Knowledge Flows and Delays in the Pharmaceutical Innovation System

Mari Jibu[1,2], Yoshiyuki Osabe[3], and Katy Börner[4]
[1]OECD, Paris, France | [2]Japan Science and Technology Agency, Tokyo, Japan
[3]Japan Patent Office, Ministry of Economy, Trade and Industry, Tokyo, Japan
[4]katy@indiana.edu, CNS, ILS, SOIC, Indiana University, Bloomington, IN, USA

OECD BETTER POLICIES FOR BETTER LIVES

CNS

Abstract

This paper presents an analysis of knowledge flows in the pharmaceutical innovation process. Backward citations, citations to non-patent literature (NPL), and forward citations that link patents, scientific publications, and pharmaceutical pipelines data on drug developments are analysed and visualized to provide a more holistic understanding. Results show that patents linked to drugs tend to be technically specialized when compared to patents without linkages to drugs. Moreover, patents linked to drugs tend to cite older patents and scientific publications and impact wider technological and scientific fields than pharmaceutical patents not linked to drugs.

Introduction

Diverse studies have been conducted to study the origin, trajectory, and destination of knowledge flows and the delays in the science and technology system. Patents and citations between patents and to non-patent literature (NPL) are analysed to understand knowledge spillovers (Lukach & Plasmans, 2002) or to measure patent quality (Squicciarni et al., 2013). The OECD Science, Technology and Industry Scoreboard 2013 (OECD, 2013) uses comprehensive and up-to-date data to report on knowledge flows via collaboration networks (e.g., derived from co-authored publications and co-inventors on patents), international migration of researchers (e.g., estimated from changes in author's addresses on publications), but also flows of royalty and licence fees for technologies. Recently, the OECD introduced a new indicator, called "Patent-Science Link," that aims to measure knowledge flows between the science base and the innovation system (OECD, 2013). According to this new indicator, patented pharmaceutical inventions account for the majority of citations made from patents to scientific publications. That is, the distance between the science base and the innovation system is much closer in pharmaceutical fields than it is in other technological fields. Pharmaceutical innovation is particularly important for drug discovery, as research and development (R&D) costs are huge and major challenges exist for arriving at cost-effective new drugs. In fact, there is a steady decrease in R&D productivity over the last number of years (Booth & Zemmel, 2004).

This paper analyses and visualized linkages between publication, patent, and drug pipeline data to increase our understanding of knowledge flows and delays in the pharmaceutical innovation system. The structure of the paper is as follows: Section 2 details data acquisition and preparation. Section 3 discusses different metrics. Section 4 presents results. The paper concludes with a discussion of key insights and their comparison to prior work.

Data Acquisition & Preparation

Five datasets by Thomson Reuters covering 1981 to 2011 are used in this analysis. (1) Publication data from the *Web of Science (WoS)* database. (2) Patent data from the *Derwent World Patents Index (DWPI)* and associated citations from the (3) *Derwent Patents Citation Index (DPCI)*. (4) Linkages between publications and patents come from the *WoS-DPCI* Linktable computed by Thomson Reuters and JST that provides information on backward citations from patents and to the non-patent literature (NPL), i.e., scholarly publications, derived from the *DPCI*. (5) Last but not least, drug pipeline data was retrieved from the *Cortellis for Competitive Intelligence* database including detailed information of exactly drugs a patent is associated with. Data was compiled on December 11, 2013.

Interested to identify patents and their linkages to the NPL in pharmaceutical fields, we extracted all 833,376 patents with the International Patent Classification (IPC) code "A61P: Specific therapeutic activity of chemical compounds or medicinal preparations" from the *DWPI* together with their citations from *DPCI*. This dataset is subsequently called "Pharma_Patents," see also Figure 1.

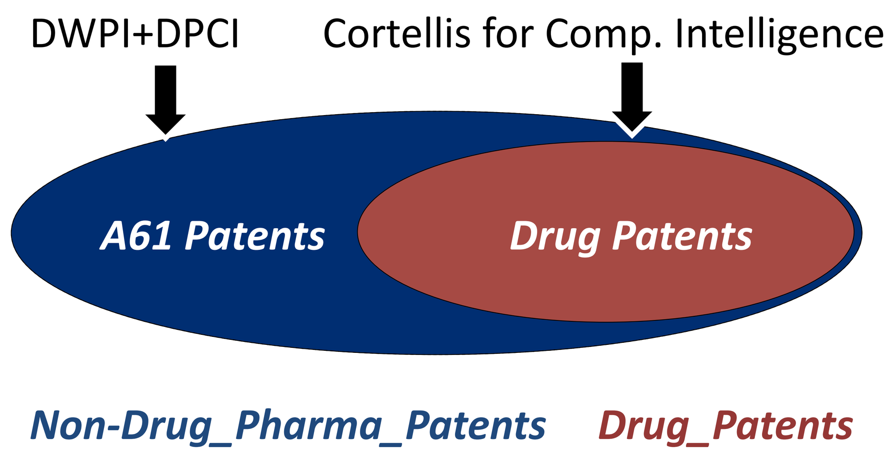

Figure 1: Data Acquisition Process

Then, we extracted 57,800 patents linked to pipeline data from the *Cortellis for Competitive Intelligence* database. Thomson Reuters Drug reports provide details of patent protection, subsequently called "Drug_Patents."

