Ex-ante Information Provision and Innovation: Natural Experiment of Herbal Patent Prior Art Adoption at the USPTO and EPO

Prithwiraj (Raj) Choudhury and Tarun Khanna[1]
February 17, 2015

We study how ex-ante information provision, in the form of codified prior art, affects innovation outcomes. Using a unique dataset of herbal patents filed on the United States Patent Office (USPTO) and European Patent Office (EPO) from 1977 to 2013, we exploit a natural experiment where the USPTO and EPO adopted a codified database of traditional herbal medicine prior art at different points in time. This database (the 'Traditional Knowledge Depository Library' or TKDL) was created by Indian state-owned R&D labs and provided the USPTO and EPO patent examiners with codified, searchable prior art on herbal formulations based on a translation of ancient Indian medicinal texts. We establish that the time lag of the USPTO adopting TKDL compared to the EPO was related to idiosyncratic differences in how the agreements were structured and negotiated, not differences in policy toward herbal patents at the USPTO and EPO. We find that the adoption of TKDL affects the level of herbal patent filing and grants. It also shifts the composition of patenting away from pure herbal formulations that are similar to prior art available in the ancient texts towards applications involving both herbs and synthetic compounds, which are more distant from the prior art and arguably less contestable. We also use unique data coded from patent image wrappers at the USPTO and validate the 'smoking gun' that prior art codification affects the search strategies of patent examiners.

[1] Harvard Business School; corresponding author: Prithwiraj (Raj) Choudhury—pchoudhury@hbs.edu.
The authors thank seminar participants at the Harvard Business School, International Economic Association Annual Conference, Oxford University and University of Minnesota for comments on a prior draft. Kanav Hasija has provided excellent research assistance on this project.

1. Introduction

Studies of innovation in economics have focused on patenting as a means of protecting intellectual property (IP). In a recent survey, Hall et al. (2014) provide a comprehensive review of the literature. An invention might be patented if the claims contained in the patent meet the twin criteria of novelty and nonobviousness. To ascertain the validity of the claims contained in a patent, the claims are to be examined against prior art contained in prior patents and non-patent sources such as publications, books, databases, etc. However, the literature has outlined several inefficiencies related to information disclosure while filing patents (Anton and Yao, 2004). Firms and inventors might disclose incomplete information while filing patents for appropriation concerns or to protect secrecy of their inventions.[2] A related issue is incomplete search strategies employed by firms filing patent applications.[3] Inventors might additionally face difficulties searching for prior art in difficult to access data sources for emerging technologies. These inefficiencies are particularly relevant for "weak" patents, i.e. patents that potentially violate publicly available prior art and have a significant probability of being overturned relatively easily (Anton, Greene and Yao, 2006).

Granting of weak patents could result in social costs related to invalidating/revising the patent through litigation and/or reexamination and could additionally result in costs to the original inventor. Allison and Lemley (1998) found that among patents challenged on validity grounds, about 46 percent of litigated patents were overturned between 1989 and 1996. Anton et al. (2006) document that prior to the creation of the Federal Circuit in 1982, this percentage was

[2] As Graham (2004) and Hedge, Mowery and Graham (2009) summarize, prior to 1999, in the U.S., patent applications remained secret until the patent is issued. By using a continuation, continuation in-part, or a division, assignees were able to possibly alter individual claims, thereby effectively extending secrecy.

[3] Cockburn and Henderson's (2003) survey data shows that only about a third of respondents conduct a prior art search before starting new R&D. Lemley (2008) documents that information technology (IT) and biotech firms in the United States purposefully ignore existing patent documents in order to avoid charges of willful infringement, and that researchers in these fields execute their research without conducting prior art search.

closer to 65 percent. Lemley (2001) estimates the cumulative annual legal costs of granting weak patents as around $2.4 billion. Earlier, Lerner (1995) estimated the patent litigation costs for litigation in 1991 to be about $1 billion.

This leads to a question of mechanisms that a social planner could employ to ensure weak patents that violate prior art are not filed and are not granted. The current paper adds to this stream of research in economics and asks the following question: *for a newly emerging technology, how does ex-ante information provision in the form of codification of prior art affect patent filing, patent examination and patent grants?* We posit that codifying prior art should affect the *level* of patent filings and patent grants, as it increases the efficiency of prior art examination, reduces possible information asymmetry between patent examiners and inventors, and weeds out 'weak patents.' Codifying prior art into a searchable database increases the chances that patent examiners access appropriate prior art to examine the claims of a patent application. This prior art might have been earlier embedded in non-patent sources that might be difficult to access during the patent examination process. Anticipating this increased scrutiny of prior art and a higher likelihood of rejection, *ex-ante* inventors and firms should be discouraged to file patents that violate prior art considerations. Additionally, codifying prior art should also impact the *composition of innovation that is patented*, as post-codification, inventors should be encouraged to file patents for innovations that are more defensible during the patent examination process. This proposition relates to Moser (2012) who studied the effect of codifying information, in the form of the Periodic Table, on patenting of chemical inventions.

In this paper, we create a unique dataset of 11,262 herbal patents filed on the United States Patent and Trademark Office (USPTO) and the European Patent Office (EPO) from 1977 to 2013. We exploit a natural experiment where the USPTO and EPO adopted a codified

database of traditional herbal prior art at different points in time. This database (the 'Traditional Knowledge Depository Library' or 'TKDL') was created by Indian state-owned R&D labs and provided the EPO and USPTO patent examiners with codified, searchable prior art on herbal formulations based on a translation of ancient Indian medicinal texts. We establish that the time lag of the USPTO in adopting the TKDL agreement compared to the EPO was related to bureaucratic differences in how the agreements were structured and negotiated, not differences in policy toward herbal patents at the EPO and USPTO.

We find that the adoption of TKDL affects the level of herbal patent filing and patent grants. Given our natural experiment, we employ a structural break test (Andrews, 1993) and find that herbal patent filings at the EPO exhibit a discontinuity in 2005, much earlier than the discontinuity in 2010 at the USPTO. We also find that TKDL shifts the direction of patent filing away from pure herbal formulations that are similar to prior art embedded in the ancient texts towards applications involving herbs and synthetic compounds, which are more distant from the prior art. We conduct a difference in differences test to document that the mix of herbal patents shifted towards herbals and synthetics at the EPO but not at the USPTO, post 2006, when the EPO had received a draft agreement related to adopting TKDL and had started using a partial and interim database. At this time, the USPTO had only made an initial request to access TKDL. Finally, to provide a direct test that ex-ante information provision affected search strategies of patent examiners, we use unique data coded from the patent image file wrappers at the USPTO and demonstrate that post TKDL, examiners are more likely to search for herbal prior art while examining a patent application. We further code the 'ethnicity' of the USPTO herbal patents as 'Indian herbal patents' or not and find that U.S. patent examiners started searching for herbal

prior art post 2009, disproportionately for Indian herbal patents. This is in accordance to the fact that TKDL codified prior art related to Indian herbs.

Our results are significant given that the prior literature in economics has mostly documented that patent reform has no effect on innovation, especially by domestic inventors (Sakakibara and Branstetter, 2001; Lerner, 2002; Branstetter et al. 2006). Our findings however suggest that conditional on a particular patenting regime, strengthening prior art affects both the level of patenting and the composition of innovation that is patented.

Our results have policy implications for several recent policy measures being adopted by the USPTO and other actors to improve the quality of prior art available to patent examiners. Two example of recent policy initiatives implemented by the USPTO include the 'peer to patent' initiative, where the USPTO, New York Law School, and IBM are collaborating to provide structured prior art to patent examiners, and the modified third party submission of prior art under the Leahy-Smith America Invents Act. Similar initiatives, explained later, are being implemented by actors such as IBM, Cisco, Rackspace, The Clearing House, and Verizon. More broadly, our results are relevant for improving the patent examination process at the USPTO and other patent examination offices. The recent literature in economics and law has documented the inefficiencies of the patent examination process at the PTOs (Cockburn, Kortum, and Stern, 2003; Alcacer and Gittelman, 2004; Sampat, 2004; Lemley and Sampat, 2012; etc.). To quote Lemley (2001: 1-2), 'the PTO has come under attack of late for failing to do a serious job of examining patents, thus allowing bad patents to slip through the system...several solutions have been proposed, but the common thread among them seems intuitively obvious: the PTO should do a more careful job of reviewing patent applications.' To quote Sampat (2010), 'there is growing concern that resource-constrained examiners face difficulties in identifying relevant

prior art, thereby contributing to the issuance of patents of questionable validity' (Sampat, 2010: 399). Our insights are relevant for patenting outcomes across technologies, but have particular relevance for new technological fields where prior art might be not be publicly available and/or codified. One reason for this is that much of the prior art in such fields tends to be in the non-patent literature, which, as suggested earlier, tends to be more difficult to search (Merges, 1999). Indeed, the USPTO may not even have access to the relevant non-patent databases in new fields.[4]

The rest of the paper is structured as follows: Section 2 outlines the theory; Section 3 describes the empirical setting and the natural experiment while Section 4 outlines the empirical questions; Section 5 describes the data and variables; Section 6 presents the results, and Section 7 concludes. References, tables and figures are at the end.

2. Codifying Prior Art - Effect on Patent Filing, Patent Examination and Patent Grants

2.1. Codifying Prior Art - Effect on Patent Filing

The broader literature on innovation in economics has long studied the effect of changing patent laws, and several recent studies have established that patent system reform has no effect on the *level* of innovation, especially by domestic inventors. Lerner (2002) studies 177 events of patent reforms in 51 countries over a 150-year period and finds that residential patent filings did not react positively to domestic patent reform. Sakakibara and Branstetter (2001) study the effect of expanding patent scope in Japan in 1988 and find no evidence of an increase in either R&D

[4] Sampat (2004) quotes former USPTO Commissioner Q. Todd Dickinson in suggesting that 'rapid progress in emerging technologies continues to challenge the USPTO's ability to access the most current information that demonstrates the state of that art' (USPTO, 1999a: 3).

spending or innovative output that could plausibly be attributed to patent reform.[5] Branstetter et al. (2006) examine how technology transfer within U.S. multinational firms changes in response to a series of patent system reforms undertaken by 16 countries from 1982 to 1999.[6] They report that relative to the pre-reform period, patenting grows for nonresidents after reform, but remains flat for domestic residents. This indicates that patent system reform had no reported effect on domestic innovation.

The recent empirical literature in economics on innovation has also looked at how patent system reform influences the *composition of innovation within a country that is patented*. Moser (2005) introduced a new dataset on innovations at two nineteenth-century fairs, and studied the effect of changing patent laws on what innovations are patented. The author collected data from the catalogues of the Crystal Palace Exhibition in London in 1851 and the Centennial Exhibition in Philadelphia in 1876 and concluded that patent laws influence the composition of innovative activity that is patented. In the nineteenth century, the absence of patent laws appears to have guided innovation toward industries where mechanisms other than patent laws protected intellectual property. Innovators in countries without patent laws concentrated in industries where secrecy was an effective alternative to patent grants, such as scientific instruments, food processing, and dye stuffs. In addition, inventors in the countries with weak patenting tended to avoid innovations in manufacturing and other machinery, which were strongly dependent on

[5] As Sakakibara and Branstetter (2001) document, until 1976, Japanese patent law allowed only one independent, single claim to be included in an invention. A 1976 amendment to the patent law allowed the inclusion of multiple dependent claims, which defined the technical ways to implement an independent claim, in the same patent application. In contrast, the 1988 reforms significantly expanded the extent to which multiple claims could be included in one patent. Patent applicants could now define the coverage of an invention with multiple claims, and those claims could be either independent of or dependent on other claims. In addition, the new law expanded the extent to which related inventions could be included in a single patent.

[6] The authors coded patent system reforms across 16 countries from 1982 to 1999 along the following five dimensions: (1) an expansion in the range of goods eligible for patent protection; (2) an expansion in the effective scope of patent protection; (3) an increase in the length of patent protection; (4) an improvement in the enforcement of patent rights; and (5) an improvement in the administration of the patent system.

patent protection. The author interprets her results as evidence that more effective patent laws might affect the *composition of innovative activity that is patented,* rather than the *level of patenting*.

In a related study, Moser (2012) argues that the ability to keep innovations secret may be a key determinant of patenting. In other words, inventors' propensity to patent depends on the effectiveness of secrecy, hence patenting is likely to increase in response to fundamental advances in science and engineering. The author tests this proposition in the context of the chemical industry and the introduction of the periodic table in 1869. As the author states, Dmitri Mendeleev's publication of the periodic table in 1869 transformed chemical analysis and made it much riskier to protect further improvements for chemical innovations through secrecy. Given this, the author finds that in 1851, none out of 32 chemical innovations in the U.S. had been patented; however by 1893, 16 percent of U.S. chemicals were protected by patents.

