R&D dynamics in the development of HIV/AIDS drugs

Jos J. Winnink[1] and Robert J.W. Tijssen[2]

[1]*winninkjj@cwts.leidenuniv.nl;* [2]*tijssen@cwts.leidenuniv.nl*

Leiden University, Centre for Science and Technology Studies (CWTS),
Wassenaarseweg 62A, 2333AL Leiden (the Netherlands)

Abstract

It is well known that the development of scientific knowledge is a complex process in which various actors operate and interact. Serendipity often plays an important role. Scientific knowledge and technological achievements are reflected in scholarly publications and patent applications. We analyse bibliographical information regarding scholarly publications and patent publications to get insight in R&D dynamics. In this study we are focusing on the R&D resulting in HIV/AIDS medicine. The development of HIV/AIDS medicines is chosen as a case study because it spans a relatively short period of time; because it shows new paradigms in medicine development and because most of the actors are still available to validate the results of our research. Experts in this research field provide information about crucial stages in the R&D process, including relevant developments that occurred even before the discovery of HIV. We show that bibliographic information can be used to analyze R&D dynamics and can be of help in identifying 'breakthrough' stages in the R&D process.

Background

The objective of our main research project is to develop a robust statistical method for data mining of bibliographical information regarding scholarly publications and patent publications. The ultimate goals are to help identifying major 'breakthrough' stages in the R&D process; to pin-point the moment the 'breakthrough' occurred and to identify general or idiosyncratic determinants and factors that played a crucial role in success or failure. There are many conceptual and methodological challenges involved; where different approaches and tools have tried in the past to tackle this detection problem with varying degrees of success (see e.g. Arbesman, 2010; Bettencourt, Kaiser, Kaur, Castillo-Chávez, Wojick, 2008; Breiner, Cuhls, Grupp, 1994; Julius, Berkoff, Strack, Krasovec, Bender, 1977; Martin, 1995).

The main research question to be answered "Which factors are deductible from bibliographical data regarding scholarly publications and patent publications that can be used to detect a 'breakthrough' as soon as possible after the event?" The following case study, on HIV/AIDS drugs research, illustrates some of the methodological work we have done in the process of developing such a detection tool.

Case study

It is known that development of new medicines is both time-consuming and costly. Several sources indicate that the time it takes from disease characterization to an approved drug is several (10–14) years (see e.g. IOM (Institute of Medicine), 2009). The costs involved in developing a new medicine from start to approval are estimated in FDA (2004) to be in the range from $ 0.8 to 1.7 billion. Furthermore it is known from various studies (e.g. Jewkes, Sawers, Stillerman, 1969; Isenson, 1969) that it can take several years and in some cases even dozens of years before scientific discoveries find their way into applications. It is therefore to be expected that medicine based on detailed knowledge of the disease causing pathogen and the molecular biological mechanisms involved start to appear several years after the identification of the pathogen. The first medicine to combat a specific disease will be based on knowledge already existing at the moment the disease or its cause is identified. Knowledge is inherited from other more or less adjacent areas and already existing medicine are tested for potential positive effects in the treatment of the disease.

