing Markush structures.

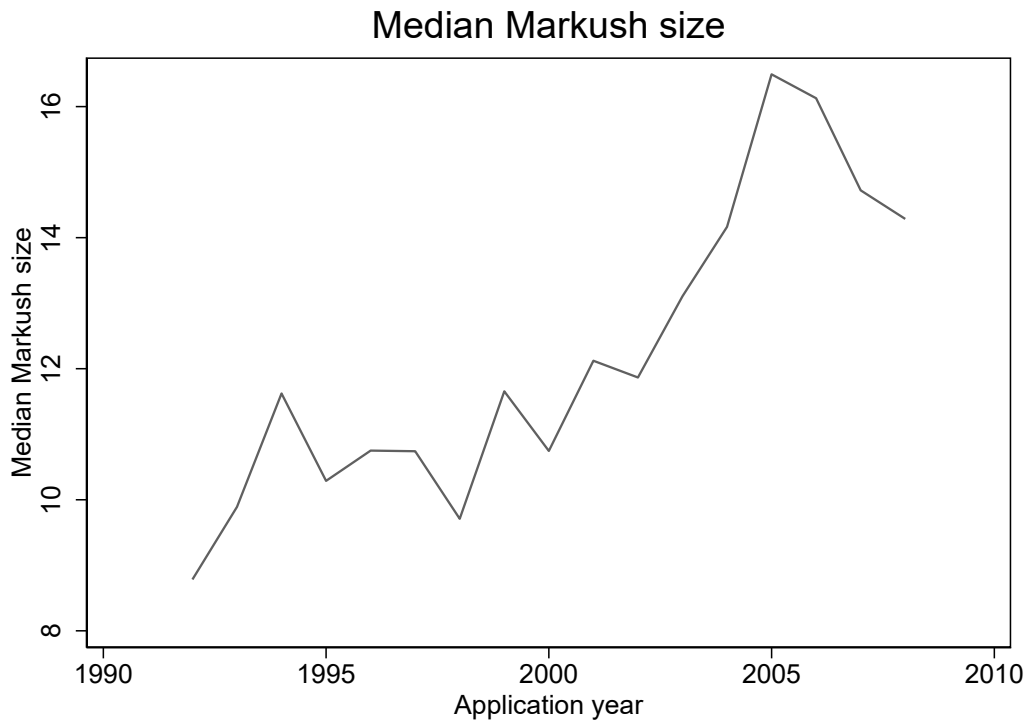


Figure 3: Median of the natural logarithm of the number of molecules claimed in EP patent applications filed between 1992 and 2008 that contain a Markush structure.

Tables

	Outcome of examination			
	Granted	**Withdrawn**	**Refused**	**Pending**
Non-Markush				
Obs.	33,731	33,758	3,125	1,042
Share (excl. pending)	47.76%	47.81%	4.43%	-
Duration	6.66 yrs	5.85 yrs	6.98 yrs	-
Markush				
Obs.	10,894	10,450	774	141
Share (excl. pending)	49.25%	47.25%	3.50%	-
Duration	6.02 yrs	5.43 yrs	6.21 yrs	-
Total				
Obs.	**44,625**	**44,208**	**3,899**	**1,183**
Share (excl. pending)	**48.12%**	**47.67%**	**4.20%**	**-**
Duration	**6.51 yrs**	**5.75 yrs**	**6.83 yrs**	**-**

Table 1: Outcomes of the patent examination procedures of pharmaceutical patents filed at the EPO between 1992 and 2008. Note: We report shares of specific examination outcomes relative to the total number of non-pending application procedures.

		All patents in our sample		
	Obs.	**Citations (5 years)**	**Type X citations (5 years)**	**At least one type X citation (5 years)**
Non-Markush	71,656	2.730	0.627	35.91%
Markush	22,259	3.300	0.712	39.40%
Total	**93,915**	**2.865**	**0.647**	**36.74%**

Table 2: Average number of citations received within 5 years, the number of Type X citations, and the likelihood of receiving at least one Type X citation for all patent documents in our sample.