Our study adds to this stream of research in economics and asks how the codification of prior art affects the *level of patent filing* and the *composition of innovation that is patented*. In so doing, we study an important antecedent of innovation that has not been studied by prior researchers—codifying prior art. We posit that the organization of prior art into a codified, searchable database for patent examiners should affect both the level of patent filing as well as the composition of innovation that is patented. We posit that post codification of prior art, inventors should not be filing patents with claims that might be closer to prior art; instead, post codification, inventors should file patents with claims that are more distant from prior art. This should be particularly true for a newly emerging technology, where the prior art is much more likely to be embedded in non-patent sources such as publications, books, etc. The reasoning for this argument is the following: post codification of prior art, inventors are aware that patent

examiners, given their access to newly available codified prior art, are likely to reject patents with claims that are similar to prior art. Knowing this, *ex-ante,* they are less likely to submit such patents for examination. Instead, post codification, they are more likely to submit patents with claims distant from prior art.

In other words, codification of prior art could affect the level of patent filing, given that assignees are ex-ante less likely to file 'weaker' patents that infringe prior art and are now likely to be rejected by patent examiners. Codification of prior art could also affect the composition of innovation that is patented – post codification, assignees are more likely to file patents for relatively more novel innovations that are more defensible through the patent examination process. However, in framing this argument, we do not wish to suggest any normative implications. As an example, we do not suggest that patents with claims more distant from prior art are 'more innovative' patents.

2.2. Codifying Prior Art - Effect on Patent Examination and Patent Grants

We next present our theoretical arguments on how codification of prior art could affect the patent examination process and patent grants. Prior to that, we summarize prior literature in economics that outlines constraints of the patent examination process.

In order for an invention to be patentable, a patent claim must be shown to be both novel and nonobvious. As Sampat (2004) outlines, under the novelty bar, an invention cannot be patented if it was previously known or used. Under the nonobviousness bar, an invention cannot be patented if 'the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which the subject matter pertains' (35 USC ξ 103a).

As Alcacer and Gittelman (2004: 6) state, 'the core value of a patent is expressed in its claims...claims cover intellectual property that is not already anticipated by existing patents or public knowledge.' The authors also describe how, at a practical level, patents contain 'prior art,' comprising both prior patented and non-patented information such that the claims of the patent are viewed as novel and nonobvious. To quote the authors, 'claims and prior art operate together to describe the invention, show its novelty over existing knowledge, and delineate the scope and the corresponding strength of the intellectual property covered by the claims' (Alcacer and Gittelman, 2004: 6-7).

Given the importance of recognizing prior art in framing the claims of a patent, it is intuitive that the efficiency of the patent-granting process hinges on the quality of prior art examination at patent offices around the world. Cockburn, Kortum, and Stern (2003) conducted a comprehensive study of the prior art examination process within one of the most important patent granting systems in the world-the USPTO. For an individual patent, the prior art search process typically begins with a review of prior U.S. patents in relevant technology classes and subclasses, either through computerized tools or by hand examination of hard copy stacks of issued patents. The search then proceeds to a word search of foreign patent documents, scientific and technical journals, and other databases. The USPTO's Scientific and Technical Information Center maintains an extensive collection of reference materials.

Patent examiners also rely, in part, on applicant disclosure of the prior art submitted with the patent application, as part of the information disclosure statements. As Cotropia, Lemley, and Sampat (2012) describe, in the United States, patent applicants have a 'duty of candor' to disclose any prior art material to patentability of an invention. However, Thomas (2001) suggests that applicants also face disincentives to search for relevant prior art. Applicants might be weary of

searching for and discovering prior art in the fear that failure to disclose may later be deemed inequitable conduct in a court of law. This implies that patent examiners play a critical role in discovering prior art related to the claims of a patent.

There exists a recent stream of empirical papers that studies the relative share of examiners and applicants in citing prior patents. Alcacer and Gittelman (2004) use a dataset of citing-cited pairs of patents of 1,500 citing patents generated from a large, random sample of patents issued over the period 2001-03, and they find that compared to the applicants, examiners are 87 percent more likely to cite patents that are in other countries than the inventors' own. Sampat (2004) uses a dataset of 502,687 utility patents issued by the USPTO from January 2001 to December 2003 and finds that at the patent level, 62 percent of the average share of citations to U.S. patents are inserted by examiners. In subsequent work, Alcacer, Gittelman, and Sampat (2009) report similar results. They examine prior art citations listed in all U.S. patents granted in 2001-03 and find that examiners add 63 percent of citations on the average patent and all citations on 40 percent of patents granted.[7] Sampat (2010) reports that there is self-sorting within technology fields and even within firms with regard to the share of applicant citations; applicants contribute more prior art for their more important inventions.[8] In a related paper, Cotropia, Lemley, and Sampat (2012) find that patent examiners did not use applicant-submitted prior art in the rejections that narrowed down the claims before the patents were issued. Instead, they relied almost exclusively on prior art that they found themselves. The authors also find that examiners do not even use applicants' late-submitted art (i.e., prior art submitted by applicants

[7] In a recent paper, Hegde and Sampat (2009) report that examiner citations to a patent are stronger predictors than applicant citations of a widely used measure of patent value-whether or not the patent will be renewed.

[8] This result builds on prior work on strategies employed by firms in acquiring patents, including work by Hall and Ziedonis (2001) and Cohen, Nelson, and Walsh (2000). The prior literature suggests that firms have stronger incentives to ensure that their patents are defensible against validity charges based on missed art in discrete product fields (such as pharmaceuticals) where patents are important for appropriating rents from R&D. In contrast, in complex product industries where many patents cover a given product, firms might have less incentives to conduct a thorough search of prior art.

later in the process of examination) in making these rejection decisions. Cotropia et al. (2012) hint at patent examiner myopia in interpreting these results. In other words, patent examiners tend to focus on references that they themselves identify.

However, examiners account for a much smaller share of references to prior art embedded in the non-patent literature and in non-U.S. patents. Sampat (2004) finds that references to non-U.S. patents and to non-patent literature are, respectively, 27 and 26 percentage points less likely to be inserted by examiners rather than by applicants, compared to references to U.S. patents. Moreover, in 69 percent of the patents citing non-patent prior art, all the citations to the non-patent prior art came from applicants. Starting with Cockburn, Kortum, and Stern (2003), the literature has also documented significant heterogeneity in the patent examination process both at the level of Art Units and at the level of individual examiners.[9] As an example, the authors document that patent examiners examining biotechnology patents rely extensively on non-patent prior art, such as scientific articles indexed by Medline. Lemley and Sampat (2012) show that more experienced examiners cite less prior art, are more likely to grant patents, and are more likely to grant patents without any rejections.

The literature on prior art examination has outlined several constraints faced by patent examiners in the prior art examination process. Lemley and Sampat (2012) describe the time constraints faced by patent examiners in searching for prior art. They cite Allison and Lemley (2000) and Lemley (2001) to describe how the patent examination process for an individual patent might take three to four years, but the examiner spends an average of only 18 hours over those years working on any given application. Merrill et al. (2004) also document that examiner

[9] As Cockburn *et al*. (2003) document, patent applications are sorted and allocated into one of 235 Art Units, a group of examiners who examine closely related technologies and constitute an administrative unit.

time constraints have increased over the years, as the number of patent applications has outpaced the growth in the number of examiners.

Starting with Cockburn, Kortum, and Stern (2003) and Jaffe and Lerner (2004), the literature has also outlined how incentives of patent examiners might be aligned against conducting a comprehensive search for prior art. As Lemley and Sampat (2012) describe, the promotion and bonus decisions of examiners is tied to the number of counts they accumulate. The counts are determined by the number of first office actions or by disposal of cases, which occurs when an application is either granted or abandoned.[10] The authors also state that this system of incentives does not reward searching for prior art.

The literature, starting with Lemley (2001), has also outlined the constraints of patent examiners in accessing relevant prior art embedded in non-patent sources, such as publications, books, and software. Lemley and Sampat (2012) also cite Thomas (2001) to argue that searching for prior art is more difficult for non-patent prior art, given that non-patent art is not readily accessible, is not conveniently identified and classified, and is not printed in a common format. Hall and MacGarvie (2010) provide interesting evidence of this from the field of software patent examination. To quote the authors, 'when the USPTO began to handle large numbers of patents in this area in 1994-95, they lacked adequate databases with software prior art' (Hall and MacGarvie, 2010: 996). The authors also document an interesting verbal exchange between a patent examiner and the editor of the *Aharonian's Patent Newsletter* concerning constraints in accessing non-patent prior art. The authors first quote the patent examiner as saying that 'the problem with obviousness is evidence…Also, a problem with ordering non-patent publications

[10] Cockburn *et al*. (2003) describe the process of issuing first office action letters as part of the patent examination process. After reviewing a patent application, examiners compose a 'first action' letter either accepting or rejecting their claims. Applicants then receive a fixed time to respond by amending the claims and/or by supplying additional evidence.

or translations of foreign documents is the time it takes to process these requests. An examiner cannot simply call a company who is making, or is believed to have made, a product which appears to infringe on a claim. At best, the examiner could ask a librarian at our library to call a company to request info, but again that takes time. With 10 hours to do a case, movement is paramount' (Hall and MacGarvie, 2010: 996). The authors then highlight the issue of resource constraints in accessing non-patent prior art such as books and publications. They quote Aharonian as saying, 'additionally for subjects like software, the cost of purchasing copies of technical papers would exceed the application fee, so I doubt many examiners would get the authority to spend such sums. Since for most software patent applications, the most relevant prior art is non-patent materials, between the statistics I cited on citing non-patent prior art (an average of two out of about 30) plus the above and below comments, one could make a good case that it is impossible for the PTO to conduct adequate novelty analyses' (Hall and MacGarvie, 2010: 997).

Given these constraints of the patent examination process, codifying prior art embedded in books, publications, and other non-patent sources should impact the level of both patent filings and patent grants. Codifying prior art in a format easily accessible and searchable by patent examiners should *ex-ante* increase the probability that patent examiners search for relevant prior art and this should reduce the time needed and costs to search for prior art. This, in turn, should impact the level of patent grants, as patent examiners are now better equipped to examine prior art and weed out 'bad patents.'

3. Empirical Setting

3.1. Indian Traditional Medicinal Knowledge

India's diversified agro-climatic zones span from the Trans Himalayas to Coastal Andaman and Nicobar. India is one of the most bio-diverse countries of the world. India comprises 16 out of the 26 possible agro-climatic zones. These diversified agro-climatic zones have resulted in India accounting for 7 to 8 percent of the recorded species of the world, with only 2.4 percent of the world's land area. The Botanical Survey of India and the Zoological Survey of India have recorded more than 47,000 plant and 81,000 animal species. This biodiversity has led to a long-standing interest in traditional medication among ancient civilizations in India. This has been formalized in several traditional medicine systems, including the *Ayurveda, Yoga, Unani,* and *Siddha* systems, collectively known as AYUSH. India officially recognizes a wide variety of plants and herbs for their medicinal values. As Dubey et al. (2004) note, more than 6,000 plants in India are used in traditional, folk, and herbal medicine. Three of the 10 most widely selling herbal medicines in the developed countries, namely preparations of Allium sativum, Aloe barbedensis, and Panax sp., have been long available in the traditional form in India. Starting in the 1990s, there has been an increase in the trend of Western firms filing patents related to medicinal herbs. A United States Federal Survey in 2008 revealed that around 38 percent of adults and 12 percent children in the United States used herbal, traditional medicine.[11]

[11] The survey revealed that the most commonly used herbal medicine in the United States included echinacea, flaxseed oil and ginseng. The survey covered about 23,300 adults and 9,400 children and was conducted by the Federal Centers for Disease Control and Prevention and the National Center for Health Statistics (source: http://www.washingtonpost.com/wp-dyn/content/article/2008/12/10/AR2008121001601.html)

3.2. Western Entities Patenting Traditional Indian Medicine

In 2000, the state owned Council of Scientific & Industrial Research (CSIR) in India conducted a study of individual plant-based medicine patents in the USPTO relating to medicinal plants of Indian origin.[12] In the view of CSIR and other domestic Indian R&D entities, several of these patents encroach on prior art documented in ancient Indian medicinal texts. Several of the patents granted to Western entities fueled growth in the health foods and herbal medicine industries in the U.S. and other Western countries. Kamboj (2000) estimated the market size of health foods and herbal medicine in the U.S. to be around $4 billion in 1996.

The filing of herbal patents by Western entities led to global debate on intellectual property protection. This is based on the provisions of the Trade Related Aspects of Intellectual Property (TRIPS) agreement signed in 1994. Article 27.1 of the TRIPS agreement laid forth three requirements for patentability: (1) the invention is new; (2) the invention involves an inventive step; and (3) the invention is capable of industrial application. Interpretation of these three requirements was left to domestic laws, which lead to different standards across the globe. Internationally, 'new' means never published, used, or known anywhere in the world (absolute novelty standards). Unlike other laws, 'new' in United States patent law means never published anywhere in the world or used/known in the United States one year before filing of the patent application (partial novelty standards).[13] Also, around the same time, many countries around the world joined a treaty called the Convention on Biological Diversity (CBD). Under this treaty, if a research team enters a CBD-contracting host country to gain access to biodiversity of the host

[12] India's 42 state-owned national laboratories are organized under an autonomous umbrella organization, the CSIR. Collectively, it has about 12,500 scientific and technical employees.
[13] United States Code (USC) 35 § 102(b).

country for their research, they have to sign an agreement with the host country detailing access terms, profit sharing, etc. The United States is the only country that has not signed the CBD.[14]

3.3. Litigation

The surge of Western patents based on Indian traditional knowledge evoked a strong reaction from the Indian scientific community. The CSIR was among the entities that initiated litigation against the Western firms securing such patents. In several instances, litigation was successful in revoking the patent. Two of the prominent examples of USPTO herbal patents being revoked include the cases of *turmeric* and *neem.*

Turmeric is one of the oldest spices used in Indian cooking for flavoring. It also has properties that make it an effective ingredient in medicines, cosmetics, and dyes. It has been used as medicine for healing wounds and rashes for centuries. In 1995, two expatriate Indians at the University of Mississippi Medical Center (Suman K. Das and Hari Har P. Cohly) were granted a U.S. patent (No. 5,401,504) on the use of turmeric to heal wounds. Soon after, the CSIR filed a reexamination request challenging the novelty of the patent. This claim was supported by the documentary evidence of traditional knowledge, including ancient Sanskrit texts and a paper published in 1953 in the *Journal of the Indian Medical Association*. This led to revocation of the patent in 1997.