In particular newly recognized diseases in the western world with high mortality rates or diseases having a great impact on society or on economy are candidates to develop medicine for. Patent publications in the area of medicine development can be seen as a sign of the application of scientific knowledge.
Since 1980 AIDS got its epidemic character. This disease with a high mortality rate started to spread in an epidemic way at first particularly under male homosexuals in the western world. Two articles (Temin, Mizutani, 1970; Baltimore,[1970]) published together in Nature are considered playing a pivotal role in the development of HIV/AIDS medicines. Their pivotal role lies in the fact that these publications contained proof for the mechanism used by retroviruses. A revision of the central dogma of molecular biology at that time was therefore necessary.
It is not until 1983 that a research group led by Montagnier discovered the retrovirus HIV as the pathogen causing AIDS (Barreé-Sinoussi et.al., 1983). HIV being a retrovirus started the testing of drugs known to act against retroviruses for their activity in curing (at least temporarily) of AIDS. De Clercq (2009) shows that in the early days of AIDS existing knowledge in the form of already known and available pharmaceuticals is used to try to cure the disease.
The development of HIV/AIDS-medicine has shown a *paradigm shift*. This paradigm shift is the "system approach". In this approach combinations of various complementary drugs are used. Currently HIV/AIDS medicines are subdivided into six distinct classes. The classification of HIV/AIDS drugs is based on the molecular biological working mechanisms the medicine acts on. In March 2010 some 28 drugs were approved in the US for human use.
We chose the area of HIV/AIDS medicine to test the hypothesis "Bibliographical data of scholarly publications and patent publications are mirroring the development of science and technology and therefore analysis of this data can help in understanding R&D." HIV/AIDS was chosen for the facts that the area is survey able; the area has shown a lot of dynamic; in a relatively short period of time (since 1980) the area has shown a lot of R&D activity resulting in serious advancements in the treatment of HIV/AIDS; the "system approach" –using combinations of various complementary drugs. For diseases other than AIDS this approach is still less common.

Method

We constructed two datasets with data from 1980 onwards. The first dataset contains bibliographical information related to 34,603 scholarly publications extracted from the version of the Thomson Reuters/CWTS *Web of Science database* (WoS)[25]. The second dataset contains bibliographical information related to 16,379 distinct patent families[26]. We extracted these data from the EPODOC[27]-database and used 2008 as a cut-off year because due to procedural regulations more recent patent data is not yet fully available. The data in both datasets was confined to scholarly publications or patent publications related to the development of HIV/AIDS medicine. In order to get a consistent focus between both data sets on the development of medicines the set of scholarly publications consists of articles and letters published in journals in the areas Immunology; Virology and Infectious diseases; the

[25] Relevant scholarly publications were select by first using the PubMed Clinical Query for AIDS. The results were transferred from PubMed into the CWTS version of the WoS. Based on the research fields we restricted the document set to publications from the fields that are most closely related to the development of medicine for HIV/AIDS. Finally we selected only publications with document types `article' and `letter'.
[26] Patent family = collection of closely related patent publications that may be considered describing the same invention.
[27] EPODOC is a database internally in use at the European Patent Office. The database contains bibliographic data for patent documents of most countries or patent authorities. This database is available to the authors because of the fact that the first author has a second affiliation with the Netherlands Patent Office and therefore has access to this database.

set of patent publications is constructed by using selective classification codes[28] for HIV/AIDS and supplemented with documents containing the terms HIV or AIDS. All patent publications had to be classified at least as medical preparation (: a medicine).
The patent publications were grouped into families of equivalent documents to prevent overestimation and to be able to identify inventions uniquely. Each patent family is counted once.
An anomaly exists in the patent data set and can therefore be seen in Figure 6 for the years 1999 and 2000. The peak is caused by patent applications related to the sole identification of DNA-sequences without specifying the functionality of the DNA-sequence. It is argued in Lawrence (2004) that due to increasing pressure for clearly defined patent applications; the completion of the Human Genome Project; and the 2000-biotech market bubble, this kind of patent applications disappeared after 2000.

Results

For both document collections we first show the number of documents per year, using the year of publication for the scholarly articles and the year the very first application of a patent family is filed (: year of oldest priority) for the patent publications. Figures 1 and 2 show the results. The first scholarly publications related to the development of HIV/AIDS medicine appear in 1981.

