

Original Paper

# Understanding Drug Repurposing From the Perspective of Biomedical Entities and Their Evolution: Bibliographic Research Using Aspirin

Xin Li<sup>1,2</sup>, PhD; Justin F Rousseau<sup>3</sup>, MD, MMSc; Ying Ding<sup>4</sup>, PhD; Min Song<sup>5</sup>, PhD; Wei Lu<sup>1</sup>, PhD

<sup>1</sup>Information Retrieval and Knowledge Mining Laboratory, School of Information Management, Wuhan University, Wuhan, China

<sup>2</sup>School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, United States

<sup>3</sup>Department of Population Health and Department of Neurology, Dell Medical School, The University of Texas at Austin, Austin, TX, United States

<sup>4</sup>School of Information, Dell Medical School, The University of Texas Austin, Austin, TX, United States

<sup>5</sup>Department of Library and Information Science, Yonsei University, Seoul, Republic of Korea

**Corresponding Author:**

Wei Lu, PhD

Information Retrieval and Knowledge Mining Laboratory

School of Information Management

Wuhan University

299 Bayi DR, Wuchang District

Wuhan, 430072

China

Phone: 86 02768752757

Email: [weilu@whu.edu.cn](mailto:weilu@whu.edu.cn)

## Abstract

**Background:** Drug development is still a costly and time-consuming process with a low rate of success. Drug repurposing (DR) has attracted significant attention because of its significant advantages over traditional approaches in terms of development time, cost, and safety. Entitymetrics, defined as bibliometric indicators based on biomedical entities (eg, diseases, drugs, and genes) studied in the biomedical literature, make it possible for researchers to measure knowledge evolution and the transfer of drug research.

**Objective:** The purpose of this study was to understand DR from the perspective of biomedical entities (diseases, drugs, and genes) and their evolution.

**Methods:** In the work reported in this paper, we extended the bibliometric indicators of biomedical entities mentioned in PubMed to detect potential patterns of biomedical entities in various phases of drug research and investigate the factors driving DR. We used aspirin (acetylsalicylic acid) as the subject of the study since it can be repurposed for many applications. We propose 4 easy, transparent measures based on entitymetrics to investigate DR for aspirin: Popularity Index ( $P_1$ ), Promising Index ( $P_2$ ), Prestige Index ( $P_3$ ), and Collaboration Index (CI).

**Results:** We found that the maxima of  $P_1$ ,  $P_3$ , and CI are closely associated with the different repurposing phases of aspirin. These metrics enabled us to observe the way in which biomedical entities interacted with the drug during the various phases of DR and to analyze the potential driving factors for DR at the entity level.  $P_1$  and CI were indicative of the dynamic trends of a specific biomedical entity over a long time period, while  $P_2$  was more sensitive to immediate changes.  $P_3$  reflected the early signs of the practical value of biomedical entities and could be valuable for tracking the research frontiers of a drug.

**Conclusions:** In-depth studies of side effects and mechanisms, fierce market competition, and advanced life science technologies are driving factors for DR. This study showcases the way in which researchers can examine the evolution of DR using entitymetrics, an approach that can be valuable for enhancing decision making in the field of drug discovery and development.

(*JMIR Med Inform* 2020;8(6):e16739) doi: [10.2196/16739](https://doi.org/10.2196/16739)

**KEYWORDS**

drug repurposing; biomedical entities; entitymetrics; bibliometrics; aspirin; acetylsalicylic acid

## Introduction

### Background

Despite recent advances in life sciences and technology, drug development is still a costly and time-consuming process with a low rate of success [1]. Discovering a new drug usually takes more than 10 years and costs around \$2 billion on average [2]. The number of targetable human genes is approximately 3000, and the identification of serious and even deadly drug side effects is ongoing [3,4]. To overcome these difficulties, many researchers have turned to drug repurposing, which is the practice of identifying novel clinical indicators for existing marketed drugs [5-7].

The past few decades have produced a few successful cases of drug repurposing. For example, sildenafil, originally developed to treat cardiovascular disease, was unexpectedly discovered to be effective against erectile dysfunction [8]. Thalidomide, once used for morning sickness, has been repurposed for the treatment of multiple myeloma [9], and metformin, originally a treatment for type 2 diabetes, has been studied for the treatment of depression, aging, obesity, and even cancer [10,11]. Beta blockers, initially indicated for hypertension, and topiramate, originally used as an antiepileptic, are both repurposed for migraineurs [12,13]. Because of its significant advantages over traditional approaches, in terms of development time, cost, and previous clinical studies, drug repurposing has attracted significant attention from pharmaceutical firms, scientists, and governments in recent years [7,14].

Methodologies for drug repurposing and their successful applications have been widely discussed. Chen et al [15] designed a system-based algorithm called the reverse gene expression score based on several large-scale publicly accessible datasets and demonstrated the potency and efficacy of vorinostat, geldanamycin, and gemcitabine for the treatment of liver cancers. Xu et al [16] found that emricasan had an inhibitory effect on the Zika virus by screening more than 6000 compounds. With the rapid development of natural language processing and deep learning techniques, robust solutions have recently been proposed and have demonstrated potential. Researchers have integrated more than 20 different datasets into a knowledge graph to predict potential drug and target pairs [17-19]. Hamilton et al [20] queried drug-gene-drug interactions within a low-dimensional embedding of biomedical knowledge graphs to predict missing or unobserved links for drug repurposing. Chang et al [21] proposed a novel deep learning model called “CDRscan” that can successfully predict the feasibility of drug repurposing and recommend the most effective anticancer agents for an individual patient. Öztürk et al [22] represented drugs and protein sequences using convolutional neural networks to predict the binding affinities of drug-target interactions.

Academic publications are produced at high volume, with around 3000 new articles currently published per day [23]. No researcher nor clinician can read and comprehend all the relevant articles in their domain [24]. The “known” knowledge has turned into “unknown known” knowledge, with hidden information and patterns waiting to be discovered. This growing body of

scholarly data opens a new era of exploiting literature and data to enable data-driven discovery [24]. Literature-based discovery, which connects disconnected entities in literature in PubMed, has been successful in identifying several cases of drug repurposing, such as fish oil for Raynaud’s syndrome, magnesium for migraine headaches, and proton pump inhibitors for atrial fibrillation [25-27]. Swanson [26] demonstrated that bibliometrics can be a useful approach to knowledge discovery and recommended that his method could be extended to other disconnected sets of scientific literature to enable cross-disciplinary innovation [28]. With entitymetrics — bibliometric indicators based on entities studied in the medical literature — researchers without domain knowledge can understand the medical function of a drug [29], identify complex undiscovered biological relationships between drugs and targets [30], and detect implicit gene-gene relationships using literature in PubMed [31]. This research demonstrates the potential of applying bibliometrics to medicine to support data-driven discovery. It represents the next generation of bibliometric studies [32] and already shows great promise [33].

### Objectives

In this research, we extended bibliometric indicators for biomedical entities mentioned in the PubMed literature to investigate drug repurposing. We used aspirin (salicylic acid) as the target drug. Aspirin is one of the most well-recognized and well-studied drugs with a history dating back to 1500 BC [34]. It was originally used as an analgesic to treat mild to moderate pain. It has been used clinically for the treatment of at least 10 diseases, including coronary artery disease, cerebrovascular disease, peripheral arterial disease, preeclampsia, diabetes, colorectal cancer, Kawasaki disease, Alzheimer’s disease, and arthritis [34,35]. New indications for aspirin are still being reported [36-38]. Aspirin has a remarkably wide range of effects and therefore provides an ideal case with which to study drug repurposing. The work described in this paper primarily aimed to identify patterns in the different repurposing phases of aspirin by analyzing the diseases, drugs, and genes related to aspirin. We propose 4 measures based on entitymetrics to identify the characteristics and patterns of drug repurposing for aspirin: Popularity Index ( $P_1$ ), Promising Index ( $P_2$ ), Prestige Index ( $P_3$ ), and Collaboration Index (CI).

### Related Work

#### Drug Repurposing

Drug repurposing has become a dynamic emerging field of drug discovery and development. According to Baker et al [39], in 2018 nearly two-thirds of 35,000 drugs or compounds described in MEDLINE were investigated as potential treatments for diseases other than those for which they were originally indicated. Nearly 200 drugs have been investigated for repurposing for more than 300 diseases. Many successfully repurposed drugs were discovered accidentally, such as the application of thalidomide for multiple myeloma [9] and sildenafil for erectile dysfunction [8].

Approaches have been proposed for the generation of hypotheses about novel drug-target interactions and have been used to develop promising directions for subsequent validation of drug

repurposing. In polypharmacology, researchers have proposed 2 types of hypotheses: (1) two drugs could be indicated for the same condition when they produce a similar gene expression profile, and (2) a disease could be one of the indications for a given drug when it has an opposite gene expression profile to that produced by the drug. The Connectivity Map (CMap; Broad Institute, Cambridge, MA), a database for more than 7000 gene-expression profiles of 1309 compounds, has been widely used in this context in previous work. Using a systematic analysis tool, L1000FWD [40], and CMap, Liu et al [41] found that the anticancer drugs KM-00927 and BRD-K75081836 can be used to inhibit histone deacetylase. Kidnapillai et al [42] used gene expression signature data and CMap to identify 10 drugs, including camptothecin, nimesulide, and rescinamine, that could be effective against bipolar disorder.

In the field of genetics, association analysis has been extensively applied to the interactions between drug targets and diseases to increase the efficiency of drug repurposing. One of the most successful cases in the field of drug repurposing was based on a genome-wide association study (GWAS) [43]. Using GWAS-driven methods, Sanseau et al [44] concluded that 15.6% of genes are the targets of marketed drugs. They found that GWAS traits can be matched with the indications of drugs and genes involved in pathogenesis have a high probability of being targets for drug repurposing. Based on a strong association between the gene TNFSF11 and Crohn's disease, the authors inferred, and subsequently confirmed, that dishubzumab, originally developed for the treatment of osteoporosis, can be used against Crohn's disease [44]. Ferrero and Agarwal [45] combined a CMap-based approach with perturbation of transcriptional profiles and disease data from GWAS for target prioritization and drug repurposing. These researchers pointed out that genetic evidence is important in maximizing the success rate of drug repurposing.

These methods in polypharmacology and genetics usually rely on the high-throughput screening of massive amounts of data related to compounds and targets. As knowledge about drug targets accumulates and computational chemistry rapidly develops, simulations of the interactions between drugs and proteins have shown the potential to replace traditional high-throughput screening. Dakshanamurthy et al [46] proposed a proteochemometric method called "train, match, fit, streamline" to conduct molecular docking of over 3000 FDA-approved compounds across the crystal structures of more than 2000 human targets. They found that mebendazole could be used for the inhibition of VEGFR2 kinase and that celecoxib was a promising therapy for malignancies because it binds an adhesion molecule, cadherin-11. Li et al [47] designed a standalone approach to dock over 30 crystal structures of MAPK<sub>14</sub> and BIM-8 with all drugs from DrugBank and found that nilotinib, as a potential inhibitor of MAPK<sub>14</sub>, could be a cure for inflammatory diseases.

Another significant source of drug repurposing is drug side effects. Typical instances of side effect-based drug repurposing include the use of sildenafil for erectile dysfunction [8] and the application of exenatide acetate for obesity [48], both of which were "happy accidents." Recently, Yang and Agarwal [49]

generated human phenotypic profiles for drugs based on over 3000 side-effect relationships extracted from PharmGKB and employed naïve Bayes methods to identify new indications for drugs according to their side effects. This study also suggested that the use of side effects is a type of clinical phenotypic assay and side effects should be rationally investigated to predict repurposing opportunities for drugs. Ye et al [50] contend that drugs with similar side effects could share the same indications because they may have the same or similar mechanisms of action. Using a side effect similarity-based drug-drug network, they transformed drug repurposing into an information retrieval issue and successfully obtained the top 5 indications of 1234 drugs approved by the FDA.

With the rise of machine learning and deep learning in computer science and bioinformatics, the problem of drug repurposing has been addressed using approaches such as classification [51,52], link prediction [53,54], entity prediction [53], and path prediction [18,55]. Liang et al [53] represented biomedical entities and their relationships in a heterogeneous network using graph2vec and knowledge2vec [56] and employed a cascade learning model to find potential interactions between drugs, genes, diseases, and treatments. They found that vitamin D could be a treatment for prostate cancer. Fu et al [55] treated drug repurposing as a binary classification problem and combined the metapath-based topological features of biomedical entities in Chem2Bio2RDF and a supervised machine learning model to predict links between drugs and targets. They found that the intrinsic feature selection Random Forest algorithm can be valuable for selecting significant topological features for the prediction of links between drugs and genes.

