HARVARD | BUSINESS | SCHOOL

Information Provision and Innovation: Natural Experiment of Herbal Patent Prior Art Adoption at the United States and European Patent Offices

Prithwiraj (Raj) Choudhury
Tarun Khanna

Working Paper 14-079

Information Provision and Innovation: Natural Experiment of Herbal Patent Prior Art Adoption at the United States and European Patent Offices

Prithwiraj (Raj) Choudhury
Harvard Business School

Tarun Khanna
Harvard Business School

Working Paper 14-079

Information Provision and Innovation: Natural Experiment of Herbal Patent Prior Art Adoption at the United States and European Patent Offices

Prithwiraj (Raj) Choudhury and Tarun Khanna[1]

We exploit a natural experiment to study how codifying information about prior innovation affects subsequent innovation. A codified database of traditional Indian herbal formulations was adopted by the European Patent Office (EPO) and the U.S. Patent Office (USPTO) at different points in time. The database, the Traditional Knowledge Depository Library (TKDL), was created by state-owned Indian R&D labs to provide patent examiners searchable "prior art" drawn from ancient Indian medicinal texts. Using a unique dataset of herbal patents filed between 1977 to 2013, we find that adoption of the TKDL affected the level of herbal patent filings and grants, and shifted the nature of patenting away from pure herbal formulations similar to those in the ancient texts toward combinations of herbs and synthetic compounds that were less similar to the prior art and apt to be less contestable.

[1] Harvard University; corresponding author: Prithwiraj (Raj) Choudhury—pchoudhury@hbs.edu. The authors thank Juan Alcacer, Ashish Arora, Nick Bloom, Lee Branstetter, Wesley M. Cohen, Michelle Gittelman, Stuart Graham, Josh Lerner, Alan Marco, Tom Nicholas, Bhaven Sampat, and reviewers and seminar participants at the American Economic Association, Harvard University, International Economic Association Annual Conference, NBER Productivity Spring Meeting, Oxford University, Searle Conference, Stanford University, University of Minnesota and the U.S. Patent and Trademark Office for comments on a prior draft. Mohit Agarwal and Kanav Hasija provided excellent research assistance.

Innovation is crucial to endogenous economic growth (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992; Jones, 1995). Recent studies of innovation in economics have explored whether patent rights facilitate or impede follow-on innovation (Williams, 2013; Galasso and Schankerman, 2015). This paper complements that literature by studying whether making information about prior innovation available in an accessible and codified way affects subsequent innovation. Our paper resembles that of Moser (2011), who studied the effect of information provision, in the form of the Periodic Table, on subsequent patenting of chemical inventions; Moser documented that, after publication of the Periodic Table in 1969, the share of chemical inventions exhibited at world fairs that were patented increased from zero percent in 1951 to 20 percent in 1893. This paper argues that, conditional on the patenting regime, providing information on prior innovation affects both the level and the nature of subsequent patenting.

Economic theory suggests that, given the public-good nature of information, competitive markets might under-incentivize information-provision initiatives (Nelson 1959; Arrow 1962). Nevertheless, several private-sector initiatives have undertaken to codify knowledge of prior art relevant to patenting. 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 association of 20 top U.S. banks, to provide non-patent prior art describing the U.S. financial infrastructure to the U.S. Patent and Trademark Office. But empirical evidence is lacking on whether codifying relevant knowledge affects the level or the nature of subsequent patenting. In an effort to close this gap in the literature, we pose the following research question: *for a newly emerging technology, how does codifying prior art—that is, information about pertinent prior innovation—affect subsequent patent filings, patent examination and the nature of subsequently*

patented innovation? In framing this research question, we assume the patenting regime to be held constant; thus we sidestep the literature on whether changes to patent rights affect the speed and nature of innovation (Sakakibara and Branstetter, 2001; Lerner, 2002; Branstetter et al. 2006).[2]

Our empirical context is patenting of traditional Chinese and Indian herbal medicinal formulations in the United States and Europe. The Nobel Prize awarded to Youyou Tu in 2015 for a novel malaria therapy using an extract of a Chinese herb, sweet wormwood, attests to Western recognition of the traditional herbal medicine of the East.[3] The U.S. market for products based on herbal remedies was estimated at $5.4 billion in 2016. Our unique dataset of herbal patents filed with the U.S. Patent and Trademark Office (USPTO) and the European Patent Office (EPO) between 1977 and 2013 demonstrates the magnitude of corporate and academic interest: it includes herbal patents filed by such multinationals as Abbott Laboratories, Bayer Bristol-Myers Squibb, Colgate Palmolive, Eli Lilly, Merck, Pfizer, Proctor and Gamble and Unilever, as well as Columbia University and the University of California. As a later section will show, studies of Chinese and Indian herbal formulations have appeared in such journals as *Science*, *Nature* and the *New England Journal of Medicine*.

We exploit a natural experiment that lends itself to a difference-in-differences specification. Our sample consists of all patents based on Chinese and Indian herbal formulations filed with the USPTO and the EPO between 1977 and 2013. At different points in time during that period, the two patenting regimes adopted a codified database of traditional *Indian* herbal prior art, with the EPO adopting the database prior to the USPTO. This database, the Traditional Knowledge Depository Library (TKDL), was created by state-owned Indian R&D labs; it provided searchable information on herbal formulations drawn from ancient Indian medicinal texts to EPO and USPTO

[2] In a recent survey, Hall et al. (2014) provide a comprehensive review of the literature.
[3] Source: https://www.nobelprize.org/nobel_prizes/medicine/laureates/2015/tu-facts.html. Accessed August 2, 2017.

patent examiners. We establish, first, that the time lag is attributable to bureaucratic differences in how agreements were structured and negotiated, not to different policies on herbal patents. We then study whether this exogenous time lag leads to statistically significant differences in patent filings and patent grants, or in the nature of patent filings between the USPTO and the EPO.

An invention can be patented if the claims presented in the patent application meet the twin criteria of novelty and non-obviousness. The validity of such claims is ascertained by examining them in light of information on related prior innovation, or *prior art*, contained in prior patents and in books, databases and the like. However, the literature has identified several inefficiencies in the process of information disclosure (Anton and Yao, 2004). Firms and inventors may disclose incomplete information to protect the secrecy of their inventions and to foil attempts at appropriation. Firms that file patent applications may also use incomplete information-search strategies (Cockburn and Henderson, 2003). These inefficiencies tend to result in "weak" patents—that is, patents that potentially violate publicly available prior art and are thus apt to be overturned (Anton, Greene and Yao, 2006). Invalidating/revising weak patents can entail social costs via litigation and/or reexamination (Lerner, 1995; Lemley, 2001).

Codifying prior art should affect the *level* of subsequent patent filings and patent grants by increasing the efficiency of patent examination, reducing information asymmetry between patent examiners and inventors, and weeding out weak patents. A large literature documents patent examiners' lack of resources, incentives and training to search for prior art; they also lack access to prior art in such non-patent sources as publications and books (Cockburn, Kortum and Stern, 2003; Alcacer and Gittelman, 2004; Jaffe and Lerner, 2004; Sampat, 2004; Lemley and Sampat, 2012). This problem is particularly acute in the case of applications based on new technologies, because the prior art is largely confined to non-patent sources. Codifying prior art within a

searchable database thus increases the likelihood that patent examiners will access pertinent prior art while examining the claims of an application. Anticipating this increased scrutiny and higher likelihood of rejection should discourage inventors and firms *ex ante* from filing patents that violate prior art. Codification of information could also affect the *nature* of innovation that is subsequently patented, as Moser (2005) has documented; similarly, we argue that codification of prior art will make applicants more likely to file patents for relatively novel innovations that are more defensible through the patent-examination process.

We find that adoption of the TKDL affects the level of herbal patent filings and grants. Using a difference-in-differences specification, we find a disproportionate decline in the level of herbal patent filings and grants at the EPO, especially for Indian herbal patents, after its receipt of the database of Indian herbal prior art. We also find that access to the TKDL shifts the nature and content of patent filings away from pure herbal formulations—similar to those in the ancient texts—toward versions containing both herbs and synthetic compounds. Finally, to validate the "smoking gun" that ex-ante information provision affects the search strategies of patent examiners, we use unique data coded from patent image file wrappers at the USPTO to demonstrate that, post-TKDL, examiners are more likely to search for prior art while examining an herbal patent application. By coding USPTO herbal patents as Indian or non-Indian, we further find that U.S. patent examiners began disproportionately searching for prior art pertinent to Indian herbal patents after 2009. We follow up our empirical analyses with field interviews of managers at the USPTO and the Indian state-owned laboratories that created the TKDL.

Our results have implications for innovation in such emerging technology fields as gene therapy and drone technology. Information about prior innovation in these fields is often available only in sources that are not readily accessible to patent examiners; this circumstance often leads to

incomplete searches for prior art and, as documented by Lerner (1995) and Lemley (2001), granting of weak patents that are litigated at high social costs. Our results indicate that codifying information on prior art could affect subsequent innovation in emerging technological fields. Our results are also pertinent to policies recently adopted by the USPTO and other actors to improve the quality of the information on prior art available to patent examiners. Examples include the "peer-to-patent" initiative whereby the USPTO, New York Law School, and IBM are collaborating to provide patent examiners structured information on prior art, and modified third-party submission of prior art under the Leahy-Smith America Invents Act. Similar initiatives are being implemented by IBM, Cisco, Rackspace, The Clearing House, Verizon and other entities. Our findings are also relevant to initiatives to improve the quality of patents by providing examiners access to information on technical standards (Bekkers et al., 2016).

Finally, our findings are pertinent to other initiatives to codify information on biological resources, including the World Economic Forum's "Amazon Third Way" initiative, which aims to codify information on the biological resources of the Amazon Basin (Nobre et al., 2016). More broadly, our paper highlights the value of the traditional knowledge of developing countries, such as Brazil, China and India (the global South) and suggests how codifying such information could affect innovation in the global North. This outlook is at odds with North-South models of innovation in economics that have characterized the global South as an imitator rather than a source of innovative ideas (Chin and Grossman, 1988; Grossman and Lai, 2004).

The paper is structured as follows: Section I describes the empirical setting and the natural experiment; Section II outlines the empirical questions; Section III describes the data and variables; Section IV presents results; and Section V concludes. References, tables and figures follow.

I. The Empirical Context

A. Western Entities and Traditional Chinese and Indian Herbal Medicine

Western firms have long filed patents using medicinal herbs from China and India; the practice has been on the increase since the 1990s. It is important here to document how central herbal remedies are to the western bio-pharma industry and to western science in general. A 2008 survey by the U.S. Centers for Disease Control and the National Center for Health Statistics reports that around 38 percent of U.S. adults and 12 percent of children use traditional herbal medicine.[4] In 2016 the U.S. herbal-remedies market was worth $5.4 billion; it is expected to reach $6.6 billion by 2021 (Mintel, 2016). Herbal and natural ingredients have also been cited as key sources for drug discovery (Doak et al. 2014); between 1981 and 2014, at least 33 percent of all new chemical entities (NCEs) were derived from natural products (Newman and Cragg, 2007). The literature attests to Western interest in the promise of natural ingredients; our search of PubMed's Dietary Supplements Subset for 499 traditional herbs found 658,488 articles, published in 11,974 unique scientific journals between 1970 and 2017; several articles appeared in such prestigious journals as *Science*, *Nature* and the *New England Journal of Medicine*.

B. Litigation

The surge in Western patents based on traditional knowledge elicited global debate on intellectual-property protection and a particularly strong reaction from the Indian scientific community. In 2000, the Indian state-owned Council of Scientific & Industrial Research (CSIR) conducted a study of patents granted by the USPTO for formulations that included medicinal plants

[4] According to the survey, which covered 23,300 adults and 9,400 children, the most commonly used herbal medicines include echinacea, flaxseed oil and ginseng. (Source: http://www.washingtonpost.com/wp-dyn/content/article/2008/12/10/AR2008121001601.html)

of Indian origin.[5] According to the CSIR and other Indian R&D entities, several of these patents encroach on prior art documented in ancient Indian medicinal texts.

The CSIR and other entities requested reexaminations of such patents, and in several instances the patent was revoked. Two prominent examples involved *turmeric* and *neem.*[6] We collected data on herbal patents litigated in U.S. federal and state courts and on those that were reexamined by the USPTO. This data is available on request; it suggests that the trend toward herbal patent filings continued unabated even after the wave of initial litigation.