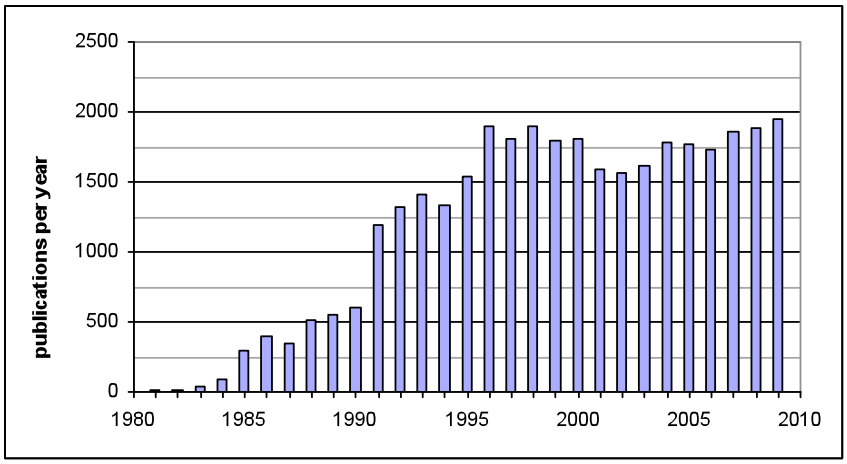

Figure 5 Scholarly publications related to the development of HIV/AIDS medicine

The patent related data shows a peak around 2000. This peak can be seen as an anomaly caused by the fact that up to 2000 regulations for patenting DNA-sequences were less restrictive. Before this changing in regulations the search for DNA-sequences and identification of DNA-sequences, e.g. in China, done at large scale led to many patent applications. We identified for three major patent applicants from the Shanghai area their patent applications and show them as a separate group in Figure 6.

[28] The following IPC/ECLA/ICO codes used (A61P31/1, C12N15/49, C07K14/15, C07K14/1,.C07K14/155, C07K14/1, C07K14/16B, C07K14/16D, C07K14/16, C07K14/81B, C07K16/10K, C07K16/10K1B, C07K16/10K1, C07K16/10K1F, C07K16/42K10A, C12N15/113A1, M12N740/03F, A61K31/00, A61K33/00, A61K35/00, A61K36/00, A61K38/0, A61K39/00, A61K41/00, A61K45/00, A61K47/00, A61K48/00) were selected. (IPC = Internal Patent Classification, ECLA = European Patent Classification, ICO = further refinement of ECLA).

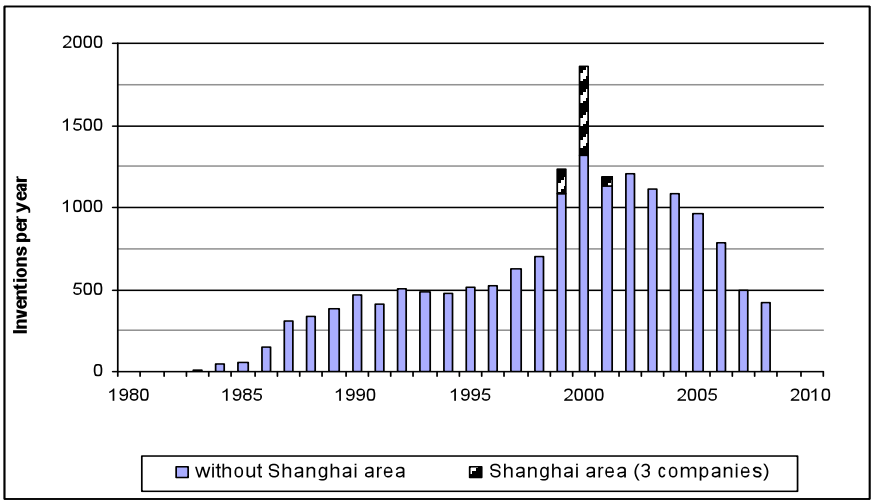

Figure 6 Number of inventions related to HIV/AIDS medicine

After changing the regulations the number of patent applications of this type dropped significantly. Also the total number of biotechnology patent applications settled down at a significant lower level (-30%) (Lawrence, 2004). These effects in combination with the incompleteness of the date from 2007 onwards result in the decline in the number of inventions after 2004.

We are interested in the geographical areas where research is done and the possible changes therein. For four geographical areas the United States, EU-15[29], China and Japan we calculated the relative involvement in scholarly publications and patent applications on a yearly basis for the period 1980 - 2009. For the patent publications we omitted the data for 1980 and 1981 because there is only 1 invention in 1981 coming from the United States. Inclusion of this data would result in a very sharp peak for the year 1981. In both cases for every publication a country of origin is only counted once. Figure 3 shows the relative involvement in the scholarly publications for authors from the four geographical areas mentioned. In figure 4 for the same geographical areas the relative involvement of the inventors is displayed.