### ***Big Scholarly Data for Medical Knowledge Discovery***

Traditionally, knowledge discovery in medical domains has relied on first-hand observation such as epidemiological statistics, follow-ups, and laboratory-generated experimental data [24]. A large number of research papers are published daily, posing significant challenges for scientists wishing to have a comprehensive understanding of their domain [24]. The "known" knowledge has turned into "undiscovered public knowledge," with patterns and information waiting to be uncovered. This large body of literature and data also provides rich opportunities for researchers to undertake data-driven knowledge discovery. The usefulness of literature-based discovery has been demonstrated in many previous research projects. For instance, the "ABC" model proposed by Swanson in 1986 [25] was used to discover relationships between biomedical entities, such as Raynaud's syndrome and fish oil [25], migraine headaches and magnesium [26], and atrial fibrillation and proton pump inhibitors [27]. The "ABC" model is co-occurrence-based and is based upon the premise that seemingly unrelated concepts A and C could be related when there is a concept B related to both A and C [27]. Since Swanson's research, various modifications of the "ABC" model have been proposed to discover hidden relationships among biomedical concepts in PubMed, such as ontology-based entity mapping [57], network-based entity extraction [58], and semantic path-based storytelling [59]. The "ABC" model and its variants indicate that bibliometrics can be a valuable method

for medical knowledge discovery in the era of big scholarly data.

Knowledge graphs of big scholarly data can contain nodes representing biomedical entities such as diseases, drugs, genes, pathways, and cell lines and non-biomedical entities such as authors, institutions, articles, journals, conferences, and keywords. Edges in the graph can represent the relationships between the biomedical entities in the literature. Lv et al [60] established a therapeutic knowledge graph for autism using drug entities and MeSH terms extracted from about 20,000 articles relating to autism published between 1946 and 2015. They proposed a novel topology-driven method incorporating various graph-analytical techniques for drug discovery and concluded that tocilizumab, sulfisoxazole, tacrolimus, and prednisone were promising for the treatment of autism. Ding et al [29] constructed an entity-entity citation graph to highlight the significance of biomedical entities embedded in literature for future knowledge discovery. Researchers have also integrated big scholarly data with other publicly accessible biomedical datasets, such as DrugBank [61], Gene Ontology [62], and SIDER [63], to form a comprehensive knowledge graph for medical knowledge discovery. A typical example is the Chem2Bio2RDF database, created by integrating more than 20 chemogenomic datasets with PubMed. Wang et al [30] proposed a novel algorithm called Bio-LDA to automatically extract latent topics in life sciences and identified relationships and patterns among compounds, genes, and diseases from Chem2Bio2RDF. He et al [64] designed a graph-mining algorithm to predict potential relationships between different biomedical entities. The case they studied demonstrated that the antidiabetic drug rosiglitazone has cardiovascular-related side effects.

Entitymetrics, an entity-driven bibliometric method, and the next generation of citation analysis [29,32] make it possible for researchers without domain knowledge to measure the impact, usage, and transfer of knowledge entities embedded in the academic literature for further knowledge discovery [32]. Ding et al [29] built an entity-entity citation graph based on articles related to metformin and detected most of the known interactions of metformin with biomedical entities. Williams et al [65] recognized and quantified relationships between academic discoveries and major advances in the domain of two new drugs, ipilimumab and ivacaftor, to enhance government support and public understanding. Zhu et al [66] established paper-entity, entity-entity co-occurrence, and entity-specific networks based

on the scientific literature to investigate the evolution of hepatic carcinoma at a granular level. Lv et al [60] discovered new indications for drugs using topology-driven trend analysis of drug-drug and drug-indication networks. These studies demonstrate the potential of the application of bibliometric methods to data-driven discovery in medical domains.

Drug repurposing, as one of the most significant issues in the field of medical knowledge discovery, has been extensively investigated [17,23,24,27,28,55-57,64]. In this research, we extended the bibliometric indicators for biomedical entities described in the PubMed literature to understand the process of drug repurposing.

## Methods

### Data Collection

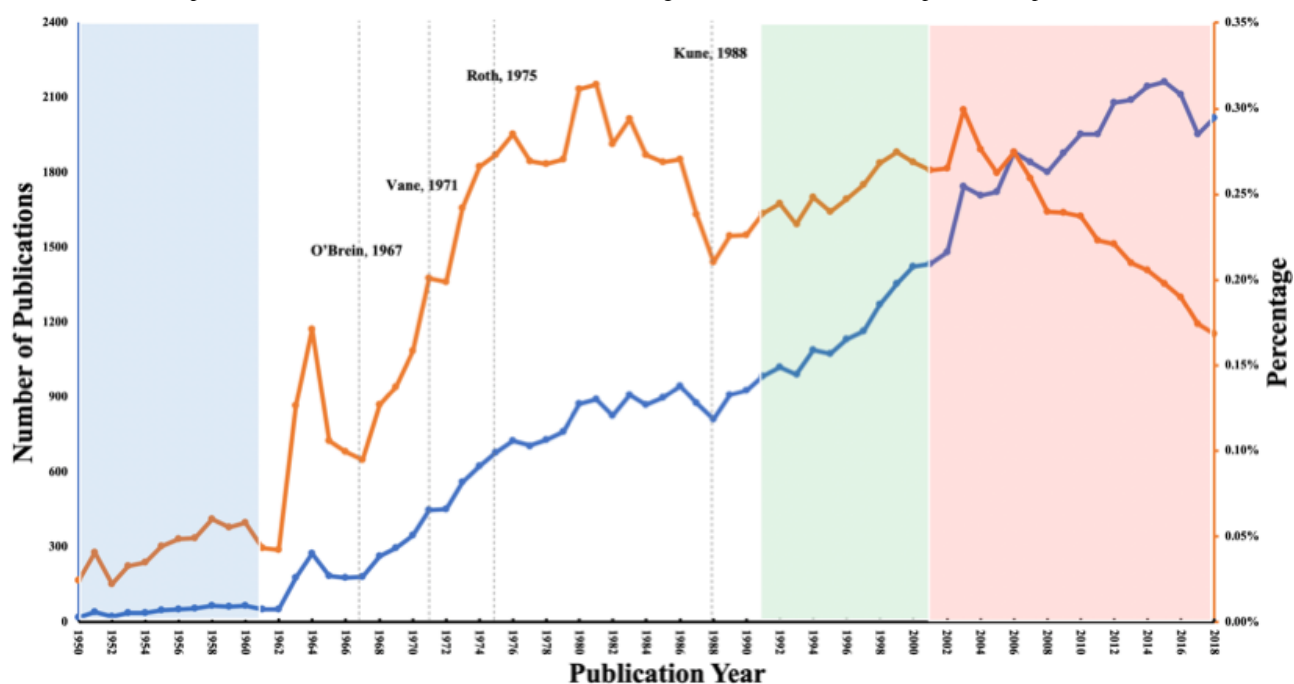
Papers on aspirin-related research published between 1951 and 2018 were collected from PubMed. Since aspirin is known by many names, the search terms were chosen from DrugBank, RxNorm, and MeSH terms [33,61]. The final search query is shown in [Textbox 1](#). Non-journal articles, non-English articles, letters, and editorial commentaries were excluded. In total, 63,387 publications from PubMed were downloaded in XML format.

To better understand the drug repurposing process of aspirin, the relevant research was divided into 4 phases based on previous studies [34,35] and expert suggestions: (1) 1951-1960, the original use; (2) 1961-1990, in-depth studies of pharmacological mechanisms and side effects; (3) 1991-2000, repurposing for cardiovascular diseases; and (4) 2001-2018, repurposing for other diseases, such as colorectal cancer and breast cancer. These phases can also be observed from the evolution and trends of the publications, as shown in [Figure 1](#) and [Table 1](#).

Before extracting biomedical entities, all articles were parsed to obtain PMIDs, publication years, titles, abstracts, authors, journals, and institutions using a dom4j XML parser written in Java. Then, we used spaCy for preprocessing (such as removing the punctuation and stop words) of titles and abstracts in the natural language processing pipeline. In addition, a novel and reliable method of author name disambiguation proposed by Lerchenmueller and Sorenson [67] was used to count distinct authors.

**Textbox 1.** Search query used for retrieving aspirin-related publications.

```
((aspirin) OR ( acetylsalicylic acid) OR (acid, acetylsalicylic) OR (“2-(acetyloxy)benzoic aci”) OR (acylpyrin) OR (aloxiprimum) OR (colfarit) OR (dispril) OR (easprin) OR (ecotrin) OR (endosprin) OR (magnecyl) OR (micristin) OR (polopiri) OR (polopiryna) OR (solprin) OR (solupsan) OR (zorprin) OR (acetysal) OR (2-acetoxybenzenecarboxylic acid) OR (2-acetoxybenzoic acid) OR (acetylsalicylate) OR (acetylsalicylsäure) OR (“acide 2-(acétyloxy)benzoïqu”) OR (acide acétylsalicylique) OR (ácido acetilsalicílico) OR (acidum acetylsalicylicum) OR (aspirina) OR (azetylsalizylsäure) OR (o-acetoxybenzoic acid) OR (o-acetylsalicylic acid) OR (o-carboxyphenyl acetate) OR (salicylic acid acetate) ) AND (“1951”[PDAT] : “2018”[PDAT]))
```

**Figure 1.** Number of aspirin-related studies in PubMed over time. The background colors indicate the 4 phases of aspirin research.**Table 1.** Descriptive statistics of the 4 phases of aspirin research.

Phases, Time span	Number of publications	Number of authors	Average number of authors	Number of journals
<b>1. Original use</b>				
1951-1955	208	318	1.76	117
1956-1960	299	498	1.88	159
1951-1960	507	794	1.83	218
<b>2. In-depth studies of pharmacological mechanisms and side effects</b>				
1961-1965	748	1310	2.01	301
1966-1970	1268	2167	2.12	418
1971-1975	2766	4880	2.40	696
1976-1980	3797	7419	2.71	895
1981-1985	4395	10,011	3.16	1033
1986-1990	4470	11,600	3.50	1101
1961-1990	17,444	31,787	2.90	2153
<b>3. Repurposing for cardiovascular diseases</b>				
1991-1995	5164	14,044	3.69	1256
1996-2000	6353	17,694	4.10	1314
1991-2000	11,517	28,818	3.91	1798
<b>4. Repurposing for other diseases</b>				
2001-2005	8099	27,784	4.22	1719
2006-2010	9366	35,313	4.94	1974
2011-2015	10,436	44,603	5.78	2410
2016-2018	6018	30,796	6.73	1881
2001-2018	33,919	118,857	5.33	3865
Total	63,387	171,559	4.39	5443

## Biomedical Entity Extraction

The biomedical entity extraction module provided by the biomedical entity search tool (BEST) [68], a dictionary-based biomedical information extraction tool based on sophisticated information retrieval approaches, was deployed to extract entities such as diseases, drugs, and genes. The BEST dictionary

is built from 12 different public sources, including NCBI Entrez Gene, DrugBank, T3DB, Animal TFDB, Therapeutic Target DataBase, PubChem, and MeSH [68]. We obtained 1472 unique disease names, 1640 unique drug names, and 3184 unique gene names from the titles and abstracts. Table 2 shows the top 10 biomedical entities of 3 different types and their frequency of appearance in PubMed articles.

**Table 2.** Top 10 biomedical entities in aspirin-related publications during 1951-2018.

Rank	Diseases	Frequency of diseases	Drugs	Frequency of drugs	Genes	Frequency of genes
1	Coronary disease	2707	Clopidogrel	6223	COX-2	3957
2	Asthma	2277	Ticlopidine	5433	CD143	1495
3	Diabetes	1840	Heparin	4391	COX-1	1179
4	Hypersensitivities, drug	1342	Indomethacin	3462	Plasminogen	1131
5	Ulcer, gastric	1146	Warfarin	3457	LDLCQ3	1081
6	Cerebral ischemia	1135	Vitamin F	2760	LPLA2	1047
7	Intracranial vascular disorder	1133	Dipyridamole	2232	GPIIb	1017
8	Ischemic heart disease	1090	Adenosine	2188	P2Y <sub>12</sub>	855
9	Carcinomas, colorectal	1085	Acetaminophen	2099	tPA	748
10	Rheumatoid arthritis	832	Prostacyclin	1498	TNF- $\alpha$	629

## Entitymetric Indicators for Biomedical Entities (P3C)

In order to quantify and visualize the academic importance of individual biomedical entities, 4 transparent and easy entitymetric indexes (P3C) were developed: Popularity Index ( $P_1$ ), Promising Index ( $P_2$ ), Prestige Index ( $P_3$ ), and Collaboration Index (CI). These indicators can be considered as the extensions of the indicators proposed by Kissin and Edwin [33] and Kissin [69] for measuring the academic interest of a drug or technique at the article level. In this study, we adapted the indicators from the perspective of biomedical entities with the goal of understanding drug repurposing. Different from Kissin's indicators, our indicators not only focus on the articles on a given drug but also consider the changes in indicators of biomedical entities (eg, diseases, drugs, and genes) and non-biomedical entities (eg, authors) that are related to the given drug. Detailed explanations of these measures are provided in the following sections.

