Technological Forecasting & Social Change 167 (2021) 120698

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

How pharmaceutical innovation evolves: The path from science to technological development to marketable drugs

Xuefeng Wang [a,*], Shuo Zhang [a], Yuqin Liu [b], Jian Du [c], Heng Huang [a]

[a] School of Management and Economics, Beijing Institute of Technology, Beijing, China
[b] School of Journalism and Publication, Beijing Institute of Graphic Communication, Beijing, China
[c] National Institute of Health Data Science, Peking University, Beijing, China

ARTICLE INFO

Keywords:
Biomedicine
Biological medicine innovation
Science-technology linkages
Interaction process
Topic analysis

ABSTRACT

Biomedical innovation is the process of transforming scientific discoveries into vaccines, biodiagnostic reagents, and genetically-engineered drugs and therapies that save or improve patients' lives. This type of process is typical of translational research, yet a great many efforts in the field of biomedical research fail to deliver the desired outcomes, and some even result in an enormous waste of time and resources. Long R&D periods and inefficient methods of transforming knowledge from basic scientific findings into practical clinical tools are the main reasons for failure. Understanding how scientific research co-evolves with technological development could provide novel and profound insights along the path of biomedical innovation. However, there are not many researches to deal with this aspect in recent years. Therefore, this paper presents a framework that traces the history of USFDA approved drugs in granular detail. Using scientific papers and patents as data sources, we use qualitative and quantitative techniques to analyze the innovation process from the inception of discovery into a marketable pharmaceutical. The focus of our analysis is the information found in science and technology documents, which can be an indicator of the interplays between discovery and development in a translational research process. Entropy statistics then provide an indication of the shared information for maximum utility in the analysis. The analysis results, which include expert judgments, could drive possible future insights into biomedical innovation with implications for policymakers.

1. Introduction

As a high-tech and highly interdisciplinary industrial sector, biomedicine can be an important part of a nation's economy (Lee et al., 2009; Burnette, 2015). Global expenditure into biomedical research and development has been rising in recent years. However, the pace of scientific discovery in biomedical research appears to have remained relatively constant in terms of both life expectancy and the number of "new molecular entities" approved by the United States Food and Drug Administration (USFDA). In other words, translating this knowledge into concrete improvements in clinical medicine has been sluggish and lagged behind the pace of discovery (Duda et al., 2014; Bowen and Casadevall, 2015; Freedman et al., 2015).

Morlacchi and Nelson (2011) argue that many policy circles still have an inadequate and overly simplistic understanding of how medical practice advances. Coupled with the growing disparity between inputs and outcomes in biomedicine, it seems clear that policymakers need finer insights into how new medical practices emerge and evolve both generally and in specific areas of medicine. One way to provide a complete but detailed picture of biomedicine would be to analyze research in many narrow areas of medicine and then combine those findings into a bigger picture (Metcalfe et al., 2005; Mina et al., 2007; Morlacchi and Nelson, 2011; Nelson et al., 2011). Although arduous, the effort may be well worth it, as the knowledge gleaned through such studies would allow us to understand how therapeutic innovations more specifically contribute to improvements in clinical practice and may illuminate the pathways to dramatically better modes of clinical practice.

Some researchers have postulated that a linear progression from medical science to medical practice is not an accurate description of the complex interactions between these two poles (Schechter et al., 2003; Zerhouni, 2003; Rees, 2004). Further, the idea that more science yields

This article belong to the special section on Advanced Tech Mining: Technical Emergence Indicators and Measurements.

* Corresponding author.
E-mail address: wxf5122@bit.edu.cn (X. Wang).

https://doi.org/10.1016/j.techfore.2021.120698
Received 2 February 2020; Received in revised form 19 February 2021; Accepted 21 February 2021
Available online 27 February 2021

better technology has been challenged by claims that the logic of basic science is fundamentally ill-equipped to solve many complex technological problems and, indeed, may even conflict with technological learning (Gittelman, 2016). For instance, Sarewitz and Nelson (2008) argue that the cause of a problem could be the technological 'fix' itself – that is, a cheap and quick technological solution in biomedicine and health care will usually create more problems than its solves. They go on to explain that research policy shaped from this view will inevitably lead to dramatic and unfortunate consequences.

For the purposes of this study, a marketable drug means a new molecular entity (NME) or a new biologic approved by the USFDA (Munos, 2009). In pharmaceuticals, the path from basic science to a marketable drug follows a very standard route. It starts with the discovery of leading compounds. Further development typically involves applying a range of techniques and technologies related to chemical manufacturing and control, pharmacokinetics, toxicology, pharmaceutical formulation, and so on. Last but not least comes testing, trials, and market approval. Therefore, we argue that biomedical innovation (i.e., pharmaceutical innovation) evolves as a result of progress along two different pathways. One is advances in scientific understanding, but the other is advances in new techniques and technologies. The latter is at least as important as the former in many cases, and sometimes more so. In this regard, we have chosen to examine pharmaceutical inventions approved by the USFDA. As Munos (2009) explains, these drugs represent genuine advances in biomedicine and should count as innovation. Through a specifically-designed framework, we trace the history of different drugs from research to development to commercialization. The results should provide policymakers and practitioners with profound insights along the path of biomedical innovation. The entire transformation process is shown in Fig. 1 below.

Most research that attempts to track and evaluate the relationships between science and technology relies on citation analysis at a document level, at least to some extent. This pioneering approach was first proposed by Carpenter et al. (1980), who used science-related references from non-patent literature as a tool to represent direct relationships between technology and scientific knowledge. In the four decades since, the links between science and technology have remained a focal point for many researchers in the fields of bibliometrics and research policy. In Glanzel and Meyer's (2003) study, scientific articles that cited patent documents revealed to the extent to which technological inventions had penetrated basic research. Sung et al. (2015) showed the reverse, using the citations from patents to science- and non-science-based references to show agency in the links between technology and science. They also identified field- and firm-specific differences in the linkages between science and technology. Subsequently, the CHI Research Institute defined a reliability indicator to measure the degree of linkages between science and technology, called "science linkage", according to the number of scientific papers cited by patents (Patrick and Steven, 2015). Du et al. (2019) also proposed an indicator, "technology linkage", as a measure of the number of key patents relating to intellectual property rights over a product. Using this indicator, they examined biomedicine as a case and found that key technologies in products have increasingly originated from scientific research.

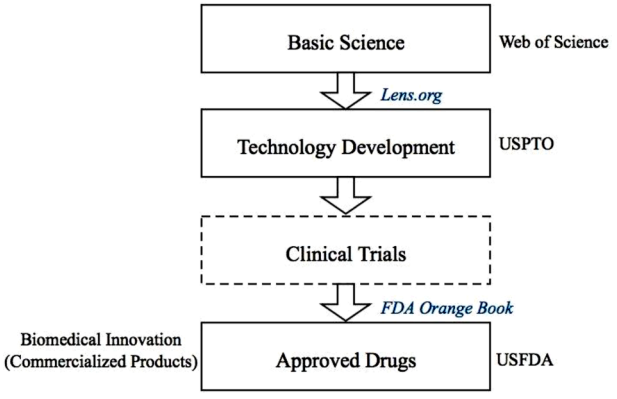

Fig. 1. The entire transformation process of drug research and development. (Source: Adapted and modified from Du et al., 2019)

Building on the literature, several studies have suggested that biomedicine relies more on public science than other industries because biomedical patents tend to include relatively more citations to basic research publications than the other sectors (Jibu, 2014; McMillan et al., 2000; Sung et al., 2015). However, while measuring citations from patents to scholarly works does provide insight into the influence of published research on invention, industry, and enterprise at the individual and institutional level, it provides a far from complete picture of translational science. Citations from an article to a patent (and vice versa) can be a measure of the degree of interaction between science and technology, but citations alone do not provide a deep analysis of content, nor do they expose the crux of how research becomes development or why some development efforts succeed while others fail.

This paper is centrally oriented to the question "How does pharmaceutical innovation evolve?" To answer the question, we propose a framework that traces the history of USFDA approved drugs in granular detail. Using scientific papers and patents as data sources, we use qualitative and quantitative techniques to analyze the innovation process from the inception of discovery into a marketable pharmaceutical. Within our analysis framework, research topics represent the elements of knowledge and topic evolution analysis show us how these topics evolve through the practice of science and technological development at a micro level. To ensure the analysis has maximum utility, entropy statistics provide an indication of the information that is shared between research activities and development activities. Experts then verify the findings and add their own finesse. The final outcome is a useful and valuable analysis, especially for policymakers, that drives insights into possible future developments within biomedicine.

The rest of this paper is organized as follows. Section 2 briefly presents the related work. Section 3 provides the proposed methodology and data sources. Section 4 contains the topic evolution analysis of biomedicine and the mutual information calculations that measure the interplay between science and technology. Sections 5 and 6 conclude the paper with a discussion on our findings and policy implications, plus the limitations of this study and directions for future work.

2. Related work

2.1. The links between science and technology in biomedicine

Throughout the entire process of researching and developing drugs, two activities are extremely critical: drug discovery and pre-clinical toxicology studies (Palucki et al., 2010; AI-Humadi, 2017). Drug discovery is part of scientific research (Bowen and Casadevall, 2015; Casadevall and Fang, 2014), and pre-clinical toxicology studies (e.g., chemical manufacturing and control, pharmacokinetic, toxicology, and pharmaceutical formulation) are part of technology development. Evidence of this type of effort is usually found in patents (Tseng, 2009; Choi and Hwang, 2014).