The extract of neem (Azadirachta indica) has been used for hundreds of years by Indians against pests and fungal diseases that attack food crops. Its oil has been used to cure colds, the flu, malaria, skin diseases, etc. In 1994, the EPO granted a patent (EPO patent No. 436257) to a U.S.-based firm, W.R. Grace Company, and the U.S. Department of Agriculture for a method for

[14] http://www.cbd.int/countries/.

controlling fungi on plants by the aid of hydrophobic neem oil. This evoked protests from the Indian farming community and led to a reexamination request filing at the EPO in 1995. The Indian activists, collectively known as the 'Neem Campaign' submitted evidence that the fungicidal properties of extracts of neem seeds had been known and used for centuries in Indian agriculture to protect crops and, therefore, were non-patentable. The patent granted on neem was revoked by the EPO in May 2000. The appendix documents a list of herbal patents at the USPTO that were litigated at Federal or state courts.

3.4. The Traditional Knowledge Digital Library (TKDL)

Given concerns of western patents violating Indian traditional medicine prior art, in the late 1990s, the Government of India moved to construct a digital library of traditional formulations. In 1999, India's Department of Ayurveda, Yoga & Naturopathy, Unani, Siddha, and Homeopathy (AYUSH) established a task force for the creation of the Traditional Knowledge Digital Library (TKDL). The mandate of the task force was to overcome the constraints of international patent examiners examining the claims of herbal patents given the lack of formal academic publications describing herbal formulations. Section 102(a) of the U.S. Patent Act states that rejection of a patent requires evidence in a printed publication describing the invention; mere proof of use in a foreign country is not good enough to reject the patent (Balasubramanian, 2012). The TKDL project began as a collaborative project between India's CSIR (of the Ministry of Science and Technology) and the Department of AYUSH (at that time, it was part of the Ministry of Health and Family Welfare).

The goal of TKDL was to codify knowledge available in the traditional Indian medicine literature on Ayurveda, Unani, and Siddha in a digitized format, translated into English, French,

German, Japanese, and Spanish. After a period of five years, the project neared its completion and totaled more than 200,000 compositions, with approximately 82,900 Ayurveda formulations, 115,300 Unani formulations, and 12,950 Siddha formulations over the span of 34 million A4-size pages. This exercise was led by the National Institute of Science Communication and Information Resources (NISCAIR), one of the R&D entities under the CSIR umbrella. These formulations were categorized through the creation of a new structured system developed for the TKDL, known as the Traditional Knowledge Resource Classification (TKRC). The TKRC consisted of approximately 25,000 subgroups and encompassed medicinal plants, minerals, animal resources, effects and diseases, methods of preparations, mode of administration, etc.

3.5. The Natural Experiment

We now outline a natural experiment where the adoption of TKDL at the USPTO and EPO was staggered based on idiosyncratic differences in how the access agreement for TKDL was structured and negotiated at the patent offices. There was a one and a half year lag between the USPTO and EPO requesting the access agreement related to TKDL and a nine month lag between the USPTO and EPO fully adopting TKDL. The EPO also adopted an interim database called the 'XPTK' database (which had a fraction of the formulations covered by the full TKDL database), created by NISCAIR around the time it requested access to the full TKDL database.

The access and use of the TKDL is subject to a restrictive non-disclosure agreement, called the 'Access Agreement' (Oguamanam, 2008). After completion of a first edition of the TKDL, the CSIR released a demo CD containing a sample of 500 herbal formulations in October 2003. This led to a request for access to the entire TKDL database by the EPO in July 2005. In August 2005, the EPO adopted an interim database called the 'XPTK' created by NISCAIR. The XPTK

database was an interim database comprising around 27,000 Indian herbal formulations; in contrast, the full TKDL database comprised more than 200,000 formulations. This interim database mostly collated prior art from two journals published by the NISCAIR – the 'Indian Journal of Traditional Knowledge' (IJTK) and the 'Medicinal and Aromatic Plants Abstracts' (MAPA). The full TKDL database additionally collated prior art from more than 90 books on the subject of Indian traditional medicine. Following the EPO adoption of the interim database, the CSIR sent a formal access agreement for the full TKDL database to the EPO for approval in July 2006. The access agreement signing by the EPO took place in February 2009.

In contrast, the USPTO began taking steps to adopt TKDL in December 2006, when it made a request to access TKDL. The USPTO formally signed the TKDL Access Agreement in November 2009 (Pappas and Byrne, 2009).

In summary, the request for access to the full TKDL database was made by the EPO in July 2005 (with the interim, partial database being implemented in August 2005) and the request for access to the full TKDL database was made by the USPTO in December 2006. We conducted qualitative analyses and interviews to establish that the lag in the USPTO in requesting access and concluding the agreement was related to bureaucratic reasons unrelated to USPTO/EPO policy toward herbal patents. In conducting this analysis, we attempt to rule out the possibility that the delay in the USPTO adopting TKDL compared to the EPO, and the subsequent difference in patent filing and patent granting for herbal patents at the USPTO compared to the EPO, were both driven by some endogenous policy difference at the EPO and USPTO that the econometrician cannot observe.

To conduct this analysis we interviewed Dominic Keating, the director of the IP Attaché Program at the USPTO, who played a key role in negotiating the TKDL Access Agreement between the USPTO and CSIR. Keating noted that the primary cause for the EPO's earlier signing was that 'the EPO had different statutory requirements than the USPTO,' which led to a number of terms in the original Access Agreement being acceptable to the former but 'nonstarters' for the latter. Keating went on to note three specific clauses that were problematic for the USPTO. First, there existed an indemnity clause in the original Agreement that would have required the USPTO to compensate the CSIR (the nodal agency for TKDL) for any lawsuits, which was quickly negotiated out of the Agreement. Second, initially only five users were allowed to access the TKDL database, which posed a problem because the USPTO did not know how many of its examiners, supervisors, and other employees would need access at a particular moment. Consequently, in the final Agreement, the user access rights were altered to a maximum of 30 IP addresses at one time. Finally, the most problematic clause for the USPTO was that the office could not provide any aspects of the TKDL to the public. This requirement did not fit with the U.S. patent system; under its statutory restrictions, the USPTO was mandated to provide documents detailing reasons for rejection to applicants. 'We can't just reject an application and give no reasons why it was rejected,' Keating noted. 'It doesn't work with the system. The USPTO has to provide all rejection documents to applicants, and this Agreement wouldn't let us do that.' As a result, several months were spent in negotiations on these points, during which time the 'USPTO's Access Agreement changed much more significantly than the EPO's.' Both parties were constrained by strictly statutory matters-the USPTO by its mandate to provide documentation for rejection and the CSIR by the necessity of preventing exploitation of the TKDL by the public. Ultimately, the Agreement was signed with the clause substantially

altered to meet the USPTO's requirements, stating that 'the USPTO shall not disclose any information of the TKDL contents to third parties unless and only to the extent that it is necessary for patent search and examination...The USPTO, whenever required may give printouts from TKDL contents to patent applicants and/or their legal representatives only for the purpose of citations of prior art, and may publicly post the search results on the USPTO's Patent Application Information Retrieval system and on other search and examination results digital access systems' ('TKDL Access Agreement').

We also interviewed Dr. V.K. Gupta, director of TKDL and senior advisor to the Indian state-owned labs that created TKDL. Dr. Gupta was, arguably, the most important architect of the TKDL database and had oversight of the process from the inception of the database until the signing of the access agreements with the USPTO and EPO. Dr. Gupta confirmed several insights from the prior interview with Keating, most importantly that the USPTO delay in signing the agreement could be attributed to its concern about not being able to make the prior art known to the public. However, to quote Dr. Gupta, 'the EPO wanted access to TKDL as early as 2005, as our efforts had become quite popular and quite well known by then at the EPO.' Dr. Gupta also outlined another reason why the signing of the USPTO Access Agreement was delayed-the formal signing was scheduled to match the schedule of the visiting Indian Prime Minister, Dr. Manmohan Singh, who visited the United States in November 2009. The USPTO Access Agreement was signed by Dr. Samir Brahmachari, head of the Indian public R&D labs, in the presence of the Indian prime minister. In contrast, the EPO Access Agreement signing was a much less publicized event and Dr. Gupta signed the agreement in one of his visits to the EPO. Figure 1 outlines the milestones of TKDL adoption by the USPTO and EPO, and Table 1 outlines differences in the access agreement clauses for the EPO and USPTO agreements.

Our analysis of secondary data and our interviews also reveal that the EPO not only adopted TKDL earlier than the USPTO, the EPO was also more proactive in using the TKDL prior art compared to the USPTO. In two cases, the EPO set aside its intention to grant the patent, both within three weeks; in 59 additional cases, the applicant withdrew the patent application due to TKDL evidence; and in 18 other cases, the applicant modified the patent's claims as a consequence (Balasubramanian, 2012).[15]

4. Empirical Questions

In this paper we study the following four empirical questions:

4.1. Effect of TKDL on Levels of Patent Filing – Structural Break Test

Our first empirical question relates to whether or not the *level* of herbal patent filing on the USPTO and EPO is affected by the introduction of the TKDL database. Codifying herbal patent prior art in a database that is easily accessible and searchable by patent examiners should *ex-ante* increase the probability that patent examiners identify relevant prior art and consequently weed out 'weak patents', i.e. patents that infringe prior art, previously not easily available to examiners. Anticipating this, *ex-ante*, patent assignees, i.e. firms, individuals and universities, should be discouraged to file such patents in the first place and incur patent filing costs. Additionally, to

[15] EPO application EP1520585 (concerning cancer treatment using Pistacia species) was one of the applications against which TKDL filed a third-party observation. In July 2009, the EPO asked the applicant to 'take position on TKDL-cited references.' In the same month, acting on the TKDL third-party submission, EPO 'set aside intention to grant patent' on an anti-vitiligo cream application (application EP1747786) and reopened the case for further examination. Almost immediately, a chain of application withdrawals began at the EPO. Unilever NV withdrew application EP1607006 for 'functional berry composition' in August 2009. Purimed Co. Ltd. (Seoul, Korea) withdrew application EP1781309 in October 2009. Similarly, applications EP2044850, EP1889638, EP1750809, EP1807098, and EP1967197 were all withdrawn in 2009 in the EPO. This series of withdrawals continued into 2010. In contrast, there was a lag in the USPTO reaction to the TKDL prior art. In June 2011, the Morinaga Milk Industry Co., Ltd. (based in Tokyo, Japan) withdrew 15 claims while retaining the application with only a single claim on their publication 20090093450, on the basis of the examiner's report. In November 2011, Laila Nutraceuticals (based in India) amended the claims of publication 20100203078 based on TKDL evidence. This was almost two years after the USPTO signed the TKDL Access Agreement.

recap Figure 1, the first version of the sample TKDL database was launched in October 2003, the request for access was made by the EPO in July 2005, the partial and interim database was implemented by the EPO in August 2005 and the EPO received the access agreement to implement the full database in July 2006. Much later, in December 2006, the USPTO initiated the process and made an official request to access the TKDL database. Given that the EPO received the access agreement and sample TKDL database prior to the USPTO, patent filing for herbal patents should decline earlier at the EPO, compared to the USPTO. In other words, the trend of herbal patent filing at the EPO should exhibit an earlier structural break compared to the USPTO.

To conduct this analysis, we perform a structural break test based on structural change test based on Andrews (1993) and Zeileis et al. (2002). The first step is to conduct a chow test between two groups of observations. For example, we have data on annual patent filings in the pre-TKDL and post-TKDL periods. The null hypothesis is that there is no difference in the patent filing function parameters for the two periods. However, the Chow test only works with a linear model with one known single break in the mean. We use the test devised by Andrews (1993) which generalizes the Chow test to work with an unknown number of unknown breaks in the mean. We employ the '*strucchange*' command in the statistical package R, to implement the Andrews test. The appendix has a detailed note on the structural break method.