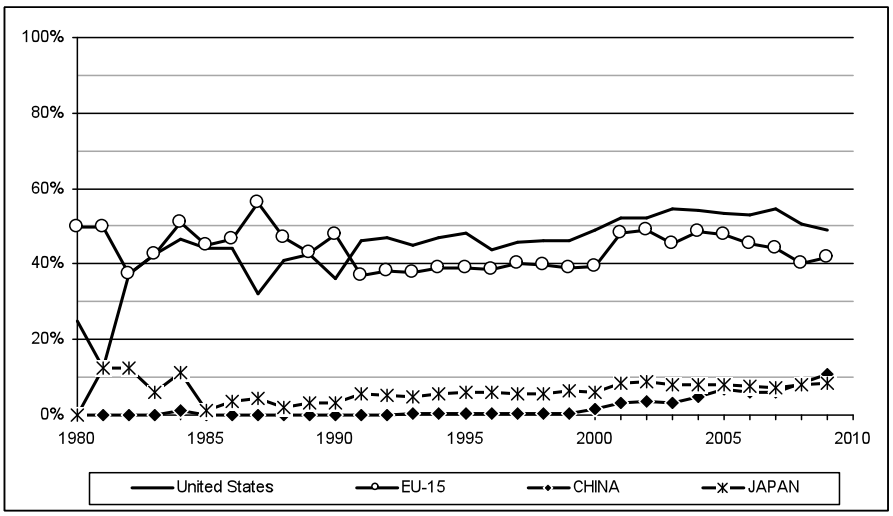

Figure 7 Relative contribution to HIV/AIDS medicine papers by country of residence of the authors

[29] EU-15 denotes the group of the 15 EU member states Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, and United Kingdom.

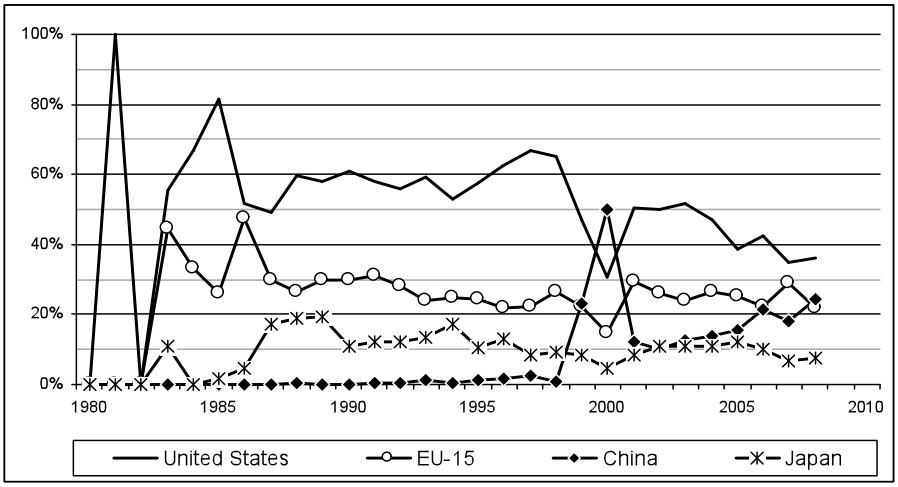

Figure 8 Relative contribution to HIV/AIDS medicine inventions by country of residence of the inventors

Discussion

Use of existing knowledge

Once the pathogen HIV was identified knowledge already around for some time was used as a basis to build medicine for this *new* disease on. It is not uncommon for existing medicine that new unforeseen utilizations are found and this is even the case for drugs already around for a considerable period of time. Figures 1 and 2 are showing the fact that scholarly publications started to appear in 1981 and patent applications for HIV/AIDS medicine were already filed from 1983 (the year of discovery of HIV) onwards. The graphs indicate the fact that at least the first medicine to be developed for HIV/AIDS were based on knowledge that already existed at the moment the AIDS causing pathogen (HIV) was identified.