Cutler and McClellan (2001) asserted that only a micro-level focus on the process of innovation has the potential to unpack the relationships that matter in the localized advancement of medical science and technology. Following this line, Mina et al. (2007) attempted to uncover the structures of medical understanding in coronary artery disease by searching for path-dependent, co-evolving scientific and technical knowledge. Morlacchi and Nelson (2011) provided additional evidence that a single-minded focus on research into a disease as a window into new therapies is incomplete at best. Their perspective restates the importance of the other two ways to advance new therapies: the development of technologies used in diagnosis or treatment and learning

in practice. Petersen et al. (2016) developed a triple helix model that can be used to trace the interplay among three key dimensions of the biomedical innovation process: disease, drugs and chemicals, and technological capabilities. In other words, analytics, diagnostics, and therapeutic techniques and equipment. Using the human papillomavirus as a case study, their analysis shows that disease and technological capabilities have the strongest link, followed by drugs and chemicals and technological capabilities. Ali and Gittelman (2016) believe that laboratory science and clinical research are fundamentally different research paradigms and, further, that invention teams that span both basic and clinical research are more effective at patent licensing than teams comprised of inventors from only one domain. They find that technology development has a strong connection with both basic and clinical research in biomedicine, which can help inform policy for successful innovation in biomedicine.

Most studies emphasize the importance of co-evolution between scientific and technical knowledge in the process of searching for new therapies and biomedical innovation more generally. However, most studies, especially the qualitative studies, only focus on a specific disease, which does not provide an overall perspective of the research field or the methods through which scientific research evolves into a successfully developed technology. Our framework operates at the micro-level, but it spans the entire field of biomedicine and is designed to reveal key points along the path of innovation.

2.2. Topic evolution analysis

Latent Dirichlet allocation (LDA) is a common topic model that represents a document as a random mixture of latent topics, where each topic consists of a group of words with specific meanings. While it has proven to be an effective method of uncovering the implicit topics in a corpus of documents (Chen et al., 2017), the classic LDA model does not consider time so it cannot capture change (Wang and McCallum 2006). Producing focused, relevant results tends to require supplementary techniques. Detecting emerging topics in various fields of research is currently an area of great interest in academic circles. It is also a critical component of resource allocation decisions in research laboratories, government institutions, and corporations. According to Mane and Borner (2004), new areas of science continually evolve, while others gain or lose importance, merge, or split. These highly dynamic changes can make retaining an overview of the structure of a specific research field difficult. Mane et al. (2004) highlight that combining co-word occurrence with graphing techniques might be an effective way of identifying both the major topics of interest but, more importantly, the trends in and between those topics. Cobo, López-Herrera, Herrera-Viedma, and Herrera (2011) presented an approach based on co-word analysis and a longitudinal framework to detect the different topics addressed in a research field during a given period of time. In the case section, they apply their approach to the research field of fuzzy set theory using publications appearing in the most prestigious journals in that area. Co-word analysis (Callon et al., 1983) has been widely used as a content analysis technique to detect emerging trends in various research fields because it is a good way to map the strength of associations between items in textual data. Hence, to measure the strength of interactions between basic and technology research, co-word analysis has been used to examine software engineering (Callon et al., 1991), information research (Ding et al., 2001), fuzzy set theory (Cobo et al., 2011), and physical chemistry (Bailón-Moreno et al., 2014).

With this framework, we propose a new topic evolution model based on co-word analysis (See details in Section 3.2- Step 2). The results provide evidence of: the topic evolution process in biomedical research; the changing interactions between scientific research and technology development over time; and insights into the priority areas and knowledge levels that are more likely to lead to new drugs.

3. Methods and data

The framework combines topic evolution analysis with entropy statistics to trace the evolution of USFDA approved drugs from discovery to market approval. Topic evolution analysis is used to track the emergence and evolution of new research and technology topics in the field of biomedicine. The entropy statistics are used to measure the mutual information between research and development as an indication of their interplay. Details of the framework and the data follow.

3.1. Data

In biomedicine, published papers are the best source of data on scientific research (Bowen and Casadevall, 2015; Casadevall and Fang, 2014), and approved patents are the best source of information on technological developments (Tseng, 2009; Choi and Hwang, 2014). These two sources of information for a given set of marketable drugs should demonstrate the translation from theoretical science into a practical technology. Data on drug approvals was sourced from the Orange Book on the USFDA website, more formally known as the "Approved Drug Products with Therapeutic Equivalence Evaluations".[1] The Orange Book is a searchable listing of key patent and exclusivity information regarding USFDA approved drugs, which is updated monthly and comprises three files. "Products" contains the drug approval applications with information like ingredients, trading names, methods of delivery, etc. "Patents" contains the particulars of the patents, such as the type, applicant, and products covered. "Exclusivity" provides details on the type and duration of a drug's protection. All three files contain application numbers, which we used to link the drug applications to the corresponding patents. From a search on 24 Feb 2019, we found 2302 drugs approvals since 1983 and 4124 corresponding USPTO patents. We then downloaded the original patents from the Derwent Innovations Index database. With the help of Lens.org,[2] we identified 38,343 unique scholarly papers cited in those patents, which we subsequently retrieved from the Web of Science.

3.2. Conceptual framework

The broad research framework is illustrated in Fig. 2.

Step 1 – Data preprocessing

Typically, medical research papers and patents have some common fields (e.g., title, abstract, author) and some unique fields (e.g., keywords, subject categories). The purpose of this step is to remove meaningless data and retrieve relevant information. We performed natural language processing techniques to obtain terms from collected records using ITGInsight software (v 1.7). Developed by the Knowledge Management and Data Analysis Laboratory at the Beijing Institute of Technology, ITGInsight[3] is based on a technical term extraction method called PC-value that was first proposed by Han et al. (2011), which considers the frequency statistics of words in documents (Wang et al., 2014). Experiments show that this method has higher accuracy for term extraction than the C-value method (Frantzi et al., 2000). The formula for calculating a term's PC-value is as follows:

$$PC-value(a)=\begin{cases} log_2^{|a|} f(a)+2^{|a|-2} g(a) & a\ is\ not\ nested, \\ log_2^{|a|}\left(f(a)-\frac{1}{|T_a|}\sum_{b\in T_a} f(b)\right)+2^{|a|-2} g(a) & otherwise \end{cases} \quad (1)$$

[1] https://www.fda.gov/Drugs/InformationOnDrugs/ucm129662.htm
[2] https://www.lens.org/lens/patcite
[3] http://en.itginsight.com/download/

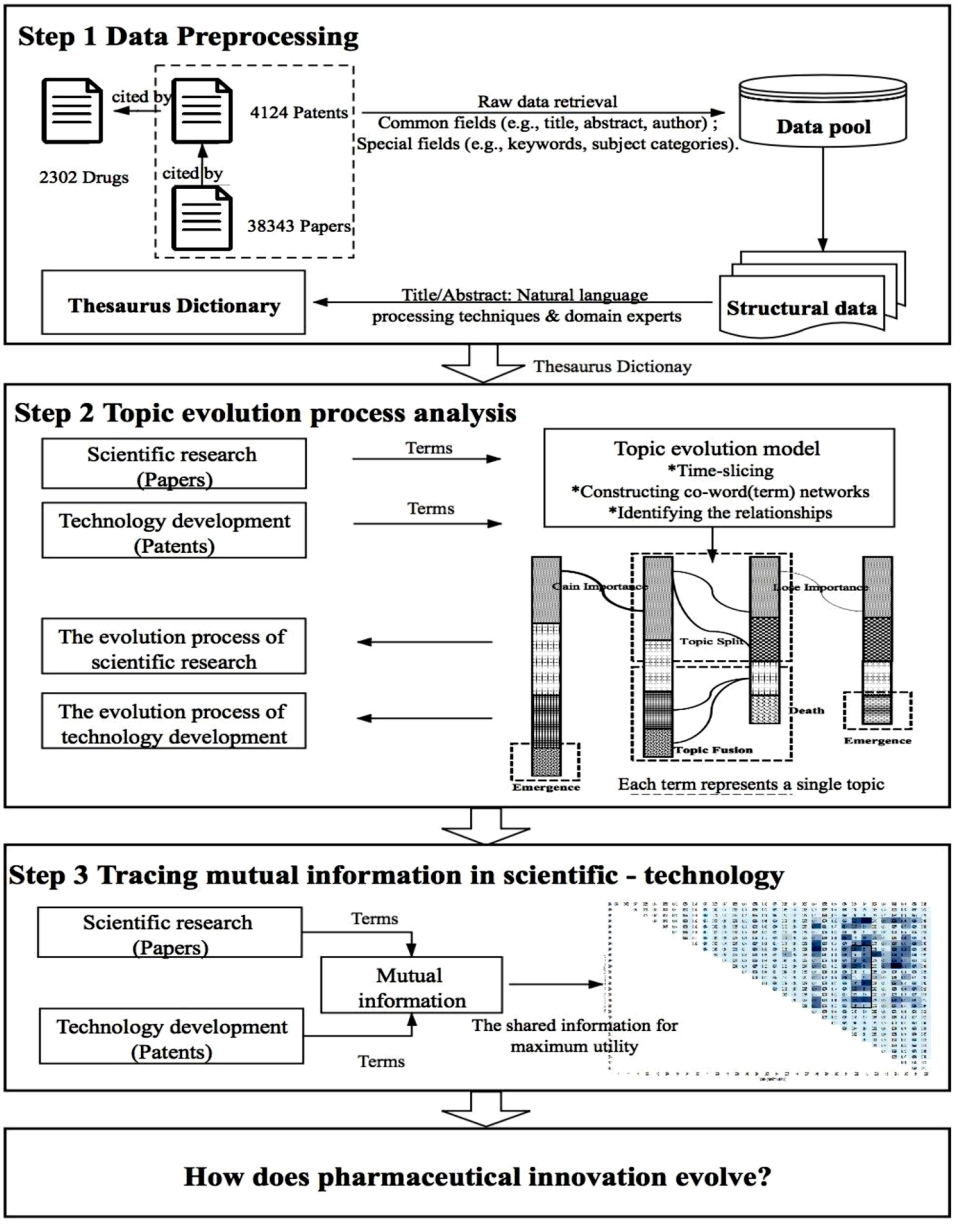

Fig. 2. Framework for the research on the biomedicine field.

where *a* is the candidate term, $|a|$ is the length of *a*, *f(a)* is its frequency of occurrence in the corpus, and *g(a)* is the document frequency of *a*. In the second part of the equation, *b* is an extracted candidate term that contains *a*, and *f(b)* is the number of times *b* appears in the corpus, i.e., its *total frequency*. T_a is the set of extracted candidate terms that contain *a*, and $|T_a|$ is the count of terms in the set.