C. The Traditional Knowledge Digital Library (TKDL)

In 1999, in response to alarm about Western patents that violated Indian traditional medicine's prior art, India's Department of Ayurveda, Yoga & Naturopathy, Unani, Siddha, and Homeopathy (AYUSH) established a task force to compile a digital library of traditional formulations, which came to be known as the Traditional Knowledge Digital Library (TKDL).[7] The task force's mandate was to eliminate constraints on international patent examiners posed by the lack of formal academic publications describing herbal formulations: according to Section 102(a) of the U.S. Patent Act, rejection of a patent requires published evidence describing prior invention; mere proof

[5] India's 42 state-owned national laboratories belong to the CSIR, an autonomous umbrella organization with about 12,500 scientific and technical employees.

[6] Turmeric is used for flavoring in Indian cooking and in cosmetics, dyes and medicines; it has been used for centuries to heal wounds and rashes. In 1995 two expatriate Indians at the University of Mississippi Medical Center were granted a U.S. patent for use of turmeric to heal wounds. The CSIR promptly challenged the novelty of the patent; the supporting documentary evidence included ancient Sanskrit texts and a 1953 paper in the *Journal of the Indian Medical Association*. The patent was revoked in 1997. The extract of neem (Azadirachta indica) has been used in India for hundreds of years to combat pests and fungal diseases that attack food crops. Its oil has also been used to cure colds, flu, malaria and skin diseases. In 1994 the EPO granted a patent to the W.R. Grace Company and the U.S. Department of Agriculture for use of hydrophobic neem oil to control plant fungi. Protests from the Indian farming community led to a reexamination request; the Indian activists submitted evidence that neem had been used for centuries to protect crops and were thus non-patentable. The patent was revoked in 2000.

[7] India is among the world's most bio-diverse countries; it encompasses 16 of the 26 agro-climatic zones and possesses 7–8 percent of all recorded species despite occupying only 2.4 percent of the world's land area. As Dubey et al. (2004) note, more than 6,000 plants are used in Indian traditional and herbal medicine. Three of the 10 best-selling herbal medicines in developed countries—preparations of Allium sativum, Aloe barbedensis, and Panax sp.—have long been available in traditional forms in India.

of use in a foreign country is insufficient (Balasubramanian, 2012). The TKDL project began as a collaboration between the CSIR (a branch of the Ministry of Science and Technology) and the Department of AYUSH (then a subdivision of the Ministry of Health and Family Welfare).

The goal of the TKDL was to codify, in a digitized format, knowledge drawn from the traditional Ayurveda, Unani, and Siddha medical literatures and translated into English, French, German, Japanese, and Spanish. After five years the project neared completion; it surpassed 200,000 formulations, translated from approximately 360 volumes of Indian traditional medicine, many written in Sanskrit.

D. The Natural Experiment

Our sample consists of all patents based on traditional Chinese and Indian herbal remedies filed with the USPTO and the EPO, but the natural experiment we exploit is limited to a codified database of solely Indian traditional prior art that was adopted by the USPTO and EPO at different points in time. Adoption of the database by the two patent offices was sequential rather than simultaneous due to differences in how they structured and negotiated the access agreement.

More than three years elapsed between when the EPO and the USPTO gained access to the 27,000 "most important" Indian herbal formulations. In 2005 the EPO adopted an interim database, consisting of around 14 percent of the most important formulations, and secured access to the complete database in 2006. The USPTO first gained access in late 2009.[8]

[8] Access to and use of the TKDL is subject to a restrictive non-disclosure agreement, called the Access Agreement (Oguamanam, 2008). In 2003, after completion of a first edition of the TKDL, the CSIR released a demo CD containing a sample of 500 herbal formulations. In 2005 the EPO adopted an interim database consisting of around 27,000 Indian herbal formulations. (The full TKDL database consisted of over 200,000 formulations.) Interviews with the CSIR confirmed that the interim database contained prior art for the "most important Indian herbal formulations likely to be patented." The EPO received a formal access agreement for the full TKDL database in 2006; signing took place in February 2009. In contrast, the USPTO first requested access to the TKDL in 2006 and signed an access agreement in November 2009 (Pappas and Byrne, 2009).

We conducted qualitative analyses and field interviews, with individuals who negotiated the TKDL access agreement at both the Indian and American ends, to rule out the possibility that the delay in the USPTO's access to the TKDL was driven by an unobservable policy difference on herbal patents; we established that the delay was for bureaucratic reasons unrelated to policy. Figure 1 lists milestones in adoption of the TKDL by the USPTO and EPO; Table 1 outlines differences in the two agencies' access agreements. The Appendix provides insights into the time lag from our field interviews.

II. Empirical Questions

This paper studies four empirical questions: whether adoption of the TKDL leads to statistically significant differences in (1) the level of patent filings, especially for Indian herbal formulations; (2) the nature of patented innovation; (3) patent grants, especially for Indian herbal formulations; and (4) whether examiners at the USPTO indeed use the TKDL during patent examination.

A. The Effect of Access to the TKDL on Levels of Herbal Patent Filings

Our first empirical question is whether the *level* of herbal patent filings is affected by adoption of the TKDL database. Codifying prior art in a database that is easily accessible and searchable by patent examiners should increase the probability that patent examiners will identify relevant prior art and thus weed out weak patents—that is, those that infringe on prior art. Anticipation of this development should *ex ante* discourage patent assignees—firms, individuals and universities—from filing such applications in the first place. As Figure 1 shows, the EPO received access to the sample database in 2003, to the partial interim database in 2005 and to the entire database in 2006. Much later, in 2009, the USPTO received access to the database. Given that the EPO accessed the TKDL database four years before the USPTO, filing for herbal patents, especially patents on

Indian herbal formulations, should have declined earlier at the EPO than at the USPTO. We view two events as potential shocks that could have negatively affected patent filings at the EPO but not at the USPTO: (1) the EPO's receipt of the demo CD containing 500 herbal formulations in 2003, and (2) its receipt of access to 27,000 formulations and then of access to the complete database in 2005–2006. Either or both of these shocks should have had a disproportionate effect on filings of patents using Indian herbal formulations at the EPO. We test for this this outcome using the following difference-in-differences specification: [9]

$$(1)\ is\_EPO\_patent = \beta_0 + \beta_1 \times post\_shock + \beta_2 \times Indian\_herbal\_patent + \beta_3 \times post\_shock \times Indian\_herbal\_patent + I + Y + X$$

Here, β_3 is the coefficient of interest. If after either or both shocks (i.e., *post2003* and/or *post2006*) there is a disproportionate decline at the EPO (as compared to the USPTO) in filings that use Indian herbal formulations, we expect β_3 to be negative and significant. We include the standard set of controls (*I*) used by the literature on patenting, including the number of claims, the level of backward and forward citations, and dummies for whether the patent is filed by a Fortune 1000 firm, a university, an individual, or a firm outside the Fortune 1000 (X). Y represents year fixed effects/time trend(s). X represents assignee fixed effects. In the base case, we run a difference-in-differences specification using a Logit model, and use a time trend and robust standard errors clustered at the level of the assignee (the individual, university or firm). In robustness checks, we use year dummies instead of the time trend; we use different time trends for the pre-shock and post-shock periods, and also use assignee fixed effects.

[9] In robustness checks, we also run the more conventional difference-in-differences specification using a count-dependent variable. The specification here is:
number of patents filed= β0+β1×EPO+β2×post_shock+β3×EPO×post_shock+I. Here, β3 is the key coefficient of interest. But we are constrained by the small number of our observations: only 73, due to the limited number of years in our sample. (We have USPTO observations for 1977–2013 and EPO observations for 1978–2013).

B. The Effect of TKDL Access on the Nature of Herbal Patent Filings

Our second empirical question is whether a change occurred in the *nature* of the innovations being patented in the post-TKDL years. We code two types of herbal patents: (1) those for pure herbal formulations, and (2) those for herbs combined with synthetic compounds. Examples of both are available on request. We create a variable *is_mixed_patent* to indicate applications for herbs combined with synthetic compounds. Such applications arguably represent a higher standard of novelty and are less likely to be rejected by examiners based on prior art codified in the TKDL database. In other words, post-TKDL, we expect the composition of herbal patent filings to shift toward more "mixed" patents. We run the following difference-in-differences specification using a Logit model, and use robust standard errors clustered at the level of the assignee to test for whether the mix of patents filings disproportionately changed at the EPO around either or both of the shocks:

$$(2)\ is\_mixed\_patent = \beta'_0 + \beta'_1 \times EPO + \beta'_2 \times post\_shock + \beta'_3 \times EPO \times post\_shock + I + Y + X$$

Here, β'_3 is the coefficient of interest. If there is a differential increase in the fraction of mixed patents filings at the EPO after either or both of the shocks (i.e., *post2003* and/or *post2006*), we expect β_3 to be positive and significant.[10]

C. The Effect of TKDL Access on Levels of Herbal Patent Grants

Our third empirical question is whether the level of herbal patent *grants* is affected by access to the TKDL database.

[10] 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 or by firms outside of the Fortune 1000 (X), etc. We also include a control for whether the patent pertains to an Indian herb. Y represents year fixed effects/time trend(s). In the base case, we run the difference-in-differences specification using a Logit model and use robust standard errors clustered at the level of assignee.

Recall that our analysis of patent filings treated two events as possible shocks: (1) the EPO's receipt of the demo CD and (2) its receipt of access to 27,000 herbal formulations and of the access agreement to the complete database. We argued that both shocks could negatively affect filings of patent filings using Indian herbal formulations at the EPO but not at the USPTO. For patent grants, however, we focus on the second shock: the EPO's receipt of access to the 27,000 formulations. Our interviews indicate that EPO examiners began searching for Indian herbal prior art only when they received that database. Though the demo CD signaled the EPO's commitment to searching for Indian herbal prior art, and could plausibly have affected filings, with regard to patent grants the first real shock occurred only when EPO examiners began searching the database of 27,000 formulations in 2005–2006: that shock should have had a disproportionate effect on grants of herbal patents at the EPO. Given that it takes three to four years to examine a patent application at the EPO—the average patent examination in our sample takes around 3.9 years—the shock of searching for Indian herbal prior art in 2005–2006 should affect patents filed three to four years earlier, or around 2002–2003. We test for this using the following difference-in-differences specification: [11]

(3) $$is\_EPO\_patent = \gamma_0 + \gamma_1 \times post\_shock + \gamma_2 \times Indian\_herbal\_patent + \gamma_3 \times post\_shock \times Indian\_herbal\_patent + I + Y + X$$

[11] In robustness checks, we also run the more conventional difference-in-differences specification using a count-dependent variable. Here, the specification is:
number of patents granted $= \beta 0 + \beta 1 \times EPO + \beta 2 \times post\_shock + \beta 3 \times EPO \times post\_shock + I$. Here β3 is the key coefficient of interest. But we are constrained by limited number of observations: only 73, due to the limited number of years in our sample. (We have USPTO observations for 1977–2013 and EPO observations for 1978–2013.)

Here, γ_3 is the coefficient of interest. If there is a disproportionate decline in grants of Indian herbal patents at the EPO after the shock identified above (i.e., *post2002*), we expect β_3 to be negative and significant.[12]

D. The Smoking-Gun Test: Analyses of Examiner Search

Our final empirical question seeks direct evidence that access to the TKDL made it easier for patent examiners to search for herbal prior art while examining a patent filing. Using the search string that USPTO patent examiners employed to search for prior art, we tested for whether they did so, and whether they did so disproportionately, after the USPTO adopted the TKDL in 2009.

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

For each patent, we create a variable *searched_herbal_priorart* to indicate that the examiner searched for herbal prior art while examining the patent filing. We then run a conditional fixed effects model (using fixed effects for individual examiners) to test whether the likelihood of searching for herbal prior art increased after adoption of TKDL. We code a variable *post_2009* to indicate whether the application was filed after the USPTO adopted TKDL in 2009; we also create a variable *indian_herb* to indicate whether the filing involved an Indian herb. We run the following specification:

[12] We include the standard set of controls (I) used by the literature on patenting, including the number of claims, the level of backward and forward citations, dummies for whether the patent is filed by a Fortune 1000 firm, a university, an individual, or a firm outside the Fortune 1000 (X). Y represents year fixed effects/time trend(s). X represents assignee fixed effects. We run the difference-in-differences specification using a Logit model, and use robust standard errors clustered at the level of assignee.

(4) $searched\_herbal\_priorart = \gamma'_0 + \gamma'_1 \times post\_2009 + \gamma'_2 \times indian\_herb + \gamma'_3 \times post\_2009 \times indian\_herb + I+X$

We expect the probability that USPTO examiners search for herbal prior art to increase after adoption of the full TKDL database in 2009; in other words, we expect γ'_3 to be positive and significant. We also expect the probability that USPTO examiners 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 that patent examiners search for Indian herbal prior art after 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, the level of backward and forward citations, and dummies for whether the patent is filed by a Fortune 1000 firm, a university, an individual, etc. We run a Fixed Effects Logit model with robust standard errors clustered by patent examiner; X represents examiner fixed effects.