The fact that patent applications for medicine against HIV/AIDS start to appear so soon after the identification of HIV as AIDS causing pathogen indicates that already existing knowledge is used in developing the first medicine and that existing medicine are tested for their activity against AIDS as might be expected from the theory.

Shift in geographical origin of authors and inventors

AIDS being first recognized as a *male homosexual* disease with an epidemic character and a mortality rate near 100% caused research activities for medicine to start in the western world. Although AIDS makes most of his victims in Africa still western companies and research institutes dominate the development of new types of medicine.

The fact that research started in the western world (United States + EU-15 in the graphs) is consistent with the ideas that AIDS was thought being a disease mainly affecting male homosexual persons in the western world. Patent data shows that research in China is becoming more and more important and may have surpassed the EU-15 in 2008. The fact that research outside the four geographical areas mentioned is quite low is a reflection of the fact that infrastructure and availability of resources are important for doing fundamental research on the development of medicine. Based on (IOM, 2009) and corresponding patent information we did not yet discover any approved HIV/AIDS drugs based on R&D executed outside the western world. Major pharmaceutical companies still have most of their research centres based in western countries; the shift towards Asia and China has become a noticeable development in global science (e.g. Tijssen, 2009).

Searching for breakthrough moments in AIDS/HIV R&D

Based on the information provided by subject experts in this field we attempted to better understand and validate the bibliographical information that reflects the developments that took place. We first looked at numbers of publications and numbers of patent filings to link this information with major steps in de development of HIV/AIDS medicine. We also looked at possible geographical shifts for authors and inventors to see if this is one of the signals we can use.

We also tried to answer the question "Do patent publications with (almost) exclusively non-patent literature references reflect the development of 'science-dependent' technologies?" More research has to be done on these patents to be conclusive. Further research during the next few months will also be done on the analysis of citation patterns, authors and inventors, co-authorships and co-inventorships. Some of the results will be presented at the conference.

References

Arbesman, S. (2010). Quantifying the ease of scientific discovery. Scientometrics, 86(2):245–250.

Baltimore, D. (1970). Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature, 226(5252):1209–1211.

Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L., (1983) "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)", Science, 220: 868–871.

Bettencourt, L., Kaiser, D., Kaur, J., Castillo-Chávez, C., and Wojick,D. (2008). Population modeling of the emergence and development of scientific fields. Scientometrics, 75(3):495–518.

Breiner, S., Cuhls, K., and Grupp, H. (1994). Technology foresight using a delphi approach - a Japanese-German cooperation. R & D management, 24(2):141–153.

De Clercq, E. (2009). The history of antiretrovirals: key discoveries over the past 25 years. Reviews in Medical Virology, 19(5):287–299.

FDA, (2004) "Innovation Stagnation: Challenge and Opportunity on the Critical Path to New Medical Technologies". March 2004. (White Paper).

IOM (Institute of Medicine) (2009) Conflict of Medical Research, Education, and Practice. Washington, DC: The National Academic Press.

Isenson, R. S. (ed.) (1969). Project hindsight (final report). Technical report, US Dept. of Defense.

Jewkes, J., Sawers, D., and Stillerman, R. (1969). The Source of Invention, 2nd edition. MacMillan.

Julius, M., Berkoff, E., C., Strack, A. E., Krasovec, F., and Bender, A. D. (1977). A very early warning system for the rapid identification and transfer of new technology. American Society for Information Science, 28(3):170–174.

Lawrence, S. (2004). Patent drop reveals pressures on industry. Nature Biotechnology, 22(8):930–931.

Martin, B. R. (1995). Foresight in science and technology. Technology Analysis & Strategic Management, 7(2):139–168.

Temin, H. M. and Mizutani, S. (1970). Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of Rous sarcoma virus. Nature, 226(5252):1211–1213.

Tijssen, R.J.W. (2009). How globalized is corporate pharmaceutical research? An analysis of Europe's multinational companies, Technology Analysis and Strategic Management, 21 (7): 859-879.