### Popularity Index ( $P_1$ )

The  $P_1$  of a certain biomedical entity reflects the percentage of publications discussing that biomedical entity among all publications in a research field during a specific period, usually 5 years. The popularity of a biomedical entity  $i$ ,  $P_1(i)$ , is given by:

$$P_1(i) = (N_i / N_T) * 100\% \quad (1)$$

where  $N_i$  is the number of publications relating to an entity  $i$  in a period, and  $N_T$  represents the total number of publications in the research field during the same period. An increase in  $P_1$  indicates growing academic interest in  $i$  in the field.

### Promising Index ( $P_2$ )

The  $P_2$  of a biomedical entity is the change in the popularity of an entity  $i$  in a research field between two continuous periods. The promising index of a specific biomedical entity  $i$ ,  $P_2(i)$ , is expressed as:

$$P_2(i) = (N_i / N_T) - (N_{pi} / N_{pT}) \quad (2)$$

where  $(N_{pi} / N_{pT})$  refers to the popularity of the entity  $i$  in the research field during a previous period of the same length as  $N_i$ .  $P_2$  reflects the change in the academic interest in a biomedical entity in a research field in two time periods. When  $P_2(i) > 0$ , it means the academic interest in  $i$  has increased and vice versa.

### Prestige Index ( $P_3$ )

$P_3$  is defined as the ratio of the number of publications about a specific biomedical entity published in the top journals compared to the number of publications about the same entity in all journals that were indexed by PubMed during the same time period. The prestige of a biomedical entity  $i$ ,  $P_3(i)$ , is calculated as:

$$P_3(i) = (N_{H20} / N_i) * 100\% \quad (3)$$

where  $N_{H20}$  represents the number of publications on  $i$  in the top 20 journals during the same period as  $N_i$ . In this study, the top 20 journals were selected based on the journal impact factor and specialty areas. These journals can be divided into two categories: multidisciplinary journals and specialty journals. Fourteen multidisciplinary journals, including JAMA, The Lancet, BMJ, and similar publications, are common for all diseases, drugs, and genes that were studied in this paper. The other 6 journals, such as Circulation, Blood, and The European Heart Journal, are highly associated with aspirin-related

specialty areas. The full list of the top 20 journals is shown in [Multimedia Appendix 1](#).  $P_3$  reflects the potential significance of a specific biomedical entity. Continuing high prestige scores could be an early signal of the success of entity-related drug discovery or repurposing [69]. We employed a threshold of 5% to indicate that  $P_3$  was of interest [69].

### Collaboration Index (CI)

The CI of a biomedical entity reflects the percentage of the number of distinct authors of articles discussing this entity out of all the distinct authors in the research domain over a period of time. The CI of a biomedical entity  $i$ ,  $CI(i)$ , is calculated by:

$$CI(i) = (N_{AI} / N_{AT}) * 100\% \quad (4)$$

where  $N_{AI}$  is the number of distinct authors of the publications relating to  $i$  in a period, and  $N_{AT}$  represents the total number of distinct authors in the field in the same period. The CI reflects the research strength of entity  $i$  in a research field, and a threshold of 5% indicates a level of interest [69].

## Results

### Overview of Aspirin-Related Studies

[Figure 1](#) shows an overview of aspirin-related research in PubMed from 1951 to 2018. The red and blue lines represent the percentage and absolute numbers, respectively, of articles in PubMed per year. The details of the publications, authors, and journals are shown in [Table 1](#). During the evolution of aspirin, Phase 1 (1951-1960) produced 507 articles, most of which were published in journals covering pharmacy-related or general medicine-related topics ([Table 1](#) and [Multimedia Appendix 2](#)). Research in Phase I focused on the anti-inflammatory and antipyretic uses of aspirin, and this phase marks the original use of aspirin.

In Phase 2 (1961-1990), a turning point can be identified in 1967, after which the number of relevant papers per year grew dramatically until 1986. Several significant pharmacological discoveries related to aspirin occurred during this period, including the antiplatelet effect [70], mechanism of inhibition of prostaglandin synthesis [71], and acetylation of the cyclo-oxygenase enzyme [72]. The percentage of aspirin-related articles in PubMed reached its peak in 1981, at about 0.32%, and then decreased. Kune et al [73] reported that aspirin could effectively reduce the incidence of colorectal cancer, after which the percentage began to rise again. After 1975, articles began to occur frequently in journals covering specialty areas, such as *Circulation* and *Thrombosis Research*. We identify this phase as the in-depth investigation of the pharmacological mechanisms and side effects of aspirin.

In Phase 3 (1991-2000), there was a steady and stable growth in the number and percentage of aspirin-related articles per year in PubMed ([Figure 1](#)). Compared to the first 10 years (1951-1960), there was a >22-fold increase in the number of articles as well as a >36-fold increase in the number of distinct authors. As shown in [Multimedia Appendix 2](#), in both 1991-1995 and 1996-2000, 4 of the top 5 journals were cardiovascular-related journals. We thus identify this phase as repurposing for cardiovascular diseases.

In Phase 4 (2001-2018), the number of articles per year grew continuously and reached its peak (2164) in 2015, but the percentage significantly reduced ([Figure 1](#)). From the information presented in [Table 1](#), we note that the numbers of articles, distinct authors, and journals in Phase 4 were all higher than those in the previous 3 periods. The average number of authors in this period had exceeded the total average (4.39). Journals covering other topics, for example *Cancer Management and Research*, *Drugs & Aging*, and *World Neurosurgery*, were increasingly represented ([Multimedia Appendix 2](#)), demonstrating that aspirin had been experimentally applied to many other diseases. We thus mark this phase as repurposing for other diseases.

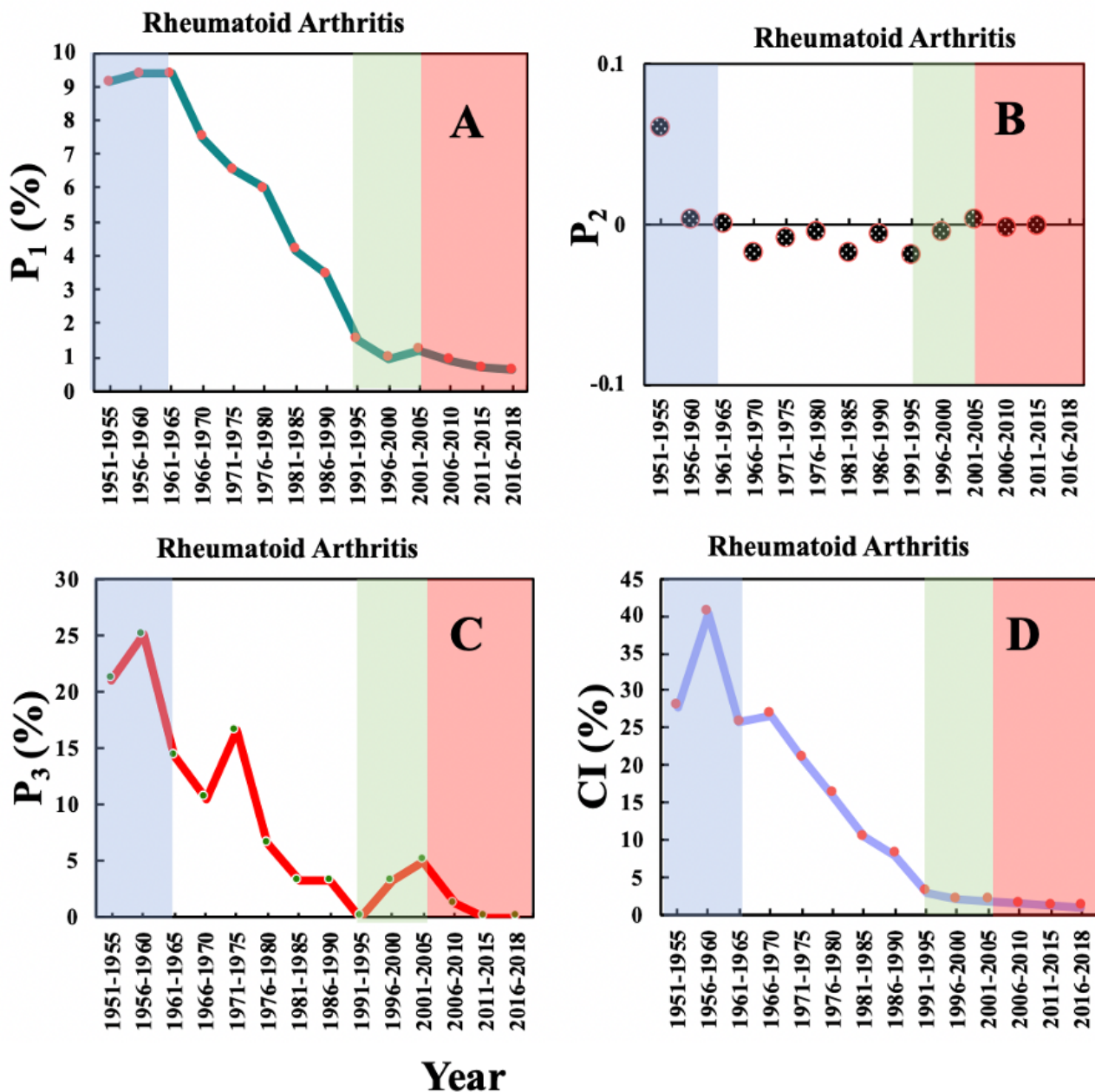
To analyze drug repurposing through all 4 phases from the biomedical entity perspective, we first computed the P3C indicators of the top 10 diseases, drugs, and genes in the cohort of aspirin articles during the period 1951-2018. The results show that there are distinct patterns of these indicators in different repurposing phases. To describe these patterns in detail, we reorganized the 30 biomedical entities (the top 10 diseases, top 10 drugs, and top 10 genes) into the 4 phases of aspirin research, according to when each achieved its maximum  $P_1$ , which indicates the focus of research in the field of aspirin. In each phase, we further analyzed the change patterns of the P3C indicators for the most popular biomedical entities, to investigate the features of different phases of drug repurposing, association between entities and P3C indicators, and possible factors driving drug repurposing at the biomedical entity level.

### Before Repurposing

Only “rheumatoid arthritis” (RA) reached its maximum  $P_1$  in Phase 1, at 9.36%, as shown in [Figure 2A](#) and then exhibited a downhill trend for the rest of the 3 phases, reaching a low of 0.63% in 2016-2018. As shown in [Figure 2B](#), for the  $P_2$  of RA, there is only one significant increase of more than 0 in all 4 phases: 0.06 in 1951-1955 (Phase 1). This observation indicates that the popularity of RA in 1951-1955 increased by 6% compared to that in 1945-1950. It can also be observed from [Figure 2C](#) that the  $P_3$  of RA was more than 5% during 1951-1980 and reached its maximum in Phase 1 (25%, 1960-1965), indicating that one quarter of the papers studying RA were published in the top 20 journals in the aspirin domain in Phase 1. In the next 3 phases, the  $P_3$  peaked twice, in Phase 2 (1971-1975) and Phase 3 (2001-2005), possibly relating to the discovery of the mechanism of anti-inflammatory and RA-induced cardiovascular diseases. Similar to  $P_1$ , as shown in [Figure 2D](#), the CI of RA peaked in 1956-1960 (40.44%), then declined to 1.02% in 2016-2018, indicating that around 40% of authors in Phase 1 were studying RA, but only about 1.02% authors still worked on the same disease in Phase 4.

In summary, in Phase 1, the  $P_1$ ,  $P_2$ ,  $P_3$ , and CI of RA reached their maxima, or showed a significant increase, indicating that RA was the disease upon which most research was focused in the aspirin domain at this time. However, the value of these indicators showed profound declines in the next 3 phases, which means that aspirin was studied in relation to other diseases and is thus an ideal example of drug repurposing.

**Figure 2.** The 4 entitymetric indexes of the biomedical entity “Rheumatoid Arthritis” over time. The background colors indicate the 4 phases of aspirin research.



**Scientific Basis for Repurposing**

As shown in Figure 3, there are 9 top biomedical entities in the aspirin domain that reached their maximum  $P_1$  in Phase 2, including 3 diseases (“asthma”; “hypersensitivities, drug”; and “ulcer, gastric”) and 6 drugs (indomethacin, acetaminophen, dipyridamole, vitamin F, adenosine, and prostacyclin). The 3 diseases can all be side effects of aspirin, while the 6 drugs can be divided into 3 categories: (1) competitors of aspirin, that is, indomethacin and acetaminophen, which are analgesic and

antipyretic drugs, respectively, with fewer side effects; (2) the antiplatelet drug dipyridamole; and (3) precursor substances in the pathway of the mechanism of action of aspirin (vitamin F, adenosine, and prostacyclin). In contrast with RA, the  $P_1$  of these biomedical entities increased from Phase 1, peaked in Phase 2, and then decreased, indicating that the side effects and mechanisms of aspirin were studied in detail in Phase 2. The  $P_1$  of indomethacin in 1976-1980 (16.75%) was the highest among these 9 entities in Phase 2, and vitamin F in 1981-1985 (11.19%) ranked second.