III. Data Collection and Data Construction

This section outlines how we created the unique dataset of herbal patents filed at the EPO and USPTO and how we coded the variables. Because there is no way to readily extract herbal patents from any of the EPO/USPTO-based datasets, we had to create our own dataset. We began with two widely used patent databases: Thomson Innovation and LexisNexis TotalPatents. To construct our database, we searched through every EPO/USPTO patent filed between 1977 and 2014 using both keyword search and patent-classification search. From approximately 12 million USPTO patent filings and 4.5 million EPO patent filings, we ended up with 7,172 at the USPTO and 4,099

at the EPO. We eliminated 2014 data from analysis due to incompleteness, leaving a dataset of 11,262 herbal patents.

A. Keyword Search

As a first step, we searched keywords in the titles and abstracts of every USPTO patent; the keywords we used were drawn from the from the database of the U.S. National Center for Complementary and Alternative Medicine (NCAM), which documents the traditional and common uses of 45 herbs. The keywords (herb name + traditional use/common use) were searched in the title/abstract/claim sections of every USPTO patent.[13]

B. Classification Search

We then used two 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), introduced in 2002 by a committee of experts at IPC Union for purposes of coding traditional-medicine formulations.[14] We also based our analysis on the USPTO classification system for herbal medicines. This was first outlined by Dominic Keating when he was first secretary for intellectual property at the U.S. Embassy, New Delhi/USPTO, and a patent attorney at USPTO, in a presentation to the World Intellectual Property Organization (WIPO).[15] The U.S. classifications 424/725 (with 55 subgroups) and 514/783 both pertain to herbal medicines.[16] After mapping these U.S. classifications with IPC, we identified the relevant IPC

[13] An example of a search string is "TITLE-ABST-CLAIM(chamomile and (skin* conditions OR ulcer OR ulcer OR diarrhea OR cancer OR sleep OR anxiety)) and DATE(>=1980-01-01 and <=2014-06-01)". This string, which refers to the herb chamomile, returned 60 patents; all were read and verified by two independent coders.

[14] The IPC patent class A61K36 refers to "Medicinal preparations of undetermined constitution containing material from algae, lichens, fungi or plants, or derivatives thereof, e.g. traditional herbal medicines."

[15] 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.

[16] For example, sub-class 424/725 pertains to "Plant Material or Plant Extract of Undetermined Constitution as Active Ingredient (e.g.., Herbal Remedy, Herbal Extract, Powder, Oil, etc."

classes. The same result was achieved using information from Georg Schiwy-Rausch's 2006 presentation at the EPO on traditional knowledge.[17]

We searched additional databases for additional herbal patents, especially traditional Chinese medicine patents to augment our search. Extracting all patents with U.S. priority (patents filed first with the USPTO)[18] gave us 400 such Chinese patents; inserting their priority numbers into the Thomson Innovation database resulted in 703 U.S. patent records relating to herbal medicines. Ninety-five percent of these records were already present in our initial sample, validating our earlier search. As a last step, we read the title and abstract of every patent record to filter out irrelevant patents.

Our two search strategies, keyword and classification search, generated 15,314 possible herbal patents at the USPTO and 8,217 at the EPO. We manually checked the title and abstract of each patent record, yielding 7,172 at the USPTO and 4,099 at the EPO. Manual checking eliminated patents that had been wrongly classified. Two independent coders were employed to collect this data; we manually cross-checked and verified the data. Eliminating incomplete 2014 data led to a dataset of 11,262 patents filed between 1977 and 2013. Given truncation concerns, we plot summary trends only through 2011. In regression analysis, we use data through 2013; we conduct robustness checks by restricting the sample to pre-2011 or pre-2009 to account for right-censoring concerns.

C. Coding the Variables

We then coded the independent variables. We first determined whether a given herbal patent was a "pure herbal formulation" ($is\_mixed\_patent = 0$) or a "mixture of herbs and other synthetic

[17] Georg Schiwy-Rausch (information manager for data acquisition at the EPO), "Traditional Knowledge at the EPO: Present & Future," available at http://pame.european-patent-office.org/pubs/hararepdf/tk_aripo_present_future.pdf.
[18] http://chmp.cnipr.cn/englishversion/advance/advance.asp

compounds/drugs" ($is\_mixed\_patent = 1$). To do so, we used the Derwent classification system, a manually curated, standardized classification system for patents maintained by Thomson Reuters that is more industry-centric than technology-centric; that is, it focuses more on usage than on patent class. Patent records belonging to Derwent classes B05, B06, and B07 pertain to mixtures of herbal medicines and synthetic compounds/drugs; other classes pertain exclusively to herbal medicines.[19]

We then coded the "ethnicity" of the USPTO herbal patents as Indian herbal patents based on the list of Indian herb names provided by the TKDL.[20] We searched the patent application for both the scientific name of an Indian herb and for its popular name in the Ayurveda, Siddha and Unani schools of medicine, as per the TKDL database.

The next independent variable is the type of assignee: we specified five assignee types: (1) a Fortune 1000 company; (2) an individual inventor; (3) a U.S. university, research organization, or governmental entity; (4) a foreign university, research organization, or governmental entity; and (5) other. The following steps were used to categorize the assignee type:

1. *Fortune 1000 companies:* We matched a partial string of each word in an assignee's name with each word in the names of Fortune 1000 companies (names available from the authors). Exact matches were created based on a manual parsing of those matches.

[19] In Derwent classification, pharmaceuticals belong to the B class. Subclass B05 consists of "other organics," including aromatics; B06 contains "inorganics," including fluorides for toothpastes, and B07 refers to "general" items, including dispensers and catheters; B04 consists of "natural products and polymers," including herbal-medicine patents but not synthetic compounds. B05, B06, and B07 are the only B classes that contain synthetic Western drugs; they consist of combinations of synthetic compounds/drugs with herbs. Fifty random abstracts including any of the three classes and 50 that included none of the three were studied to confirm the effectiveness of using Derwent classes to code the "is mixed patent" variable. The result was independently verified by two coders and checked by the researchers.

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

2. *Individuals:* Patent records with no assignee were labeled *individual*. We used fuzzy computational methods to check for whether the names of the assignee and the inventor matched; such patent records were also labeled as individual.[21]
3. *University/research organization:* We matched a partial string of the assignee's name with the keywords *council, board, college, center, centre, university, research, organization, school, laboratoire,* and *institut*. Universities and research organizations named in the filtered results were labeled U.S. or foreign depending on the country named in the assignee column.
4. *Others:* All other patents were labeled *others*. Most were firms outside the Fortune 1000.

We next coded the search string employed by the examiner of each USPTO patent. This data was collected from the "image file wrapper": in 2003 the USPTO implemented the Image File Wrapper (IFW) system—an image technology system for storage and maintenance of patent-application records (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)*. The image file wrapper is an electronic record of the patent-examination process. Each USPTO patent application now has a publicly available compressed image file folder containing several documents. The document whose name ends with the abbreviation SRNT contains the search strings used by the examiner to search prior art.[22] Using an automated program,

[21] In the USPTO and EPO data, names could be listed as "Firstname Middlename Lastname", "Lastname, Firstname Middlename," or "Firstname (Alias/Preferred Name) Middlename Lastname"; we checked all naming conventions to extract first, middle and last names. Our algorithm then 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 so 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, we accept two names as the same.

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

we downloaded the pair data from USPTO Bulk Downloads (https://www.google.com/googlebooks/uspto-patents-pair.html). We were able to download 6,921 image file wrappers (out of the 7,172 USPTO patents) and found at least one SRNT file for 1,932 patents. (We could not find an SRNT file for every USPTO patent in our sample because the USPTO Bulk downloads data was still incomplete.) We coded the variable *searched_herbal_priorart* as 1 if the SRNT file mentioned at least one of the herbal patent classes listed in Appendix Table A1. The SRNT files were coded by two independent coders. Finally, we coded family patents filed with the EPO using INPADOC (International Patent Documentation), an EPO database.

IV. Results

A. Summary Trends

Figure 2 illustrates trends in the number of USPTO and EPO herbal patent filings between 1977 and 2011. The figure shows an earlier break in patent filings at the EPO than at the USPTO: specifically, it suggests a discontinuity at the EPO around 2003, when the demo CD was made available to the EPO.

We conducted additional analyses of the summary trend data. To empirically verify the common-trends assumption in the pre-treatment data (relevant to the difference-in-differences specification discussed in a later section), we conducted two robustness checks. First, we employed a Logit specification with year dummies, where the dependent variable is a dummy variable "is_EPO," indicating that the patent was filed with the EPO. We then calculated marginal effect for each year, which is the probability of "is_EPO" across year. Results available from the authors indicate that the probability of "is_EPO" declines after 2000, but the confidence intervals between neighboring years overlap until 2003, the year when the demo CD was made available to the EPO.

Next, we regressed the number of patent applications on the EPO dummy interacted with a linear trend, which is the application year in the data set. We performed this analysis separately for the pre-treatment period (prior to 2003) and the post-treatment period (2003 and later). The objective of this exercise was to see whether the linear trend is identical for the two groups. Results available from the authors indicate that, in the pre-treatment period, the predicted counts of applications to the EPO and USPTO always overlap in the confidence interval and are indistinguishable. However, in the post-treatment period the predicted counts of applications to the EPO and USPTO do not overlap in the confidence interval. This pattern indicates that, prior to 2003, the two trends are indistinguishable; after 2003, the two trends diverge. Summary statistics for EPO and USPTO patent applications and for the combined sample appear in Table 2. Figure 3 plots the trend in herbal patents granted at the USPTO and EPO between 1977 and 2011.

B. Empirical Question 1: How Does the Level of Herbal Patent Filings React to TKDL

Did the 2003 shock, the 2009 shock or both disproportionately affect patent filings at the EPO? Results reported in Table 3 are in line with the difference-in-differences specification (1). Table 3 reports on the results of the EPO's receipt of a demo CD of the TKDL in 2003. (The demo CD, which included only 500 formulations, was not used by EPO examiners to search for prior art, but it signaled the EPO's commitment to the TKDL project. The USPTO did not receive access to the TKDL until 2009.) The dependent variable is "Is EPO patent," or whether a filing represents an herbal formulation filed with the EPO. The results indicate that the EPO's receipt of the demo CD in 2003 resulted in a disproportionate decline in filings, especially for Indian herbal patents, at the EPO as compared to the USPTO. The variable of interest is the interaction between indian_herb and post2003 (that is, the variable *indian_post2003*); the coefficient for this variable is negative and significant across all models. The results are robust to adding year fixed effects instead of the

time trend, adding separate time trends for the pre- and post-shock periods, adding fixed effects for the assignee, etc. As Figure 4A shows, the average predicted probabilities of a non-Indian and an Indian herbal patent being filed at the EPO prior to the demo CD are 0.54 and 0.46 respectively. The average probability of a non-Indian herbal filing at the EPO after the CD is 0.41; that of an Indian herbal filing is 0.20.

Results reported in Table 4 are in line with the difference in differences specification (1). Table 4 reports on the results of the EPO's receipt of access to the most important codified Indian herbal formulations in 2005–2006. (The USPTO did not receive access to the same prior art until 2009.) Models 1–7 include the years 1977–2009; models 8–14 include the years 2004–2009. In other words, models 8–14 report the incremental effects on filings of receipt of access to the database in 2005–2006, over and beyond the effect of receipt of the demo CD in 2003. (Unlike the demo CD, the database was actually used by EPO examiners to search for prior art.) The dependent variable is "Is EPO patent," or whether a filing represents an herbal formulation filed with the EPO. The results indicate that the EPO's receipt of access to the database in 2005–2006 resulted in a disproportionate decline in filings, especially for Indian herbal patents, at the EPO as compared to the USPTO. The variable of interest is the interaction between indian_herb and post2006 (in other words, the variable *indian_post2006*). The coefficient for the variable is negative and significant across all models. The results are robust to adding year fixed effects instead of the time trend, adding separate time trends for the pre- and post-shock periods, adding fixed effects for the assignee, etc. Figure 4B plots the average predicted probabilities of an EPO filing using Table 4, Models 7 and 14.

C. Test for Mechanism - Family Patents for USPTO Patents

Next we explore a mechanism that could plausibly explain the disproportionate decline in Indian herbal patent filings at the EPO after the codification shock: that U.S. inventors were less likely after the shock to seek a family patent from the EPO for an Indian herbal patent. A *family patent* protects a given invention in different patent systems; the country where the invention is first patented is termed "priority"; the patent is then filed with other patent systems.