4.2. Effect of TKDL on Composition of Herbal Patent Filings

Our second empirical question relates to whether or not there was a change in the *composition* of herbal patents in the post-TKDL years. We code herbal patents as comprising two types: (1) pure herbal patents that are exclusively based on herbal formulations; and (2) herbs combined with synthetic compounds. We create a variable *is_mixed_patent* to indicate

patent applications comprising herbs combined with synthetic compounds. Given the theoretical arguments presented earlier, herbal patents comprising both herbs and synthetic compounds arguably hold a higher standard of novelty and are less likely to be rejected by examiners based on an examination of prior art codified by the TKDL database. In other words, post TKDL, we expect the composition of herbal patent filing to shift towards more 'mixed' patents.

We then exploit our natural experiment of the EPO adopting TKDL earlier than the USPTO to test for whether there is a differential change in the mix of the herbal patents being filed at the EPO and the USPTO before and after July 2006, when the EPO received the access agreement from CSIR and had already implemented the interim and partial database, while USPTO still had not even made a request for access. We run the following difference in differences specifications using a Logit model, and use robust standard errors clustered at the level of assignee (individual, university or firm filing patent) to test for whether the mix of patents disproportionately changed at the EPO (compared to the USPTO) around the 2006:

$$(1)\ is\_mixed\_patent = \beta_0 + \beta_1 \times EPO + \beta_2 \times post2006 + \beta_3 \times EPO \times post2006 + I$$

Here, β_3 is the coefficient of interest. If there is a differential increase in the fraction of mixed patents at the EPO compared to the USPTO post July 2006, we expect β_3 to be positive and significant. We include the standard set of controls (*I*) used by the literature on patenting including the number of claims, level of backward and forward citations, dummies for whether or not the patent is filed by a Fortune 1000 firm, university, individual, etc.[16] We also included a control for whether or not the herbal patent pertained to an Indian herb.

[16] In robustness checks, we also test for whether or not the composition of herbal patent grants disproportionately shifts towards mixed patents at the EPO around 2006. We additionally ran robustness checks for whether or not the composition of both filings and grants of herbal patents disproportionately moved towards mixed patents at the EPO after 2009, when the TKDL was fully adopted by both the EPO and USPTO, with the EPO adopting the full version of the database nine months earlier than the USPTO. To test this, we run the following specification: is_mixed_patent= β0+β1×EPO+β2×post2009+β3×EPO×post2009+I. Here β3 is the key coefficient of interest

4.3. Effect of TKDL on Levels of Patent Grants – Structural Break Test

Our third empirical question relates to whether or not the *level* of herbal patent grants on the USPTO and EPO is affected by the introduction of the TKDL database. To conduct this analysis, we perform a structural break test based on structural change test based on Andrews (1993) and Zeileis et al. (2002). As earlier, we employ the '*strucchange*' command in the statistical package R, to implement the Andrews test. We analyze the trend of herbal patent grants based on the year of application. In other words, if a_t represents the number of patents filed in year 't' and g_t represents the number of patents granted filed in year 't' that are later granted based on examination, we analyze the trend of g_t. Our interviews suggest that there is a three to five year lag between when a patent is filed at the EPO/USPTO, until the time it may be granted. Given this we expect the structural break to appear a few years prior to the EPO/USPTO adopting the full TKDL database in 2009. Given the three to five year timeframe needed to examine patents, we expect the structural break of patent grants (based on application year) to appear three to five years prior to 2009. In other words, we expect the structural break in patent grants based on application year to be in 2004, 2005 or 2006.

4.4. Test of Effect of Ex-ante Information Provision – Analyses of Examiner Search Strings from Patent Image File Wrappers

Our final empirical question is focused on providing direct evidence that TKDL made it easier for patent examiners to search for herbal prior art while examining a patent. To conduct this analysis, we used the search string employed by patent examiners at the USPTO to search for prior art while examining a patent and test for whether or not examiners searched for herbal prior art post the USPTO adopting TKDL in November, 2009. Given that the TKDL database

codified prior art related to Indian herbal patents, we also tested for whether or not examiners disproportionately searched for herbal prior art for Indian herbal patents, post 2009.

The examiner search string was coded from the 'image file wrapper' associated with each USPTO patent. Section 5.3 explains the data collection and coding process related to patent image file wrappers, in detail.

For each patent, we create a variable *searched_herbal_priorart* to indicate that the USPTO examiner searched for herbal prior art while examining the patent. We then run a conditional fixed effects model (using fixed effects for individual examiners) to test whether or not the likelihood of searching for herbal prior art increased post adoption of TKDL. We code a variable *post_2009* to indicate whether or not the application was filed after the USPTO adopted TKDL in 2009. We also created a variable *indian_herb* to indicate whether or not the herbal patent related to an Indian herb and run the following specification:

$$(2)\ searched\_herbal\_priorart = \beta_0 + \beta_1 \times post\_2009 + \beta_2 \times indian\_herb + \beta_3 \times post\_2009 \times indian\_herb + I$$

We expect the probability of USPTO examiners to search for herbal prior art to increase after the USPTO adopted the full TKDL database in 2009; in other words, we expect β_1 to be positive and significant. Given that the TKDL adoption process for the USPTO allowed the USPTO examiners access to prior art related to Indian herbs, we expect the probability of USPTO examiners to search for herbal prior art to be higher for Indian herbs; in other words, we expect β_2 to be positive and significant. However β_3 is the coefficient of interest. If there is a differential increase in the likelihood of patent examiners searching for herbal prior art for Indian herbs, post TKDL adoption in 2009, we expect β_3 to be positive and significant. We include the standard set of controls (*I*) used by the literature on patenting including the number of claims,

level of backward and forward citations, dummies for whether or not the patent is filed by a Fortune 1000 firm, university, individual, etc., and run a patent examiner Fixed Effects Logit model with robust standard errors clustered by patent examiner.

5. Data Collection and Coding of Variables

This section outlines how we created the unique dataset of herbal patents filed on the EPO and USPTO and how we coded the variables. Given that there is no easily identifiable way of extracting herbal patents from any of the EPO/USPTO-based datasets, we had to create our own dataset.

Our data sources comprise two widely used patent databases-Thomson Innovation and LexisNexis TotalPatents. We searched through every EPO/USPTO patent filed from 1977 to 2014 to construct our database, and we used both 'keyword search' and 'patent classification search' to identify the patents of interest. We started with a search space of around 12 million USPTO patents and around 4.5 million EPO patents with a publication period of January 1, 1977 till April 30, 2014 and finally end up with 7172 herbal patents at the USPTO and 4099 herbal patents at the EPO. In actual analyses, we do not use data for 2014 given data incompleteness concerns for 2014. This led to a dataset of 11,262 herbal patents filed between 1977 and 2013.

5.1. Keyword Search

In the first step, we used keyword search in the title and abstracts of every USPTO patent to identify herbal patents of interest. We used keywords from the U.S. National Center for Complementary and Alternative Medicine (NCAM). In this database, 45 herbs are documented, along with their traditional and common use. In the first step, the keywords (herb name +

traditional use/common use) were searched in the title/abstract/claim sections of every USPTO patent.

5.2. Classification Search

We then used two relevant patent classification systems-the International Patent Classification (IPC) and the U.S. Patent Classification (USPC) to search for herbal patents. The most relevant IPC was A61K36+ (with 207 subgroups) that was introduced in 2002 by a Committee of Experts at IPC Union for coding traditional medicine formulations.

Table A1 in the appendix lists the specific IPC classifications relating to herbal medicine that were used in our classification based search. We also based our analysis on the USPTO classification system focused on herbal medicines. This was first outlined by Dominic Keating, first secretary for intellectual property at the U.S. Embassy, New Delhi/USPTO and patent attorney at USPTO, in a presentation made to the World Intellectual Property Organization (WIPO).[17] He revealed that major U.S. classifications 424/725 (having 55 subgroups) and 514/783 relate to herbal medicines. After mapping these U.S. classifications with IPC, we identified the relevant IPC classes. The same result was achieved using information from Georg Schiwy-Rausch's presentation at the EPO on Traditional Knowledge.[18] The corresponding U.S. Patent Classifications are outlined in Table 4.

We also used additional databases to augment our search and used the Traditional Chinese Medicine (TCM) database to extract more patents. All patents having priority as the United States (i.e. patents that were filed at the USPTO prior to being filed elsewhere) were extracted

[17] Dominic Keating, *Defensive Protection of Traditional Knowledge at the United States Patent & Trademark Office* (available at www.wipo.int/edocs/mdocs/tk/en/wipo_tkdl_del_11_ref_t7_2.pdf).
[18] Georg Schiwy-Rausch, information manager, data acquisition at the EPO, has the presentation on *Traditional Knowledge at the EPO: Present & Future* (13-15 December 2006) available at http://pame.european-patent-office.org/pubs/hararepdf/tk_aripo_present_future.pdf.

from the TCM database.[19] This gave us 400 Chinese patents with U.S. priority. Once these 400 priority numbers were inserted in the Thomson Innovation database, 703 U.S. patent records relating to herbal medicines were extracted. Ninety-five percent of these patents records were already present in our initial sample of patent records, thus validating our earlier search. In the last step, we read every patent record to filter out irrelevant patents. We read the title and abstract of each of the patent records to complete the extraction process.

At the end of keyword and classification search, we came up with 15,314 possible herbal patents at the USPTO (out of 12 million) and 8217 possible herbal patents at the EPO (out of 4.5 million); we then manually checked each one of these patent records (title and abstract) to further curate herbal patents. This yielded 7172 herbal patents at the USPTO and 4099 herbal patents at the EPO. Two independent coders were employed to collect this data and the researchers cross-checked and verified the data. In actual analyses, we do not use data for 2014, given data incompleteness concerns for 2014. This led to a dataset of 11,262 patents filed between 1977 and 2013.

5.3. Coding the Variables

After creating the database of herbal patents, we coded the independent variables. We first categorized whether or not the herbal patents were based on 'pure herbal formulations' ($is\_mixed\_patent = 0$) or whether or not they were a 'mixture of herbs and other synthetic compounds/drugs' ($is\_mixed\_patent = 1$). To do this categorization, we used the Derwent classification for each patent. The Derwent Patent class is a manually curated standardized classification system for patents maintained by Thomson Reuters and the classification is more

[19] http://chmp.cnipr.cn/englishversion/advance/advance.asp

industry centric than technology centric. After analyzing various mixes of Derwent classes in herbal medicine patent records, we concluded that herbal medicine patent records containing Derwent classes B05, B06, or B07 comprised a mix of herbal medicines and other synthetic compounds/drugs. In the absence of any of these classes, the composition is purely made up of herbal medicines.[20]

We then coded the 'ethnicity' of the USPTO herbal patents as 'Indian herbal patents' based on the list of Indian herb names provided by TKDL.[21] In summary, we searched the patent application for the scientific name of the Indian herb, as well as the name of the herb in the Ayurveda, Siddha and Unani schools of medicine.

The next independent variable relates to the type of assignee for each herbal patent. For both EPO and USPTO herbal patents, we coded patents based on five possible assignee types: (1) Fortune 1000 companies; (2) individual inventors; (3) U.S. university, research organization, or government organization; (4) foreign university, research organization, or government organization; and (5) others. The following steps were used to categorize the assignee type:

1. *Fortune 1000 companies:* In the first step, a partial string matching of each word in the assignee name was carried out with words in the names of each Fortune 1000 company name listed in the appendix (Table A2). In the next step, exact matches were created based on a manual parsing of the matches from the previous step.

[20] In Derwent classification, the B class refers to 'pharmaceuticals.' Subclass B05 refers to 'other organics,' B06 to 'inorganics,' and B07 to 'general.' B04 refers to 'natural products and polymers,' which is also the most popular class in the herbal medicine patent group. B05, B06, and B07 are the only three classes in B (pharmaceuticals) that contain synthetic Western drugs. Thus, a presence of these three classes signifies a combination of synthetic compounds/drugs with herb. Fifty random abstracts of patent records having any of these three classes and 50 random abstracts of patent records with absence of all of these three classes were studied and this result was independently verified by two different coders.

[21] The list of Indian herb names is available at the following website: http://www.tkdl.res.in/TKDL/LangDefault/Common/Utility/KeywordDemo/F-Plant-Name_Tips.asp

2. *Individuals:* All patent records with no assignee were labeled as '*individual*.' We also used fuzzy computational methods to check for whether or not the assignee name and the inventor names on a patent matched.[22] Such patent records were also labeled as '*individual*.'
3. *University/research organization:* We carried out partial string matching of the assignee name with any of the following keywords: 'council,' 'board,' 'college,' 'center,' 'centre,' 'university,' 'research,' 'organization,' 'school,' '*laboratoire*,' and '*institut*.' After carefully examining the filtered results, university or research organizations were labeled as a U.S. or foreign university/research entity depending on the country mentioned under the assignee column.
4. *Others:* All other patents were labeled as having assignee type equal to 'others.' This mostly comprises firms outside the Fortune 1000 list.