	Highest development stage reported				
	Preclinical trials	Phase I	Phase II	Phase III	Approval
No Markush patents					
Drugs (drug-indication)	10,988	7,219	5,674	2,694	1,666
Share	100.0%	65.7%	51.6%	24.5%	15.2%
Fence	5.83	6.87	8.96	10.00	9.39
At least on Markush patent					
Drugs (drug-indication)	15,078	11,263	9,287	4,567	3,056
Share	100.0%	74.7%	61.6%	30.3%	20.3%
Fence	16.67	21.53	32.47	44.48	68.26
Total					
Drugs (drug-indication)	26,066	18,482	14,961	7,261	4,722
Share	100.0%	70.9%	57.4%	27.9%	18.1%
Fence	11.39	15.30	23.63	30.95	48.51

Table 3: Number of drug-indication-level drug development projects in our data broken down by highest stage of development. Note, shares refer to the share of projects that entered a specific stage relative to all projects that entered preclinical trials. Note further, fence is referring to the number of unique patent families associated with a drug development project in Cortellis.

Variable	Duration of patent examination					
	Grant		Withdrawal		Refusal	
	(1)	(2)	(3)	(4)	(5)	(6)
Markush characteristics						
Markush (0/1)		-0.0745***		-0.0686***		-0.0249
		[0.00698]		[0.00897]		[0.0250]
Markush size (log)		0.000576*		-0.00170***		0.00332***
		[0.000337]		[0.000406]		[0.00127]
Procedural information						
PCT application (0/1)	-0.164***	-0.162***	0.136***	0.140***	0.0580***	0.0563***
	[0.00490]	[0.00489]	[0.00608]	[0.00607]	[0.0152]	[0.0152]
DOCDB family size	-0.00420***	-0.00413***	0.0331***	0.0333***	0.0228***	0.0228***
	[0.000186]	[0.000186]	[0.000358]	[0.000358]	[0.000938]	[0.000938]
Number of claims	0.00908***	0.00912***	-0.00407***	-0.00397***	-0.00154***	-0.00157***
	[0.000173]	[0.000173]	[0.000116]	[0.000116]	[0.000208]	[0.000210]
4-digit IPC classes	0.00992***	0.0109***	0.0302***	0.0315***	0.0458***	0.0449***
	[0.00145]	[0.00146]	[0.00186]	[0.00186]	[0.00508]	[0.00507]
Bibliographic information						
# backward references to patent documents	0.00587***	0.00588***	-0.00218***	-0.00198***	0.000124	0.0000718
	[0.000269]	[0.000268]	[0.000237]	[0.000237]	[0.000648]	[0.000647]
Share X references	0.239***	0.234***	0.0414***	0.0327***	-0.0366	-0.0334
	[0.00659]	[0.00659]	[0.00813]	[0.00813]	[0.0223]	[0.0224]
Share Y references	0.175***	0.171***	0.0317***	0.0259***	0.00367	0.00608
	[0.00802]	[0.00801]	[0.00974]	[0.00973]	[0.0270]	[0.0271]
# backward references to non-patent literature	0.00944***	0.00918***	-0.000492	-0.000851***	-0.00200***	-0.00192***
	[0.000301]	[0.000300]	[0.000327]	[0.000327]	[0.000766]	[0.000769]
Share X references	0.114***	0.117***	-0.0297***	-0.0233***	-0.0105	-0.0142
	[0.00685]	[0.00684]	[0.00810]	[0.00809]	[0.0218]	[0.0218]
Share Y references	0.102***	0.100***	-0.0440***	-0.0463***	-0.0356	-0.0350
	[0.00803]	[0.00802]	[0.00942]	[0.00940]	[0.0256]	[0.0256]
Applicant characteristics						
Annual # of appls (00s)	-0.121***	-0.113***	-0.0701***	-0.0582***	-0.151***	-0.155***
	[0.00507]	[0.00511]	[0.00623]	[0.00627]	[0.0148]	[0.0149]
Government	0.00136	-0.00317	0.112***	0.105***	0.0815**	0.0841**
	[0.00943]	[0.00941]	[0.0121]	[0.0121]	[0.0326]	[0.0326]
Individual	0.0275***	0.0205**	0.0241**	0.0135	-0.000273	0.00239
	[0.0100]	[0.0100]	[0.0122]	[0.0122]	[0.0313]	[0.0313]
University	0.0695***	0.0661***	0.122***	0.116***	0.102***	0.104***
	[0.00735]	[0.00734]	[0.00841]	[0.00840]	[0.0231]	[0.0231]
Applicant country ind.	YES	YES	YES	YES	YES	YES
Time fixed indicators	YES	YES	YES	YES	YES	YES
Constant	1.347***	1.370***	1.550***	1.581***	2.395***	2.392***
	[0.0128]	[0.0129]	[0.0192]	[0.0193]	[0.0444]	[0.0450]
Observations	92,121	92,121	92,121	92,121	92,121	92,121