Table 5 reports on applications for family patents at the EPO for herbal patents previously granted by the USPTO. The data on EPO family patents for USPTO herbal patents was collected from the EPO's INPADOC database. Specifically, Table 5 reports on whether the 2005–2006 codification shock affected the probability of filing an EPO family-patent application for a USPTO herbal patent. The dependent variable is "*is EPO family patent*," or whether the USPTO herbal patent has a corresponding family family-patent application filed with the EPO. The results indicate that the shock of 2005-2006 affected filings for EPO family patents, especially for Indian herbal patents. The variable of interest is the interaction between indian_herb and post2006 (in other words, the variable *indian_post2006*). The coefficient for this variable is negative and significant across all models.

D. Empirical Question 2: How Does the Nature of Herbal Patenting React to TKDL

The results reported in Table 6 are in line with the difference-in-differences specification (2) and exploit the fact that the TKDL was implemented earlier at the EPO than at the USPTO. The dependent variable is "Is Mixed Patent," or whether the patent is a mixed patent. [The results indicate a disproportionate shift toward mixed patents after partial adoption of the TKDL by the EPO in 2006. Table 6 tracks the post-2006 effect until 2009, when the TKDL had been fully adopted by both the EPO and the 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. The results are robust to adding year fixed effects instead of the time trend, adding separate time trends for the pre- and post-shock periods, adding fixed effects for the assignee, etc. Figure 5 plots average predicted probabilities of observing a mixed patent using Model 7 of Table 6. The average predicted probabilities of observing of applications for a mixed patent at the EPO and USPTO prior to 2006 are 0.16 and 0.24 respectively; those after 2006 are 0.25 and 0.28 respectively.

E. Empirical Question 3: How Do Herbal Patent Grants React to TKDL

Results reported in Table 7 are in line with the difference-in-differences specification (3). Given the EPO's three-to-four-year time frame to examine patents, the shock of 2005–2006 would have affected patents filed in 2002–2003. Thus the main independent variable is a dummy variable for whether a patent was filed after 2002 (post2002). The dependent variable is "Is EPO patent," or whether a patent is an herbal patent formulation filed with the EPO. The results in Table 7 indicate that the shock of 2005–2006 did affect grants of patents filed three to four years earlier, especially Indian herbal patents filed with the EPO. The variable of interest is the interaction between indian_herb and post2002 (in other words, the variable indian_post2002). The coefficient for this variable is negative and significant across all models. The results are robust to adding year fixed effects instead of the time trend, adding separate time trends for the pre- and post-shock periods, adding fixed effects for the assignee, etc. Figure 6 plots the average predicted probability of an EPO patent being granted before and after the shock. The average predicted probabilities of a granted patent being an EPO patent before the filing year of 2002 are 0.41 and 0.38 for non-Indian and Indian herbal patents respectively; after the filing year of 2002, the average predicted probabilities are 0.35 and 0.24 for non-Indian and Indian herbal patents respectively.

F. Empirical Question 4: Smoking Gun Test Using Examiner Search Strings from Patent Image File Wrappers

Finally, we used unique data from the USPTO to validate that codification of prior art affects the search strategies of patent examiners. The results reported in Table 8 indicate that U.S. patent examiners are disproportionately likely to search for herbal prior art post-2009, the year the TKDL database was adopted by the USPTO. Here we employ a fixed effects Logit model (specification 4) 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 finding suggests that, after 2009, U.S. patent examiners started searching the TKDL for prior art, disproportionately in the case of patent applications using Indian herbal formulations.

To follow up these empirical findings, we conducted field interviews with patent examiners at the USPTO. The supervisor of the unit responsible for examining U.S. patent applications related to herbal extracts confirmed in 2015 that examiners inserted references from the TKDL database in around 10 percent of herbal-extract patent applications. The supervisor also confirmed that the references collected from the TKDL were not available from any other prior-art database. Nor did examiners refer to search reports from the EPO as a matter of standard practice while examining herbal patents.

G. Additional Robustness Checks

We also investigated contemporaneous policy shocks that might have affected our results. One such policy shock was the World Health Organization's Traditional Medicine Survey, conducted

in 2002–2005;[23] another was a directive on traditional medicine issued by the European Union Parliament.[24] Such policy changes could have affected filings and/or grants of *all* herbal patents. But, given our findings that filings and grants of Indian herbal patents were disproportionately affected by the EPO's adoption of the TKDL, we maintain confidence in our results.

To rule out the possibility that publicity about adoption of the TKDL by the EPO could have affected patent filings/grants in both jurisdictions, we collected media references to herbal medicine, the TKDL, etc., around the shock period in both European and U.S. media outlets using Factiva and LexisNexis. The results, available from the authors, indicate far more media mentions in Europe than in the United States: 361 articles in European media in 2003 (when the TKDL CD was adopted at the EPO) versus 63 U.S. articles.

To investigate the possibility of strategic patenting in response to the EPO policy change, we examined patent-filing patterns in Germany and Spain (the two European countries in our sample with the most patent filings). Our worry was that assignees filing patents with the USPTO might bypass the EPO by filing patents in individual European countries. Our analysis indicates that this was not the case. We first grouped the patents in our sample into Derwent World Patents Index (DWPI) patent families (a set of simple patent families available in the Thomson Innovation database and very similar to the DOCDB patent families used in the master database at the EPO). When we analyzed German and Spanish patent filings for DWPI families with USPTO priority application and without EPO patent filings, we observed that they represent a miniscule share of overall patent filings. In other words, individual European patent filings are highly correlated with EPO patent filings; we do not see any evidence of strategic patenting.

[23] Source: http://www.wpro.who.int/health_technology/book_who_traditional_medicine_strategy_2002_2005.pdf. Website accessed on July 20 2017.
[24] Source: https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-1/dir_2004_24/dir_2004_24_en.pdf. Website accessed on July 20, 2017.

V. Conclusion

We study how provision of information, in the form of codified prior art, affects innovation. The innovation literature offers very few opportunities to study an exogenous variation in patent policy across countries. As Lerner (2002) argues, a challenge facing cross-sectional studies of the impact of IP on patenting/innovation is that unobserved factors can affect both, leading to incorrect inferences. This paper has introduced a new dataset of herbal patents filed at the EPO and USPTO and exploited a natural experiment characterized by an exogenous time lag between the EPO and the USPTO in adoption of a database of codified prior art. We study the effect of prior-art codification on patent filing, patent grants and the nature of the innovation being patented. We find, first, a disproportionate decline in the level of herbal patent filings and grants at the EPO, as compared to the USPTO, especially for Indian herbal patents, after the shock of prior-art codification. Second, we find a disproportionate shift in filings of mixed patent applications (those consisting of both an herb and a synthetic compound) at the EPO after the codification shock. Finally, we find that examiners were more likely to search for Indian herbal prior art after adoption of the TKDL by the USPTO in 2009. Our findings contribute to the literatures on intellectual property rights (IPR) and innovation, on quality of patents in standards, on patent litigation, on information codification and on the role of developing countries in global innovation.

Our results also contribute to the longstanding debate on whether changes in the IPR regime affect patenting and innovation (Lerner 2002, Moser 2005). We sidestep one of the main findings of this literature: that IPR reform has no effect on domestic innovation. Our results suggest that, conditional on a particular IPR regime, strengthening prior art codification affects both the direction and the composition of patenting. More broadly, our findings contribute to the literature on whether IP rights to existing technologies affect subsequent innovation (Williams, 2013). Our

findings also have implications for improvements to patent examination, particularly for new technologies. Much prior art in technical fields resides in the non-patent literature, which tends to be more difficult to search.[25]

An important limitation of our study is external validity. Future research will need to study the effect of prior-art codification in other contexts and other patenting classes. In this regard, our results are relevant to several new policy initiatives. Two such initiatives being implemented by the USPTO are the Peer to Patent initiative and new provisions on third-party submission of prior art under the new America Invents Act. The Peer to Patent project is a joint initiative of the USPTO, New York Law School and IBM to gather publicly available prior art in a structured manner; it pertains to technologies such as software and business methods, telecommunications, speech recognition, translation, biotechnology, bioinformatics and biopharmaceuticals. Similar projects are being piloted by patent offices in Australia, Japan, Korea and the UK.[26] The Leahy–Smith America Invents Act (AIA), a U.S. federal statute enacted in 2011, has improved the process whereby third parties submit relevant prior art to the Patent Office. The AIA makes this process anonymous for third parties, and allows them to comment on the prior art submitted. Actors other than the USPTO, including firms, have also 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 of 20 top U.S. banks, to provide the USPTO with non-patent prior art describing the U.S. financial infrastructure. Merges (2004) documents the role of ex-ante information disclosure in two cases: the Merck Gene

[25] Sampat (2004) quotes former USPTO Commissioner Q. Todd Dickinson's assertion 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).
[26] Source: http://en.wikipedia.org/wiki/Peer-to-Patent (website accessed on February 16, 2015).

Index and IBM's investment in Linux.[27] The Appendix summarizes current initiatives to strengthen and codify prior art in multiple industries.

The literature on setting standards has delineated issues pertaining to the quality of patents with regard to science and engineering process standards (SEPs) (Willingmyre, 2012). The EPO is pursuing agreements with standard-setting organizations (SSOs) to provide examiners access to information on technical standards. For example, Bekkers et al. (2016) examine a policy change aimed at including information revealed during the standard-setting process in the official definition of prior art. Several innovations in mobile telecommunications had already been discussed at standard-setting organizations before being applied for as patents; the EPO thus collaborated with several SSOs to implement a platform ensuring examiners easy and prompt access to all relevant documents (Willingmyre, 2012). Willingmyre reports a negative and strongly significant reduction in the patent-granting rate after this shock, suggesting that the patent-granting process became more careful and selective after the policy implementation. To our knowledge, the USPTO has not implemented similar initiatives. Our study suggests that access to such information helps Patent Offices improve patent quality (by denying patents that should not be granted), the issue that standards regimes complain most about.

Our findings also have implications for the literature on patent litigation. Facilitating patent examiners' access to searchable codified prior art might discourage grants of legally contestable patents and thus reduce litigation and social costs. Lemley (2001) documents that around 1600 lawsuits involving about 2000 patents were filed annually as of 2001. Though the number that was

[27] In 1995, the pharmaceutical company Merck created a public database that made gene sequences publicly available, arguably to prevent patenting and thus to protect Merck, which uses gene sequences as an input. Similarly, IBM's investment in Linux created an open-source alternative to Microsoft's Windows platform and thus lowered the costs of the operating system, an essential input for IBM. Bhaskarabhatla and Hegde (2014) provide an in-depth study of patenting policies at IBM.

litigated is relatively small compared to the number of patents granted, the costs of litigation were very high: Lemley (2001) found the cost of patent litigation for each side to be around $799,000 through end of discovery and $1.5 million through trial and appeal.[28]

Our findings are also relevant to other initiatives to codify information on biological resources, including those of the Amazon. Debate continues on whether the biological resources of the Amazon could be codified as a global public good (Nobre et al., 2016). The Amazon Third Way initiative being pioneered by the World Economic Forum (WEF) is attempting to design and deploy an Amazonian Bank of Codes, an open global public-good digital platform that will map the biological assets of the Amazon and provide a global marketplace that reduces search and transaction costs for providers and users of IP in the Amazonian countries and globally.[29]

Finally, our results have implications for the literature in economics that has distinguished between the innovative North and the imitative South. The underlying premise is that most patented products and processes consumed in the South are developed in the North. Stronger patent protection in the South would protect the North against imitation in its export markets, but the South would have to pay higher prices for those products. Thus it is in the interests of the South to maintain weak patent protection to facilitate imitation, benefitting consumers in the South via lower prices (Chin and Grossman, 1988). Grossman and Lai (2004) consider the incentives and benefits of local innovation. However, the realm of herbal patents is one in which entities from the developed countries have patented herbal formulations from China and India, where prior art has been public knowledge for decades, if not for centuries.

[28] The literature has also documented the social costs of granting bad patents. Lemley (2001) outlines the in terrorem effects of bad patents: potential competitors and follow-on innovators might be deterred from entering a field by the existence of bad patents. Lanjouw and Lerner (2001) show that large firms tend to use preliminary injunctions to impose financial distress on smaller rivals.

[29] Source: https://www.weforum.org/agenda/2017/06/bio-inspired-design-amazon-technology

References

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth through Creative Destruction." Econometrica, 60(2): 3

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.

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.

Bekkers, Rudi, Arianna Martinelli, and Federico Tamagni. The causal effect of including standards-related documentation into patent prior art: evidence from a recent EPO policy change. No. 2016/11. Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy, 2016.

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

Bhaskarabhatla, Ajay, and Deepak Hegde. "An organizational perspective on patenting and open innovation." Organization Science 25.6 (2014): 1744-1763.

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).

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.