Next, we coded the search string employed by patent examiners for each USPTO patent. This data was collected and coded using the 'image file wrapper' associated with each USPTO patent. In late 2003, the USPTO implemented the Image File Wrapper (IFW) system which is an image technology system for storage and maintenance of records associated with patent applications (based on *Notification of United States Patent and Trademark Office Patent Application Records being Stored and Processed in Electronic Form, 1271 Off. Gaz. Pat. Office 100 dated June 17, 2003).* In other words, the image file wrapper for a patent is an electronic record of the patent examination process related to the patent. Each USPTO patent now has a publicly available

[22] Our fuzzy matching algorithm first splits names in first name, last name, and middle name. Then it tries to conduct exact matches of first names with first names and last names with last names. Then it tries to find an approximate match score between first names and first names. We do that by using a combination of levenshtein and n-gram (N=2) distance. The levenshtein distance between two strings A and B = Number of operations (addition, deletion, or substitution of a single character) to reach from string A to string B. The n-gram distance = Probability of two strings matching based on n-grams (here 2-gram character) tuples between two strings matching. We code the First Name - First Name Match score = (ngram between names + (1 - levenshtein distance / maximum string length)) / 2. If (first name - first name match + last name - last name match) / 2 >0.7, then we accept two names to be same.

compressed image file folder with several documents. The document of interest ends with the abbreviation 'SRNT'. This document contains search strings used by examiner while searching prior art.[23] We were able to download 6921 image file wrappers (out of the 7172 USPTO patents) and found at least one SRNT file for 1932 patents. We coded the variable *searched_herbal_priorart* as '1' if the SRNT file contained at least one of the herbal patent classes listed in appendix Table A1. The SRNT files were downloaded and coded by two independent coders in November 2014.

6. Results

6.1. Summary Trends

Figures 2-5 outline several summary trends of herbal patents. Figure 2 shows the trend of herbal patents filed on the USPTO and EPO between 1977 and 2012. Figure 3 outlines the trend between 2002 and 2012, with the starting point being around the time the sample database was released. Both of these figures indicate an earlier break in patent filing at the EPO compared to the USPTO. Figures 4 and 5 outline the trend of herbal patents granted at EPO and USPTO, based on application year, for time periods 1977-2012 and 2002-2012, respectively. Summary statistics for EPO and USPTO patents are reported in Table 2.

6.2. Empirical Question 1: How Herbal Patent Filing Reacts to TKDL

We now report results of the structural breaks test implemented using the Andrews (1993) methodology for patent filings between years 2002 till 2012. We chose 2002 as the starting point of this analysis as this is the year prior to the first major TKDL milestone – the release of the sample TKDL database. For patent filings at the EPO, the test detected a structural break in

[23] A sample image file wrapper is available at the following link: http://storage.googleapis.com/uspto-pair/applications/12102391.zip

2005. For the USPTO, the test detected a structural break in 2010. This confirms the hypothesis that filings at the EPO experienced an earlier break compared to the USPTO. To recap, the EPO made a request for access to TKDL in July 2005, implemented the partial and interim database in August 2005, and received the access agreement to implement the full database in July 2006. Later, in December 2006, the USPTO initiated the process and made an official request to access the TKDL database. Given that the EPO received the access agreement and sample TKDL database prior to the USPTO, patent filing for herbal patents should decline earlier at the EPO, compared to the USPTO.

6.3. Empirical Question 2: How Composition of Herbal Patents Reacts to TKDL

Results reported in Tables 3 and 4 are in line with the difference in differences specification (1) and exploit the fact that TKDL was implemented earlier at the EPO compared to the USPTO. The results indicate that with the TKDL being partially adopted by the EPO in 2006, there is a disproportionate shift in the mix of herbal patenting towards 'mixed patents' (herbs added to synthetic compounds) at the EPO compared to the USPTO.

Table 3 only considers patents filed until 2009, given that the EPO and USPTO fully implement TKDL in 2009 while Table 4 conducts a robustness check and considers all years beyond 2006. The variable of interest is the interaction between *is_EPO* and *post2006* (in other words, the variable *isEP_times_post06*). The coefficient for this variable is positive and significant across all models in Tables 3 and 4.[24]

6.4. Empirical Question 3: How Herbal Patent Grants Reacts to TKDL

[24] We repeat this analysis with the shock being considered in 2009 (the year TKDL is fully adopted by EPO and USPTO) and finds similar results. In other words, both shocks of 2006 (partial implementation of TKDL at EPO) and 2009 (full implementation of TKDL at EPO), there is a disproportionate shift in the mix of herbal patenting towards 'mixed patents' (herbs added to synthetic compounds) at the EPO compared to the USPTO. We also repeat the analyses with patent grants and finds similar results. Results available with authors.

We now report results of the structural breaks test implemented using the Andrews (1993) for patent grants using application year. For patent grants based on application year, the test indicated a structural break in 2004 for both the EPO and USPTO. Given the four to five year lag between year of patent filing and the actual year of patent grant, this indicates that both at the EPO and USPTO, there was a structural break in grants around 2009. This coincides with the year the EPO and USPTO signed the access agreements and adopted the full TKDL database.

6.5. Empirical Question 2: Direct Test of Effect of Ex-ante Information Provision Using Examiner Search Strings from Patent Image File Wrappers

Results reported in Table 5 indicates that U.S. patent examiners are disproportionately likely to search for herbal prior art post 2009, the year the TKDL database was adapted by the USPTO. Here we employ a patent examiner fixed effects Logit model (specification 2) and find robust evidence across all models 1-8 that patent examiners are likely to search for herbal prior art post 2009.

As the coefficient of the interaction term (post 2009 times Indian herb) in Model 8 indicates, we also find evidence that patent examiners are disproportionately likely to search for herbal prior art for Indian herbs, post 2009. This is suggestive that U.S. patent examiners started searching for herbal prior art post 2009, disproportionately for Indian herbal patents. This is in accordance with the fact that TKDL codified prior art related to Indian herbal formulations.

7. Discussion

In the innovation literature, there are very few opportunities to study an exogenous variation in patent policy across countries. This paper has both introduced a new dataset of herbal patents filed at the EPO and USPTO and exploits a natural experiment where there is a

exogenous time lag in the USPTO adopting a database of codified prior art related to herbal patents, compared to the EPO. Given this exogenous variation in the timing of adopting prior art, we study the effect of prior art codification on patent filing, patent examination and patent grants. We also use unique data from patent image wrappers to validate the 'smoking gun', that prior art codification affects patent examiner search.

In summary, we have two sets of findings. Firstly, we study the effect of prior art codification on patent filings and find an earlier structural break in the trend of herbal patent filing at the EPO compared to the USPTO. We also find a disproportionate shift in the filing of 'mixed' patents (i.e. patents comprising both a herb and a synthetic compound) at the EPO compared to the USPTO around 2006, the year in which the EPO received the draft access agreement from TKDL and had already implemented the partial and interim database (while the USPTO was only beginning to start the process). We interpret these results as possible evidence that patent assignees, i.e. firms, individuals and universities filing patents, were reacting strategically to the EPO being more proactive in adopting TKDL. It is conceivable that the EPO draft access agreement in July 2006 and its adoption of the partial and interim database in August 2005 acted as a signal of EPO's commitment towards searching for codified herbal prior art and ex-ante deterred assignees to file 'weak' patents at the EPO. Our next set of results relate to the effect of prior art codification on patent examination and patent grants. At both the EPO and USPTO, we find a structural break in patent grants based on application year in 2004. This is in accordance with the fact that patents filed in 2004 would have come up for examination around 2008-2009, around the time the full TKDL database was adopted by the EPO and USPTO. We also use unique data of patent examiner search strings embedded within patent image file wrappers and provide direct evidence of the 'smoking gun' hypothesis that examiners are more

likely to search for herbal prior art post TKDL being adopted by the USPTO in 2009. We also code the 'ethnicity' of the herbs and find that USPTO examiners are disproportionately likely to search for herbal prior art for Indian herbal patents, post adoption of TKDL. This is in accordance to the fact that TKDL codified prior art related to Indian herbs.

Our results make a contribution to the research by economists and legal scholars on how to make patent examination more efficient.[25] An important white space in this literature is that prior empirical research in economics has not adequately studied the implications of codifying and improving the quality of prior art provided to patent examiners. Our results are relevant for several new policy initiatives being implemented by the USPTO and other actors. Two such initiatives being implemented by the USPTO include the 'Peer to Patent' initiative and the new provisions related to third party submission of prior art under the new America Invents Act. The 'Peer to Patent' project is an initiative of the USPTO, the New York Law School and IBM to gather publicly available prior art in a structured manner and relates to technologies such as software and business methods, telecommunications, speech recognition, translation, biotechnology, bioinformatics and biopharmaceuticals. Similar projects are being piloted by the patent offices in Australia, Japan, Korea and the UK.[26] In addition, the Leahy–Smith America Invents Act (AIA), a United States federal statute that was passed by Congress and was signed into law by President Barack Obama on September 16, 2011 has now improved the process for

[25] Prior literature has studied several solutions to increase the efficiency of the patent examination process at the PTO. Lemley (2001) discusses the implications of increasing the time that patent examiners spend in searching prior art from the average of 18 hours to 36 hours. Farrell and Merges (2004) discuss solutions related to administrative reform at the PTO, including improving incentives for patent examiners. The key insight here is that the patent examination process should not only reward examiners who grant patents, but also reward examiners who reject patents based on a careful examination of prior art. The issue of patent examiner incentives also has been studied extensively by Cockburn, Kortum, and Stern (2003). Cockburn et al. (2003) also discuss implications of processes such as patent examiner supervision, selection of examiners, and training. Yet another solution is 'sorting.' As Sampat (2004) points out, patent examiners should conduct a more thorough prior art search for patent applications with the greatest potential social cost, i.e., those that would impinge on many future inventive efforts. An example of this is provided by Bessen and Meurer (2004): the case of the USPTO's 'second pair of eyes' policy meant to improve the patent examination process for business method patents belonging to patent class 705.

[26] Source: http://en.wikipedia.org/wiki/Peer-to-Patent (website accessed on February 16, 2015).

third parties to submit relevant prior art to the Patent Office for review during the patent examination process. The AIA has made this process anonymous for third parties and has allowed a provision for third parties to comment on the prior art submitted. In addition to the USPTO, other actors including firms have implemented initiatives to codify prior art. Examples include the 'IBM Technical Disclosure Bulletin', a searchable source of prior art published between 1958 and 1998; initiatives by Cisco, Rackspace and Verizon to publish product and technological documentation; and an initiative by The Clearing House, an industry association comprising 20 of the top banks in the U.S., to provide the USPTO with non-patent prior art describing the U.S. financial infrastructure, etc. Merges (2004) documents the role of ex ante information disclosure plays in two specific examples: the Merck Gene Index and IBM's investment in Linux.[27]

Our findings have several implications for patenting innovation in both new technological fields and older technological fields that are newly patent eligible, like software. Sampat (2004) summarizes the constraints in searching for prior art in such areas. He quotes Merges (1999) in stating that prior art in such fields tends to be in the non-patent literature, which is more difficult to search. Patent examiners at the USPTO, EPO, and other patent offices might not have access to the relevant non-patent databases in such fields. Sampat also cites Thomas (2001) and other authors to outline the learning curve in searching for prior art in new technological fields and cites Popp and Johnson (2001) and Thomas (2001) in describing the difficulties in recruiting and retaining patent examiners in new technological areas. In his own work, Sampat (2004) analyzes prior art searching in an emerging technology field-

[27] In 1995, the pharmaceutical company Merck created a public database in which it disclosed gene sequences and made them publicly available, arguably to prevent patenting in this area and protect Merck which uses gene sequences as an input. On the other hand, IBM's investment in Linux created an open-source alternative to Microsoft's Windows platform and lowered the costs of the operating system, an essential input for IBM.

nanotechnology. From 2001 to 2003, the USPTO granted 3748 patents in this field. To compare prior art searches in nanotechnology with those in other technological fields, the author constructs a comparison sample with the same primary patent examiners as those on the nanotechnology patents and finds that any given reference in a nanotechnology patent is approximately five percentage points less likely to be inserted by an examiner than in other fields. He also finds that in this emerging technology, 30 percent of the references are to non-patent prior art, almost 10 percentage points greater than the control set.

Our findings also have several implications for the literature on patent litigation. Codifying prior art from books, publications, and other non-patent sources and enabling patent examiners access to easily searchable codified prior art might lead to lower grants of legally contestable patents and might reduce litigation and social costs. The literature in economics and law has documented both the growing preponderance and the legal and social costs of granting 'bad patents,' i.e., patents that do not hold ground upon later scrutiny of the twin premises of novelty and nonobviousness. As Lemley and Sampat (2009) state, 75 percent of all patent applications lead to a patent being granted. Allison and Lemley (1998) report that nearly half the granted patents that are later litigated turn out to be invalid. As Lemley and Sampat (2012) state, the PTO seems positioned to narrow claims in patent applications, but generally not to reject applications. The literature has also outlined several facts related to patent litigation. Lemley (2001) documents that of the roughly 2 million patents in force as of 2001, around 1600 patent lawsuits involving about 2000 patents were filed every year. Though the number of patents that were litigated is relatively small compared to the number of patents granted, the costs of litigation were very high. Lemley (2001) also states that the cost of patent litigation to each side is around $799,000 through end of discovery and $1.5 million through trial and appeal. In

subsequent work, Cotropia, Lemley, and Sampat (2012) cite the AIPLA survey (2009) to state that bad patents impose substantial attorney fees on defendants, a median of $5 million per case. The literature has also documented the social costs of granting bad patents (Farrell and Merges, 2004). Cotropia, Lemley, and Sampat (2012) cite Chien (2009) and Lemley and Shapiro (2007) to document that bad patents might lead smaller firms to drop their products rather than defend the legality of the contested patents and might lead to inefficient outcomes in licensing technologies. Related to this, Lemley (2001) outlines the *in terrorem* effects of bad patents, where potential competitors or follow-on innovators in a field might be deterred from entering the field by the existence of bad patents owned by their competitors and the prohibitive costs of litigating these patents. In prior related work, Lerner (1995) found that small biotechnology firms avoided R&D in fields where the threat of litigation from larger firms was high. In subsequent work, Lanjouw and Lerner (2001) show that preliminary injunctions in patent cases tend to be used by larger firms hoping to impose financial distress on smaller rivals.