Next, the *Drug-Patents* were subtracted from the *A61P-Patents* resulting in a dataset of 325,576 "Non-Drug Pharma Patents" that have the A61P code but are not linked to drugs.

Finally, all 115, 252 NPL for *Drug_Patents* and 718,269 *Non-Drug_Pharma_Patents* were retrieved using the *WoS-DPCI Linktable*.

Methodology

In order to understand the citation lag, generality, subject, and scope for the two patent-NPL datasets, four indexes were computed. All four are explained and defined here.

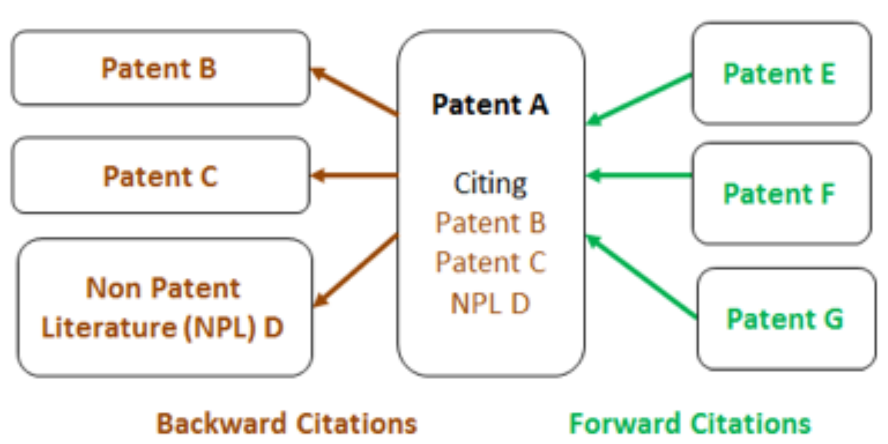

Figure 2. Patent backward citations to patents and NPL (green arrows) and forward citations to patents (red arrows).

The citation lag can be computed for backward citations to identify the promptness by which patents cite existing works or for forward citations to identify the speed by which patents are cited by future patens, see Figure 2. It is defined as the mean of the difference of the publication year of the focus patent (patent A in Fig. 2) minus the publication years of all cited works (e.g., patents B, C and NLP D in Fig. 2). Analogously, the citation lag of all forward citations is defined as the mean of the difference of the publication years from all citing works (e.g., patents E-G in Figure 2) to the publication year of the focus patent.

The generality index G_x is a quantitative measure that reflects the diversity of patents that are cited by a given focal patent as well as the diversity of patents that are citing the focal patent. The number of different IPCs associated with cited and citing patents is used as a proxy for technology diversity. The index is high if cited and citing patents cover a wide range of technology fields. Let x be the focal patent with y_i patents citing the focal patent x, with $i=1, \ldots, N$ then G_x for all 4-digit IPC subclasses can be calculated as follows:

$$G_X = 1 - \sum_{j=1}^{M_i} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{T_{ji}^n}{T_i^n} \right)^2$$

where

T_i^n is the total number of IPC n-digit classes in y_i,

T_{ji}^n is the total number of IPC n-digit classes in the j^{th} IPC 4-digit class in y_i, and

$j=1 \ldots M_i$ is the cardinal of all IPC4-digit classes in y_i

The index was calculated for all 4-digit and analogously for all 6-digit IPC subclasses for all patents in *Drug_Patents* and *Non-Drug_Pharma_Patents*.

The subject index S_x is a new indicator which is based on the generality index but computed for NPL. Let x be the focal patent which cites y_i, $i=1 \ldots, N$ scientific publications (NLP) then S_x can be computed as follows:

$$S_X = 1 - \sum_{j=1}^{J} \left(\frac{N_{ij}}{N_i} \right)^2$$

where

N_i is the total number of subject code in y_i and

N_{ij} is the total number of subject code in the j^{th} subject code in y_i.

The subject codes in each scientific publication are counted, using basic element 1. The subject index was calculated for all scientific publications (NLP) cited by patents in *Drug_Patents* and *Non-Drug_Pharma_Patents*.

The patent scope $SCOPE_p$ is often associated with the technological and economic value of patents with broad scope patents having a higher value (Lerner, 1994). For each patent P, its scope is defined as follows:

$$SCOPE_p = n_p ;\ n \in \{ IPC_1^4; \ldots IPC_i^4, \ldots IPC_j^4; \ldots; IPC_n^4 \} \text{ and } IPC_i^4 \neq IPC_j^4$$

where n_p denotes the number of distinct 4-digit IPC subclasses listed in the patent P and it is normalized by dividing each individual value by the maximum value found in the dataset. The patent scope was calculated for all distinct 4-digit and 6-digit IPC subclasses for all patents in *Drug_Patents* and *Non-Drug_Pharma_Patents*.