Pre-processing the vocabulary involved word segmentation, tagging parts of speech, manually deleting meaningless and extreme words, and combining words with more than 95% similarity, e.g., system and systems. All words were then sorted according to their PC-value to form the preliminary results. The most important aspect of this step is to build a thesaurus from the preliminary results, which is usually done better with the help of domain experts. The thesaurus is subsequently used to reduce noise, consolidate related terms and provide more refined terms. The details are shown in Table 1. From here, all further analyses were based on these terms, not on the full text.

Step 2 – Topic evolution process analysis

Topic evolution analysis can provide a rich picture of the technical information in both papers and patents to reveal how topics emerge as a subject of interest and how they develop over time. In other words, it can expose both the science and the technology. We used ITGInsight (v 1.7) to conduct our topic evolution analysis. The specific methods are outlined below.

We divided scientific paper collection into time periods $D_t = \{d_{t1}, dt_2, dt_3, \ldots, d_{tm}\}$ (*d*: document; *m*: document number; *t*: time). To avoid interference from meaningless terms, we also took the thesaurus generated in Step 1 and performed further word segmentation on the set of papers. Thus, each paper corresponds to many terms, $d = \{w_1, w_2, w_3, \ldots, w_z\}$ (*w*: term) with each term representing a single topic. The next step was to construct a co-word network for each time period based on term (topic) co-occurrence. The same process was then repeated for the patents. More specifically, if Topic *a* and Topic *b* appeared in the same document, they were deemed to have a co-occurrence relationship. Further, the number of times *a* and *b* co-occurred in any document was considered a reflection of the strength of the connection and was weighted according to the number of co-occurrences within a certain

Table 1
Overview of the natural language processing.

Step	Description
Pre-processing	
1	Title/Abstract (4124 USPTO patents and 38,343 unique scholarly papers cited in these patents) – apply natural language processing techniques embedded in ITGInsight (v 1.7)
2	Data cleaning–remove meaningless words (e.g. one-day therapy, healthy young subjects, absorption rate, conscious rats, normotensive subjects, f forms, direct label, healthy control subjects), and extreme words (e.g. occurrence in only one record, word length less than 2)
3	Fuzzy matching–combine words with similar structures based on pattern commonality, such as stemming and text similarity)
4	Sequencing–Sort the words according to their PC-value, forming the preliminary results
Expert knowledge	
5	Consolidation and modification (preliminary results) – words that indicate the same meaning (e.g., abbreviations, synonyms and different representing forms of chemicals) will be merged to improve the integration level after consulting the domain experts, forming a thesaurus
Term extraction	
6	Title/Abstract (4124 USPTO patents and 38,343 unique scholarly papers cited in these patents) –apply natural language processing techniques embedded in ITGInsight (v1.7) and the thesaurus for term extraction

period. The differences between each co-word network for each time period reveals how topics have evolved. For instance, some topics might emerge during a time period, while others might disappear. Some gain or lose importance; others might fuse or split. We identified the following evolutionary relationships. Fig. 3 illustrates some examples of these relationships.

- Ø **Emergence**: a topic is mentioned for the first time. It could be entirely new or loosely related to something known.
- Ø **Death**: a topic that has been mentioned in previous time periods is not mentioned at all.
- Ø **Gaining importance**: mentions of the topic increase over two or more consecutive periods.
- Ø **Losing importance**: mentions of the topic decrease over two or more consecutive periods.
- Ø **Topic fusion**: in time period *T-1*, two topics appear in different documents, but, in time period *T*, they begin to appear in the same document.

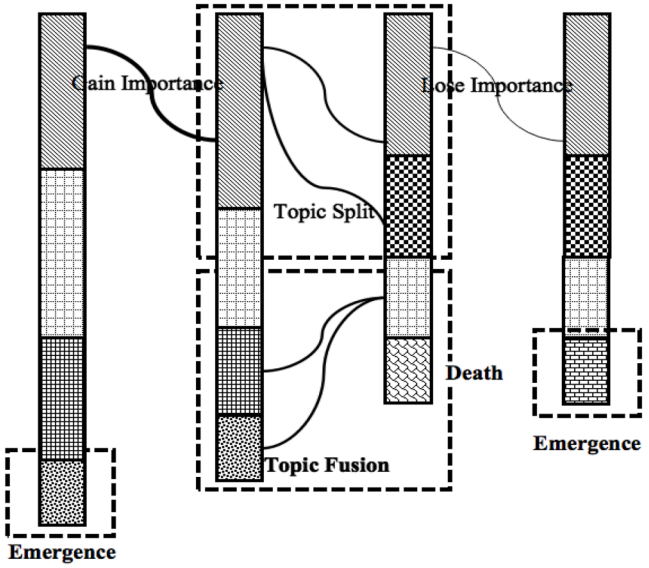

Fig. 3. The process of topic evolution analysis.

- Ø **Topic split**: in time period *T-1*, two topics appear in the same document, but, in time period *T*, they begin to appear in different documents.

Step 3 – Tracing the mutual information between science and technology

This step is where we measure the information shared between scientific research and technological development. The mutual information can be calculated according to Shannon's information entropy formula (Shannon, 1948). Shannon (1948) defines information entropy as the frequency of random events. When only a discrete random variable x exists, the information entropy is calculated as:

$$H_x = -\sum_i P(x_i) log_2 P(x_i) \quad (2)$$

In the above formula, H_x is the value of information entropy, that is, the size of uncertainty in units of bits, and $P(x_i)$ is the probability of variable x. On the basis of this definition, the information entropy when two variables x and y exist is calculated as:

$$H_{xy} = -\sum_{ij} P(x_i, y_j) log_2 P(x_i, y_j) \quad (3)$$

$$T_{xy} = H_x + H_y - H_{xy} \quad (4)$$

In formula (3), H_{xy} is the two-dimensional information entropy and $P(x_i, y_j)$ is the joint probability distribution of two variables. The transfer amount T_{xy} in formula (4) represents the two-dimensional mutual information, i.e., the uncertainty of information transmission between the innovation process representing the strength of interaction between scientific research and technology development and is the overlapping part of them (Lee and Kim, 2016).

Therefore, a decreasing mutual information value (T_{xy}) is consistent with the outward expansion of an emerging research field and indicates that scientific research and technology innovation has become more distinctly decoupled rather than becoming locked in its initial combination. By contrast, an increasing mutual information value (T_{xy}) is an indication of increasing dyadic coupling, possibly originating from the interactive integration of two research fields (Petersen et al., 2016).

4. Results

In pharmaceutical development, there is a set path from basic research to a marketable drug. It starts with the discovery of leading compounds, then moves through a series of technologies related to chemical manufacturing and control, pharmacokinetics, toxicology, pharmaceutical formulations, and others until it is fully developed. The last stage is approval for market, which involves tests, trials, and other hoops. According to Sternitzke (2010), it takes an average of 8–14 years to move from compound synthesis to commercialization. For policymakers, understanding why it takes this amount of time to bring a drug to market could be useful. Further, without knowledge of how and what takes time, little can likely be done to shorten the period of development. As mentioned before, we argue that biomedical innovation evolves as a result of progress along two different pathways – one being advances in scientific understanding, and the other being technological advances. In this regard, understanding how scientific research co-evolves with technological development could provide novel and profound insights along the path of translation process. Therefore, our analysis draws from the information found in both scientific papers and technology patents, to reveal the interplays between research and development in a translational research process.

Following the method outlined in Section 3, Sections 4.1 and 4.2 provide the results of topic evolution analysis for research and development, respectively, and Section 4.3 presents the entropy statistics for the shared information.

4.1. The evolution process of scientific research

To begin assembling a picture of how interest in these topics has emerged and evolved, we first took the 38,343 scientific papers and divided them into years. We then selected the top-30 most frequently mentioned terms for each year and conducted topic evolution analysis (Callon et al., 1991) as outlined in Section 3.2 Step 2. Fig. 4 illustrates how the topics have gained or lost importance, merged, or split over the period of study. Each topic has a different color, and the thickness of the connection represents the strength between topics.

This analysis reveals insights on two levels – first, some broad trends in the field and, second, a host of micro-level translations. There are too many specific evolutions to discuss individually. Hence, these next few sections discuss the three main overarching trends we discovered through a select few micro-level examples. These examples are good illustrations of how scientific research co-evolves with technological development along the path of biomedical innovation. They also show the power of the framework to capture big picture and small picture insights concurrently.

(1) The mechanisms of a drug's action mostly fall into particular categories

Therapeutic mechanisms are the specific biochemical interactions through which a drug produces its pharmacological effects. Over the period of study, we find that these mechanisms mostly fall into two categories: receptor mechanisms and non-receptor mechanisms. Hypertension/blood pressure is perhaps the strongest example of the difference between the two. Blood pressure has been a dominant topic in scientific research since 1991, but, in 1997, a strong connection between blood pressure, losartan potassium, and ACE inhibitors emerged. Both losartan potassium and ACE inhibitors have been studied as treatments for hypertension, but they have different therapeutic mechanisms. ACE inhibitors prevent angiotensin II from being produced (a non-receptor mechanism), while losartan potassium blocks the angiotensin receptors (a receptor mechanism). There is no clear evidence yet for which treatment is better (Ng et al., 2012). However, ACE inhibitors are more likely to be the first treatment of choice, while losartan potassium is likely to be prescribed to patients who have not responded to treatment with ACE inhibitors.