Finally, our results have implications for the literature in economics that has characterized Western research entities as the innovative 'North' and has labeled emerging market entities as the imitating 'South.' The underlying premise here is that most of the patented products or processes that are consumed in the South are developed in the North. While stronger patent protection in the South would protect the North against imitation in their export market, the South would have to pay higher prices for those products. As a result, it is in the interest of the South to choose weak patent protection to facilitate imitation, thus benefitting domestic consumers in the South through lower prices (Chin and Grossman, 1988). In addition to considering consumer surplus in the South, Grossman and Lai (2004) also consider the

incentives and benefits of local innovation.[28] However, the example of herbal patents is one where Western entities patented herbal formulations from China and India, where prior art was public knowledge for decades, if not centuries.

[28] In their model, the South tries to balance between two opposing priorities. On the one hand, there is the sum of extra deadweight loss that results from extending patents for domestic firms and the loss in consumer surplus that results from monopolistic pricing by foreign firms. On the other hand, by better protecting IP in the South, there is the benefit of greater innovation by firms in both the countries. They also find that the incentives to strengthen IP protection increase as the relative capacity for research grows in the South.

References - INCOMPLETE - NEW CITATIONS TO BE ADDED

Alcacer, Juan, and Michelle Gittelman. 'Patent citations as a measure of knowledge flows: The influence of examiner citations.' The Review of Economics and Statistics 88.4 (2006): 774-779.

Alcacer, Juan, Michelle Gittelman, and Bhaven Sampat. 'Applicant and examiner citations in US patents: An overview and analysis.' Research Policy38.2 (2009): 415-427.

Allison, John R. and Mark A. Lemley (1998), 'Empirical Evidence on the Validity of Litigated Patents.' American Intellectual Property Law Association (AIPLA) Quarterly Journal 26: 185-268.

Anton, James J., and Dennis A. Yao. "Little patents and big secrets: managing intellectual property." RAND Journal of Economics (2004): 1-22.

Anton, James, Hillary Greene, and Dennis Yao. "Policy implications of weak patent rights." Innovation Policy and the Economy, Volume 6. The MIT Press, 2006. 1-26.

Balasubramanian, Namita (2012), 'Extent of Use of TKDL in Patent Offices: Trends and Concerns.' Working Paper Series, SSRN.

Baum C. (2001) 'Stata: The language of choice for time series analysis', The Stata Journal.

Bessen, J., and Meurer M., J. (2004), 'Lessons for patent policy from empirical research on Patent litigation', Boston University Working Paper.

Bhagwati, Jagdish N., and Hamada, Koichi (1974). 'The Brain Drain International Integration of Markets for Professionals and Unemployment: A Theoretical Analysis,' Journal of Development Economics 1: 19-24.

Branstetter L., Fisman R., and Foley, F. (2006), 'Do stronger intellectual property rights increase international technology transfer? Empirical evidence from U.S. firm-level data.' The Quarterly Journal of Economics.

Cockburn, I., S. Kortum, and S. Stern. 'Are all patent Examiners Equal? Examiners, Patent Characteristics and Litigation Outcomes in Cohen, W. and Merrill, S.(eds.) Patents in the Knowledge $ Based Economy.' (2003).

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. *Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)*. No. w7552. National Bureau of Economic Research, 2000.

Cotropia, Christopher A., Mark A. Lemley, and Bhaven Sampat. 'Do applicant patent citations matter?.' *Research Policy* 42.4 (2013): 844-854.

Farrell, Joseph and Robert P. Merges (2004), 'Incentives to Challenge and Defend Patents: Why Litigation Won't Reliably Fix Patent Office Errors and Why Administrative Patent Review Might Help.' Berkeley Technology Law Journal 19: 943-970.

Gilbert, Richard, and Carl Shapiro. 'Optimal patent length and breadth.' *The RAND Journal of Economics* (1990): 106-112.

Graham, Stuart J. H. 2004. “Hiding in the Patent’s Shadow: Firms’ Uses of Secrecy to Capture Value from New Discoveries.” https://smartech.gatech. edu/xmlui/handle/1853/10725

Grossman, G., M., and Lai, E. (2004), 'International Protection of Intellectual Property', The American Economic Review 94: 1635-1653.

Hall, Bronwyn H., and Megan MacGarvie. 'The private value of software patents.' *Research Policy* 39.7 (2010): 994-1009.

Hall, Bronwyn H., and Rosemarie Ham Ziedonis. 'The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1979-1995.'*RAND Journal of Economics* (2001): 101-128.

Hall, Bronwyn, Christian Helmers, Mark Rogers, and Vania Sena. "The choice between formal and informal intellectual property: a review." Journal of Economic Literature 52, no. 2 (2014): 375-423.

Hegde, Deepak, and Bhaven Sampat (2009), 'Examiner Citations, Applicant Citations, and the Private Value of Patents.' Economics Letters 105: 287-289.

Hegde, Deepak, David C. Mowery, and Stuart J. H. Graham. 2009. “Pioneering Inventors or Thicket Builders: Which U.S. Firms Use Continuations in Patenting?” Management Science 55 (7): 1214–26.

Jaffe, Adam B. and Josh Lerner (2004), Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It. Princeton: Princeton University Press.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993), 'Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.' The Quarterly Journal of Economics 108: 577-598.

Kanwar S, and Evenson R. (2009), 'On the strength of intellectual property protection that nations provide.' Journal of Development Economics.

Kerr, W., R. (2008), 'Ethnic scientific communities and international technology diffusion', The Review of Economics and Statistics.

Kerr W and Lincoln W. (2010), 'The Supply side of innovation: H-1B visa reforms and U.S. Ethnic invention.' Journal of Labor Economics.

Lanjouw J. (1998), 'The introduction of pharmaceutical product patents in India: Heartless exploitation of the poor and suffering?' National Bureau of Economic Research.

Lanjow J and Schankerman M. (2001), 'Characteristics of Patent Litigation: A Window on Competition.' The RAND Journal of Economics.

Lemley, Mark A. (2001), 'Rational Ignorance at the Patent Office.' Northwestern University Law Review 95: 1-34.

Lemley, Mark A., and Bhaven Sampat. 'Examiner characteristics and patent office outcomes.' *Review of Economics and Statistics* 94.3 (2012): 817-827.

Lemley, Mark A., Doug Lichtman, and Bhaven N. Sampat (2005), 'What to Do about Bad Patents?' Regulation 28: 10-13.

Lerner, Josh. "Patenting in the Shadow of Competitors." Journal of law and economics (1995): 463-495.

Lerner Josh (2002), 'Patent protection and innovation over 150 years.' National Bureau of Economic Research.

Lerner, Josh. 'The empirical impact of intellectual property rights on innovation: Puzzles and clues.' *The American Economic Review* (2009): 343-348.

Merges, R.P. 1999. As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform. *Berkeley Technology Law Journal* 14.

Merrill, S.A., R.C. Levin & M.B. Myers (eds) 2004. *A Patent System for the 21st Century*. Washington, DC: The National Academies Press.

Moser, Petra, "How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World Fairs," The American Economic Review, Volume 95, Number 4, September 2005, pp. 1214-1236.
Moser, Petra. *Why Don't Inventors Patent?*. No. w13294. National Bureau of Economic Research, 2007.

Moser, Petra. "Innovation without patents: Evidence from World's Fairs." Journal of Law and Economics 55.1 (2012): 43-74.

Nordhaus, William D. 'The optimum life of a patent: reply.' *The American economic review* (1972): 428-431.

Oguamanam, Chidi (2008), 'Patents and Traditional Medicine: Digital Capture, Creative Legal Interventions, and the Dialectics of Knowledge Transformation.' Indiana Journal of Global Legal Studies 15: 489-528.

Perron P. (1990), 'Testing for a unit root in a time series with a changing mean.' Journal of business and economic statistics.

Rosenkopf L, and Almeida P. (2003), 'Overcoming local search through alliances and mobility.' Management Science 49: 751-766.

Sakakibara M, and Branstetter L. (1999), 'Do stronger patents induce more innovation? Evidence from the 1998 Japanese patent law reforms.' National Bureau of Economic Research.

Sampat, Bhaven N. *Examining patent examination: an analysis of examiner and applicant generated prior art*. Diss. University of Michigan, 2004.

Sampat, Bhaven N. 'When do applicants search for prior art?.' *Journal of Law and Economics* 53.2 (2010): 399-416.

Saxenian A. (2005), 'From brain drain to brain circulation: Transnational communities and regional upgrading in India and China.' Studies in Comparative International Development.

Thomas, J.R. 2001. Collusion and Collective Action in the Patent System: A Proposal for Patent Bounties. *University of Illinois Law Review* , 316-322.

Table 1
Comparison of the final TKDL 'Access Agreements' signed by the USPTO and EPO with the CSIR

USPTO	**EPO**
No general restrictions on purpose of use of TKDL	General restriction on use of TKDL 'for the purposes of European patent grant procedures'
Explicit access permissions to both USPTO examiners/staff and 'contractors engaged in search of Patent Cooperation Treaty (PCT) applications'	Use limited to EPO only
Information use allowed for both patent examination and other 'internal purposes such as statistical and technical analysis, training, developing classification schedules, definitions and planning, etc.'	Information use permitted only for 'the purposes of the European patent grant procedure in all its phases'
TKDL information transfer is permitted to third parties for patent search and examination, to patent applicants/legal representatives for prior art purposes, and to the public via the USPTO's Patent Application Information Retrieval System	Sharing of TKDL information is limited only for patent grant procedures and to patent applicants, although form of information provision is not limited
User access restrictions to 30 IP addresses simultaneously	Options for user ID-based access restrictions or IP address-based access

Notes: Based on analyses of researchers of the Traditional Knowledge Digital Library (TKDL) Access Agreements concluded by the Council of Scientific and Industrial Research with the USPTO and EPO and interviews.

Table 2

Summary Statistics for EPO and USPTO Herbal Patents

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Claim count	11,260	17.16	57.83	0	5,929
Backward citations	11,262	6.697	22.13	0	1,175
Forward citations	11,262	3.419	9.328	0	177
Application year	11,262	2,003	6.864	1,977	2,013
Granted	11,262	0.445	0.497	0	1
Indian herb	11,262	0.0690	0.253	0	1
Is mixed patent (herb + synthetic)	11,247	0.268	0.443	0	1
Is EPO	11,262	0.364	0.481	0	1
Assignee type1(Fortune1000 firms)	9,172	0.0341	0.182	0	1
Assignee type2 (Individuals)	9,172	0.135	0.342	0	1
Assignee type3 (Other firms)	9,172	0.731	0.443	0	1
Assignee type4 (Universities)	9,172	0.0997	0.300	0	1

Notes:

- We started with a search space of around 12 million USPTO patents and around 4.5 million EPO patents with a publication period of January 1 1977 till April 30, 2014 and finally end up with 7172 herbal patents at the USPTO and 4099 herbal patents at the EPO. We do not use data for 2014 given data incompleteness concerns for 2014. This led to a dataset of 11,262 patents filed between 1977 and 2013.
- We coded patents as being filed for a pure herb or for an herb added to a synthetic compound (is_mixed=1). We used the Derwent classification for each patent and coded patents as mixed if part of Derwent classes B05, B06, or B07
- We coded patents as pertaining to an Indian herb based on data of herb names listed by TKDL (http://www.tkdl.res.in/TKDL/LangDefault/Common/Utility/KeywordDemo/F-Plant-Name_Tips.asp). In summary, we searched the patent application for the scientific name of the Indian herb, as well as the name of the herb in the Ayurveda, Siddha and Unani schools of medicine
- Assignee information was available for 9172 out of the 11262 patents. We coded patents filed by Fortune 1000 firms based on the list of firms in Appendix Table A2. We used fuzzy computational methods to check for whether or not the assignee name and the inventor names on a patent matched. Such patent records were also labeled as 'individual.' We searched the assignee name for the following keywords: 'council,' 'board,' 'college,' 'center,' 'centre,' 'university,' 'research,' 'organization,' 'school,' 'laboratoire,' and 'institut' to code the assignee as a University. All other patents were labeled as having assignee type equal to 'others.' This mostly comprises firms outside the Fortune 1000 list