**Figure 3.** The Popularity Index ( $P_1$ ) of the biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the 4 phases of aspirin research.

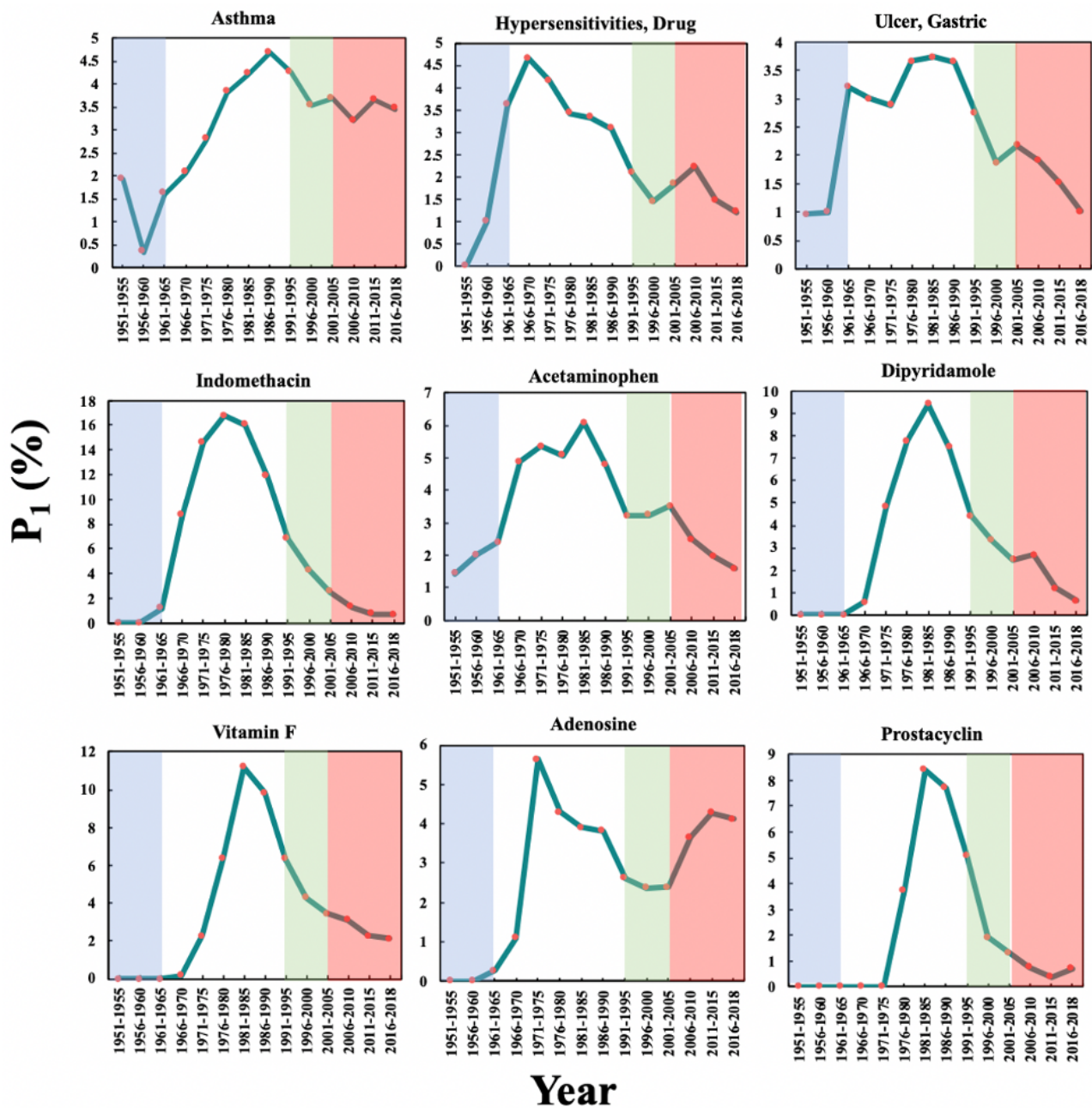


Figure 4 shows the  $P_2$  of these 9 biomedical entities in the aspirin domain over time. The  $P_2$  of the 3 side effects had a significant increase of more than zero in Phase 2, indicating that interest in the side effects of aspirin increased sharply: 1961-1965 and 1976-1980, for “asthma”; 1961-1965 for “hypersensitivities, drug”; and 1961-1965 for “ulcer, gastric.” The time periods in which the  $P_2$  of the 6 drugs showed significant increases are generally later than those for the side effects, such as 1971-1975 for indomethacin and 1981-1985 for prostacyclin. This observation indicates that the discovery and in-depth study of side effects may have positive effects on the discovery of the mechanism of action of aspirin as well as the development of alternatives with fewer side effects.

Figure 5 shows the  $P_3$  of these 9 biomedical entities in the aspirin domain, demonstrating a feature common to all 9 entities:

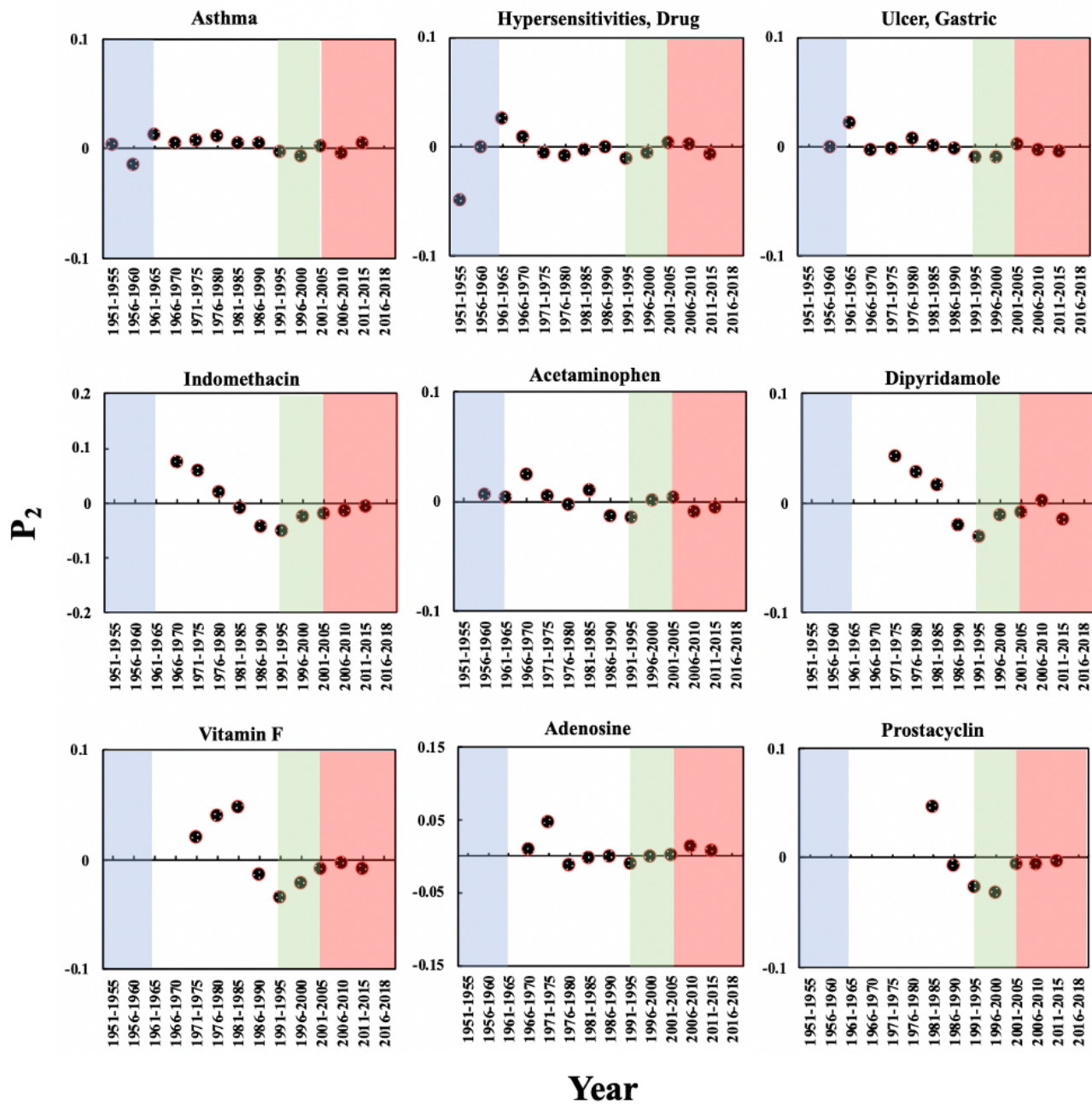
a gradual decline with a fluctuation in  $P_3$  after reaching a maximum in Phase 1 or Phase 2. The highest initial  $P_3$  values of “hypersensitivities, drug” and “ulcer, gastric” occurred in Phase 1, revealing that both side effects had been taken seriously by researchers in Phase 1. The  $P_3$  of “hypersensitivities, drug” in 1956-1960 (33.33%) was higher than that of RA in 1956-1960 (25.00%). In 2011-2015, the  $P_3$  of only 2 entities are over the 5% threshold: 5.82% for adenosine and 10.00% for prostacyclin. In the aspirin domain, papers studying these 2 entities published in the top 20 journals comprised more than 5% of papers published in all of the journals indexed by the PubMed in 2011-2015. This observation indicates that the 2 entities were still important foci of research in the aspirin domain.

It can be observed from Figure 3, Figure 5, and Table 3 that  $P_3$ , on average, achieved its maxima 10.7 years earlier than  $P_1$ . In

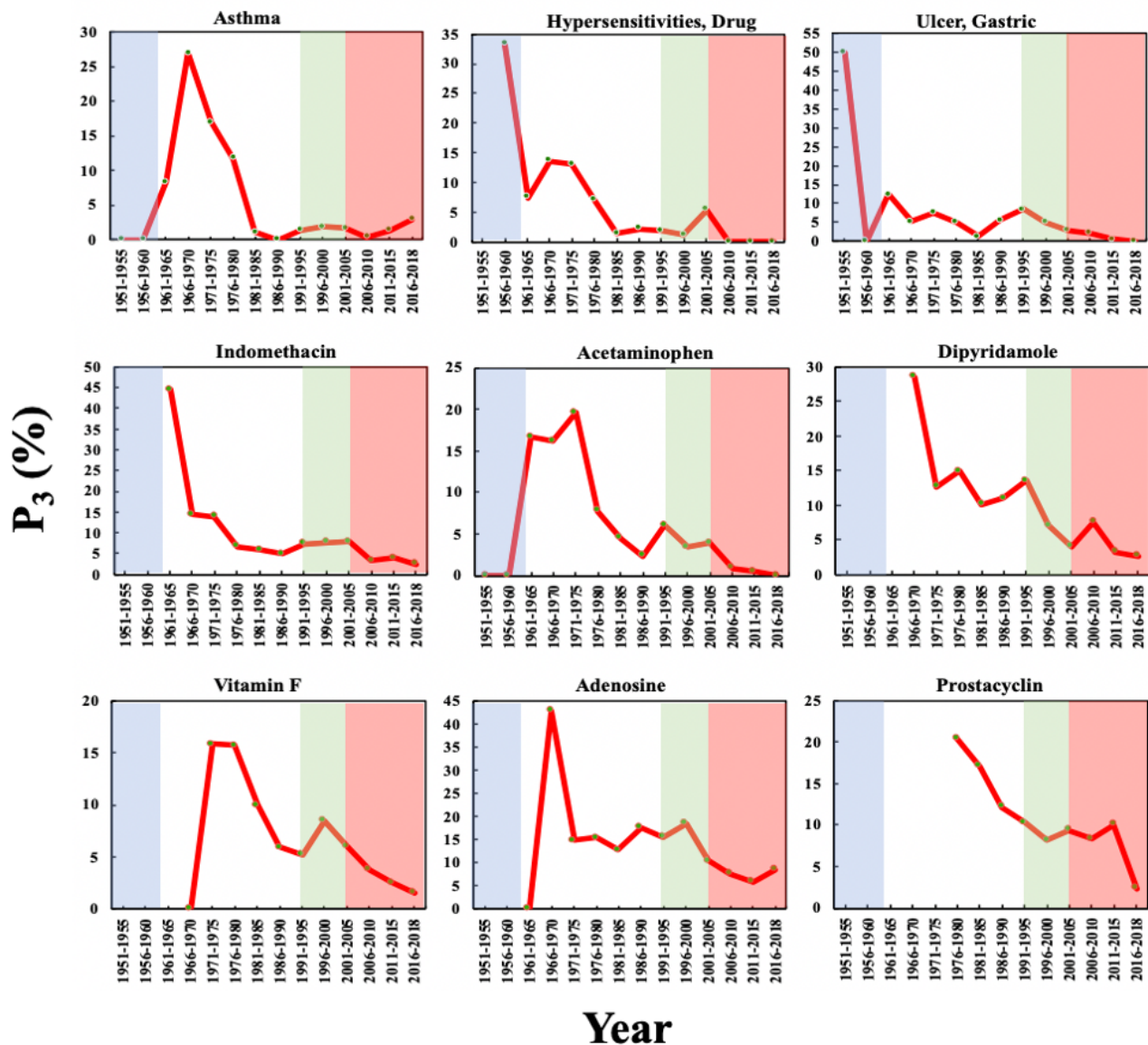
particular, for “hypersensitivities, drug” and “ulcer, gastric,” the intervals can be as long as 20 years. This observation indicates that  $P_3$  can reflect an early sign of academic interest

into biomedical entities, a phenomenon that could be potentially valuable for tracking the research frontiers of a drug.