Table 4: Coefficient estimates from log-logistic AFT duration models.
Note: *** 1% significant, ** 5% significant, * 10% significant.

	Type X citations (5 yrs)		Type X citations (5 yrs) > 1			
			Coef.	dy/dx	Coef.	dy/dx
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Markush characteristics						
Markush (0/1)		0.0592***			0.0507***	0.0192***
		[0.0198]			[0.0155]	[0.00589]
Markush size (log)		0.00321***			0.00156**	0.000587**
		[0.000901]			[0.000724]	[0.000273]
Procedural information						
PCT application (0/1)	-0.0523***	-0.0576***	0.0116	0.00436	0.00849	0.00320
	[0.0131]	[0.0131]	[0.0104]	[0.00393]	[0.0104]	[0.00394]
DOCDB family size	0.0236***	0.0235***	0.0185***	0.00696***	0.0184***	0.00692***
	[0.000478]	[0.000477]	[0.000463]	[0.000175]	[0.000463]	[0.000175]
Number of claims	0.00528***	0.00515***	0.00415***	0.00156***	0.00407***	0.00154***
	[0.000262]	[0.000261]	[0.000213]	[0.0000805]	[0.000214]	[0.0000806]
4-digit IPC classes	0.0447***	0.0428***	0.0211***	0.00795***	0.0198***	0.00744***
	[0.00402]	[0.00403]	[0.00320]	[0.00121]	[0.00321]	[0.00121]
Bibliographic information						
# backward references to patent documents	0.00276***	0.00267***	0.0117***	0.00442***	0.0116***	0.00438***
	[0.000493]	[0.000493]	[0.000621]	[0.000234]	[0.000620]	[0.000234]
Share X references	0.222***	0.235***	0.144***	0.0544***	0.152***	0.0573***
	[0.0187]	[0.0188]	[0.0143]	[0.00538]	[0.0143]	[0.00540]
Share Y references	-0.0111	-0.00302	-0.00156	-0.000588	0.00346	0.00130
	[0.0234]	[0.0234]	[0.0175]	[0.00661]	[0.0175]	[0.00661]
# backward references to non-patent literature	-0.000342	-0.0000439	0.00108*	0.000408*	0.00136**	0.000511**
	[0.000576]	[0.000576]	[0.000592]	[0.000223]	[0.000593]	[0.000224]
Share X references	0.0336*	0.0244	-0.00401	-0.00151	-0.00894	-0.00337
	[0.0184]	[0.0184]	[0.0145]	[0.00545]	[0.0145]	[0.00546]
Share Y references	-0.00924	-0.00459	-0.0144	-0.00542	-0.0125	-0.00470
	[0.0224]	[0.0224]	[0.0171]	[0.00644]	[0.0171]	[0.00644]
Applicant characteristics						
Annual # of applications (00s)	0.141***	0.129***	0.0588***	0.0222***	0.0500***	0.0188***
	[0.0135]	[0.0135]	[0.0108]	[0.00407]	[0.0109]	[0.00409]
Government	-0.333***	-0.322***	-0.147***	-0.0543***	-0.141***	-0.0521***
	[0.0301]	[0.0301]	[0.0216]	[0.00778]	[0.0216]	[0.00780]
Individual	-0.381***	-0.369***	-0.234***	-0.0850***	-0.226***	-0.0821***
	[0.0321]	[0.0321]	[0.0227]	[0.00782]	[0.0227]	[0.00786]
University	-0.219***	-0.210***	-0.0877***	-0.0328***	-0.0828***	-0.0310***
	[0.0209]	[0.0209]	[0.0156]	[0.00575]	[0.0156]	[0.00577]
Applicant country indicators	YES	YES	YES	YES	YES	YES
Time fixed indicators	YES	YES	YES	YES	YES	YES
Constant	-1.391***	-1.423***	-0.989***	n.a.	-1.012***	n.a.
	[0.0428]	[0.0432]	[0.0311]		[0.0313]	
Observations	92,122	92,122	92,122	92,122	92,122	92,122