Table 3
Difference in Differences Test – Patent Filing Mix Post 2006 until 2009

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5	(6) Model 6	(7) Model 7
is_EPO	-0.543***	-0.499***	-0.519***	-0.433***	-0.523***	-0.603***	-0.418***
	(0.0808)	(0.0809)	(0.0824)	(0.0843)	(0.0816)	(0.0865)	(0.0913)
post2006	0.448***	0.428***	0.444***	0.548***	0.445***	0.432***	0.499***
	(0.0970)	(0.0978)	(0.0970)	(0.100)	(0.0967)	(0.0983)	(0.101)
isEPO_times_post2006	0.397**	0.366**	0.402**	0.305*	0.400**	0.423***	0.315**
	(0.157)	(0.158)	(0.157)	(0.159)	(0.156)	(0.158)	(0.160)
claim_count		0.0147***					0.0135***
		(0.00234)					(0.00228)
backward_citations			0.00373**				0.00289*
			(0.00161)				(0.00149)
forward_citations				0.0122***			0.0106***
				(0.00305)			(0.00293)
indian_herb					0.398***		0.349***
					(0.121)		(0.126)
assgn_type1						0.168	0.0493
						(0.146)	(0.142)
assgn_type2						-0.218*	-0.160
						(0.112)	(0.113)
assgn_type4						-0.194	-0.187
						(0.126)	(0.125)
Observations	7,238	7,236	7,238	7,238	7,238	7,238	7,236

Notes:

- Results reported in Table 3 are in line with the difference in differences specification (1) and exploit the fact that TKDL was implemented earlier at the EPO compared to the USPTO. The EPO had made a request for the TKDL access agreement in July 2005, implemented the partial and interim database in August 2005 and was sent a copy of the draft agreement in July 2006. The USPTO however set the ball rolling one and a half years later than the EPO and requested for the access agreement in December 2006
- The results indicate that with the TKDL being partially adopted by the EPO in 2006, there is a disproportionate shift in the mix of herbal patenting towards 'mixed patents' (herbs added to synthetic compounds) at the EPO compared to the USPTO.
- Table 3 conducts this analysis for the *post2006* effect until 2009, when TKDL was fully adopted by both the EPO and USPTO
- The variable of interest is the interaction between *is_EPO* and *post2006* (in other words, the variable *isEPO_times_post2006*). The coefficient for this variable is positive and significant across all models
- Robust standard errors clustered by assignee in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 4
Difference in Differences Test – Patent Filing Mix Post 2006 (all subsequent years till 2013 included)

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5	(6) Model 6	(7) Model 7
is_EPO	-0.543***	-0.496***	-0.519***	-0.433***	-0.525***	-0.579***	-0.391***
	(0.0808)	(0.0809)	(0.0820)	(0.0842)	(0.0815)	(0.0829)	(0.0877)
post2006	0.681***	0.657***	0.678***	0.790***	0.682***	0.686***	0.747***
	(0.0765)	(0.0768)	(0.0763)	(0.0808)	(0.0770)	(0.0750)	(0.0799)
isEPO_times_post2006	0.318***	0.255**	0.324***	0.218*	0.315***	0.334***	0.195*
	(0.109)	(0.110)	(0.109)	(0.112)	(0.109)	(0.108)	(0.112)
claim_count		0.0158***					0.0151***
		(0.00213)					(0.00210)
backward_citations			0.00372***				0.00312**
			(0.00130)				(0.00122)
forward_citations				0.0122***			0.0101***
				(0.00303)			(0.00292)
indian_herb					0.367***		0.321***
					(0.104)		(0.108)
assgn_type1						0.145	0.0469
						(0.131)	(0.128)
assgn_type2						-0.0717	-0.00977
						(0.0789)	(0.0799)
assgn_type4						-0.259**	-0.241**
						(0.113)	(0.110)
Observations	9,149	9,147	9,149	9,149	9,149	9,149	9,147

Notes:

- Results reported in Table 4 are in line with the difference in differences specification (1) and exploit the fact that TKDL was implemented earlier at the EPO compared to the USPTO. The EPO had made a request for the TKDL access agreement in July 2005, implemented the partial and interim database in August 2005 and was sent a copy of the draft agreement in July 2006. The USPTO however set the ball rolling one and a half years later than the EPO and requested for the access agreement in December 2006
- The results indicate that with the TKDL being partially adopted by the EPO in 2006, there is a disproportionate shift in the mix of herbal patenting towards 'mixed patents' (herbs added to synthetic compounds) at the EPO compared to the USPTO.
- Table 4 conducts this analysis for the *post2006* effect for all subsequent years for which data is available (i.e. till 2013)
- The variable of interest is the interaction between *is_EPO* and *post2006* (in other words, the variable *isEPO_times_post2006*). The coefficient for this variable is positive and significant across all models
- Robust standard errors clustered by assignee in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 5
Fixed Effects Logit – Examiner Search String Analyses for all USPTO Herbal Patents

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5	(6) Model 6	(7) Model 7	(8) Model 8
post2009	1.556*	1.843**	1.610**	1.625*	1.337	1.960***	2.031**	1.992**
	(0.833)	(0.860)	(0.807)	(0.888)	(0.814)	(0.657)	(0.824)	(0.848)
claim_count		0.0530***					0.0545***	0.0545***
		(0.0195)					(0.0198)	(0.0199)
backward_citations			0.0101*				0.00217	0.00225
			(0.00556)				(0.00458)	(0.00455)
forward_citations				0.0225			-0.0352	-0.0346
				(0.0620)			(0.0436)	(0.0437)
is_mixed_herbal_patent					0.488		0.396	0.411
					(0.755)		(0.700)	(0.696)
Indian herb						0.973*	1.058*	1.029*
						(0.523)	(0.577)	(0.570)
post2009_indian herb								12.42***
								(0.987)
Observations	977	977	977	977	977	977	977	977
Examiner FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes:

- In Table 5 we employ a patent examiner fixed effects Logit model (specification 2); the dependent variable here is *searched_herbal_priorart*
- We find robust evidence across all models 1-8 that patent examiners are likely to search for herbal prior art post 2009
- The coefficient of the interaction term (*post2009 times Indian herb*) in Model 8 indicates that U.S. patent examiners are disproportionately likely to search for herbal prior art for Indian herbs post 2009; this is in accordance with the fact that TKDL codified herbal prior art related to Indian herbs.
- Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Figure 1
Milestones for TKDL Adoption by the European Patent Office and the United States Patent Office

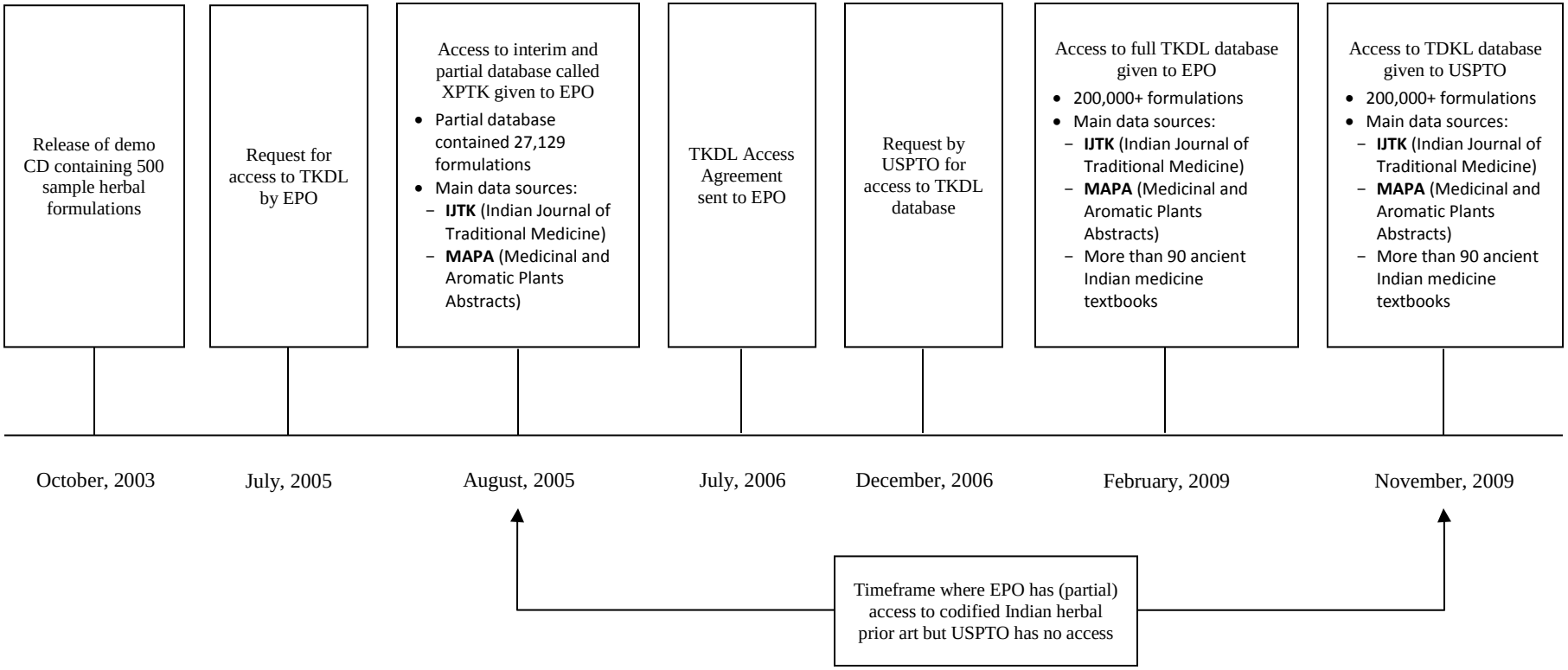


Figure 2
Herbal Patent Filings at EPO and USPTO – 1977 till 2012

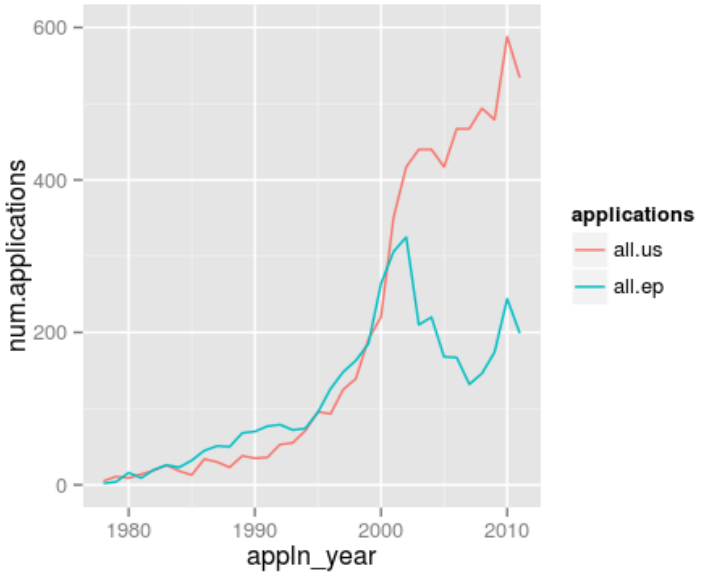
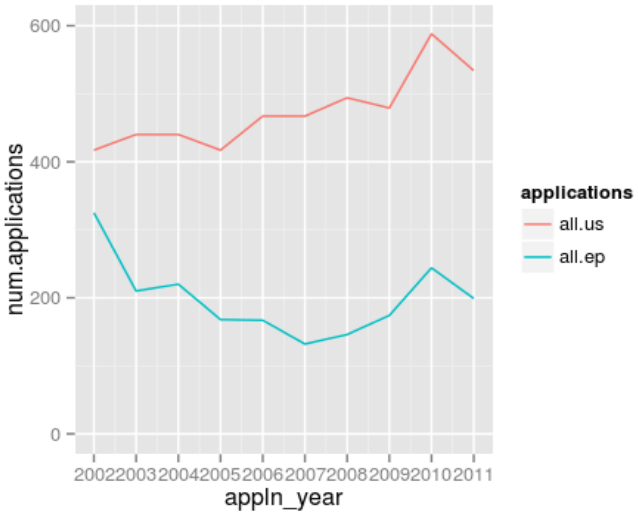


Figure 3
Herbal Patent Filings at EPO and USPTO – 2002 till 2012 (release of demo CD in 2003, first TKDL milestone)



Notes – Figures 2 and 3 plot the trend of herbal patent filings at the EPO and USPTO. Figure 4 plots this trend for all years between 1977 and 2012 while Figure 5 plots this trend for 2002-2012. We chose 2002 as the starting year in Figure 3 given the fact that the first major milestone for TKDL (the release of the CD comprising the sample database) happened in 2003.In addition to plotting the trend of herbal patent filings, we also run the structural breaks test based on Andrews (1993). We find a structural break in 2005 at the EPO and in 2010 at the USPTO. This is in accordance with the fact that the EPO requested for the draft TKDL agreement in July 2005, implemented the partial and interim database in August 2005 and received the draft TKDL agreement in July 2006 respectively, while the USPTO started the process of requesting for access much later, in December 2006.