**Figure 4.** The Promising Index ( $P_2$ ) of the biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the 4 phases of aspirin research.



**Figure 5.** The Prestige Index ( $P_3$ ) of the biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the 4 phases of aspirin research.



**Table 3.** Intervals between the time periods of the maxima of  $P_1$  and  $P_3$ .

Biomedical entity	Time period of the maximum of $P_1$ (T1)	Time period of the maximum of $P_3$ (T2)	T1-T2 (years)
Asthma	1986-1990	1966-1970	20
Hypersensitivities, drug	1966-1970	1956-1960	10
Ulcer, gastric	1976-1980	1956-1960	20
Indomethacin	1976-1980	1961-1965	15
Acetaminophen	1981-1985	1971-1975	10
Dipyridamole	1981-1985	1966-1970	15
Vitamin F	1981-1985	1971-1975	10
Adenosine	1971-1975	1966-1970	5
Prostacyclin	1981-1985	1976-1980	5

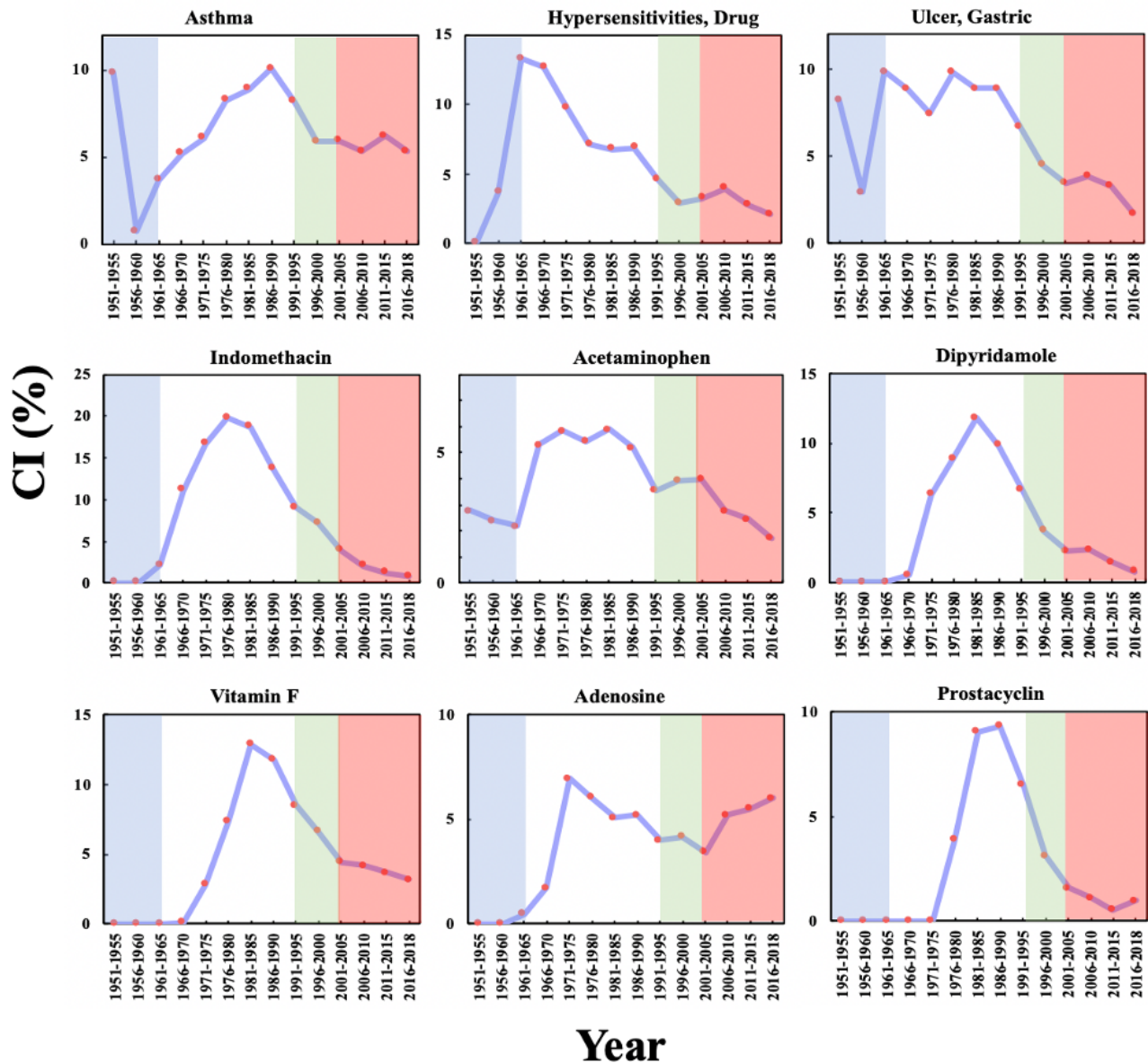
The results of the CI of these 9 biomedical entities in the aspirin domain are presented in Figure 6, which shows that the CIs for these biomedical entities have similar trends to those of  $P_1$  over

time. Among all 9 biomedical entities during 1951-2018, indomethacin achieved the highest maximum CI in 1976-1980 (19.79%), indicating that it became a strong competitor to aspirin

as an analgesic agent in Phase 2. This result also demonstrates that during the last 5-year period (2011-2015), the CIs of only 2 of the 9 entities were >5%, indicating that the 2 entities were still the subject of research by a considerable number of scientists (>2230) in the aspirin research community in 2011-2015. The 2 biomedical entities include “asthma” (6.21%) and adenosine (5.50%).

Based on the observation of P3C in Phase 2 and previous studies on aspirin [34,35], we can conclude that, on one hand, the in-depth investigation of the side effects and mechanism of action of aspirin provided the knowledge basis and research direction for drug repurposing. On the other hand, due to the market competition from other drugs, as well as the serious side effects, pharmaceutical companies attempted to discover new indicators for aspirin, in order to maintain the sales volume of aspirin.

**Figure 6.** Collaboration Index (CI) of the biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the 4 phases of aspirin research.



### Repurposing Aspirin for Cardiovascular-Related Diseases

In Phase 3, 5 top biomedical entities comprising 4 diseases and 1 drug reached their maximum  $P_1$ , as shown in Figure 7A. The 4 diseases were all cardiovascular-related, including “coronary disease” ( $P_1$  of 18.88% in 1996-2000), “cerebral ischemia” ( $P_1$  of 2.57% in 1996-2000), “intracranial vascular disorder” ( $P_1$  of 5.73% in 1991-1995), and “ischemic heart disease” ( $P_1$  of 3.01%

in 1996-2000). Compared with Figures 2 and 3, the  $P_1$  of the previous 10 biomedical entities that peaked in the Phase 1 or Phase 2 were considerably lower than that of coronary disease, indicating that cardiovascular-related disease was the focus of the aspirin domain in that time. Coronary disease is often referred as ischemic heart disease and is the most common cardiovascular-related disease worldwide; similarly, cerebral ischemia and intracranial vascular disorder represent the same condition, commonly known as stroke. These conditions were reportedly the first and second most common causes of death

worldwide in the early 21st century [74]. The demand for the prevention and treatment of such fatal diseases could be one of the factors driving the repurposing of aspirin for cardiovascular-related diseases.

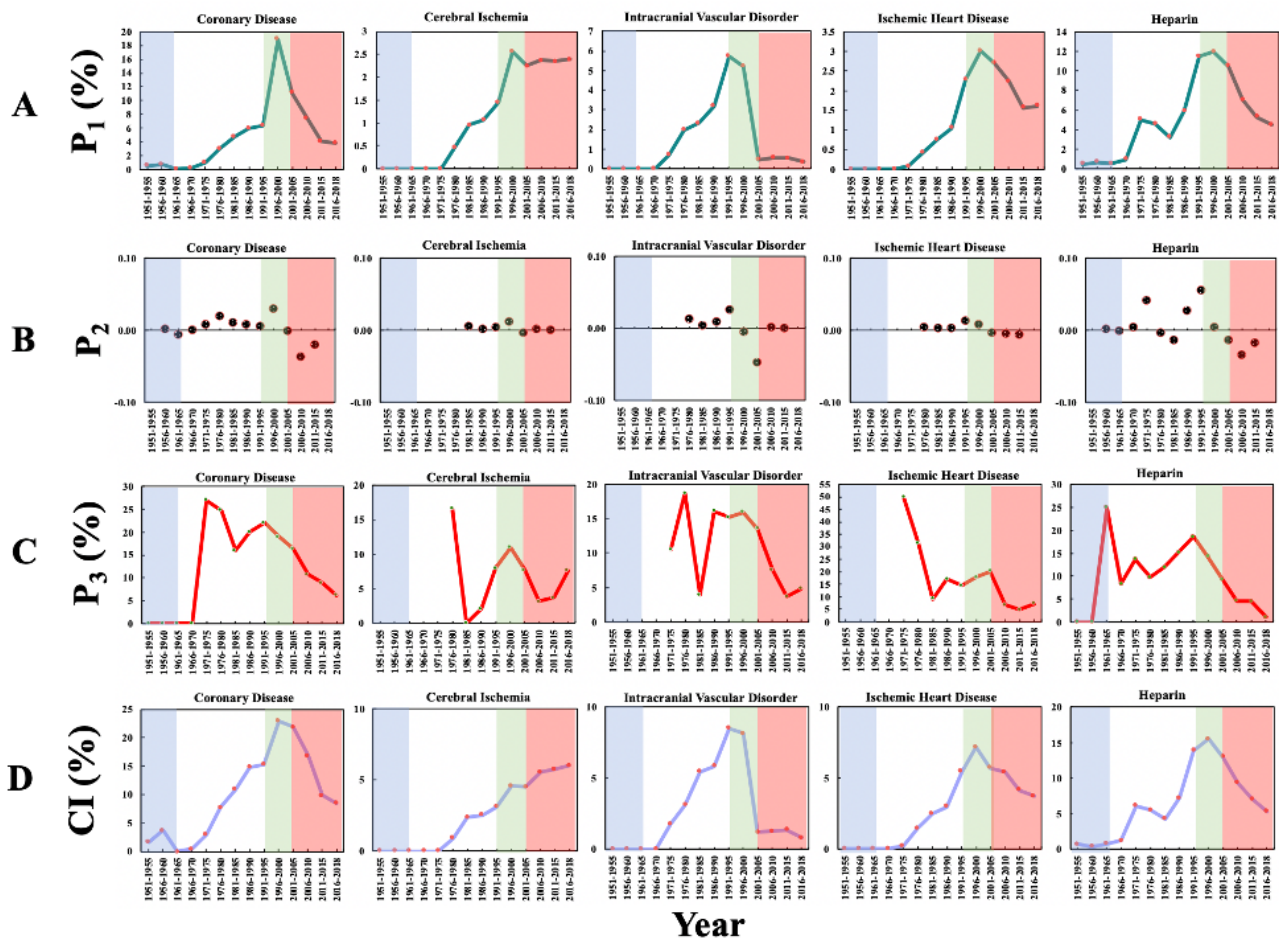
The only drug that reached its maximum  $P_1$  in Phase 3 is heparin (11.92% in 1996-2000). As one of the most common anticoagulant drugs, heparin has always been the reference drug for repurposing aspirin to treat cardiovascular-related diseases, which could be the reason for the increase in the academic interest in heparin in the aspirin domain. There was another peak of heparin in Phase 2 (5.03%, 1971-1975), which could be related to an increase in research into the mechanisms of the antiplatelet effect of aspirin in Phase 2.

Figure 7B shows the changes in  $P_2$  of these 5 biomedical entities over time. All 5 biomedical entities demonstrated a significant increase in Phase 3. “Coronary disease” and “cerebral ischemia” increased in 1991-1995, and “intracranial vascular disorder”, “ischemic heart disease,” and heparin increased in 1991-1995. The  $P_2$  of the 2 entities also showed significant increases in Phase 2, consistent with the fact that aspirin was clinically used for coronary disease before the discovery of its antiplatelet effect: 0.02 in 1976-1980 for “coronary disease” and 0.10 in 1971-1975 for heparin.