Most quick-acting drugs are based on receptor mechanisms. And, further, most quick-acting drugs tend to achieve their therapeutic effects by binding antagonists to either endogenous ligands or substrate competitive receptors. For example, paliperidone palmitate, which had high word-frequency from 2010 to 2011, is the main ingredient of atypical long-acting antipsychotics (INVEGA TRINZA and INVEGA SUSTENNA[4]). They are fast-acting in acute phases of mania and continuously improve symptoms in maintenance phases. They also have solid efficacy and convenient administration, offering a new treatment option for schizophrenia that could help patients stick to their regimen, preventing relapse. All are based on joint antagonism of the central dopamine 2 (D2) and 5-hydroxytryptamine 2 (5HT2A) receptors. Another quick-acting drug that acts directly on receptors/transmitters of the central nervous system is ZOHYDRO ER.[5] Its main ingredient is hydrocodone, which had high word-frequency from 2015 to 2016 (hydrocodone er [2015], hydrocodone er tablet prototypes [2015], hydrocodone [2016]). Specifically, hydrocodone is a phenanthrene derivative commonly used in combination with acetaminophen to relieve moderate to severe pain.

(2) Reducing complications from treatments cannot be overlooked

Drugs are often accompanied by side effects that endanger patients' health during treatment. Therefore, ways to reduce potential side effects has become a vital part of the drug research and development process. Diabetes mellitus (DM) is a good example of this phenomena. DM is a common, non-communicable, chronic disease that has been thoroughly studied, with particular dominance in 2003, 2009, and 2018. Intensive glycemic control (IGC) (intensive therapy [1993]) has been studied to treat DM for some time. However, landmark clinical studies have demonstrated that, although IGC might reduce the risk of microvascular events, macrovascular events, and mortality are either unaffected or elevated (Gerstein et al., 2008). In addition, the most common side effect of IGC is hypoglycemia, which increases the risk of death, dementia, and depression. As Fig. 4 shows, several strong connections between terms that related to blood sugar control research emerged around 2008 and 2009: hazard ratio ↔ intensive therapy ↔ standard therapy ↔ intensive glucose control ↔ intensive control; intensive-therapy group ↔ standard-therapy group ↔ normal glycated hemoglobin levels ↔ glycated hemoglobin value, etc. These findings have roused heated debate among researchers, as can be found in Kelly et al. (2009), who summarizes the purported clinical benefits and harms of intensive versus conventional glucose control for adults with type 2 diabetes. In their overall analyses, intensive glucose control reduced the risk of cardiovascular disease and increased the risk of severe hypoglycemia. Additionally, Ray et al. (2009) undertook a meta-analysis of randomized controlled trials to determine whether intensive treatment was beneficial, believing that intensive compared to standard glycemic controls could reduce coronary events without increasing the risk of death. Nowadays, lifestyle modifications, symptom reduction, and sustained inner immune homeostasis are keys to preventing and treating DM because, when followed, these measures of controlling blood sugar levels can result in significant quality of life improvements with little to no side effects (Saito et al., 2011; Nguyen et al., 2017).

(3) Drug delivery pathways are an important research turning point

Inhalation is an important drug delivery pathway that is widely used in clinical medicine. As is well known, asthma, chronic obstructive pulmonary disease, and allergic rhinitis are all commonly treated by inhaling a drug into the body's blood circulation through a drug delivery device. However, differences in morphology, molecular weight, charge, and the moisture absorption of drug particles can lead to dramatically different therapeutic effects. Fig. 4 shows that pulmonary hypertension and nitric oxide had a strong connection in 2000, and inhaled insulin and glycemic control were strongly connected in 2005. In the following years, both were widely researched because they were new types of inhaled treatments with inherent advantages, such as their safety and fast-acting effects. Exubera and Afrezz are prime examples, which were approved by the USFDA as inhaled treatments for DM in 2006 and 2014, respectively. Exubera was developed by Pfizer as a combination drug and device product for treating diabetes. It consists of technosphere insulin and an oral inhaler device (the Gen2 inhaler). Inhaled technosphere insulin is an effective and generally well-tolerated agent for the prandial treatment of hyperglycemia in T1DM and T2DM patients and was thought to provide a solution to insulin initiation barriers, such as injection phobia, which concerns weight gain and hypoglycemia. The glycemic efficacy of technosphere insulin is lower than that of subcutaneous insulin, but inhaled insulin has a lower risk of severe hypoglycemia and weight gain. Exubera was not particularly successful at market, which the manufacturer attributes to competition and mispricing, not issues with the innovation process (Munos, 2009). However, the MannKind Corporation subsequently developed Afrezza as a better

[4] INVEGA TRINZA and INVEGA SUSTENNA (Main ingredient: paliperidone palmitate) were approved by USFDA on 31/6/2009 and 18/5/2015 respectively.

[5] ZOHYDRO ER (Main ingredient: hydrocodone) was approved by USFDA on 25/10/2013. HYDROCODONE BITARTRATE AND ACETAMINOPHEN (Main ingredient: acetaminophen and hydrocodone bitartrate) was approved by USFDA on 29/10/2018.

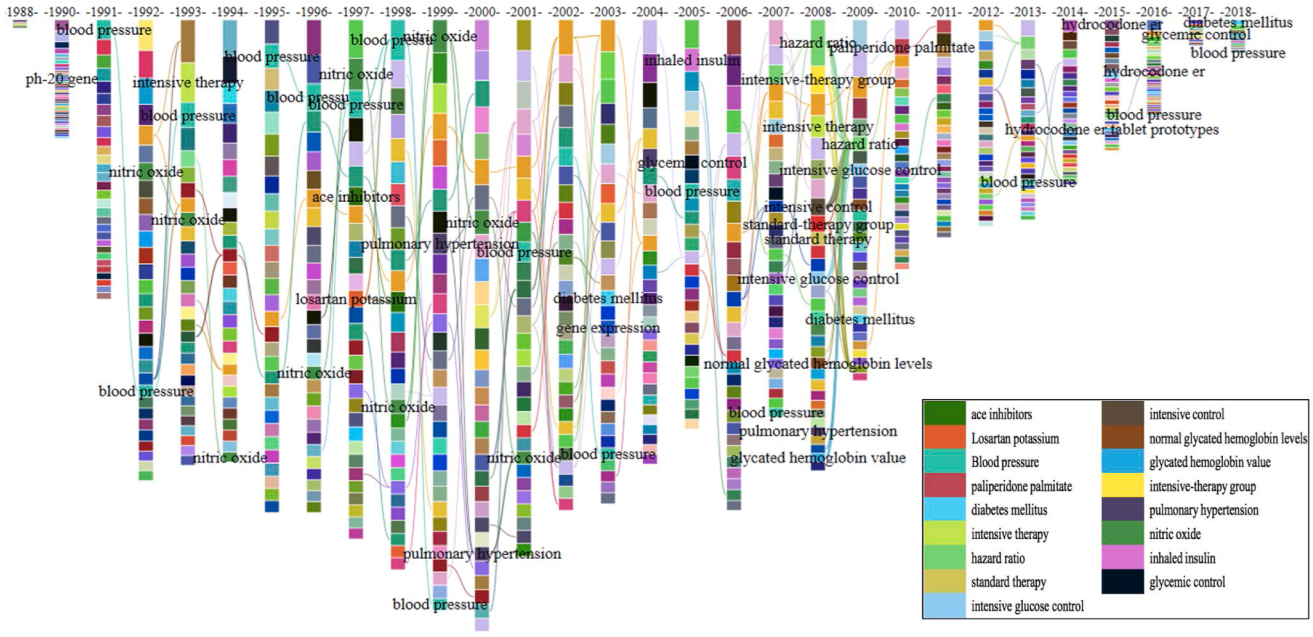

Fig. 4. The evolution of scientific research in biomedicine.
(Note: Fig. 4 is drawn by ITGInsight v 1.7)

version of Exubera. The inhalation powder is human insulin, and the dosing system and delivery device is simpler.

4.2. The evolution process of technology development

Next, we analyzed technological development. As we know, technology is a broad concept. Among other things, it involves the development and application of tools, machines, materials and processes that help to solve human problems (McOmber, 1999). In the field of bibliometrics and text mining, researchers use IPCs or keywords that relate to technical attributes to define technologies (Song et al., 2017). Hence, for the purpose of this study, all the terms related to technologies about chemical manufacturing and control, pharmacokinetic, toxicology, pharmaceutical formulation, etc. reflect elements of technology development (Tseng, 2009; Choi and Hwang, 2014).

Similar to Section 4.1, we took the 4124 patents and divided them into years. We then selected the top-30 most frequently mentioned terms for each year and conducted a topic evolution analysis (Callon et al., 1991) (as outlined in Section 3.2 Step 2). The results are shown in Fig. 5, followed by our key observations.

(1) The therapeutics mechanism is changing

The most apparent element of the results is the numerous connections between terms that relate to different organic groups in different time slices, e.g., 1–6c haloalky [2001] ↔ 1–6c alkyl [2001] ↔ 1–6c alkoxy [2001], 1–6c alkoxy [2005] ↔ 1–6c alkyl [2006], 4–9c heterocycloalkylalkyl [2004] ↔ mixed aryl [2004] ↔ 4–9c cycloalkyl [2004] ↔ 1–6c alkyl [2005]. All these terms represent important components related to the receptor mechanism of drugs. For example, azabicyclic compounds are antibacterial agents that can be used therapeutically or as disinfectants (Lampilas et al., 2006). Phenethanolamine compounds are beta 2 adrenoreceptor antagonists for treating respiratory diseases, skin diseases, depression, and congestive heart failure, among other things (Box, Coe, and Looker, et al., 2005). Notably, 4-oxoquinoline compounds, pharmaceutical salts, anti-HIV agents, and 1–4c alkyl appear together in 2003. 4-oxoquinoline compounds (or their pharmaceutical salts) have found use as anti-HIV agents in the treatment and prevention of AIDS. They are particularly efficacious when administered in combination with protease inhibitors and reverse transcriptase inhibitors due to their synergies. Plus, they are safe and have few side effects. The connection suggests a period of transition, where experts first began to focus on how to use these organic groups to synthesize specific enzyme inhibitors. Fig. 6 shows that, by 2017, chemically-synthesized enzyme inhibitors had become the main sources of new drugs, e.g., the strong connections between 3 and 8c cycloalkyl, 2–6c alkenyl, and 1–4c alkyl. These terms represent components of enzyme inhibitors, that can inhibit specific enzymes related to certain diseases in organisms for therapeutic effect. Enzyme inhibitors mainly come from plants, microorganisms, and chemical synthesis. Combining high throughput screening technology with combinatorial chemistry and combinatorial biosynthesis technology, large-scale screening of enzyme inhibitors has been realized, which is the main channel for many large pharmaceutical companies in the world to screen new drugs based on enzyme inhibitors.