Figure 4
Herbal Patent Grants by Application Year at EPO and USPTO – 1977 till 2012

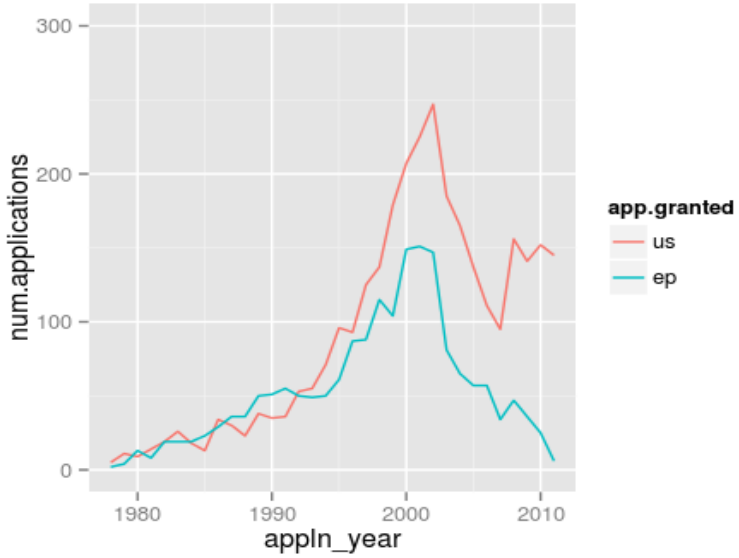
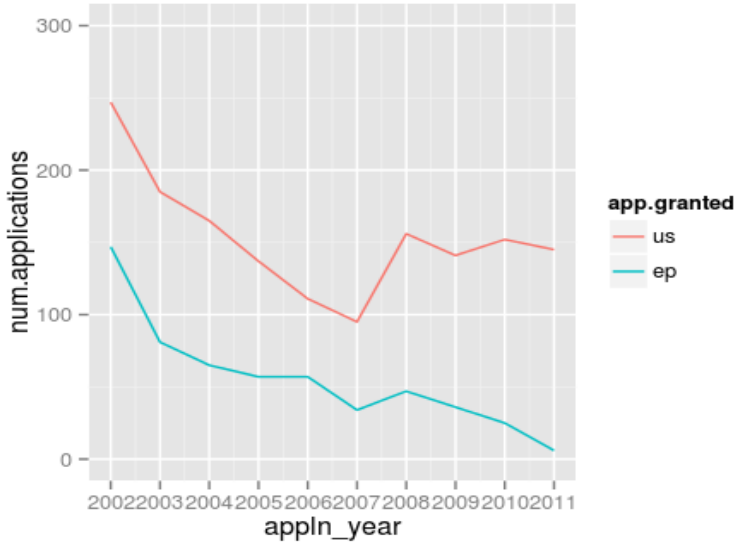


Figure 5
Herbal Patent Grants by Application Year at EPO and USPTO – 2002 till 2012 (release of demo CD in 2003, first TKDL milestone)



Notes – Figures 4 and 5 plot the trend of herbal patent grants at the EPO and USPTO based on year of application. Figure 4 plots this trend for all years between 1977 and 2012 while Figure 5 plots this trend for 2002-2012. We chose 2002 as the starting year in Figure 5 given the fact that the first major milestone for TKDL (the release of the CD comprising the sample database) happened in 2003.Patents filed in year 't' would have come up for examination four to five years later, in years 't+4' or 't+5'.In addition to plotting the trend of herbal patent grants based on application year, we also run the structural breaks test based on Andrews (1993). At both the EPO and USPTO, we find a structural break in patent grants based on application year in 2004. This is in accordance with the fact that patents filed in 2004 would have come up for examination around 2008-2009, around the time the full TKDL database was adopted by the EPO and USPTO.

Appendix: Details on the Structural Break Test

We employ the structural change test based on Andrews (1993) and Zeileis et al. (2002). Here, the first step is to conduct a chow test between two groups of observations. For example, we have data on patent filings in the pre-TKDL and post-TKDL periods. The null hypothesis is that there is no difference in the patent filing function parameters for the two periods.

In other words, for the pre-TKDL period (suppose there are n observations):

$$y_1 = \beta_0$$

For post-TKDL period (suppose there are m observations):

$$y_1 = \gamma_0$$

The null hypothesis of no structural break is

$$H_0: \beta_0 = \gamma_0$$

We run OLS on the two equations separately and have two sums of squares of error (SSE_1 and SSE_2). The unrestricted error sum of squares is

$$SSE_U = SSE_1 + SSE_2$$

We then run the regression on the stacked sample of $n + m$ observations and get SSE_R.

We then compute the F statistic as:

$$\frac{(SSE_R - SSE_U)/k}{SSE_U/(n+m-2k)} \sim F_{k,n+m-2k}.$$

k is number of restrictions; in this example, it is 1.

A traditional Chow (1960) test reports the F test statistic based on the equation above. The major drawback of this is that the structural break change point has to be known in advance. In this paper we implement a generalized Chow test (implemented by the command "strucchange" in R). The idea here is to do a Chow test at every possible break point (i.e. every possible year, in our case). Break points are detected based on the F statistics at every possible break point.[29]

[29] Details available in A. Zeileis, F. Leisch, K. Hornik, C. Kleiber (2002), strucchange: An R Package for Testing for Structural Change in Linear Regression Models. Journal of Statistical Software 7(2), 1–38; D. W. K. Andrews. Tests for parameter instability and structural change with unknown change point. Econometrica, 61:821–856, 1993; D. W. K. Andrews and W. Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica, 62:1383–1414, 1994.

Appendix - *Litigation related to USPTO Herbal Patents*

Publication Number	Patent Description	Filing Date	Publication Year	Assignee	Case	Litigation Start Date
5087623A	Reducing hyperglycemia and stabilizing the level of serum glucose in humans	1989	1992	Nutrition 21	Nutrition21, LLC v. GNC Corporation (Eastern District of Texas)	23-Jun-05
5716625A	Formulations and methods for reducing skin irritation	1994	1998	Cosmederm Technologies	Sensory Systems v Clovermore International (Californian Southern District)	23-Aug-04
6469012B1	Pyrazolopyrimidinones for the treatment of impotence	1995	1999	Pfizer Inc.	Pfizer v. Teva Pharmaceuticals USA, Inc	24-Mar-10
5804203A	Formulations and methods for reducing skin irritation	1996	1998	Cosmederm Technologies	Sensory Systems v Clovermore International (Californian Southern District)	23-Aug-04
5804596A	A method of promoting lean body mass in an individual and treating mood disorder	1997	1998	Sabinsa Corporation	Sabinsa Corporation v. Alchem International (New Jersey District Court)	5-May-03
6139850A	Formulation of a strontium compound as antiallergens	1997	2000	Cosmederm Technologies	Sensory Systems v. Clovermore Intl (Californian Southern District)	23-Aug-04
6039949A	Plant extracts obtained by filtration, centrifuging and dialysis	1997	2000	CampaMed Inc	Optigenex, Inc. v. Cytodyne, LLC (Southern District New York)	13-Apr-05
6106838A	Using herbal-based active ingredients for pharmaceutical purposes	1997	2000	Fotios A. Nitsas	Echopharm USA LLC v. Ralco Nutrition Inc. (District of Delaware)	28-Oct-11
5945107A	herbal plant extracts for weight reduction	1998	1999	Natural Medio Tech A/S	Zoller Lab v. NBTY (District of Utah Central Division)	16-Jan-04
6264994B1	Plant matter from Uncaria tomentosa for the treatment of alzheimers and others	1998	2001	University of Washington, Seattle	Proteotech Inc v. Unicity International Inc (Western District of Washington)	8-Sep-06
6177122B1	Cancer chemoprotective food products	1998	2001	Johns Hopkins School of Medicine	Caudill Seed & Warehouse Co.. v. Brassica Protection Products LLC (Western District of Kentucky)	4-Sep-07
6238696B1	Process for providing non-lipid, liquid form herbal extracts in a vegetable gelatin, HPMC	2000	2001	GAIA Herbs Inc	Gaia Herbs, Inc. v. Aquacap Pharma., Inc (Western District of North Carolina)	17-Sep-03
6475530B1	Weight loss treatment in mammals using natural compounds	2000	2002	Eric H Kurts	Iovate Health Sciences Inc. v. Masuda (Western District of New York)	31-Oct-08

Publication Number	Patent Description	Filing Date	Publication Year	Assignee	Case	Litigation Start Date
7435424B1	Plant extracts for use in the unsaponification of products	2000	2008	International Flora Technologies Ltd.	International Flora Technologies Ltd. v. Desert Whale Jojoba Company, Inc (District of Arizona	8-Apr-09
6475532B2	Calcium channel antagonist used for inhibiting the secretion of parathyroid hormones	2000	2002	Fujix Inc	United States of America v. Lane Labs-USA Inc (Californian Central District)	18-Feb-11
6410062B1	Using extracts of feverfew for the treatment of inflammatory disorders	2000	2002	Johnson & Johnson	Patent Research Institute LLC v. Johnson & Johnson (Eastern District of Texas)	9-Mar-11
6482432B2	Process for providing non-lipid, liquid form herbal extracts in a vegetable gelatin, HPMC	2001	2002	GAIA Herbs Inc	Gaia Herbs, Inc. v. Vitality Works, Inc. (District of Columbia)	31-Oct-03
6383508B1	A deer and geese repellent formulation and method for warding off geese or deer from a shrub or plant	2001	2002	James Messina	Messina, et al. v. Meyhoeffer (District of New Jersey)	26-Nov-03
6660297B2	A nutritional or dietary supplement composition that strengthens and promotes retinal health	2001	2003	Bach & Lomb Inc	Bausch & Lomb Incorporated v. Rexall Sundown, Inc (Western District of New York)	9-Dec-03
6361805B2	A water soluble extract of the plant species uncaria for healthcare purposes	2001	2002	Ronald W Pero	Optigenex, Inc. v. Cytodyne (Southern District of New York)	13-Apr-05
6419963B1	Pharmaceutical composition including the herb Coptis chinesis Franch for the treatment of diaper rash	2001	2002	Sarfaraz K Niazi	PBN Pharma v. Niazi (Northern District of Illinois)	5-Sep-14
6730333B1	Nutraceutical compositions derived from the fruit of the *Garcinia mangostana L.*	2002	2004	DBC LLC. (now Xango)	Xango v. New Vision USA (District of Utah)	4-May-04
6716459B2	Using plant extract; corosolic acid	2002	2004	Futoshi Matsuyama	Use Techno Corporation et al v. Kenko USA, Inc (Northern District of California)	24-Apr-06
6552206B1	Compositions from Lepidium plant useful for the prevention of	2002	2003	Pure World Botanicals Inc.	Naturex, Inc. v. Phyto Tech Corp (District of New Jersey)	17-Aug-06

Publication Number	Patent Description	Filing Date	Publication Year	Assignee	Case	Litigation Start Date
	cancer and sexual dysfunction					
6830765B2	Green tea extract for the treatment of obesity	2002	2004	Laboratoires Arkopharma	Iovate Health Sciences, Inc v. Allmax Nutrition, Inc (District of Massachusetts)	18-Dec-07
7175859B1	Extracts from plant varieties; *Cissus*, *Vernonia* and *Brillantasia* for weight control	2002	2007	Gateway Health Alliances Inc.	Iovate Health Sciences International, Inc. v. USP Labs, LLC (Western District of New York)	21-Nov-08
6759063B2	Use of plant extract from *Serenoa* for treatment of physical conditions e.g. obesity	2002	2004	Anthony L. Almada	Iovate Health Sciences International, Inc. v. Ultralab Nutrition, Inc. (Western District of New York)	8-Dec-08
7824706B2	Dietary supplement and nutritional aid for tract ulcers in horses and other animals	2003	2010	Freedom Health LLC	Freedom Health, LLC v. Figuerola Group, Inc (Northern District of Ohio)	14-Jun-12
7202220B2	Compositions of plant carbohydrates as dietary supplements	2004	2007	Mannatech Inc.	Mannatech, Inc v. Techmedica Health Inc (Northern District of Texas)	5-May-06
6875891B2	Process of preparing highly water soluble salts including extracts from fruits from the *Garcinia* species	2004	2005	Laila Impex	Public Patent Foundation, Inc v. Iovate Health Science Research Inc (Southern District of New York)	6-May-09
7094433B1	Livestock-grafting aid containing peppermint oil and other natural ingredients.	2005	2006	Kent A Bowers	American Soil Technologies, Inc. v. Hansen (District of Arizona)	21-Sep-06
8025907B2	Use of plant extract from *Acmella oleracea*, for the accelerated reparation of functional wrinkles	2009	2011	William A. Belfer	Belfer Cosmetics, LLC v. Milestone Cosmetics LLC (Southern District of Texas)	23-Apr-12
7947312B2	Use of extract of *Uncaria tomentosa* for medical anti-inflammatory purposes	2009	2011	Ronald W Pero	Optigenex Inc. v. Jeunnesse Global Holdings (Eastern District of Texas)	24-Aug-12
8435321B2	Water-dispersible pellets	2012	2013	The Andersons Inc.	The Andersons, Inc. v. Enviro Granulation, LLC (Middle District of Florida)	27-Nov-13

Notes –To collect this litigation data, we used 'Docket Navigator' (http://home.docketnavigator.com), an online subscription service that searched legal dockets at both State and Federal courts in the U.S. for corresponding patent publication numbers. However it must be noted that any disputes that were settled prior to the filing of a complaint at a court would not be recorded in the database.