The pattern of  $P_3$  for these 5 entities over time is displayed in Figure 7C. All 5 biomedical entities reached their maxima in Phase 2, earlier than the maximum of  $P_1$ . “Coronary disease” reached a maximum in 1971-1975, and heparin reached a maximum in 1961-1965. The difference from the previous phases is that the  $P_3$  of these 5 biomedical entities peaked again in Phase 3. For instance, “coronary disease” peaked in 1991-1995, and heparin peaked in 1991-1995, indicating that these biomedical entities were important topics of research in both Phase 1 and Phase 3.

Figure 7D shows the CI of the 5 biomedical entities during 1951-2018, in which the CI demonstrated a dynamic trajectory very similar to that of  $P_1$ . The maximum of “coronary disease” in Phase 3 is highest at 22.91% in 1996-2000, indicating that “coronary disease” attracted the greatest share of the authors in the aspirin domain. “Coronary disease” and “cerebral ischemia” in Phase 4 and heparin in Phases 2 and 4 surpassed the threshold value of 5%. The CI of “cerebral ischemia” steadily grew after Phase 3, showing a different trend from the other 4 biomedical entities, which increased in Phase 1 and Phase 2, peaked in Phase 3, and then dramatically decreased. This observation may illustrate that “cerebral ischemia,” unlike the other biomedical entities, is still increasing in popularity and collaboration, so additional increases are still expected.

Figure 7. The 4 entimetric indexes of the biomedical entities on cardiovascular diseases in the aspirin domain over time. The background colors show the 4 phases of aspirin research.



### Repurposing Aspirin for Other Diseases

In Figure 8, there are 15 biomedical entities that reached their maximum  $P_1$  in Phase 4. Unlike the previous phases, most of the biomedical entities were genes and can be divided into 3 categories according to the diseases to which they are related: (1) inflammatory-related genes (eg, COX-2, LPLA2, and TNF- $\alpha$ ), (2) cardiovascular-related genes (eg, COX-1, CD143, plasminogen, LDLCQ3, GPIIb, P2Y<sub>12</sub>, and tPA), and (3) cancer-related genes (eg, TNFa, COX-2, COX-1, and LPLA2). These observations indicate that aspirin was actively studied for these 3 aspects of diseases from the perspective of genes in Phase 4. In particular, the maximum  $P_1$  of COX-2 was the highest among these 15 biomedical entities at 21.97% in 2001-2005, revealing that COX-2 was considered to be very important in the aspirin domain at that time.

Figure 8 also shows that the  $P_1$  of 2 diseases peaked in Phase 4. One is “diabetes,” whose  $P_1$  in 2006-2010 was 6.83%. In fact, as early as 1875, Ebstein and Müller [75] discovered that

aspirin had the effect of lowering blood glucose levels. Inspired by this observation, scientists have since been trying to use aspirin for the treatment of diabetes [75]. There are several peaks in the  $P_1$  of “diabetes” in previous phases. In the 21st century, it has been recommended that patients with diabetes who have an increased risk of cardiovascular disease take aspirin as a primary preventative [5,76]; this could be the reason why the academic interest in “diabetes” in the aspirin domain increased again. The other disease is “carcinomas, colorectal.” Its  $P_1$  peaked in 2001-2005 and then increased significantly after a small decline in 2006-2010, a pattern which is very different from other diseases in the aspirin domain. Repurposing aspirin for the treatment of colorectal carcinomas appears to be a focus of research in the aspirin domain today. The  $P_1$  of 3 drugs also peaked in Phase 4, including the antiplatelet drugs clopidogrel and ticlopidine, which are competitors of aspirin as antiplatelet drugs [35], and warfarin, which is an anticoagulation drug that is similar to heparin and has been found to be superior to aspirin for secondary prevention of ischemic stroke with nonvalvular atrial fibrillation [77,78].

Figure 8. The Popularity Index ( $P_1$ ) of the biomedical entities on repurposing aspirin for other diseases over time. The background colors show the 4 phases of aspirin research.

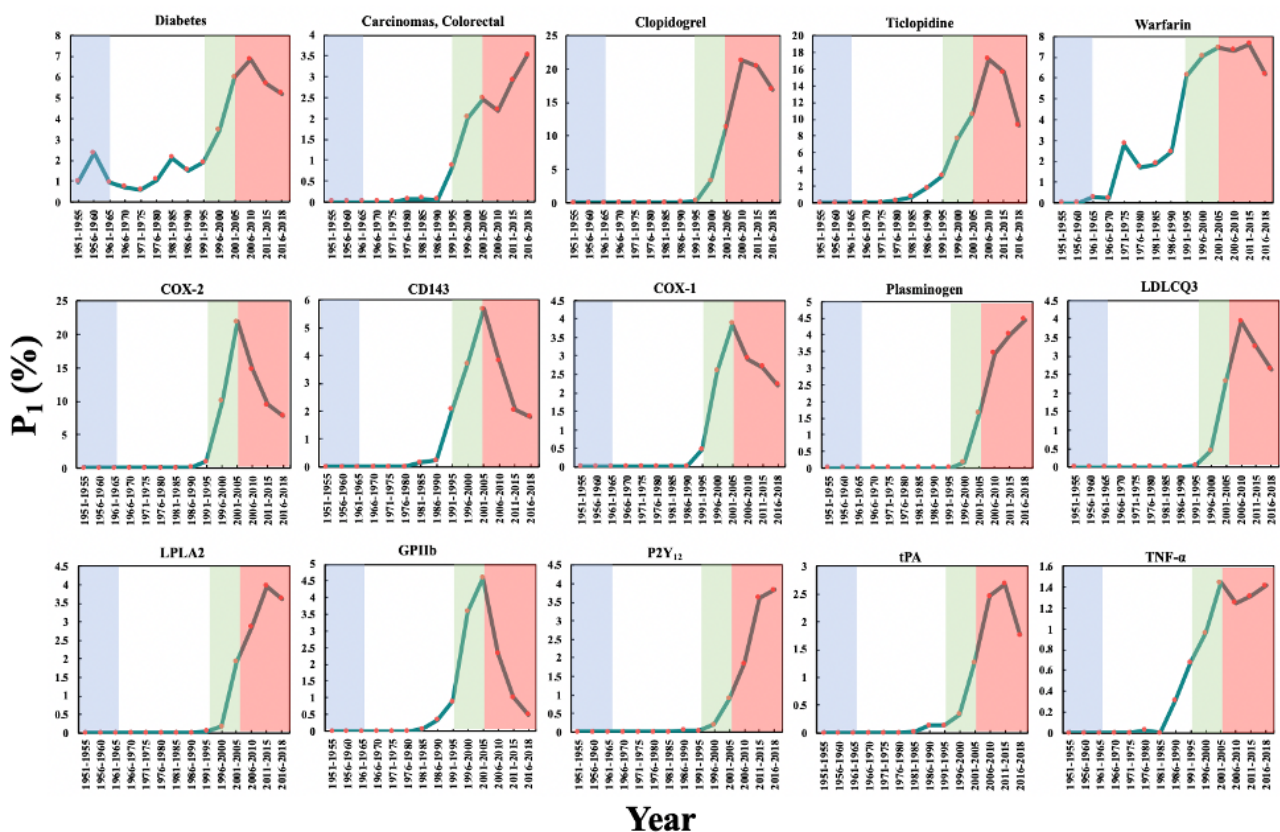


Figure 9 presents the changes in  $P_2$  of these 15 biomedical entities over time. All of the genes demonstrate an increase of more than 0 in Phase 4. Unlike these genes, the diseases and drugs showed several significant increases of more than 0 in different phases, which reflects a longer history of research in the aspirin domain. For example, the increases occurred in 1956-1960, 1996-2000, and 2001-2005 for “diabetes”; 1996-2000, 2001-2005, and 2006-2010 for clopidogrel; and 1971-1975 and 1991-1995 for warfarin.

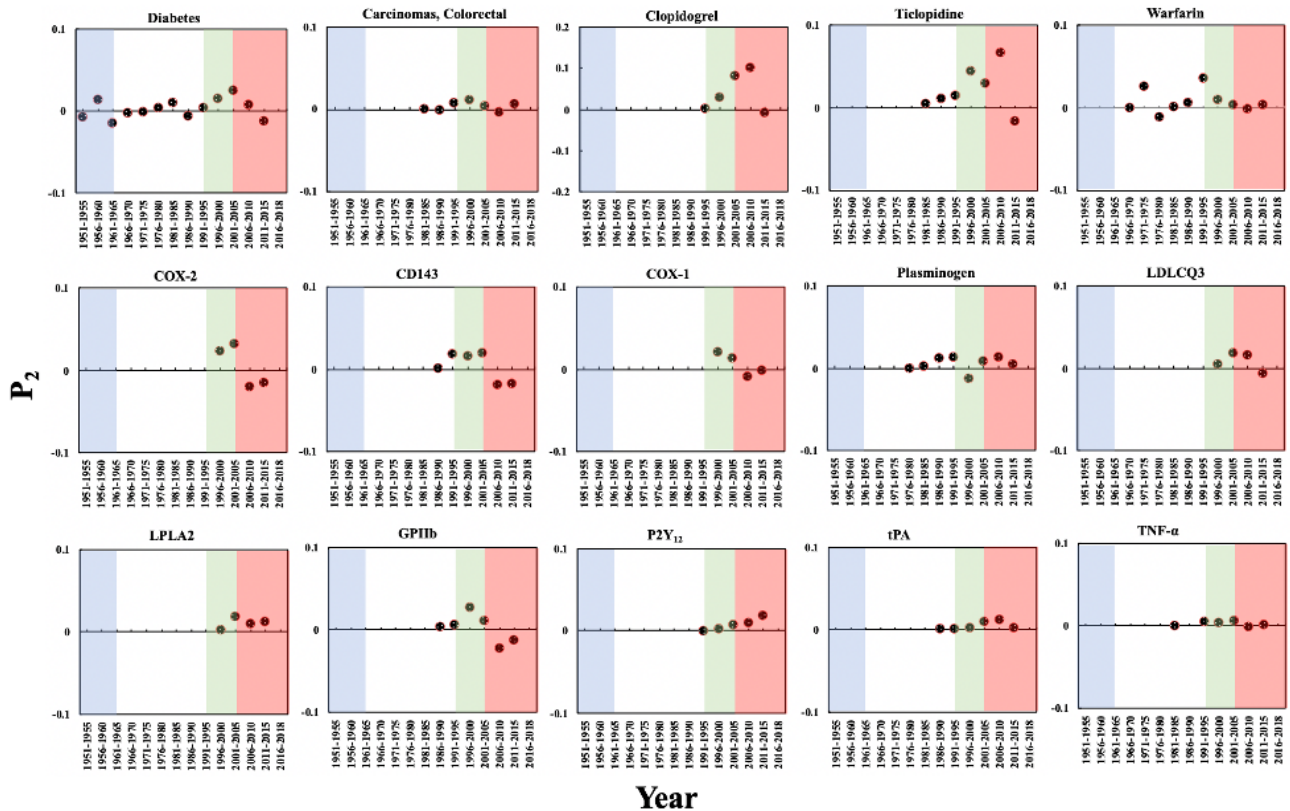
The changes in  $P_3$  of these 15 biomedical entities over time are shown in Figure 10, from which we can make two observations. First, the  $P_3$  of these biomedical entities demonstrated that the time period of the maximum of  $P_3$  was much earlier than that of the maximum of  $P_1$ . Second, unlike the biomedical entities noted in previous sections, the diseases and drugs had  $\geq 2$  significant peaks in different phases. For instance, “diabetes” had peaks of 42.86% in 1956-1960, 25.00% in 1971-1975, and 14.22% in 1996-2000, and “carcinomas, colorectal” had peaks

of 33.33% in 1981-1985, 15.91% in 1991-1995, and 14.15% in 2006-2010. These numbers indicate that these entities attracted considerable interest in the field of aspirin research and high-impact papers on these conditions were published. However, the genes usually had only one peak in  $P_3$  in Phase 3 or 4, illustrating that these genes are relatively new topics in the aspirin domain.