Table 5: Coefficient estimates from negative binomial models on citation frequency in general as well as Type X citations, probit models and their marginal effects at the mean of all variables for patent documents receiving at least one Type X citation.
Note: *** 1% significant, ** 5% significant, * 10% significant

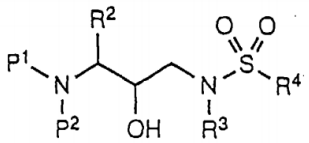
Variable	Fence (number of patent documents per drug) (1)	(2)	(3)	(4)
Markush characteristics				
At least one Markush patent (0/1)		1.177 ***	0.700***	0.636***
		[0.0156]	[0.0139]	[0.0170]
Number of Markush patents			0.0707***	0.0677***
			[0.00100]	[0.00109]
Average size of Markush structure				0.00357***
				[0.000544]
Highest development stage				
Phase I (0/1)	0.180***	0.154***	0.108***	0.106***
	[0.0226]	[0.0211]	[0.0183]	[0.0185]
Phase II (0/1)	0.499***	0.441***	0.329***	0.323***
	[0.0182]	[0.0170]	[0.0147]	[0.0148]
Phase III (0/1)	0.833***	0.770***	0.621***	0.619***
	[0.0249]	[0.0232]	[0.0201]	[0.0203]
Approval (0/1)	1.492***	1.344***	1.069***	1.066***
	[0.0224]	[0.0209]	[0.0180]	[0.0181]
Drug Characteristics				
First indication of drug (0/1)	-0.956***	-0.806***	-0.622***	-0.617***
	[0.0151]	[0.0142]	[0.0122]	[0.0123]
Biologic drug (0/1)	0.0798***	0.549***	0.742***	0.746***
	[0.0154]	[0.0159]	[0.0134]	[0.0135]
Company country fixed effects	YES	YES	YES	YES
ICD9 fixed effects	YES	YES	YES	YES
Time fixed effects	YES	YES	YES	YES
Constant	2.846***	2.865***	2.584***	2.427***
	[0.474]	[0.438]	[0.378]	[0.945]
Observations	26,066	26,066	26,066	26,066

Table 6: Coefficient estimates from negative binomial models of the number of patent documents linked to a drug in Cortellis on the drug-indication level.
Note: *** 1% significant, ** 5% significant, * 10% significant

Appendix

Claims

1. A compound represented by the formula:



40

EP 0 810 209 A2

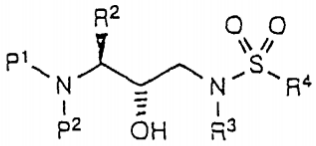
wherein:

P^1 and P^2 independently represent hydrogen, alkoxycarbonyl, aralkoxycarbonyl, alkylcarbonyl, cycloalkylcarbonyl, cycloalkylalkoxycarbonyl, cycloalkylalkanoyl, alkanoyl, aralkanoyl, aroyl, aryloxycarbonyl, aryloxycarbonylalkyl, aryloxyalkanoyl, heterocyclylcarbonyl, heterocyclyloxycarbonyl, heterocyclylalkanoyl, heterocyclylalkoxycarbonyl, heteroaralkanoyl, heteroaralkoxycarbonyl, heteroaryloxycarbonyl, heteroaroyl, alkyl, alkenyl, cycloalkyl, aryl, aralkyl, aryloxyalkyl, heteroaryloxyalkyl, hydroxyalkyl, aminocarbonyl, aminoalkanoyl, and mono- and disubstituted aminocarbonyl and mono- and disubstituted aminoalkanoyl radicals wherein the substituents are selected from alkyl, aryl, aralkyl, oycloalkyl, cycloalkylalkyl, heteroaryl, heteroaralkyl, heterocycloalkyl, heteroocycloalkyalkyl radicals, or where said aminoalkanoyl radical is disubstituted, said substituents along with the nitrogen atom to which they are attached form a heterocycloalkyl or heteroaryl radical;

R^2 represents alkyl, aryl, cycloalkyl, cycloalkylalkyl and aralkyl radicals, which radicals are optionally substituted with a group selected from alkyl and halogen radicals, $-NO_2$, -C≡N, CF_3, $-OR^9$, $-SR^9$, wherein R^9 represents hydrogen and alkyl radicals;

R^3 represents hydrogen, alkyl, haloalkyl, alkenyl, alkynyl, hydroxyalkyl alkoxyalkyl, cycloalkyl, cycloalkylalkyl, heterocycloalkyl, heteroaryl, heterocycloalkylalkyl aryl, aralkyl, heteroaralkyl, aminoalkyl and mono- and disubstituted aminoalkyl radicals, wherein said substituents are selected from alkyl, aryl, aralkyl, cycloalkyl, cycloalkylalkyl, heteroaryl, heteroaralkyl, heterocycloalkyl, and heterocycloalkylalkyl radicals, or in the case of a disubstituted aminoalkyl radical, said substituents along with the nitrogen atom to which they are attached, form a heterocycloalkyl or a heteroaryl radical; and

R^4 represents radicals as defined by R^3 except for hydrogen,
except the compounds of the formula



wherein P^1 is tert-butyloxycarbonyl (boc)
P^2 is hydrogen
R^2 is cyclohexylmethyl, R^3 is hydrogen
and R^4 is pyridyl; thiophenyl; phenyl; 8-methyl-chinolinyl; N-n-propyl- or N-n-butyl-morpholinyl; dimethylaminopropyl; methyl or n-butyl.

Figure A.1: First claim of EP0810209 (A2) retrieved from Espacenet, https://worldwide.espacenet.com/publicationDetails/originalDocument?CC=EP&NR=0810209A2&KC=A2&FT=D&ND=3&date=19971203&DB=EPODOC&locale=, latest visit 4th of April 2021.