(2) Drug release patterns are gradually changing in the drug discovery process

The word “tranexamic acid” occurred frequently from 2008 to 2012 and was often accompanied with the term “dosage form”. In the field of medicine, tranexamic acid (or its pharmaceutical salt) is useful for treating patients with menorrhagia, conization of the cervix, epistaxis, hyphema, or hereditary angioneurotic edemas. It can also lessen heavy menstrual bleeding. In 2008, an oral form of tranexamic acid was patented with a release time somewhere in between the immediate and controlled release times of its predecessors (Greiwe et al., 2008). This modified release form both increases the in-vitro dissolution rate of tranexamic acid and reduces the common side effects associated with the immediate release forms, such as headaches, nausea, vomiting, diarrhea, constipation, cramping, and bloating. In 2010, a tablet formulation of tranexamic acid was developed (Facemire et al., 2010). This release mechanism can prevent a large bolus of tranexamic acid forming in the stomach, which means patients should see fewer adverse effects from tranexamic acid therapy. Over the next two years, researchers attempted to adjust the proportions of the tranexamic acid with its pharmaceutical salt to increase bioavailability levels. Success appears to have been achieved in 2013, as indicated by the

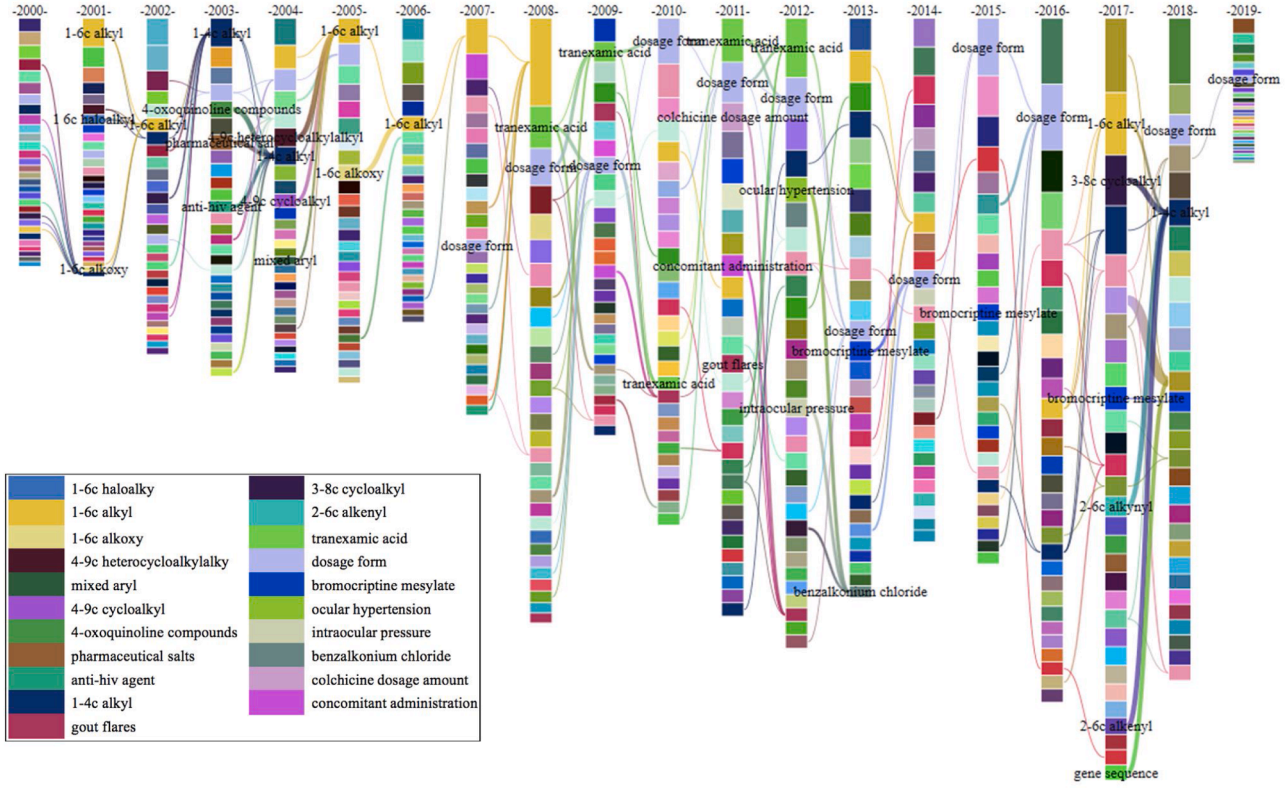

Fig. 5. Technology evolution pathways in biomedicine.
(Note: Fig. 5 is drawn by ITGInsight v 1.7)

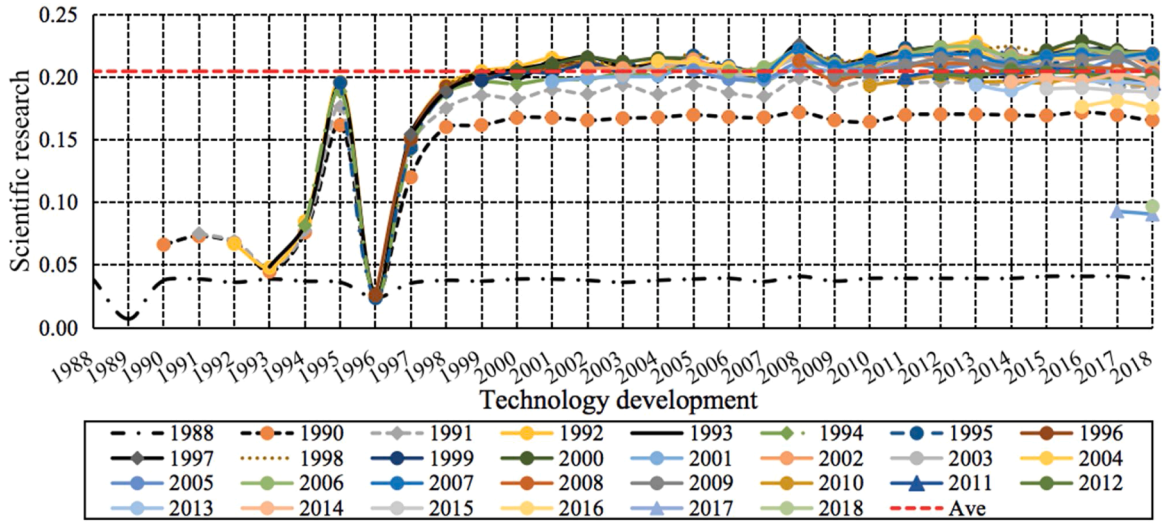

Fig. 6. The knowledge penetration from scientific research into technology development.

co-occurrence between bromocriptine mesylate ↔ dosage form that grew in dominance in 2013, 2015, 2017 and 2018, when experts found that micronized bromocriptine mesylate in the form of an oral tablet could allow the gastric and intestinal mucosa to absorb a substantial amount of bromocriptine. Ever since, researchers have been exploring new and different release modes for bromocriptine mesylate in attempts to find effective treatments for a range of diseases. For example, normal-release bromocriptine mesylate is useful for treating hyperprolactinemia-associated dysfunction, acromegaly, and Parkinson's disease. It also prevents or mitigates intolerance to levodopa therapy for Parkinson's disease, physiological lactation, insulin resistance, hyperinsulinemia, and hyperglycemia. Immediate-release bromocriptine mesylate (cycloset) is useful for improving glycemic control in Type II diabetes patients (Scranton et al., 2008).

(3) The proportion of pharmaceutical composition plays an important role in medicinal effects

It was already known that an ophthalmic composition of bimatoprost and benzalkonium chloride in aqueous form was helpful for treating glaucoma or intraocular hypertension in mammals. In 2012, a strong connection between "ocular hypertension", "intraocular pressure", and

"benzalkonium chloride" emerged as an evolution of these previous findings. The 2012 composition shows higher bioavailability and higher permeability than only bimatoprost (Chang et al., 2006, 2009). Moreover, the proportion of bimatoprost and benzalkonium chloride also affects the permeability of bimatoprost across the corneal epithelial cell layers and results in less hyperemia. Currently, 0.01–0.015 wt./vol.% of bimatoprost and 0.02 wt./vol.% of benzalkonium chloride are considered as the most effective formulation ratio (Chang et al., 2013). In addition, there are numerous studies on pharmaceutical composition in patents. For instance, a pharmaceutical composition which can be used for drugs that are unstable in polyethylene glycol containing compositions is proposed by Ishida et al. (2009). The experiment results show that it can be used as stable pharmaceuticals for administering drugs especially (i) amide compounds for treating and preventing e.g. sleep disorders (e.g. circadian rhythm sleep disorder or jet lag), senile dementia, Alzheimer's disease, osteopathies, cerebral circulatory disorders, head trauma, stress, depression, convulsions, anxiety, epilepsy, Parkinson's disease, hypertension, cataracts, cancer and diabetes, or (ii) amine compounds useful as amyloid beta protein production inhibitors for treating and preventing e.g. cerebral vascular disorders, head trauma and spinal disorders.