Results

Using the metrics introduced in the previous section, a number of novel results can be computed.

Technology Delays: Citation Lag

Comparing citation lag data for *Drug_Patents* and *Non-Drug_Pharma_Patents* reveals the temporal dynamics of knowledge flows. As can be seen in Table 1, forward citations from *Non-Drug_Pharma_Patents* come from patents that were published on average 2.17 years later while *Drug_Patents* are cited faster—after 1.89 years on average.

Backward citations from *Non-Drug_Pharma_Patents* go to patents that were published on average 3.4 years earlier and they go to much more recent NPL—published only 1.69 years earlier on average.

Interestingly, *Drug_Patents* cite older works than *Non-Drug_Pharma_Patents*: Cited patents are 5.64 years old and cited NPL are 2.5 years old on average.

All values plotted in Table 1 are statistically significant at the 1% level. In sum, they show that *Drug_Patents* cover larger temporal ranges and are cited more quickly than *Non-Drug_Pharma_Patents*.

Table 1. Forward and Backward Citation Lags for *Non-Drug_Pharma_Patents* and *Drug_Patents*

		Non-Drug_ Pharma_Patents	Drug_ Patents
Citation Lag in yrs	Forward Citations by Patents	2.17	1.89
	Backward Citations to Patents	3.40	5.64
	Backward Citations to NPL	1.69	2.50

Technology Diversity: Generality & Subject Index

The generality index was calculated for 4- and 6-digit IPCs for forward and backward citations for *Non-Drug_Pharma_Patents* and *Drug_Patents*, see Table 2. As can be seen, *Drug_Patents* have higher generality index than *Non-Drug_Pharma_Patents*. As for the subject index, *Drug_Patents* also have a higher value than *Non-Drug_Pharma_Patents*. That is, on average, *Drug_Patents* draw on more diverse technology "base knowledge" and are cited by a more diverse set of patents that have more varied IPCs. All values plotted in Table 2 are statistically significant at the 1% level.

Table 2. Generality Index for Forward and Backward Citations for *Non-Drug_Pharma_Patents* and *Drug_Patents*

		Non-Drug_ Pharma_Patents	Drug_ Patents
Generality Index (4-Digits)	Forward Citations	0.36	0.37
	Backward Citations	0.40	0.54
Generality Index (6-Digits)	Forward Citations	0.46	0.50
	Backward Citations	0.52	0.73
Subject Index	Backward Citations to NPL	0.22	0.28

Technology Value: Scope

The patent scope was computed for *Non-Drug_Pharma_Patents* and *Drug_Patents*, see Table 3. Interestingly, in the pharmaceutical fields, the scope of *Drug_Patents* tends to be lower than that of *Non-Drug_Pharma_Patents*. This is unexpected as patents linked to drugs are presumably more valuable than those not linked to drugs.

Table 3. Scope for *Non-Drug_Pharma_Patents* and *Drug_Patents*

		Non-Drug_ Pharma_Patents	Drug_ Patents
Scope	4-Digits	0.13	0.11
	6 Digits	0.16	0.15

Conclusions

In our prior work, we introduced new drug-patent indicators for identifying patents related with pharmaceutical entities' R&D progress ("Pre-clinical" → "Phase 1" → "Phase 2" → "Phase 3" → "Filed" → "Approved" → "Marketed") (Jibu & Osabe, 2014). In that work, we also showed that IPC count, forward citations, and citations to NPL are efficient drug-patent-indicators. However, this work is novel in that it examines backward citations and their value for computing the quality of drug-patent linkages. Results indicate that citation lags and the generality of backward citations are statically significantly different for *Non-Drug_Pharma_Patents* and *Drug_Patents*.

Addendum

Note that the opinions expressed in this paper are those of the authors and do not represent those of the institutions that the authors belong to.

Acknowledgments

We would like to thank Fernando Galindo-Rueda for his expert comments and support of this research. This work was partially funded by the National Institutes of Health under awards NIA P01AG039347 and U01 GM098959.

References

Booth, Bruce, and Rodney Zemmel. 2004. "Prospects for Productivity." *Nature Reviews Drug Discovery* 3, 451-456.

Jibu, Mari and Yoshiyuki Osabe. 2014. "Refined R&D Indicators for Pharmaceutical Industry." *Future Information Technology, Lecture Notes in Electrical Engineering*, Vol. 309, pp 549-554.

Lerner, Joshua. 1994. "The Importance of Patent Scope: An Empirical Analysis." *The RAND Journal of Economics*. 25 (2), 319-333.

Lukach, Ruslan and Joseph Plasmans. 2002. "Measuring knowledge spillovers using patent citations: evidence from the Belgian firm's data" CESifo Working Paper NO.754 Category 9: Industrial organization,

OECD. 2013. *OECD Science, Technology and Industry Scoreboard 2013: Innovation for Growth*. Paris, France: OECD Publishing.

Squicciarni, Mariagrazia, Helene Dernis and Chiara Criscuolo. 2013. "Measuring Patent Quality: Indicators of Technological and Economic Value" *OECD/DSTI/DOC* 3.