Galasso, Alberto, and Mark Schankerman, Patents and Cumulative Innovation: Causal Evidence from the Courts', The Quarterly Journal of Economics, Volume 130, Issue 1, 1 February 2015, Pages 317–369, https://doi.org/10.1093/qje/qju029

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, Gene M., and Elhanan Helpman. "Trade, knowledge spillovers, and growth." European economic review 35.2-3 (1991): 517-526.

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.

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.

Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." Journal of Political Economy, 103(4): 759-84.

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.

Nobre, Carlos A., et al. "Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm." Proceedings of the National Academy of Sciences 113.39 (2016): 10759-10768.

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.

Romer, Paul M. "Endogenous technological change." Journal of political Economy 98.5, Part 2 (1990): S71-S102.

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.

Williams, Heidi, Intellectual Property Rights and Innovation: Evidence from the Human Genome, 2013, Journal of Political Economy 121(1): 1-27.

Willingmyre, G. T. (2012): Cooperation between Patent Offices and Standards Developing Organizations, Washington, DC: National Academies of Science.

TABLE 1. COMPARISON OF TKDL ACCESS AGREEMENTS SIGNED BY THE USPTO AND THE EPO

	USPTO	**EPO**
Restrictions on use	No general restrictions on use of the TKDL	General restriction on use of the TKDL for purposes other than "European patent grant procedures"
Restrictions on access	Explicit access permitted to both USPTO examiners/staff and "contractors engaged in search of Patent Cooperation Treaty (PCT) applications"	Use limited to EPO staff
Permitted users	Information use permitted 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 "purposes of the European patent grant procedure in all its phases"
Permitted information transfer	TKDL information may be shared with third parties for purposes of patent search and examination, with patent applicants/legal representatives for prior-art purposes, and with the public via the USPTO's Patent Application Information Retrieval System	TKDL information may be shared only for purposes of patent grant procedures and only with patent applicants; the form of information provision is not limited
User access	User access is restricted to 30 IP addresses simultaneously	ID-based access and IP address-based access may be restricted

Source: Interviews and analyses of the TKDL access agreements signed by the Indian Council of Scientific and Industrial Research with the USPTO and the EPO.

TABLE 2 . EPO AND USPTO HERBAL PATENTS, SUMMARY STATISTICS

	EPO & USPTO Combined					EPO Only					USPTO Only				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max	N	mean	sd	min	max	N	mean	sd	min	max
Claim count	11,260	17.16	57.83	0	5,929	4,099	14.15	15	0	203	7,161	18.88	71.56	0	5,929
Backward citations	11,262	6.697	22.13	0	1,175	4,099	3.51	4.867	0	91	7,163	8.522	27.34	0	1,175
Forward citations	11,262	3.419	9.328	0	177	4,099	0.619	4.757	0	177	7,163	5.021	10.81	0	164
Application year	11,262	2003	6.864	1977	2013	4,099	2001	7.171	1978	2013	7,163	2005	6.287	1977	2013
Granted	11,262	0.445	0.497	0	1	4,099	0.445	0.497	0	1	7,163	0.446	0.497	0	1
Indian herb	11,262	0.069	0.253	0	1	4,099	0.0373	0.19	0	1	7,163	0.0871	0.282	0	1
Is mixed patent (herb + synthetic)	11,247	0.268	0.443	0	1	4,085	0.193	0.395	0	1	7,162	0.31	0.463	0	1
Is EPO patent	11,262	0.364	0.481	0	1	4,099	1	0	1	1	7,163	0	0	0	0
Assignee type1(Fortune1000 firms)	9,172	0.0341	0.182	0	1	4,088	0.024	0.153	0	1	5,084	0.0423	0.201	0	1
Assignee type2 (Individuals)	9,172	0.135	0.342	0	1	4,088	0.002	0.0442	0	1	5,084	0.242	0.428	0	1
Assignee type3 (Other firms)	9,172	0.731	0.443	0	1	4,088	0.914	0.281	0	1	5,084	0.584	0.493	0	1
Assignee type4 (Universities)	9,172	0.0997	0.300	0	1	4,088	0.0602	0.238	0	1	5,084	0.131	0.338	0	1

Notes:

- From a search space of around 12 million USPTO patents and 4.5 million EPO patents published between January 1, 1977, and April 30, 2014, we ended up with 7172 herbal patents at the USPTO and 4099 herbal patents at the EPO. We do not use 2014 data for reasons of incompleteness. The final dataset consisted of 11,262 patent applications filed between 1977 and 2013.
- We coded patent applications as having been filed for a pure herb or for an herb combined with a synthetic compound (is_mixed patent=1). We used the Derwent classification system and coded patent applications as mixed if they belonged to Derwent classes B05, B06, or B07
- We coded a patent application as pertaining to an Indian herb if it contained an herb name (scientific name or popular name) listed in the TKDL. (http://www.tkdl.res.in/TKDL/LangDefault/Common/Utility/KeywordDemo/F-Plant-Name_Tips.asp). We searched the application for the scientific name of the Indian herb and for its name in the Ayurveda, Siddha and Unani schools of medicine.
- Assignee information was available for 9172 of the 11262 patent applications. We coded those filed by Fortune 1000 firms using a list available from the authors. We used fuzzy computational methods to determine whether the names of the assignee and the inventor were identical; such filings were labeled *individual*. We searched the assignee name for the keywords *council, board, college, center, centre, university, research, organization, school, laboratoire, and institut* to code the assignee as a university. All other assignees were labeled *others*. Most were firms outside the Fortune 1000.

TABLE 3. EFFECT OF THE EPO'S RECEIPT OF A PRIOR-ART DEMO CD ON FILINGS OF HERBAL PATENTS: DIFFERENCE-IN-DIFFERENCES TEST USING LOGIT SPECIFICATION

	Dependent Variable: *Is_EPO_patent (Based on Patent Filings)*						
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
post2003	-0.663***	-0.555***	-0.520***	-0.568***	-0.692***	-0.355***	-0.600***
	(0.0733)	(0.0775)	(0.0779)	(0.0833)	(0.0796)	(0.0983)	(0.105)
indian_herb		-0.662***	-0.623***	-0.572***	-0.579***	-0.645***	-0.412***
		(0.122)	(0.124)	(0.125)	(0.151)	(0.124)	(0.158)
indian_post2003		-0.992***	-0.981***	-1.066***	-1.035***	-0.997***	-1.114***
		(0.199)	(0.200)	(0.213)	(0.194)	(0.201)	(0.206)
Assignee type1					4.111***		4.502***
					(0.527)		(0.544)
Assignee type3					5.317***		5.410***
					(0.416)		(0.417)
Assignee type4					3.801***		3.806***
					(0.431)		(0.438)
claim_count			-0.0151***				-0.0165***
			(0.00266)				(0.00289)
backward_citations				-0.0897***			-0.0928***
				(0.00774)			(0.00837)
Time trend	No	No	No	No	No	Yes	Yes
Observations	7,252	7,252	7,250	7,252	7,252	7,252	7,250

Notes:

- The results reported in Table 3 are in keeping with the difference-in-differences specification (1). The shock is *the EPO's receipt of a demo CD of the TKDL in 2003*. We use the shock to test whether *filings* of patents decline disproportionately at the EPO, especially for patents based on Indian herbs. This demo CD consisted of only 500 Indian herbal formulations; it was not used by EPO examiners to search for prior art, but signaled the EPO's commitment to the TKDL project; the USPTO received access to the TKDL only in November 2009. The dependent variable is "Is EPO patent," or whether the patent application is an herbal patent filed with the EPO.
- The results indicate a disproportionate decline in filings of patents, especially Indian herbal patents, at the EPO. The variable of interest is the interaction between *indian_herb* and *post2003* (that is, *indian_post2003*). The coefficient for this variable is negative and significant across all models.
- The results are robust to adding year fixed effects instead of the time trend, adding separate time trends for the pre- and post-shock periods, adding fixed effects for the assignee, etc. Robust standard errors clustered by assignee appear in parentheses; *** p<0.01, ** p<0.05, * p<0.1

TABLE 4. EFFECT ON FILINGS OF HERBAL PATENTS OF THE EPO'S RECEIPT OF A PRIOR-ART DATABASE, DIFFERENCE-IN-DIFFERENCES TEST USING LOGIT SPECIFICATION

	Dependent Variable: *Is_EPO_Patent (Based on Patent Filings)*													
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 13	Model 14
post2006	-0.765***	-0.682***	-0.640***	-0.687***	-0.802***	-0.412***	-0.524***	-0.403***	-0.325***	-0.304***	-0.306***	-0.366***	-0.219	-0.126
	-0.0834	-0.0866	-0.0871	-0.0918	-0.0895	-0.0984	-0.105	-0.0926	-0.095	-0.0949	-0.0972	-0.0973	-0.185	-0.204
indian_herb		-0.833***	-0.782***	-0.748***	-0.774***	-0.812***	-0.583***		-0.915***	-0.850***	-0.853***	-0.953***	-0.916***	-0.744***
		-0.118	-0.119	-0.123	-0.146	-0.118	-0.149		-0.22	-0.225	-0.222	-0.238	-0.22	-0.248
indian_post2006		-0.698***	-0.701***	-0.802***	-0.783***	-0.704***	-0.938***		-0.668***	-0.677***	-0.741***	-0.709***	-0.666***	-0.827***
		-0.235	-0.232	-0.243	-0.247	-0.235	-0.255		-0.239	-0.238	-0.244	-0.259	-0.24	-0.268
assgn_type1					4.032***		4.454***					2.268***		2.355***
					-0.526		-0.545					-0.615		-0.629
assgn_type3					5.268***		5.394***					3.852***		3.772***
					-0.414		-0.416					-0.511		-0.511
assgn_type4					3.739***		3.802***					2.465***		2.320***
					-0.432		-0.44					-0.551		-0.554
claim_count			-0.0155***				-0.0163***			-0.0140***				-0.0171***
			-0.00262				-0.00281			-0.00484				-0.00491
backward_citations				-0.0881***			-0.0907***				-0.0598***			-0.0584***
				-0.00753			-0.00814				-0.00948			-0.00964
Time Trend	No	No	No	No	No	Yes	Yes	No	No	No	No	No	Yes	Yes
Observations	7,252	7,252	7,250	7,252	7,252	7,252	7,250	2,609	2,609	2,608	2,609	2,609	2,609	2,608

Notes:

- The results reported in Table 4 are in keeping with the difference-in-differences specification (1). The shock is *the EPO's receipt in 2005–2006 of access to codified prior art for the most important Indian herbal formulations.* We use the shock to test whether filings of patents decline disproportionately at the EPO, especially for patents based on Indian herbs. Models 1–7 include the years 1977–2009; models 8–14 include the years 2004–2009; in other words, models 8–14 report the incremental effects on filings of access to the database in 2005–2006, over and above the effect of receipt of the demo CD in 2003. The CD signaled EPO's commitment to the TKDL project; the database was actually used by EPO examiners to search for prior art.
- The results indicate a disproportionate decline in the filing of patents, especially for Indian herbal patents, at the EPO. The variable of interest is the interaction between *indian_herb* and *post2006* (that is, the variable *indian_post2006*). The coefficient for this variable is negative and significant across all models. The results are robust to adding year fixed effects instead of the time trend, adding separate time trends for the pre- and post-shock periods, adding fixed effects for the assignee, etc. Robust standard errors clustered by assignee appear in parentheses; *** p<0.01, ** p<0.05, * p<0.1

TABLE 5. EFFECT ON FILINGS OF FAMILY PATENTS OF THE EPO'S RECEIPT OF A PRIOR-ART DATABASE, DIFFERENCE-IN-DIFFERENCES TEST USING LOGIT SPECIFICATION

	Dependent Variable: *Is_EPO_family_patent (Based on Patent Filings)*				
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 4
indian_herb	-0.147	-0.171*	-0.157*	-0.233**	-0.263**
	(0.0950)	(0.0955)	(0.0953)	(0.115)	(0.116)
indian_post2006	-0.275***	-0.275***	-0.259**	-0.252**	-0.240*
	(0.102)	(0.102)	(0.102)	(0.128)	(0.129)
Assignee type1				1.473***	1.380***
				(0.164)	(0.166)
Assignee type3				0.779***	0.743***
				(0.0735)	(0.0742)
Assignee type4				0.171*	0.144
				(0.103)	(0.103)
claim_count		0.00813***			0.00896***
		(0.00172)			(0.00227)
backward_citations			0.00417***		0.00185
			(0.00156)		(0.00142)
Observations	7,157	7,155	7,157	5,078	5,076
Year FE	YES	YES	YES	YES	YES

Notes:

- Table 5 reports the effect on filing of family patents at the EPO for herbal patents previously granted by the USPTO. Data on EPO family patents was collected from the EPO's INPADOC database. We exploit the fact that the EPO received access to prior art on 27,000 Indian herbal formulations in 2005–2006; the USPTO did not receive access to the TKDL until2009.
- The dependent variable is "Is EPO family patent" (that is, whether a USPTO herbal patent has a corresponding family patent filed with the EPO). The results indicate that the shock of 2005–2006 affected the filing of EPO family patents, especially for Indian herbal patents.
- The variable of interest is the interaction between *indian_herb* and *post2006* (that is, the variable *indian_post2006*). The coefficient for this variable is negative and significant across all models
- Robust standard errors clustered by assignee appear in parentheses; *** p<0.01, ** p<0.05, * p<0.1

TABLE 6. EFFECT ON THE COMPOSITION OF HERBAL PATENT FILINGS OF THE EPO'S RECEIPT OF A PRIOR-ART DATABASE, DIFFERENCE-IN-DIFFERENCES TEST USING LOGIT SPECIFICATION

	Dependent Variable: *Is_mixed_patent (Based on Patent Filings)*						
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
is_EPO	-0.543***	-0.499***	-0.519***	-0.523***	-0.603***	-0.512***	-0.486***
	(0.0808)	(0.0809)	(0.0824)	(0.0816)	(0.0865)	(0.0829)	(0.0891)
post2006	0.448***	0.428***	0.444***	0.445***	0.432***	0.183	0.199*
	(0.0970)	(0.0978)	(0.0970)	(0.0967)	(0.0983)	(0.116)	(0.117)
isEPO_times_post2006	0.397**	0.366**	0.402**	0.400**	0.423***	0.364**	0.366**
	(0.157)	(0.158)	(0.157)	(0.156)	(0.158)	(0.156)	(0.157)
claim_count		0.0147***					0.0132***
		(0.00234)					(0.00234)
backward_citations			0.00373**				0.00277*
			(0.00161)				(0.00146)
indian_herb				0.398***			0.342***
				(0.121)			(0.125)
Assignee type1					0.386**		0.165
					(0.167)		(0.166)
Assignee type3					0.218*		0.120
					(0.112)		(0.113)
Assignee type4					0.0245		-0.111
					(0.149)		(0.150)
Time trend	No	No	No	No	No	Yes	Yes
Observations	7,238	7,236	7,238	7,238	7,238	7,238	7,236

Notes:

- Results reported in Table 6 are in line with the difference-in-differences specification (2). The table exploits the 2006 shock to study whether the nature of the inventions being patented disproportionately changed at the EPO.
- The dependent variable is "Is mixed patent," or whether or not the patent is a mixed patent
- The results indicate a disproportionate shift toward mixed patents (herbs added to synthetic compounds) at the EPO after its partial adoption of the TKDL in 2006.
- Table 6 analyzes the *post2006* effect until 2009, when the TKDL was fully adopted by both the EPO and the USPTO. The variable of interest is the interaction between *is_EPO* and *post2006* (that is, the variable *isEPO_times_post2006*). The coefficient for this variable is positive and significant across all models. The results are robust to adding year fixed effects instead of the time trend, adding separate time trends for the pre- and post-shock periods, adding fixed effects for the assignee, etc. Robust standard errors clustered by assignee appear in parentheses; *** p<0.01, ** p<0.05, * p<0.1

TABLE 7. EFFECT ON GRANTS OF HERBAL PATENTS OF THE EPO'S RECEIPT OF A PRIOR-ART DATABASE, DIFFERENCE-IN-DIFFERENCES TEST USING LOGIT SPECIFICATION

	Dependent Variable: *Is_EPO_patent (Based on Patent Grants)*						
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
post2002	-0.718***	-0.640***	-0.650***	-0.381***	-0.676***	-0.319**	-0.301**
	(0.0946)	(0.0995)	(0.0997)	(0.101)	(0.104)	(0.124)	(0.137)
indian_herb		-0.395***	-0.380***	-0.312**	-0.312**	-0.360***	-0.159
		(0.132)	(0.133)	(0.144)	(0.156)	(0.134)	(0.171)
indian_post2002		-0.629***	-0.615***	-0.628***	-0.561**	-0.654***	-0.552**
		(0.207)	(0.207)	(0.226)	(0.238)	(0.208)	(0.251)
Assignee type1					4.796***		5.516***
					(0.795)		(0.817)
Assignee type3					5.975***		6.182***
					(0.711)		(0.711)
Assignee type4					4.473***		4.427***
					(0.722)		(0.730)
claim_count			-0.00872**				-0.00480
			(0.00342)				(0.00456)
backward_citations				-0.118***			-0.138***
				(0.00760)			(0.00877)
Time trend	No	No	No	No	No	Yes	Yes
Constant	-0.241***	-0.216***	-0.0960	0.444***	-5.681***	0.396**	-4.836***
	(0.0642)	(0.0649)	(0.0852)	(0.0809)	(0.709)	(0.179)	(0.730)
Observations	4,544	4,544	4,542	4,544	4,544	4,544	4,542

Notes:

- Results reported in Table 7 are in keeping with the difference-in-differences specification (3).
- Given the three-to-four-year time frame to examine patents at the EPO, the shock of 2005–2006 would have affected patents filed in 2002–2003. Accordingly, the shock (the main independent variable) is a dummy variable for whether the patent was filed after 2002 (*post2002*). We use the shock to test whether *grants* of patents declined disproportionately at the EPO, especially for patents based on Indian herbs.
- The dependent variable is "Is EPO patent," or whether the patent is an herbal patent filed with the EPO. The results indicate that the shock of 2005–2006 affected the granting of patents filed three to four years earlier (in 2002–2003), especially Indian herbal patents.
- The variable of interest is the interaction between *indian_herb* and *post2002* (that is, *indian_post2002*). The coefficient for this variable is negative and significant across all models. The results are robust to adding year fixed effects instead of the time trend, adding separate time trends for the pre- and post-shock periods, adding fixed effects for the assignee, etc.
- Robust standard errors clustered by assignee appear in parentheses; *** p<0.01, ** p<0.05, * p<0.1

TABLE 8 . EFFECT ON EXAMINER SEARCH STRATEGY OF THE USPTO'S RECEIPT OF A PRIOR-ART DATABASE, FIXED EFFECTS LOGIT

	Dependent Variable: *Searched_herbal_priorart (Based on Patent Filings)*							
VARIABLES	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	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:

- Table 7 employs a patent-examiner fixed effects Logit model (specification 4); the dependent variable here is *Searched_herbal_priorart.*
- We find robust evidence across models 1–8 that USPTO patent examiners were likely to search for herbal prior art post-2009. We use the 2009 shock to test whether USPTO patent examiners searched for herbal prior art after 2009, especially for patents based on Indian herbs.
- The coefficient of the interaction term (*post2009 times Indian herb*) in Model 8 indicates that U.S. patent examiners were disproportionately likely to search for prior art for Indian herbs post-2009.
- Robust standard errors appear in parentheses; *** p<0.01, ** p<0.05, * p<0.1

FIGURE 1. MILESTONES IN ADOPTION OF CODIFIED INDIAN HERBAL PRIOR ART BY THE EPO AND THE USPTO

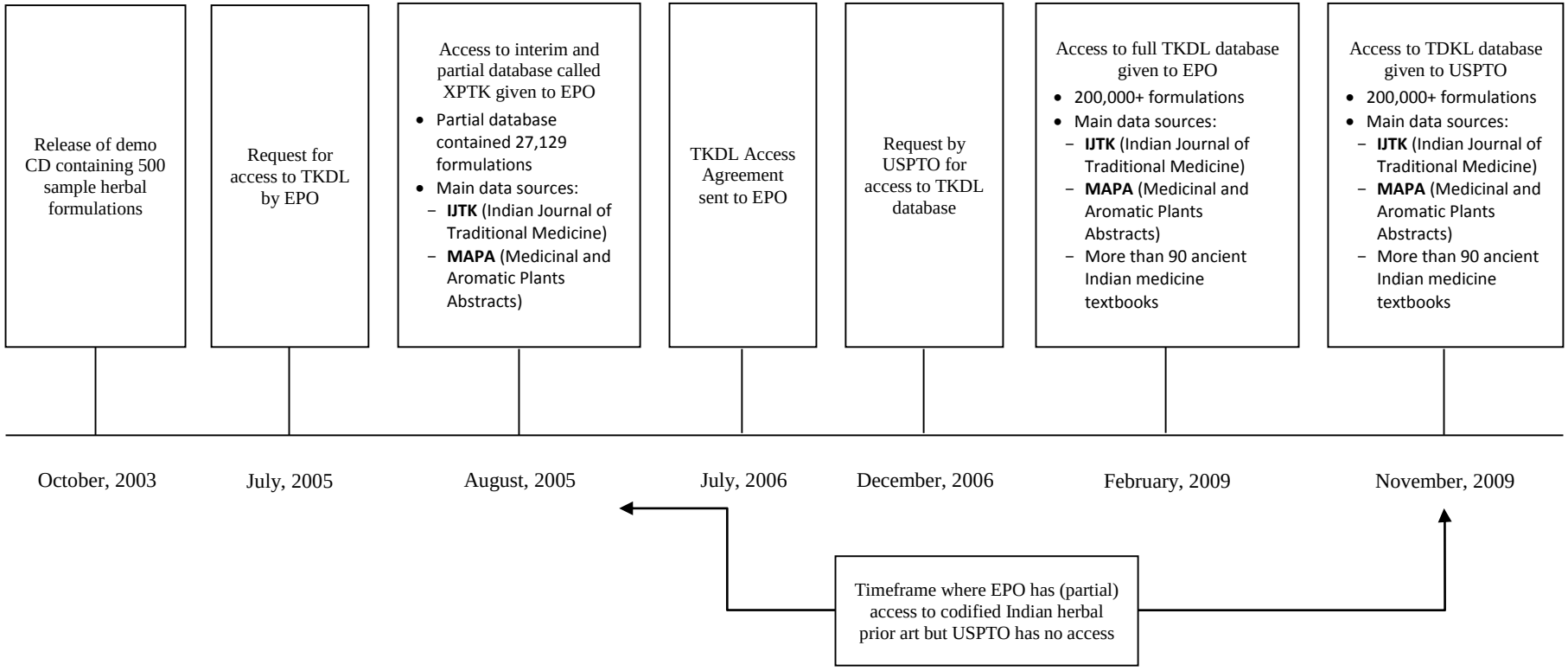
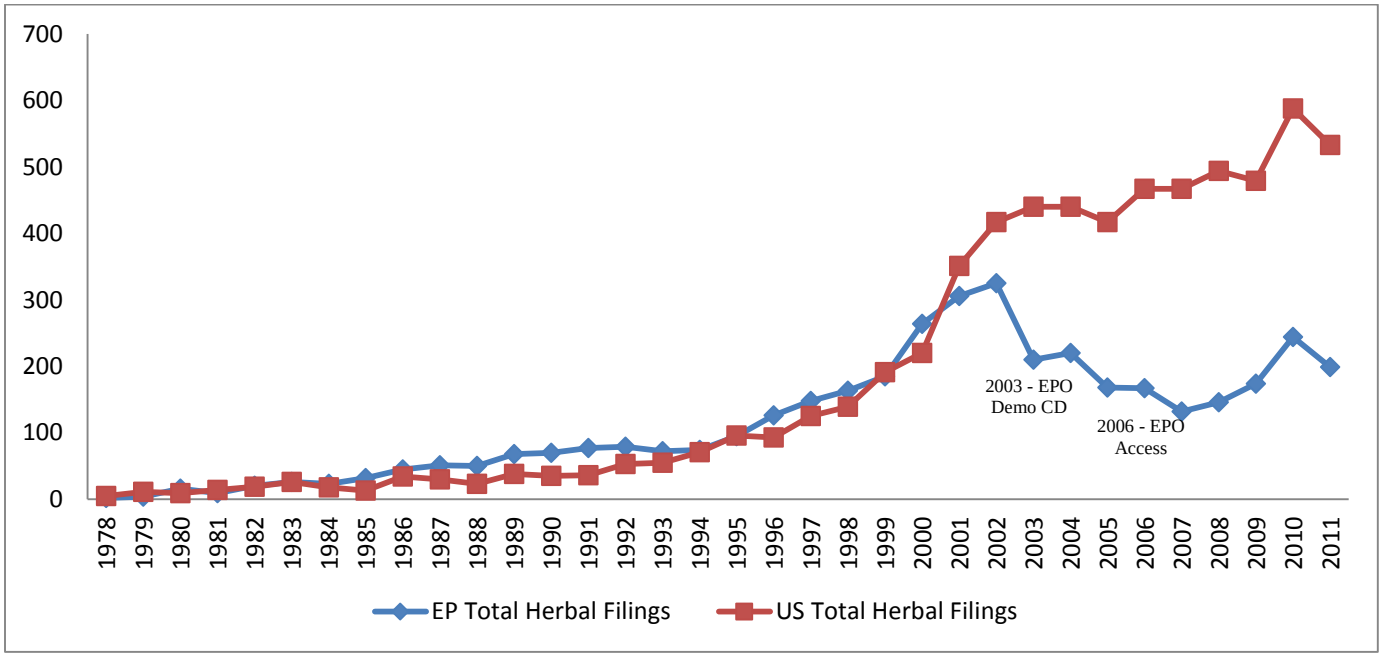
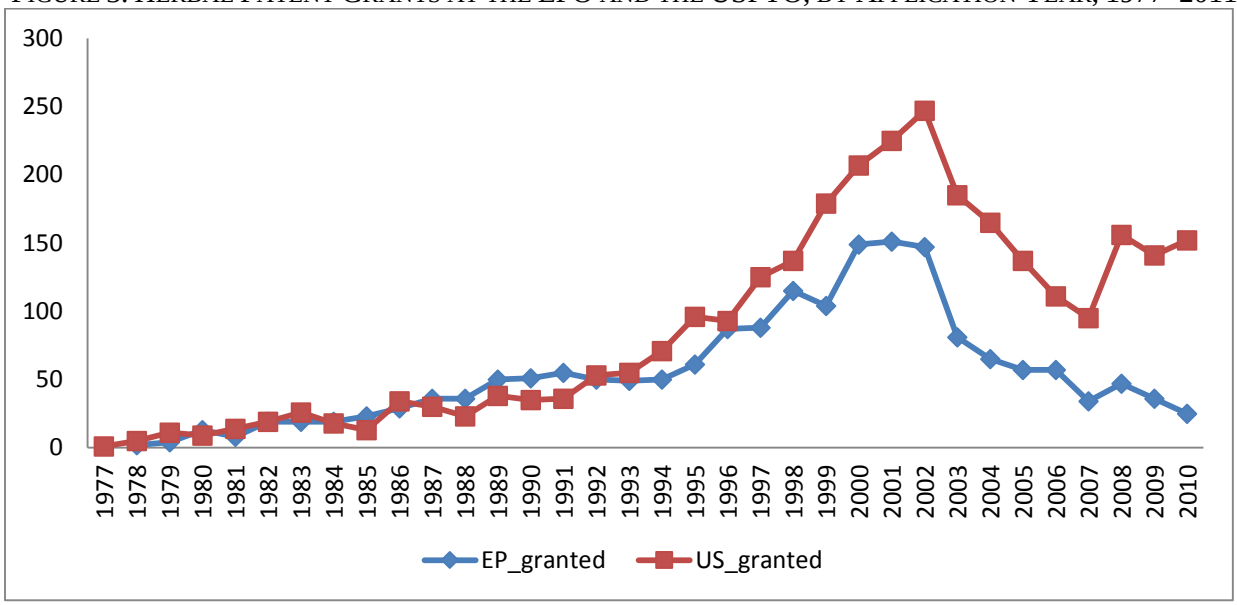