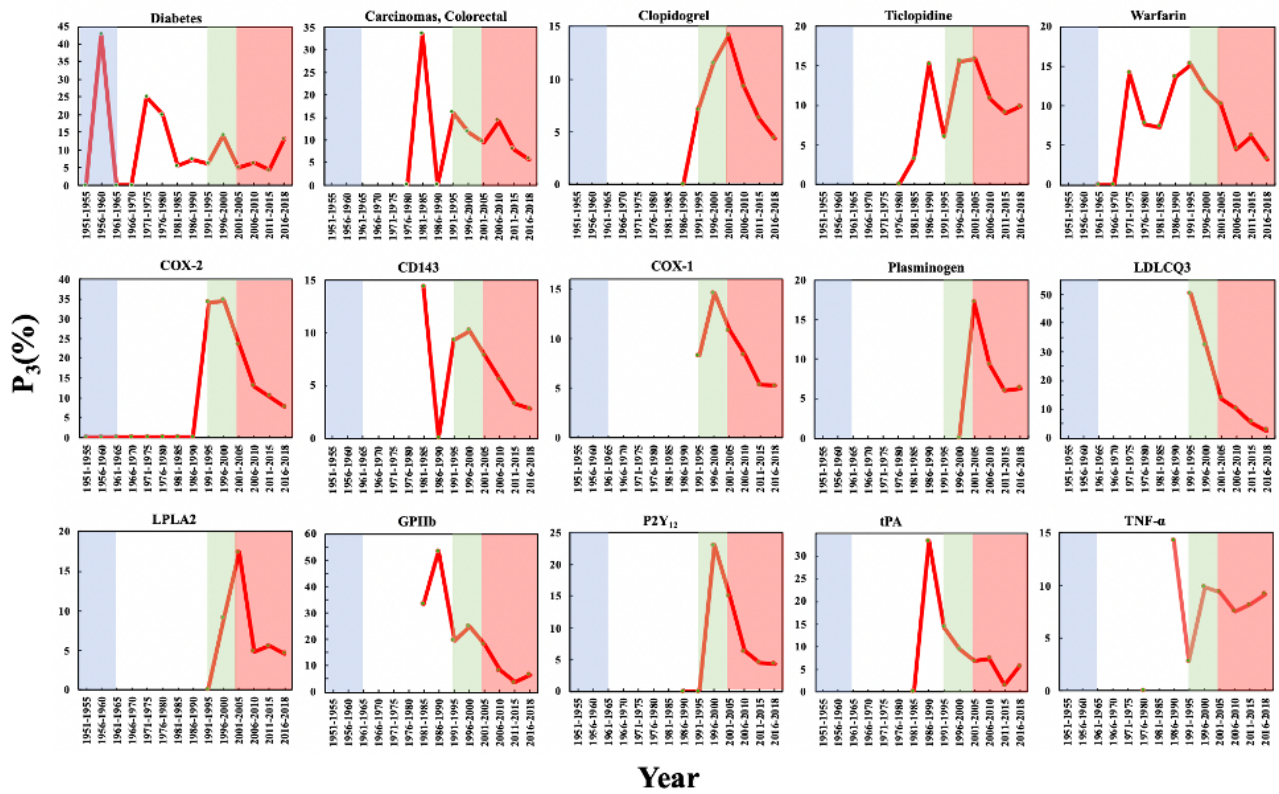
The CI data for these 15 biomedical entities are presented in Figure 11, which shows that the maximum CI for COX-2 is the

highest, at 34.37%, in 2001-2015, denoting that COX-2 was the focus of aspirin research in Phase 4; the research and development of Vioxx, a selective COX-2 inhibitor with fewer side effects, may be one of the reasons [79]. The CI of 2 drugs, clopidogrel (25.54% in 2001-2015) and ticlopidine (20.74% in 2006-2010), reveals fierce competition between aspirin and these alternative antiplatelet drugs. This competition could have driven the repurposing of aspirin for other diseases, especially cancers, that have an urgent demand for effective treatment.

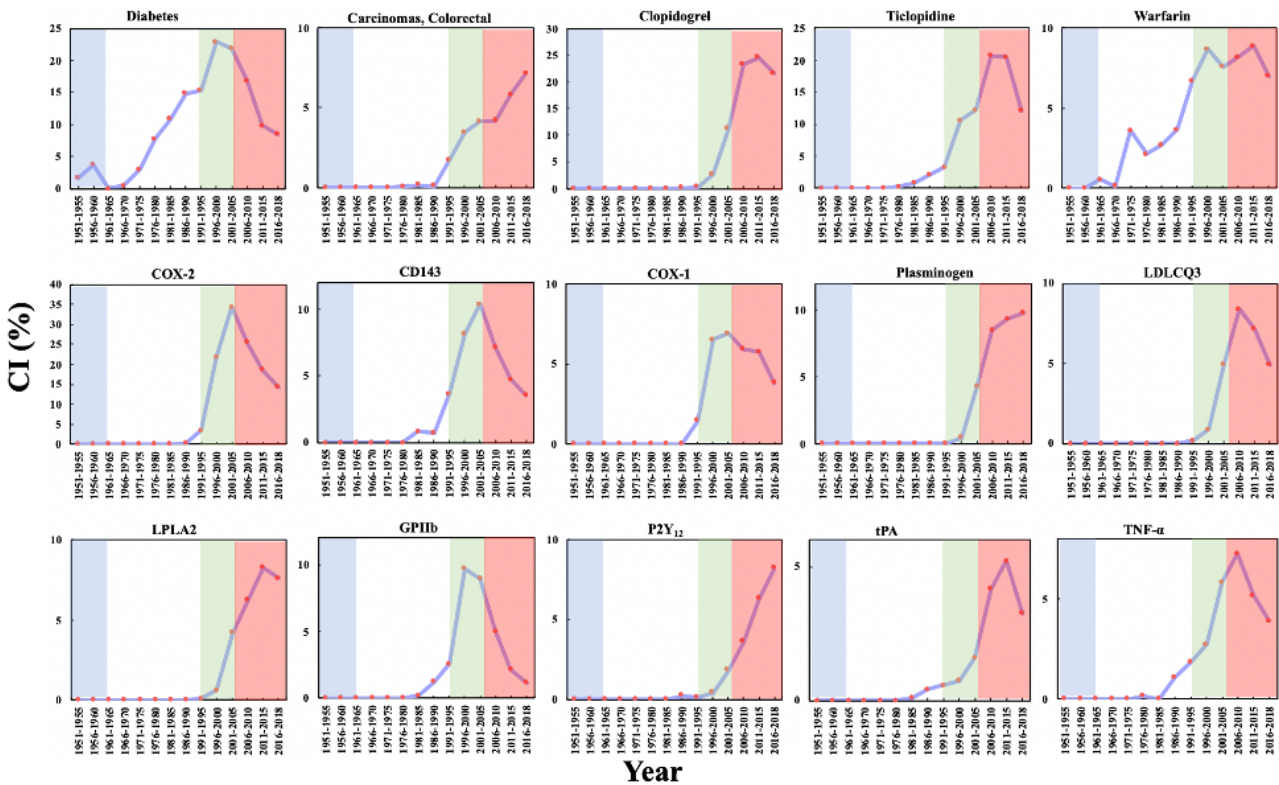
**Figure 9.** The Promising Index ( $P_2$ ) of the biomedical entities on repurposing aspirin for other diseases over time. The background colors show the 4 phases of aspirin research.



**Figure 10.** The Prestige Index ( $P_3$ ) of the biomedical entities on repurposing aspirin for other diseases over time. The background colors show the 4 phases of aspirin research.



**Figure 11.** The Collaboration Index (CI) of the biomedical entities on repurposing aspirin for other diseases over time. The background colors show the 4 phases of aspirin research.





## Discussion

### Principal Findings

This study examines drug repurposing from the perspective of the evolution of biomedical entities, using aspirin as the study subject. It is of paramount importance for drug discovery to identify the factors that drive repurposing as well as potential patterns among biomedical entities in various phases of the drug research timeline. The main contribution of this paper is twofold. First, we proposed 4 entitymetric indices (P3C) to quantify changes in academic interest in biomedical entities and to reveal the granular process of drug repurposing. Second, we divided aspirin research into 4 phases, including original use (1951-1960), in-depth studies of pharmacological mechanisms and side effects (1961-1990), repurposing for cardiovascular-related diseases (1991-2000), and repurposing for other diseases (2001-2018), taking into consideration 3 granular perspectives—disease, drug, and gene—that contribute to a comprehensive understanding of the features of the repurposing process.

Our entitymetric results indicate that aspirin is representative of the process of drug repurposing. The research findings can be summarized as follows. In Phase 1, aspirin was routinely used to ease pain, fever, and inflammation and was often used in the treatment of RA [34], with a P3C that peaked in 1951-1960. Despite the widespread use of aspirin, at this stage, its mechanism of action was not well understood [34]. In Phase 2, the side effects and mechanisms of actions of aspirin were studied extensively, as shown by the maxima of  $P_1$  and CI, as well as a significant increase in  $P_2$  for the relevant biomedical entities in 1961-1990. The anti-platelet effect [70], inhibition of prostaglandin synthesis [71], and acetylation effect on the enzyme cyclo-oxygenase [72] were uncovered. These discoveries provided a solid knowledge foundation for the successful repurposing of aspirin. The highest  $P_1$  in 1961-1990 was for indomethacin (16.75%), denoting fierce competition with aspirin for its original use. This could be one of the factors contributing to the repurposing of aspirin.

In Phase 3, aspirin was successfully used for several cardiovascular-related diseases because of its antiplatelet effects [80]. The related diseases and drugs achieved their highest values of  $P_1$  and CI as well as significant increases in  $P_2$  in 1991-2000. As these diseases are the most common diseases worldwide, according to data from the World Health Organization [74], the demand for the prevention and treatment of fatal diseases is also another potential factor driving drug repurposing. In the last phase, there was a large number of studies suggesting the use of aspirin for other diseases, especially colorectal cancer [36]. The greatest difference from previous phases is that aspirin was studied at the genetic level. Ten genes reached their maxima of  $P_1$  and CI as well as an apparent increase in  $P_2$  in 2001-2018. This observation could indicate that the development of modern science and technology,

such as gene sequencing, molecular simulation, and deep learning, accelerates the process of drug repurposing of aspirin. Meanwhile, 2 fatal diseases — diabetes and colorectal carcinoma — as well as 3 competitive drugs of aspirin as an antiplatelet agent — clopidogrel, ticlopidine, and warfarin (an anticoagulant and competitor with aspirin for stroke prevention) — also had peak  $P_1$  and CI values and a great increase in  $P_2$ .

Methodologically, in this study, we developed 4 entitymetrics and demonstrated how to use them to investigate the process of drug repurposing. The results demonstrate that the maxima of  $P_1$ ,  $P_3$ , and CI are closely associated with the different phases of research of aspirin repurposing. The  $P_1$  and CI metrics can indicate dynamic trends in academic interest in a given biomedical entity over a long time period. For instance, long-lasting increases in  $P_1$  and CI signal interest in repurposing, while  $P_2$  is more sensitive to immediate changes in academic interest in a specific biomedical entity, since it takes into consideration data from the two most recent periods. Moreover,  $P_3$  can reflect a research focus far earlier than the other 3 indices, which means that a continuously high  $P_3$  may be valuable as an early signal of the emergence and transfer of research topics in drug research. If  $P_3$  does indeed have predictive power, it could be due to the involvement of top domain experts in the peer review of manuscripts in top journals with high impact factors [81,82]. Additionally, due to their easy implementation and interpretability, these indices can be applied in multiple domains, such as drug assessment, drug discovery, and pharmacovigilance.

### Limitations and Future Directions

There are several limitations in the current paper. First, the data included in our analysis are limited to articles indexed in PubMed. Some real-world data, such as electronic health records, clinical trials, and social media, in which aspirin and its related biomedical entities were mentioned, should be included. In our future work, we will use different types of data sources for studying drug repurposing and take into account other entities related to drugs, including other biomedical entities, such as pathways, proteins, and cells, and non-biomedical entities, such as authors, institutions, and countries. The landscape of collaborations between academic institutions and pharmaceutical companies could affect the drug repurposing process. Second, there are several ways of measuring the impact of a journal, such as the impact factor and relative citation ratio. Third, this study mainly focused on investigating the repurposing journey of aspirin, but we did not test whether it can be used to predict future drug repurposing. In future studies, we will evaluate the different impact measures of a journal and choose a proper measure better fitted to the chosen drug. Furthermore, we will also aim to test the proposed metrics on other drugs to understand their repurposing journeys (eg, metformin) to see whether generalized patterns exist in different repurposing processes.

## Acknowledgments

This study was supported by the Major Project of the National Natural Science Foundation of China (71673211). The support provided by the China Scholarship Council (CSC) during a visit by Xin Li to Indiana University Bloomington (No. 201806270047) is acknowledged. This work was also partly supported by the Bio-Synergy Research Project (NRF-2013M3A9C4078138) of the Ministry of Science, ICT, and Future Planning through the National Research Foundation. The authors are also grateful to the anonymous referees and editors for their invaluable and insightful comments.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Top 20 journals related to aspirin research.

[\[DOCX File , 13 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Changes in the number of journals with aspirin-related publications during 1951-2018. The top 5 journals and their frequencies are indicated using heat maps for every 5-year period.

[\[DOCX File , 587 KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

Table S2 summarizes the peaks of  $P_1$ ,  $P_3$ , and CI as well as the increase in  $P_2$  for all the top 30 bioentities. The details can be found in the supplementary information section, including Phase 2 (1961-1990, the scientific basis for repurposing), Phase 3 (1991-2000, repurposing aspirin for cardiovascular-related diseases), and Phase 4 (2001-2018, repurposing aspirin for other diseases).