4.3. Tracing mutual information in scientific research and technology development

As described in the Methodology (Step 3), we calculated the mutual information between scientific research and technology development on the basis of Shannon's definition of information. After term extraction, we calculated the mutual information for all pairwise terms in every year. The cumulative frequency of all terms was considered to be the benchmark for calculating the weight of every term. More specifically, two-dimensional mutual information is a kind of quantitative analysis method that represents dynamic evolution trend on the information interaction between science–technology. The unit of information entropy is expressed in bits to facilitate the research results.

From Figs. 6 and 7, we can see that the mutual information between each science-technology pair has a tendency to increase as time passes, which indicates the connections between them are growing stronger. However, this does not last forever – all eventually reach a steady state. Fig. 6 further shows that, as more technology is developed, the amount of knowledge penetration from scientific research into technological innovation gradually increases in the periods following. But nothing in Fig. 7 reveals the promotion effect of technology development on scientific research.

The analysis results show that problems with information interaction between scientific research and technological development still exist, and that overall interaction efficiency is low. In recent years, researchers and policymakers have paid considerable attention to the transfer of basic science research achievements by significantly promoting the flow of knowledge across institutions. However, this move is in its infancy and requires further steps in due course.

Fig. 8 shows a more in-depth analysis on the information interaction between scientific research and technology development. These illustrations show the time periods over which the knowledge transformation process occurs. The relevant numbers are marked in Fig. 8, when the bilateral mutual information between scientific research and technology development is greater than the average. We can see that the period between 1997 and 2007 was dominated by scientific research and technology development is concentrated around 2012–2013. Therefore, it can be concluded that there is 5–16 year delay to generate an adaptive technology that can help convert theoretical research into practical clinical medicine.

5. Discussions and policy implications

Over the past few decades, innovation in biomedicine has been an important part of the policy agenda for many countries. However, biomedical pursuits can impose a serious burden on financial resources because they typically demand enormous levels of investment, they are risky, and the return on investment is a very long term. The focus of our analysis so far has been on the topic evolution paths of scientific research and technological development over time at a micro level using as our subject of analysis drugs approved for use in the world's largest market for pharmaceutical products – the United States. A deeper understanding of how pharmaceutical innovation evolves may lead to increased confidence in science by society and additional support for future research. Our discussion here concentrates on three related issues: research trends in biomedicine, how traditional concepts of treatment are transforming, and the cooperative supervision of both scientific research and technology development.

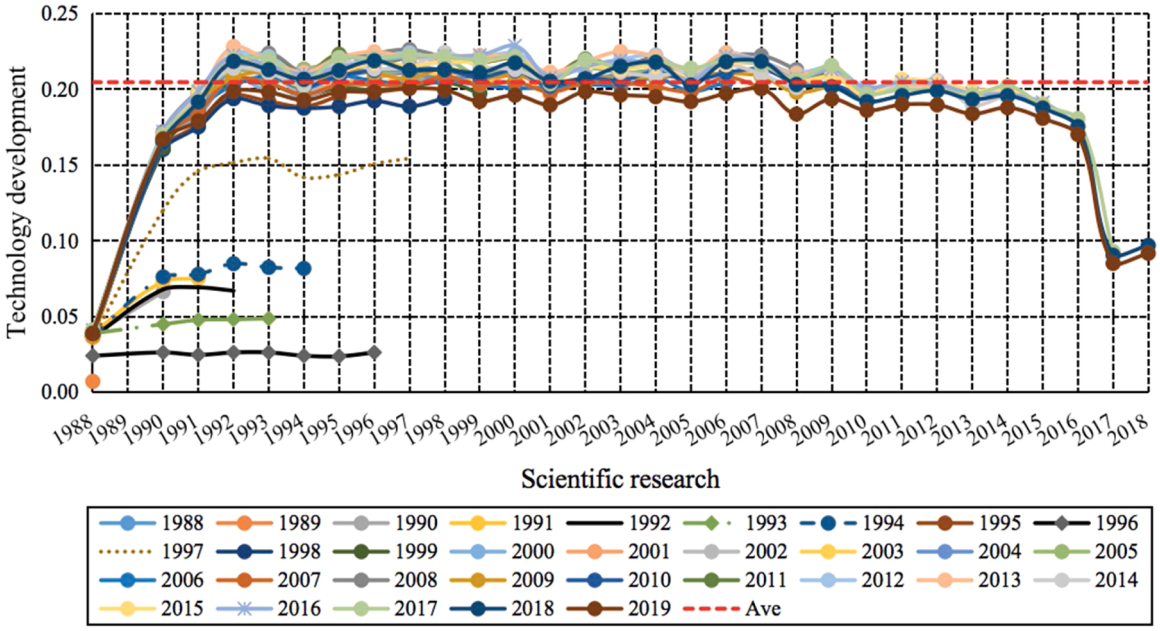

Fig. 7. The promotion effect of technology development on scientific research.

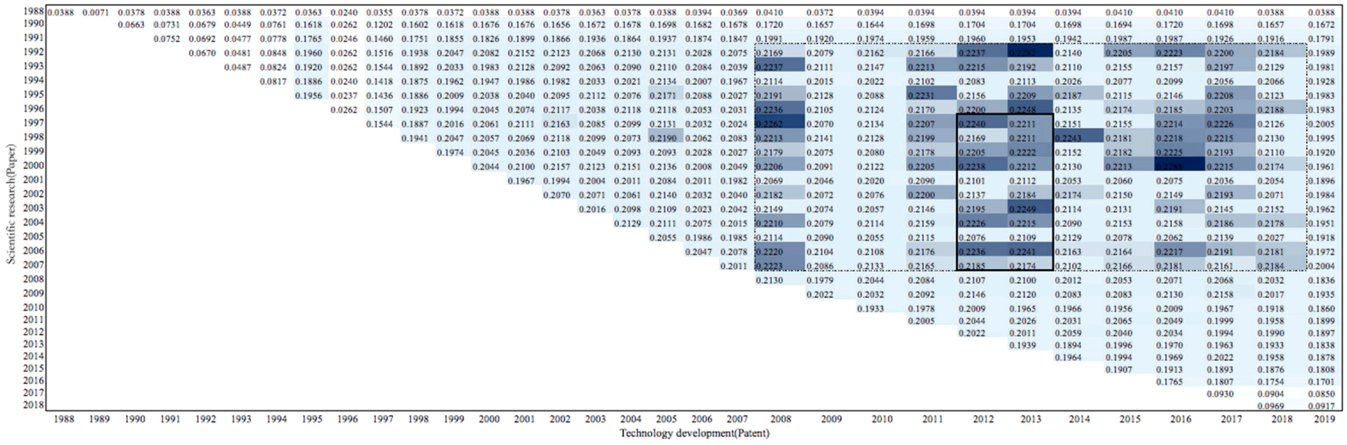

Fig. 8. The shared information between scientific research and technological innovation.

In terms of research trends, we two main drug action mechanisms: those that operate on receptors and those that do not (*scientific research 1*). However, our topic evolution analysis revealed gaps between science and technology that imply the following path: receptor mechanism → targeted therapy (non-receptor mechanism) → gene therapy → stem cell therapy. Today, targeted therapy is the main treatment for many diseases, and it would be natural to assume that gene-based therapies are the next therapeutic mechanism. For example, Luxturna was approved by the USFDA in 2017. Unlike other approved drugs, it is an adeno-associated virus vector-based gene therapy indicated for the treatment of patients with confirmed biallelic RPE65 mutation-associated retinal dystrophy.[6] However, terms like 'gene expression', 'cart gene', and similar have been topics of research in scientific papers for decades. Hence, researchers have been focusing on genes for a long time, but most work remained theoretical until 2017 when terms like 'gene sequencing' became more prevalent. Gene sequencing is a technique that identifies and analyzes entire gene sequences from blood or saliva. In recent years, researchers have been exploring ways of using gene sequences to predict the chances of someone developing a disease and treating it in advance. It is also interesting to note that "graft versus host disease pathogenesis" (GvHD), which relates to stem cell therapy, has seen high word-frequency in recent years (Kurtzberg et al., 2014). Therefore, we can infer that stem cell therapy may be another fledgling branch of biomedical research. The key here is that, usually, changing a therapeutic mechanism does not occur without the support of a technological development. Receptor mechanisms and targeted therapies rely on age-old organic synthesis (*technology development 1*), but gene therapy was stuck in theory until gene sequencing became viable. Similarly, stem cell therapy also required some advances in supporting technology to kindle resurging interest.

In addition, the analysis shows that drug release patterns are already changing (*technology development 2*), although gradually, which also relates to the changing proportions of drugs versus salts in pharmaceutical composition (*technology development 3*). Both these technology trends are attempts to reduce the complications associated with treatments – a motivation that has dominated biomedicine (*scientific research 2*). Generally, survival and a functional cure have been the first consideration in all clinical treatments. But, more recently, how to effectively reduce potential side effects becomes a vital part of drug research and development process. Our analysis shows that "qt prolongation gene sequence" and "qt interval" had a strong connection in 2017. Cardiac QT interval prolongation is one of the important risk factors in clinical malignant arrhythmias and sudden cardiac death. After identifying an individual's gene sequences and analyzing their gene expression products, one can determine whether a compound is capable of prolonging a qt interval in that person. Meanwhile, it is also one of the significant side effects of taking antidepressants. Interestingly, a nasal spray formulation of esketamine (Spravato) was licensed as adjunctive therapy for the management of adults with treatment-resistant depression (TRD) two years after the appearance of "qt prolongation gene sequence" and "qt interval". Given its novel mechanism and the current paucity of approved pharmacotherapy options for TRD, esketamine nasal spray in conjunction with an oral antidepressant provides an important treatment option for this difficult-to-treat high-risk patient population. This phenomenon indicates that reducing complications from treatment is indeed an important consideration in drug discovery and exploring new drug delivery pathways is also worth further researching (*scientific research 3*).