FIGURE 2. HERBAL PATENT FILINGS AT EPO AND USPTO, 1977–2011



Notes: Figure 2 plots trend in herbal patent filings at the EPO and the USPTO. We also run the structural-breaks test (Andrews, 1993) and find structural breaks at the EPO in 2005 and at the USPTO in 2010. To empirically verify a common-trends assumption in the pre-treatment data (relevant to the difference-in-differences specification), we conduct two robustness checks. First, we employ a Logit specification with year dummies, where the dependent variable is a dummy variable *is_EPO*, indicating that the patent was filed with the EPO. We then calculate marginal effect for each year, which is the probability of *is_EPO* across years. Results available from the authors indicate that the probability of *is_EPO* declined after 2000, but the confidence intervals overlapped between neighboring years until 2003. Second, we regress the number of patent applications on the EPO dummy interacted with a linear trend, which is the application year in the data set. We performed this analysis separately for the pre-treatment period (prior to 2003) and the post-treatment period (after 2003). The objective of this exercise is to see whether the linear trend is the same for the two groups. Results available from the authors indicate that, in the pre-treatment period, the predicted counts of applications to the EPO and the USPTO always overlap in the confidence interval and are not distinguishable. However, in the post-treatment period, the predicted counts of applications to the EPO and the USPTO do not overlap in the confidence interval. This finding indicates that, prior to 2003, the two trends are indistinguishable; after 2003, they diverge.

FIGURE 3. HERBAL PATENT GRANTS AT THE EPO AND THE USPTO, BY APPLICATION YEAR, 1977–2011



Notes: Figure 3 plots trends in herbal patent grants at the EPO and USPTO, based on year of application. Patents filed in year t would have come up for examination three to four years later, in years t+3 or t+4.

FIGURE 4A. PREDICTED PROBABILITIES OF AN EPO PATENT BEING FILED BEFORE AND AFTER RECEIPT OF THE DEMO CD

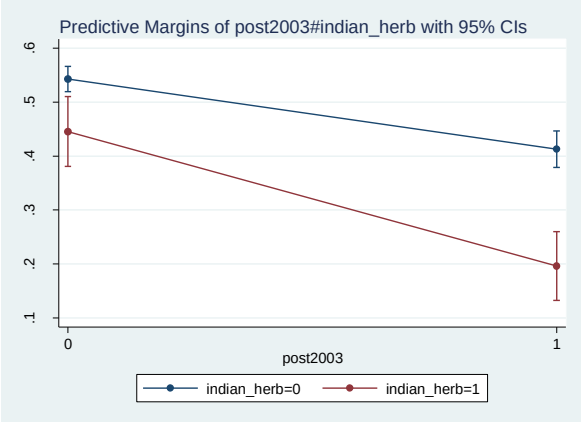
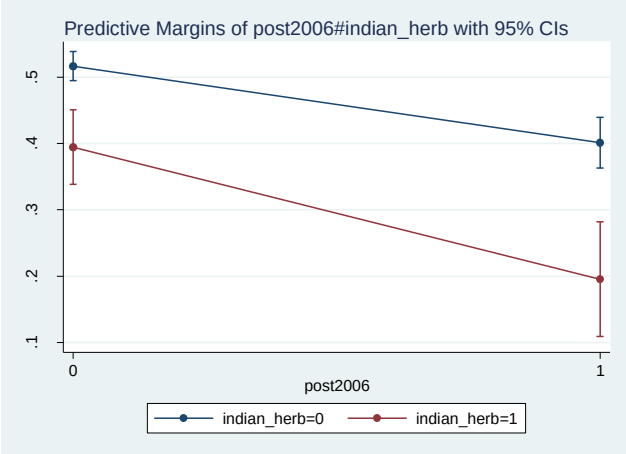
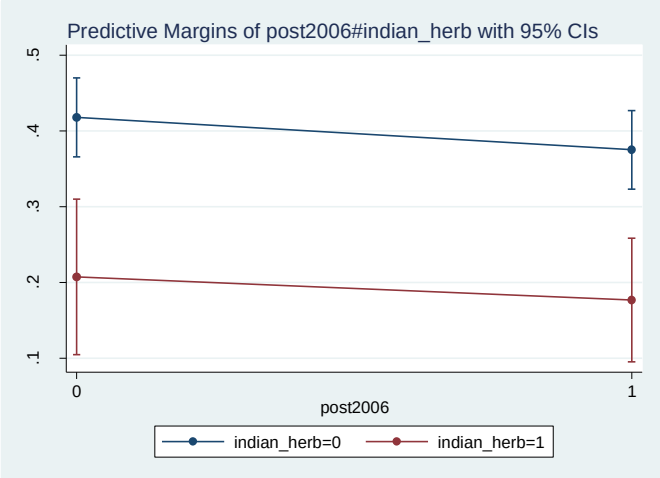


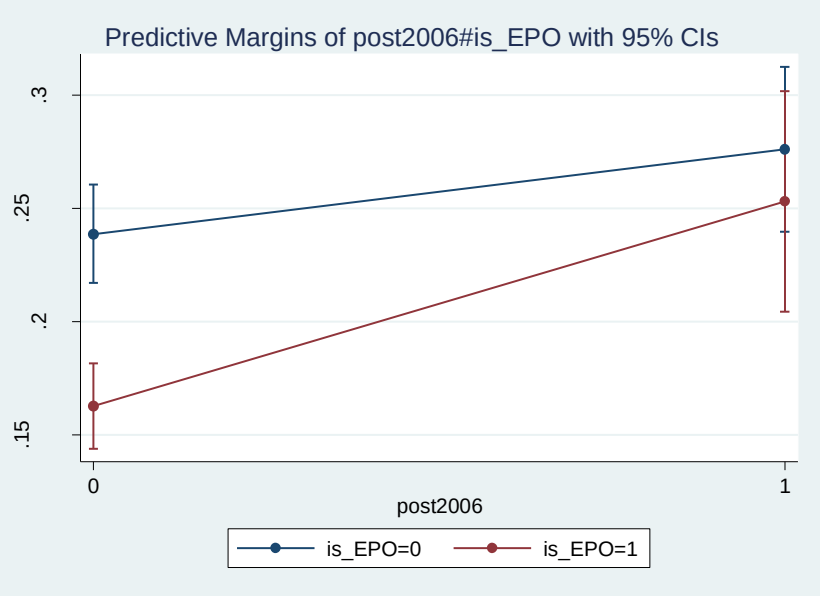
FIGURE 4B. PREDICTED PROBABILITIES OF AN EPO PATENT BEING FILED BEFORE AND AFTER RECEIPT OF ACCESS TO THE DATABASE





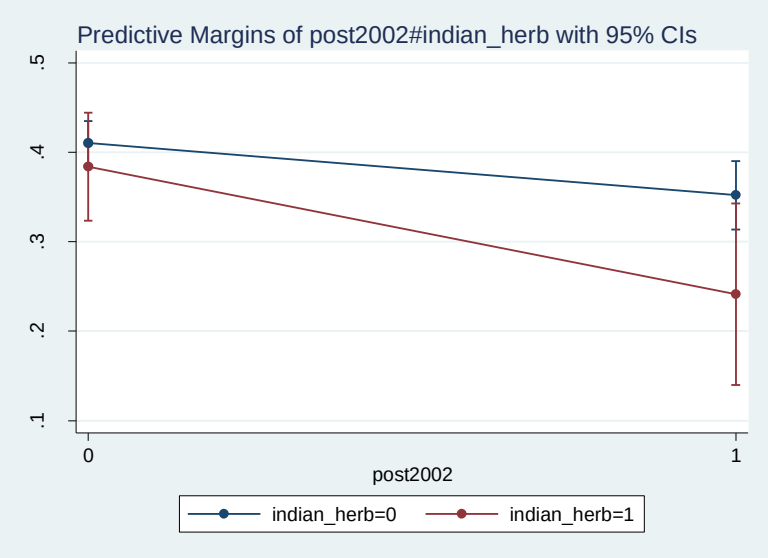
Notes: Figure 4A plots the average predicted probability of an EPO patent filing before and after receipt of the demo CD in 2003, using Table 3, Model 7. Figure 4B plots the average predicted probability of an EPO patent **[patent grant?]** before and after the EPO began using the database in 2005–2006, using Table 4, Models 7 and 14. In other words, the upper panel in Figure 4B does not account for receipt of the demo CD; the lower panel does account for the incremental effect over and above the effect of the demo CD.

FIGURE 5. PREDICTED PROBABILITIES OF FILINGS OF MIXED PATENTS BEFORE AND AFTER THE EPO'S RECEIPT OF ACCESS TO THE DATABASE



Notes: Figure 5 plots the average predicted probability of a mixed patent being filed before and after the EPO began using the database in 2005–2006, using Table 6, Model 7.

FIGURE 6. PREDICTED PROBABILITIES OF GRANTS OF PATENTS BEFORE AND AFTER THE EPO'S RECEIPT OF ACCESS TO THE DATABASE



Notes: Figure 6 plots the average predicted probability of an EPO patent being granted before and after the EPO began using the database in 2006, computed using Table 7, Model 7.

Appendix A

Insights from Field Interviews

We conducted qualitative analyses and interviews to rule out the possibility that the delay in the USPTO's adoption of the TKDL was driven by endogenous policy differences between the EPO and the USPTO that econometricians cannot observe.