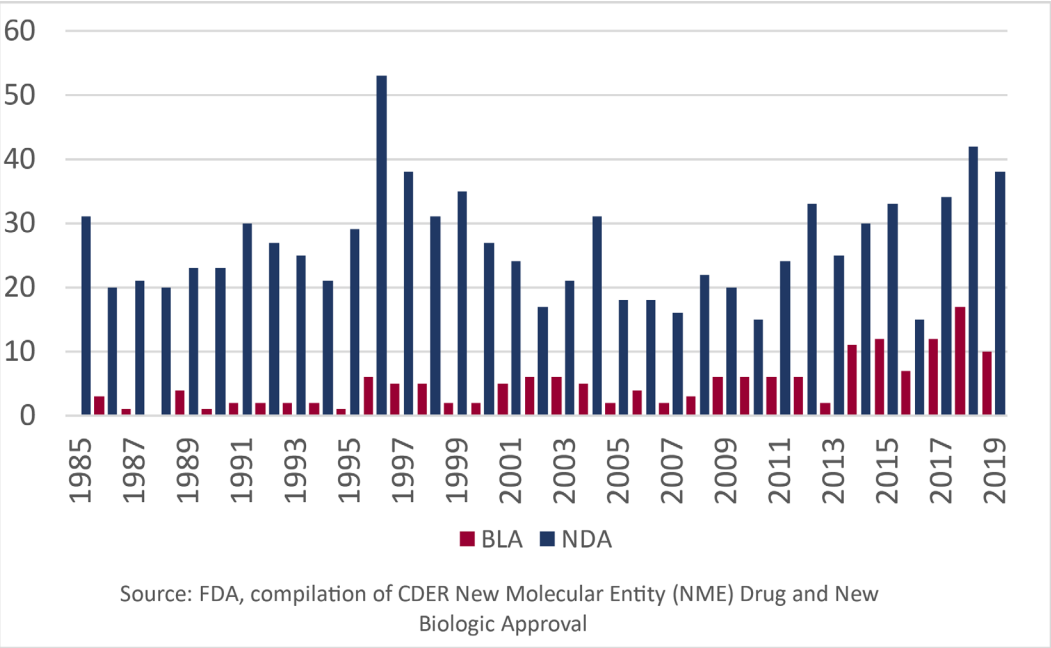


Figure A.2: Approvals of New Molecular Entities (NMEs) and Biologics License Applications (BLAs) at the FDA between 1985 and 2019

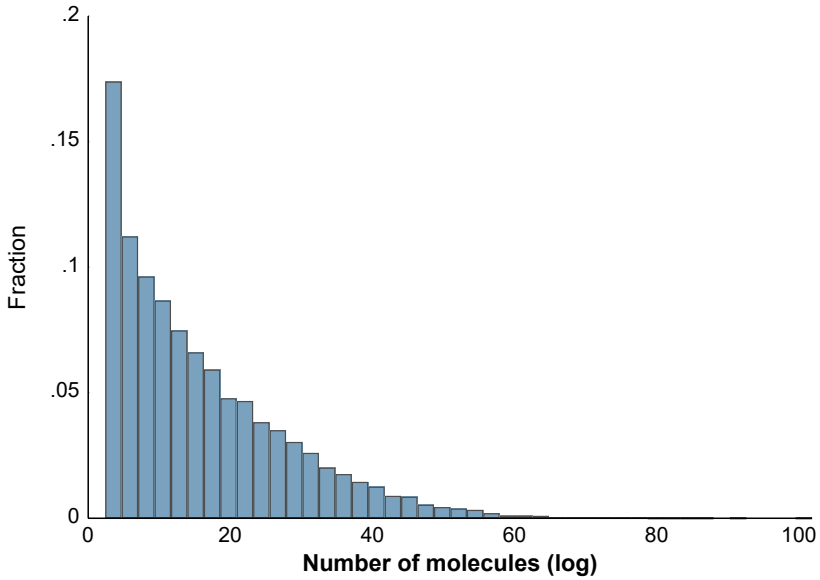


Figure A.3: Distribution of the natural logarithm of the number of molecules contained in the Markush structures in our sample of pharmaceutical patent documents filed at the EPO between 1992 and 2008.

Recent ESMT Working Papers

	ESMT No.
Decertification in quality-management standards by incrementally and radically innovative organizations Joseph A. Clougherty, University of Illinois at Urbana-Champaign Michał Grajek, ESMT European School of Management and Technology	22-04
Do decision makers have subjective probabilities? An experimental test David Ronayne, ESMT European School of Management and Technology Roberto Veneziani, Queen Mary University of London William R. Zame, University of California at Los Angeles	22-03
Is your machine better than you? You may never know. Francis de Véricourt, ESMT Berlin Huseyin Gurkan, ESMT Berlin	22-02
Mismanaging diagnostic accuracy under congestion Mirko Kremer, Frankfurt School of Finance and Management Francis de Véricourt, ESMT Berlin	22-01
The economics of dependence: A theory of relativity Hans W. Friederiszick, ESMT Berlin and E.CA Economics Steffen Reinhold, E.CA Economics and ECARES –Université Libre de Bruxelles	21-02