From the above analysis, we can conclude that the relationship between scientific research and technological development is crucial, but a time lag does exist in the knowledge transformation process (Collins, 2011; Reidenberg, and Erle, 2012). Our analysis shows it takes 5–16 years to generate an adaptive technology that acts as a catalyst for translating theoretical research into clinical practice. Therefore, we suppose scientific research institutes and pharmaceutical enterprises should be encouraged to cooperate, lest they evolve independently as suggested by Gittelman (2016). Research institutes should focus on pre-clinical processes and concept verification (usually clinical trial phase II). They should seek to verify concepts as fast as possible, and then sell their evidence to large pharmaceutical enterprises. Pharmaceutical companies should focus on managing patents and organizing external partners to conduct drug R&D. Only in this way can the biomedical industry achieve a more efficient business model. In addition, we find that integrating different types of knowledge will help speed up the diffusion and application of scientific research in biomedicine. However, it is also important to understand that the paradigms and beliefs about the very nature of medical discovery in different research fields are not same, and even conflict in some cases. For example, the problems with "permeation" in different disciplines in biomedical innovation is very complex, and involves understanding, integrating, and disambiguating. For policymakers, it is necessary to develop demand-oriented strategies and form a new research paradigm, promoting common development should be the focus of future policy efforts and initiatives.

We are sure that this research has explored trajectories of innovation progress and not just purely of inventions. We also believe that the research framework and the software developed by our laboratory is capable of generating robust evidence of innovation at a micro level that yields interesting possibilities for applications in other research domains. We hope that this paper has demonstrated that, for most researchers and policymakers in the sciences, focusing on the relationships

[6] https://www.drugs.com/pro/luxturna.html

between science and technology is a valuable and important direction for future research.

6. Limitations and future research

This study has some limitations that are worth discussing. First, we measured the interplay between scientific research and technology development in terms of mutual information. Yet, different forms of uncertainty exist, which impact the results of word segmentation and selection. In future studies, we intend to explore these impacts and develop solutions to refine the analysis process and results. Second, the specific focus on the Orange Book and USFDA approved drugs limits the generalizability of this study. In future work, we intend to conduct a more systematic analysis of additional medical areas to generalize our findings.

This framework allowed us to trace the evolution of topics in biomedicine at a content level. We were also able to explore the interplays between scientific research and technological development using mutual information as an indicator. The knowledge represented through this process provides novel quantitative insights into the biomedical ecosystem, with practical implications for inventors, investors, and policymakers. Our next undertaking is a heterogeneous information network that contains scientific research, technological development, marketable drugs, and drug side effects with the aim of forecasting the frontier research points in the field of biomedicine.

Author Statement

Xuefeng Wang: Conceptualization, Writing - Review & Editing
Shuo Zhang: Writing - Original Draft, Visualization
Yuqin Liu: Methodology, Software
Jian Du: Resources
Heng Huang: Data Curation, Visualization

Acknowledgments

This work is partly supported by the General Program of the National Natural Science Foundation of China (Grant no. 72074020 and 71774012) and the strategic research project of the Development Planning Bureau of the Chinese Academy of Sciences (Grant no. GHJ-ZLZX-2019-42).The findings and observations present in this paper are those of the authors and do not necessarily reflect the views of the supporters or the sponsors. The authors would like to thank the anonymous reviewers for their constructive input into this paper.

References

Al-Humadi, N., 2017. Pre-clinical toxicology considerations for vaccine development. Vaccine 35 (43), 5762–5767. https://doi.org/10.1016/j.vaccine.2017.09.021.
Ali, A., Gittelman, M., 2016. Research paradigms and useful inventions in medicine: patents and licensing by teams of clinical and basic scientists in academic medical centers. Res. Policy 45 (8), 11–23. https://doi.org/10.1016/j.respol.2016.03.015.
Bailón-Moreno, R., Jurado-Alameda, E., Ruiz-Baños, R., 2014. The scientific network of surfactants: structural analysis. J. Am. Soc. Inf. Sci. Technol. 57 (7), 949–960. https://doi.org/10.1002/asi.20362.
Box, P.C., et al. New Phenethanolamine Compounds are Beta 2 Adrenoreceptor Agonists Used for Treating e.g., Respiratory diseases, Skin diseases, Depression and Congestive Heart Failure. database: U.S. Patent 2,005,075,394 [P]. 2005.
Bowen, A., Casadevall, A., 2015. Increasing disparities between resource inputs and outcomes, as measured by certain health deliverables, in biomedical research. Proc. Natl. Acad. Sci. U.S.A. 112 (36), 11335–11340. https://doi.org/10.1073/pnas.1504955112.
Burnette, M.H., 2015. The “research audit” model: a prototype for data-driven discovery of interdisciplinary biomedical research. Portal Libr. Acad. 15 (4), 645–659. https://doi.org/10.1353/pla.2015.0052.
Callon, M., Courtial, J.P., Laville, F., 1991. Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemsitry. Scientometrics 22 (1), 155–205. https://doi.org/10.1007/bf02019280.
Callon, M., Courtial, J.P., Turner, W.A., Bauin, S., 1983. From translations to problematic networks: an introduction to co-word analysis. Soc. Sci. Inf. Sci. Soc. 22 (2), 191–235. https://doi.org/10.1177/053901883022002003.
Carpenter, M.P., Cooper, M., Narin, F., 1980. Linkage between basic research literature and patents. Res. Manag. 23 (2), 30–35. https://doi.org/10.1080/00345334.1980.11756595.
Casadevall, A., Fang, F.C., 2014. Causes for the persistence of impact factor mania. MBio 5 (3). https://doi.org/10.1128/mBio.01342-14. Article e00064-14.
Chang, C., et al. Composition Useful for Treating Glaucoma and Intraocular Hypertension Comprises Bimatoprost and Benzalkonium Chloride in Specified Higher Amount. Database: U.S. Patent 2,006,211,770 [P]. 2006.
Chang, C., Chang, J.N., Schiffman, R.M., et al. Ophthalmic Composition Useful for Treating Glaucoma Or Ocularhypertension, Comprises bimatoprost, and Benzalkonium Chloride. Database: U.S. Patent 2,014,100, 287[P]. 2013.
Chang, C., et al. Ophthalmic Composition in the Form of an Aqueous Liquid Useful for Treating Glaucoma Or Intraocular Hypertension in mammals, Comprises Bimatoprost and Benzalkonium Chloride. Database: U.S. Patent 2,009,149,546[P]. 2009.
Chen, H.S., Zhang, G.Q., Zhu, D.H., Lu, J., 2017. Topic-based technological forecasting based on patent data: a case study of Australian patents from 2000 to 2014. Technol. Forecast. Soc. Change 119, 39–52. https://doi.org/10.1016/j.techfore.2017.03.009.
Choi, J., Hwang, Y.S., 2014. Patent keyword network analysis for improving technology development efficiency. Technol. Forecast. Soc. Change 83, 170–182. https://doi.org/10.1016/j.techfore.2013.07.004.
Cobo, M.J., Lopez-Herrera, A.G., Herrera-Viedma, E., Herrera, F., 2011. An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the Fuzzy Sets Theory field. J. Informetr. 5 (1), 146–166. https://doi.org/10.1016/j.joi.2010.10.002.
Collins, F.S., 2011. Reengineering translational science: the time is right. Sci. Transl. Med. 3 (90) https://doi.org/10.1126/scitranslmed.3002747. Article 90cm17.
Cutler, D.M., McClellan, M., 2001. Is technological change in medicine worth it? Health Aff. Millw. 20 (5), 11–29.
Ding, Y., Chowdhury, G.G., Foo, S., 2001. Bibliometric cartography of information retrieval research by using co-word analysis. Inf. Process. Manag. 37 (6), 817–842. https://doi.org/10.1016/s0306-4573(00)00051-0.
Duda, G.N., et al., 2014. Changing the mindset in life sciences toward translation: a consensus. Sci. Transl. Med. 6 (264) https://doi.org/10.1126/scitranslmed.aaa0599. Article 264cm12.
Du, J., Li, P.X., Guo, Q.Y., Tang, X.L., 2019. Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis. J. Informetr. 13 (1), 132–148. https://doi.org/10.1016/j.joi.2018.12.004.
Facemire, J.W., Greiwe, J.S., Heaslley, R.A., Modest, J.D., Moore, K.A. Tranexamic Acid Tablet Formulation Useful for Treating Patient Suffering From Menorrhagia Comprises Mixture of Tranexamic Acid and Modified Release Material Such That Formulation Provides In-Vitro Dissolution Release Rate of Tranexamic acid. database: U.S. Patent 2,010,143,468 [P]. 2010.
Frantzi, K., Ananiadou, S., Mima, H., 2000. Automatic recognition of multi-word terms: the c-value/nc-value method. Int. J. Digit. Libr. 3 (2), 115–130. https://doi.org/10.1007/s007999900023.
Freedman, L.P., Cockburn, I.M., Simcoe, T.S., 2015. The economics of reproducibility in preclinical research. PLoS. Biol. 13 (6) https://doi.org/10.1371/journal.pbio.1002165. Article e1002165.
Gerstein, H.C., et al., 2008. Effects of intensive glucose lowering in type 2 diabetes. N. Engl. J. Med. 358 (24), 2545–2559. https://doi.org/10.1056/nejmoa0802743.
Gittelman, M., 2016. The revolution re-visited: clinical and genetics research paradigms and the productivity paradox in drug discovery. Res. Policy 45 (8), 82–97. https://doi.org/10.1016/j.respol.2016.01.007.
Glanzel, W., Meyer, M., 2003. Patents cited in the scientific literature: an exploratory study of ’reverse’ citation relations. Scientometrics 58 (2), 415–428. https://doi.org/10.1023/a:1026248929668.
Greiwe, J.S., Heaslley, R.A., Moore, K.A., Facemire, J.W., Modest, J.D. Oral Dosage Form, Useful e.g. for Treating Human Patient Suffering From menorrhagia, epistaxis, Hyphema and Hereditary Angioneurotic edema, Comprises Tranexamic Acid and Modified Release Material e.g. Vinyl Polymers. Database: U.S. Patent 2,009,048,341 [P]. 2008.
Han, H., Zhu, D.H., Wang, X.F., 2011. Technical term extraction method for patent document. J. Chin. Soc. Sci. Tech. Inf. 30 (12), 1280–1285.
Ishida, H., Fukuta, M. Pharmaceutical Composition Comprises Drug Unstable in Polyethyleneglycol Containing Preparations Coated With Copolyvidone. Database: U.S. Patent 2,009,274,733-A1[P]. 2009-11-5.
Jibu, M., 2014. Mapping of scientific patenting: toward the development of ’J-GLOBAL foresight’. Technol. Anal. Strateg. Manag. 26 (4), 485–498. https://doi.org/10.1080/09537325.2013.877129.
Kelly, T.N., Bazzano, L.A., Fonseca, V.A., Thethi, T.K., Reynolds, K., He, J., 2009. Systematic review: glucose control and cardiovascular disease in type 2 diabetes. Ann. Intern. Med. 151 (6) https://doi.org/10.7326/0003-4819-151-6-200909150-00137, 394-W130.
Kurtzberg, J., et al., 2014. Allogeneic human mesenchymal stem cell therapy (remestemcel-l, prochymal) as a rescue agent for severe refractory acute graft-versus-host disease in pediatric patients. Biol. Blood Marrow Transplant. 20 (2), 229–235. https://doi.org/10.1016/j.bbmt.2013.11.001.
Lampilas, M., et al. New Azabicyclic Compounds, Useful as Antibacterial Agents for Therapeutic Use or as Disinfectants, Also Newintermediate. Database: U.S. Patent 7,112,592[P]. 2006-9-26.
Lee, Y.H., Kim, Y., 2016. Analyzing interaction in R&D networks using the Triple Helix method: evidence from industrial R&D programs in Korean government. Technol. Forecast. Soc. Change 110, 93–105. https://doi.org/10.1016/j.techfore.2015.10.017.