At the USPTO we interviewed Dominic Keating, director of the IP Attaché Program, and Karin Ferriter of the Office of Policy & International Affairs, who served as an observer at negotiations on the TKDL Access Agreement between the USPTO and CSIR. Keating asserted that "the EPO had different statutory requirements than the USPTO," which led a number of terms in the original access agreement to be acceptable to the former but "nonstarters" for the latter. The most problematic clause for the USPTO prohibited it from revealing any of the content of the TKDL to the public. This requirement was out of keeping with the U.S. patent system, whose statutory restrictions mandated the USPTO to document its reasons for rejecting applications. "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." Both parties were constrained by statutory considerations—the USPTO by its mandate to provide documentation of rejections and the CSIR by its mandate to prevent exploitation of the TKDL by the public. Ultimately, it was agreed that the USPTO would disclose TKDL contents to third parties only to the extent necessary for patent search and examination.

Karin Ferriter further explained the delay in the USPTO's adoption of the TKDL: "India wanted to know examiner-entered search terms for each and every herbal patent application, irrespective of whether the application was published or not, and that was not acceptable to the USPTO. Negotiations on this went on for years." Also, the USPTO waited for the original Sanskrit

books that had been transcribed to create the TKDL to arrive at its library so that the source material would be publicly available at the USPTO, a statutory requirement.

We also interviewed Dr. V.K. Gupta, director of the TKDL and senior advisor to the Indian state-owned labs that created it. Dr. Gupta, arguably the main architect of the TKDL database, had oversight of the process from its inception until the signing of the access agreements. He confirmed several of Keating's assertions, notably that the delay in signing was due to the USPTO's concern about not being able to make prior art known to the public.

Initiative	Details	Current Status & Other Information
Peer to Patent[30]	• Initiative by the USPTO, run under New York Law School's Center for Patent Innovations • Online system that enables members of the public to supply the USPTO with information relevant to assessing the claims of pending patent applications. • Connects open network for community input to the legal decision-making process, with the patent examiner making the final determination on the basis of legal standards. • Peer-to-Patent involves "1) review and discussion of posted patent applications, 2) research to locate prior art references 3) uploading prior art references relevant to the claims, 4) annotating and evaluating submitted prior art, and 5) top ten references, along with commentary, forwarded to the USPTO."[31]	• First one-year pilot launched with USPTO on June 15, 2007; Extended by the USPTO for another year, ending June 15, 2009 [32]; Extension came with an increase in the maximum number of applications and expansion in the scope of applications eligible • Over the course of two years, the USPTO used Peer-to-Patent submitted prior art references to reject one or more claims in 18 patent applications.[33] • On October 25, 2010, a new 1-year pilot commenced with the USPTO; aimed to test scalability; New pilot ended Sep 30, 2011; expanded scope, increased subject matter classes, increased maximum no. of applications[34] • Now deemed a "historic initiative" by the Peer to Patent website and latest brochure [35] [36]
Third party submission of prior art[37]	• The Third-Party Preissuance Submissions is a new provision made effective on September 16, 2012 under the America Invents Act • 35 U.S.C. 122(e) provides "a mechanism for third parties to submit patents, published patent applications, or other printed publications of potential relevance to the examination of a patent application..." • A third party can search for prior art references, submit the reference with additional documents and a description of relevance; which is received by the USPTO and upon compliance, is considered by the examiner during examination	• From its launch in September 2012 till April 2015, about 10,000 documents were submitted by third parties • Out of all documents submitted, 32% documents were non-patent literature, 29% documents were patents, 22% documents were published US apps, while 17% were foreign references. • Out of the submissions made, 78% of all submissions were deemed proper[38] • A survey of examiners found that 23% examiners found the submission to be greatly helpful, 29% found it moderately helpful, 35% found it limited, while 13% found it not useful at all[39]

[30] http://www.peertopatent.org/; Accessed 7th September, 2015
[31] http://www.peertopatent.org/about-the-project/; Accessed 7th September, 2015
[32] http://www.uspto.gov/about-us/news-updates/uspto-extends-and-expands-peer-review-pilot; Accessed 7th September, 2015
[33] http://www.peertopatent.org/wp-content/uploads/sites/2/2013/11/CPI_P2P_YearTwo_lo.pdf; Accessed 7th September, 2015
[34] http://www.uspto.gov/patent/initiatives/peer-review-pilot-fy2011; Accessed 7th September, 2015
[35] http://www.peertopatent.org/; Accessed 7th September, 2015
[36] http://www.peertopatent.org/wp-content/uploads/sites/2/2013/11/P2P_Brochure.pdf; Accessed 7th September, 2015
[37] http://www.uspto.gov/patent/initiatives/third-party-preissuance-submissions; Accessed 7th September, 2015
[38] http://www.uspto.gov/patent/initiatives/third-party-preissuance-submissions; Accessed 7th September, 2015
[39] http://www.uspto.gov/patents/init_events/preissuance_examiner_survey_statistics.pdf; Accessed 7th September, 2015

IBM Technical Disclosure Bulletin[40]	• IBM Technical Disclosure Bulletin was a technical publication produced by IBM between 1958 and 1998[41]. The Technical Disclosure Bulletin functioned as a defensive publication – the aim was to disclose inventions that IBM did not want competitors to get patents on. [42] • IBM utilized the concept of "searchable prior art" - by publishing details of how to make and use the invention, IBM gave patent examiners a searchable source of prior art that they could cite against patent applications.[43] • IBM Technical Disclosure Bulletin's articles have been cited over 48,000 times in U.S. patents.[44] • IBM's Manny Schecter (Chief Patent Counsel) noted that IBM has "long cooperated with the USPTO in providing technical education for patent examiners and will continue to do so."[45]	• IBM no longer publishes the Technical Disclosure Bulletin, though it still continues to publish a variety of periodic research reports[46] • The Technical Disclosure Bulletins are available to search on IP.com (added in 2005[47]) and are kept on file with Patent & Trademark Offices and U.S. Government Depository Libraries [48] [49]. • Currently, the Prior Art Database at IP.com is the "exclusive location for IBM Technical Disclosure Bulletin (TDB) documents on the web."[50] • IBM was a sponsor of the Peer to patent project
Cisco[51]	• On February 20, 2014, Cisco committed to "assembling their own public product documentation, converting it to electronic form, and making it readily searchable by (patent) examiners." • Committed to "continue to electronically publish many invention submissions that are not internally approved to be patent filings so that these too can be used by examiners" • "Continuing to provide examiners access to senior Cisco technical talent in the form of training on current technology and the prior art in the networking field." • Cisco is a member of the United for Patent Reform[52] [53]	Mark Chandler (Senior Vice President, General Counsel and Secretary, Cisco) noted IEEE's Decision on Patent Clarifications on Feb 8, 2015, committing that "Cisco will work with the IEEE and other stakeholders to ensure that the new clarifications are implemented in a fair and equitable manner."[54] Mark Chandler also stated on June 4, 2015 about the Senate Judiciary Committee approving patent litigation reform: "…we will work with Senators on and off the Judiciary Committee to refine the legislation as it moves to the floor, especially to assure that there continues to be an effective Inter Partes Review process…" [55]
Rackspace [56]	• Pledged "a new level of transparency with the USPTO that will make finding non-patent prior art much easier for patent examiners" • On March 6, 2014, Rackspace committed to "share with the USPTO complete technical information about all of prior products, including documents that were previously considered confidential." • Also pledged to provide "ongoing information to the USPTO detailing each product that the company develops and releases." • Rackspace is a member of the United for Patent Reform[57]	Beyond support, Van Lindberg (Vice President and Associate General Counsel for Rackspace) also noted: • "At Rackspace we have pushed hard for Congress to enact patent litigation reform. We have made it one of our core missions to battle patent trolls"[58] • "We are glad to see the new focus on patent validity. We challenge the validity of every patent that is asserted against us, and we encourage all others to do the same."[59]

[40] http://ibmip.com/2014/03/14/why-ibm-supports-uspto-patent-quality-initiatives/; Accessed 7th September, 2015
[41] http://qed.econ.queensu.ca/faculty/ware/848/PDFs/Disclosure_paper.pdf; Accessed 7th September, 2015
[42] https://en.wikipedia.org/wiki/IBM_Technical_Disclosure_Bulletin#cite_note-1; Accessed 7th September, 2015
[43] Scott Baker and Claudio Mezzetti detail this strategy in their paper "Disclosure as a Strategy in the Patent Race", published in the Journal of Law and Economics, Vol. 48, No. 1 (April 2005), pp. 173-194 [http://www.jstor.org/stable/10.1086/426879]
[44] http://qed.econ.queensu.ca/faculty/ware/848/PDFs/Disclosure_paper.pdf; Accessed 7th September, 2015
[45] http://ibmip.com/2014/03/14/why-ibm-supports-uspto-patent-quality-initiatives/; Accessed 7th September, 2015
[46] http://qed.econ.queensu.ca/faculty/ware/848/PDFs/Disclosure_paper.pdf; Accessed 7th September, 2015
[47] http://www.marketwired.com/press-release/ipcom-adds-ibm-technical-disclosure-bulletin-tdb-back-file-ipcom-prior-art-database-656208.htm; Accessed 7th September, 2015
[48] http://www.intellogist.com/wiki/IBM_Technical_Disclosure_Bulletins; Accessed 7th September, 2015
[49] http://web.archive.org/web/20090326055039/http://www.ibm.com/ibm/licensing/patents/disclosures.shtml; Accessed 7th September, 2015
[50] https://ip.com/prior-art-database/; Accessed 7th September, 2015
[51] http://blogs.cisco.com/gov/statement-of-mark-chandler-on-need-for-patent-system-that-rewards-innovation-and-discourages-speculation; Accessed 7th September, 2015
[52] http://www.unitedforpatentreform.com/patent-reform; Accessed 7th September, 2015
[53] United for Patent Reform is a coalition of businesses who are working together to fight wanton abuse of the patent litigation system by patent assertion entities
[54] http://blogs.cisco.com/gov/statement-of-mark-chandler-on-the-ieee-decision-on-patent-clarifications; Accessed 7th September, 2015
[55] http://blogs.cisco.com/gov/statement-by -mark-chandler-on-senate-judiciary-committee-passage-of-patent-reform; Accessed 7th September, 2015
[56] http://www.rackspace.com/blog/transparency-the-ultimate-weapon-in-the-fight-against-patent-trolls/; Accessed 6th September, 2015
[57] http://www.unitedforpatentreform.com/patent-reform; Accessed 7th September, 2015
[58] http://www.rackspace.com/blog/a-sad-day-for-patent-reform-a-bad-day-for-innovation/; Accessed 7th September, 2015
[59] http://www.rackspace.com/blog/a-change-in-the-patent-troll-landscape/; Accessed 7th September, 2015

Verizon[60] [61]	• On Feb 19, 2014, Verizon announced its support for the White House's new patent quality initiative[62], detailing support in two dimensions – training to patent examiners, and sharing new developments • Verizon noted that "for a number of years, Verizon had been pleased to participate, at the PTO's invitation, in the agency's annual Technology Fairs. Through those Technology Fairs, Verizon engineers had traveled to PTO headquarters to give technical presentations on network architecture, media delivery and other issues." • Verizon committed to "continuing their work with the PTO to update examiners on state-of-the-art developments in communications technology, including on network architecture and media delivery" • Verizon is a member of the United for Patent Reform[63]	In a statement on April 29, 2015, Verizon commended senators involved in "crafting a balanced, bipartisan bill that preserves innovators' ability to defend their patents, while cutting down on abusive patent litigation tactics that hurt inventors big and small."[64] In a statement on Feb 5, 2015, Verizon stated its support for Patent Reform Legislation - "Verizon applauds Chairman Goodlatte's introduction of the Innovation Act, which passed the House last year with strong bipartisan support"[65]. On Dec 16, 2014, Verizon signed a patent licensing agreement with Google to "lower the risk of frivolous patent litigation", stating that "they look forward to striking similar deals with other high-tech companies also concerned with the innovation tax that patent trolls often collect".[66]

[60] http://publicpolicy.verizon.com/blog/entry/verizon-supports-the-white-house-patent-quality-initiative; Accessed 7th September, 2015
[61] http://www.uspto.gov/blog/director/entry/calling_on_the_crowd_to1; Accessed 7th September, 2015
[62] https://www.whitehouse.gov/the-press-office/2014/02/20/fact-sheet-executive-actions-answering-president-s-call-strengthen-our-p; Accessed 7th September, 2015
[63] http://www.unitedforpatentreform.com/patent-reform; Accessed 7th September, 2015
[64] http://publicpolicy.verizon.com/blog/entry/verizon-supports-patent-litigation-reform-legislation; Accessed 7th September, 2015
[65] http://publicpolicy.verizon.com/assets/docs/VZ_NR_--_Verizon_Statement_on_Innovation_Act_Intro_2.5.15.pdf; Accessed 7th September, 2015
[66] http://publicpolicy.verizon.com/blog/entry/verizon-signs-patent-licensing-deal-with-google-to-promote-innovation; Accessed 7th September, 2015