Lee, E.S., McDonald, D.W., Anderson, N., Tarczy-Hornoch, P., 2009. Incorporating collaboratory concepts into informatics in support of translational interdisciplinary biomedical research. Int. J. Med. Inf. 78 (1), 10–21. https://doi.org/10.1016/j.ijmedinf.2008.06.011.

Mane, K.K., Borner, K., 2004. Mapping topics and topic bursts in PNAS. Proc. Natl. Acad. Sci. U.S.A. 101, 5287–5290. https://doi.org/10.1073/pnas.0307626100.

McMillan, G.S., Narin, F., Deeds, D.L., 2000. An analysis of the critical role of public science in innovation: the case of biotechnology. Res. Policy 29 (1), 1–8. https://doi.org/10.1016/s0048-7333(99)00030-x.

McOmber, J.B., 1999. Technological autonomy and three definitions of technology. J. Commun. 49 (3), 137–153. https://doi.org/10.1111/j.1460-2466.1999.tb02809.x.

Metcalfe, J.S., James, A., Mina, A., 2005. Emergent innovation systems and the delivery of clinical services: the case of intra-ocular lenses. Res. Policy 34 (9), 1283–1304. https://doi.org/10.1016/j.respol.2005.01.015.

Mina, A., Ramlogan, R., Tampubolon, G., Metcalfe, J.S., 2007. Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge. Res. Policy 36 (5), 789–806. https://doi.org/10.1016/j.respol.2006.12.007.

Morlacchi, P., Nelson, R.R., 2011. How medical practice evolves: learning to treat failing hearts with an implantable device. Res. Policy 40 (4), 511–525. https://doi.org/10.1016/j.respol.2011.01.001.

Munos, B., 2009. Lessons from 60 years of pharmaceutical innovation. Nat. Rev. Drug Discov. 8 (12), 959–968. https://doi.org/10.1038/nrd2961.

Nelson, R.R., Buterbaugh, K., Perl, M., Gelijns, A., 2011. How medical know-how progresses. Res. Policy 40 (10), 1339–1344. https://doi.org/10.1016/j.respol.2011.06.014.

Ng, T.M.H., Ackerbauer, K.A., Hyderi, A.F., Hshieh, S., Elkayam, U., 2012. Comparative effects of nesiritide and nitroglycerin on renal function, and incidence of renal injury by traditional and RIFLE criteria in acute heart failure. J. Cardiovasc. Pharmacol. Ther. 17 (1), 79–85. https://doi.org/10.1177/1074248411406441.

Nguyen, B., Clements, J., 2017. Obesity management among patients with type 2 diabetes and prediabetes: a focus on lifestyle modifications and evidence of antiobesity medications. Expert Rev. Endocrinol. Metab. 12 (5), 303–313. https://doi.org/10.1080/17446651.2017.1367285.

Palucki, M., Higgins, J.D., Kwong, E., Templeton, A.C., 2010. Cheminform abstract: strategies at the interface of drug discovery and development: early optimization of the solid state phase and preclinical toxicology formulation for potential drug candidates. J. Med. Chem. 53 (46), 5897–5905. https://doi.org/10.1002/chin.201046236.

Patrick, T., Steven, G., 2015. Using science and technology indicators to manage R&D as a business. EMJ Eng. Manag. J. 13 (3), 9–14. https://doi.org/10.1080/10429247.2001.11415121.

Petersen, A.M., Rotolo, D., Leydesdorff, L., 2016. A triple helix model of medical innovation: supply, demand, and technological capabilities in terms of medical subject headings. Res. Policy. 45 (3), 666–681. https://doi.org/10.1016/j.respol.2015.12.004.

Ray, K.K., et al., 2009. Effect of intensive control of glucose on cardiovascular outcomes and death in patients with diabetes mellitus: a meta-analysis of randomised controlled trials. Lancet 373 (9677), 1765–1772. https://doi.org/10.1016/s0140-6736(09)60697-8.

Rees, J., 2004. The fundamentals of clinical discovery. Perspect. Biol. Med. 47 (4), 597–607. https://doi.org/10.1353/pbm.2004.0068.

Reidenberg, M.M., Erle, H., 2012. U.S. basic research: delayed drug development. Science 337 (6102), 1605.

Saito, T., et al., 2011. Lifestyle modification and prevention of type 2 diabetes in overweight Japanese with impaired fasting glucose levels a randomized controlled trial. Arch. Intern. Med. 171 (15), 1352–1360. https://doi.org/10.1001/archinternmed.2011.275.

Sarewitz, D., Nelson, R., 2008. Three rules for technological fixes. Nature 456 (7224), 871–872. https://doi.org/10.1038/456871a.

Scranton, R.E., Farwell, W., Ezrokhi, M., Gaziano, J.M., Cincotta, A.H., 2008. Quick release bromocriptine (cycloset (TM)) improves glycaemic control in patients with diabetes failing metformin/sulfonylurea combination therapy. Can. J. Diabetes 32 (4), 324. https://doi.org/10.1016/S1499-2671(08)24099-2.

Schechter, A.N., Perlman, R.L., Rettig, R.A., 2003. Why is revitalizing clinical research so important, yet so difficult? Perspect. Biol. Med. 47 (4), 476–486. https://doi.org/10.1353/pbm.2004.0070.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Techn. J. 27 (3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Song, K., Kim, K.S., Lee, S., 2017. Discovering new technology opportunities based on patents: text-mining and f-term analysis. Technovation. https://doi.org/10.1016/j.technovation.2017.03.001, 60-61 1-14.

Sternitzke, C., 2010. Knowledge sources, patent protection, and commercialization of pharmaceutical innovations. Res. Policy 39 (6), 810–821. https://doi.org/10.1016/j.respol.2010.03.001.

Sung, H.Y., Wang, C.C., Huang, M.H., Chen, D.Z., 2015. Measuring science-based science linkage and non-science-based linkage of patents through non-patent references. J. Informetr. 9 (3), 488–498. https://doi.org/10.1016/j.joi.2015.04.004.

Tseng, C.Y., 2009. Technology development and knowledge spillover in Africa: evidence using patent and citation data. Int. J. Technol. Manag. 45 (1–2), 50–61. https://doi.org/10.1504/ijtm.2009.021519.

Wang, X., Mccallum, A., 2006. Topics over time: a non-Markov continuous-time model of topical trends. SIGKDD 424–433.

Wang, X.F., et al., 2014. Collaboration network and pattern analysis: case study of dye-sensitized solar cells. Scientometrics 98 (3), 1745–1762. https://doi.org/10.1007/s11192-013-1180-8.

Zerhouni, E., 2003. The NIH roadmap. Science 302 (5642), 63–72. https://doi.org/10.1126/science.1091867.

Xuefeng Wang is a professor in the School of Management and Economics, Beijing Institute of Technology, China. His-specialty is technology innovation management, data mining and science and technology evaluation. His-current research emphasises identifying disruptive technology and forecasting innovation pathways.

Shuo Zhang is a PhD candidate in the School of Management and Economics at Beijing Institute of Technology, China. Her research interests are technology forecasting and innovation management, particularly the study of text mining.

Yuqin Liu is a assistant professor in the School of Journalism and Publication, Beijing Institute of Graphic Communication, China. His-main academic research fields include patent analysis, technology forecasting and innovation management, particularly the study of technology roadmapping.

Jian Du is a assistant professor in the National Institute of Health Data Science, Peking University, China. His-main academic research fields include information science and library science.

Heng Huang is a PhD candidate in the School of Management and Economics at Beijing Institute of Technology, China. His-research interests are innovation management, technology forecasting, and data mining.