[\[DOCX File , 26 KB-Multimedia Appendix 3\]](#)

## References

1. Schneider G. Automating drug discovery. *Nat Rev Drug Discov* 2018 Feb;17(2):97-113. [doi: [10.1038/nrd.2017.232](https://doi.org/10.1038/nrd.2017.232)] [Medline: [29242609](https://pubmed.ncbi.nlm.nih.gov/29242609/)]
2. Parrish MC, Tan YJ, Grimes KV, Mochly-Rosen D. Surviving in the Valley of Death: Opportunities and Challenges in Translating Academic Drug Discoveries. *Annu Rev Pharmacol Toxicol* 2019 Jan 06;59:405-421. [doi: [10.1146/annurev-pharmtox-010818-021625](https://doi.org/10.1146/annurev-pharmtox-010818-021625)] [Medline: [30208282](https://pubmed.ncbi.nlm.nih.gov/30208282/)]
3. Deftereos SN, Andronis C, Friedla EJ, Persidis A, Persidis A. Drug repurposing and adverse event prediction using high-throughput literature analysis. *WIREs Syst Biol Med* 2011 Feb 16;3(3):323-334. [doi: [10.1002/wsbm.147](https://doi.org/10.1002/wsbm.147)]
4. Cha Y, Erez T, Reynolds IJ, Kumar D, Ross J, Koystiger G, et al. Drug repurposing from the perspective of pharmaceutical companies. *British Journal of Pharmacology* 2017 May 18;175(2):168-180. [doi: [10.1111/bph.13798](https://doi.org/10.1111/bph.13798)]
5. Strittmatter S. Overcoming Drug Development Bottlenecks With Repurposing: Old drugs learn new tricks. *Nat Med* 2014 Jun;20(6):590-591 [FREE Full text] [doi: [10.1038/nm.3595](https://doi.org/10.1038/nm.3595)] [Medline: [24901567](https://pubmed.ncbi.nlm.nih.gov/24901567/)]
6. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, et al. The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat Med* 2017 Apr 07;23(4):405-408 [FREE Full text] [doi: [10.1038/nm.4306](https://doi.org/10.1038/nm.4306)] [Medline: [28388612](https://pubmed.ncbi.nlm.nih.gov/28388612/)]
7. Sheard S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov* 2019 Jan;18(1):41-58 [FREE Full text] [doi: [10.1038/nrd.2018.168](https://doi.org/10.1038/nrd.2018.168)] [Medline: [30310233](https://pubmed.ncbi.nlm.nih.gov/30310233/)]
8. Fenig DM, McCullough A. Sildenafil in the treatment of erectile dysfunction. *Aging Health* 2007 Jun;3(3):295-303.
9. Singhal S, Mehta J, Desikan R, Ayers D, Roberson P, Eddlemon P, et al. Antitumor Activity of Thalidomide in Refractory Multiple Myeloma. *N Engl J Med* 1999 Nov 18;341(21):1565-1571. [doi: [10.1056/nejm199911183412102](https://doi.org/10.1056/nejm199911183412102)]
10. Lee A, Morley J. Metformin decreases food consumption and induces weight loss in subjects with obesity with type II non-insulin-dependent diabetes. *Obes Res* 1998 Jan;6(1):47-53 [FREE Full text] [doi: [10.1002/j.1550-8528.1998.tb00314.x](https://doi.org/10.1002/j.1550-8528.1998.tb00314.x)] [Medline: [9526970](https://pubmed.ncbi.nlm.nih.gov/9526970/)]
11. Berstein LM. Metformin in obesity, cancer and aging: addressing controversies. *Aging (Albany NY)* 2012 May;4(5):320-329 [FREE Full text] [doi: [10.18632/aging.100455](https://doi.org/10.18632/aging.100455)] [Medline: [22589237](https://pubmed.ncbi.nlm.nih.gov/22589237/)]
12. Pascual J, Rivas MT, Leira R. Testing the combination beta-blocker plus topiramate in refractory migraine. *Acta Neurol Scand* 2007 Feb;115(2):81-83. [doi: [10.1111/j.1600-0404.2006.00772.x](https://doi.org/10.1111/j.1600-0404.2006.00772.x)] [Medline: [17212609](https://pubmed.ncbi.nlm.nih.gov/17212609/)]

13. Diener H, Tfelt-Hansen P, Dahlöf C, Láinez MJA, Sandrini G, Wang S, MIGR-003 Study Group. Topiramate in migraine prophylaxis--results from a placebo-controlled trial with propranolol as an active control. *J Neurol* 2004 Aug;251(8):943-950. [doi: [10.1007/s00415-004-0464-6](https://doi.org/10.1007/s00415-004-0464-6)] [Medline: [15316798](https://pubmed.ncbi.nlm.nih.gov/15316798/)]
14. Simsek M, Meijer B, van Bodegraven AA, de Boer NK, Mulder CJ. Finding hidden treasures in old drugs: the challenges and importance of licensing generics. *Drug Discov Today* 2018 Jan;23(1):17-21 [FREE Full text] [doi: [10.1016/j.drudis.2017.08.008](https://doi.org/10.1016/j.drudis.2017.08.008)] [Medline: [28867540](https://pubmed.ncbi.nlm.nih.gov/28867540/)]
15. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua M, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun* 2017 Jul 12;8(1):16022 [FREE Full text] [doi: [10.1038/ncomms16022](https://doi.org/10.1038/ncomms16022)] [Medline: [28699633](https://pubmed.ncbi.nlm.nih.gov/28699633/)]
16. Xu M, Lee EM, Wen Z, Cheng Y, Huang W, Qian X, et al. Identification of small-molecule inhibitors of Zika virus infection and induced neural cell death via a drug repurposing screen. *Nat Med* 2016 Oct 29;22(10):1101-1107 [FREE Full text] [doi: [10.1038/nm.4184](https://doi.org/10.1038/nm.4184)] [Medline: [27571349](https://pubmed.ncbi.nlm.nih.gov/27571349/)]
17. Mons B, van Haagen H, Chichester C, Hoen P, den Dunnen JT, van Ommen G, et al. The value of data. *Nat Genet* 2011 Mar 29;43(4):281-283. [doi: [10.1038/ng0411-281](https://doi.org/10.1038/ng0411-281)] [Medline: [21445068](https://pubmed.ncbi.nlm.nih.gov/21445068/)]
18. Gao Z, Fu G, Ouyang C, Tsutsui S, Liu X, Yang J, et al. edge2vec: Representation learning using edge semantics for biomedical knowledge discovery. *BMC Bioinformatics* 2019 Jun 10;20(1):306 [FREE Full text] [doi: [10.1186/s12859-019-2914-2](https://doi.org/10.1186/s12859-019-2914-2)] [Medline: [31238875](https://pubmed.ncbi.nlm.nih.gov/31238875/)]
19. Chen B, Dong X, Jiao D, Wang H, Zhu Q, Ding Y, et al. Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC Bioinformatics* 2010 May 17;11:255 [FREE Full text] [doi: [10.1186/1471-2105-11-255](https://doi.org/10.1186/1471-2105-11-255)] [Medline: [20478034](https://pubmed.ncbi.nlm.nih.gov/20478034/)]
20. Hamilton W, Bajaj P, Zitnik M, Jurafsky D, Leskovec J. Embedding Logical Queries on Knowledge Graphs. In: the proceedings of 32nd International Conference on Neural Information Processing Systems (NIPS'18). 2018 Presented at: NIPS18; Dec 3 – Dec 8; Montreal, Canada p. 2030-2041.
21. Chang Y, Park H, Yang H, Lee S, Lee K, Kim TS, et al. Cancer Drug Response Profile scan (CDRscan): A Deep Learning Model That Predicts Drug Effectiveness from Cancer Genomic Signature. *Sci Rep* 2018 Jun 11;8(1):8857 [FREE Full text] [doi: [10.1038/s41598-018-27214-6](https://doi.org/10.1038/s41598-018-27214-6)] [Medline: [29891981](https://pubmed.ncbi.nlm.nih.gov/29891981/)]
22. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. *Bioinformatics* 2018 Sep 01;34(17):i821-i829 [FREE Full text] [doi: [10.1093/bioinformatics/bty593](https://doi.org/10.1093/bioinformatics/bty593)] [Medline: [30423097](https://pubmed.ncbi.nlm.nih.gov/30423097/)]
23. Jinha AE. Article 50 million: an estimate of the number of scholarly articles in existence. *Learn. Pub* 2010 Jul 01;23(3):258-263. [doi: [10.1087/20100308](https://doi.org/10.1087/20100308)]
24. Ding Y, Stirling K. Data-driven Discovery: A New Era of Exploiting the Literature and Data. *JDIS* 2016 Nov 03;1(4):1-9. [doi: [10.20309/jdis.201622](https://doi.org/10.20309/jdis.201622)]
25. Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect Biol Med* 1986;30(1):7-18. [doi: [10.1353/pbm.1986.0087](https://doi.org/10.1353/pbm.1986.0087)] [Medline: [3797213](https://pubmed.ncbi.nlm.nih.gov/3797213/)]
26. Swanson DR. Undiscovered Public Knowledge. *The Library Quarterly* 1986 Apr;56(2):103-118. [doi: [10.1086/601720](https://doi.org/10.1086/601720)]
27. Swanson DR. Two medical literatures that are logically but not bibliographically connected. *J. Am. Soc. Inf. Sci* 1987 Jul;38(4):228-233. [doi: [10.1002/\(sici\)1097-4571](https://doi.org/10.1002/(sici)1097-4571)]
28. Cory KA. Discovering Hidden Analogies in an Online Humanities Database. *Comput Hum* 1997;31(1):1-12. [doi: [10.1023/A:1000422220677](https://doi.org/10.1023/A:1000422220677)]
29. Ding Y, Song M, Han J, Yu Q, Yan E, Lin L, et al. Entitymetrics: measuring the impact of entities. *PLoS One* 2013;8(8):e71416 [FREE Full text] [doi: [10.1371/journal.pone.0071416](https://doi.org/10.1371/journal.pone.0071416)] [Medline: [24009660](https://pubmed.ncbi.nlm.nih.gov/24009660/)]
30. Wang H, Ding Y, Tang J, Dong X, He B, Qiu J, et al. Finding complex biological relationships in recent PubMed articles using Bio-LDA. *PLoS One* 2011 Mar 23;6(3):e17243 [FREE Full text] [doi: [10.1371/journal.pone.0017243](https://doi.org/10.1371/journal.pone.0017243)] [Medline: [21448266](https://pubmed.ncbi.nlm.nih.gov/21448266/)]
31. Song M, Han N, Kim Y, Ding Y, Chambers T. Discovering implicit entity relation with the gene-citation-gene network. *PLoS One* 2013;8(12):e84639 [FREE Full text] [doi: [10.1371/journal.pone.0084639](https://doi.org/10.1371/journal.pone.0084639)] [Medline: [24358368](https://pubmed.ncbi.nlm.nih.gov/24358368/)]
32. Ding Y, Zhang G, Chambers T, Song M, Wang X, Zhai C. Content-based citation analysis: The next generation of citation analysis. *J Assn Inf Sci Tec* 2014 Jun 06;65(9):1820-1833. [doi: [10.1002/asi.23256](https://doi.org/10.1002/asi.23256)]
33. Kissin I, Edwin LB. Top Journals Selectivity Index: is it acceptable for drugs beyond the field of analgesia? *Scientometrics* 2011 May 5;88(2):589-597. [doi: [10.1007/s11192-011-0403-0](https://doi.org/10.1007/s11192-011-0403-0)]
34. Montinari MR, Minelli S, De Caterina R. The first 3500 years of aspirin history from its roots - A concise summary. *Vascul Pharmacol* 2019 Feb;113:1-8. [doi: [10.1016/j.vph.2018.10.008](https://doi.org/10.1016/j.vph.2018.10.008)] [Medline: [30391545](https://pubmed.ncbi.nlm.nih.gov/30391545/)]
35. Bordons M, Bravo C, Barrigón S. Time-tracking of the research profile of a drug using bibliometric tools. *J. Am. Soc. Inf. Sci* 2004 Jan 16;55(5):445-461. [doi: [10.1002/asi.10397](https://doi.org/10.1002/asi.10397)]
36. Gilligan MM, Gartung A, Sulciner ML, Norris PC, Sukhatme VP, Bielenberg DR, et al. Aspirin-triggered proresolving mediators stimulate resolution in cancer. *Proc Natl Acad Sci U S A* 2019 Mar 26;116(13):6292-6297 [FREE Full text] [doi: [10.1073/pnas.1804000116](https://doi.org/10.1073/pnas.1804000116)] [Medline: [30862734](https://pubmed.ncbi.nlm.nih.gov/30862734/)]

37. Volz J, Mammadova-Bach E, Gil-Pulido J, Nandigama R, Remer K, Sorokin L, et al. Inhibition of platelet GPVI induces intratumor hemorrhage and increases efficacy of chemotherapy in mice. *Blood* 2019 Jun 20;133(25):2696-2706. [doi: [10.1182/blood.2018877043](https://doi.org/10.1182/blood.2018877043)] [Medline: [30952674](https://pubmed.ncbi.nlm.nih.gov/30952674/)]
38. Jackson S, Zhu B, Pfeiffer R, Liu Z, Gadalla S, Koshiol J. Aspirin may extend biliary tract cancer survival: Results from population-based cohort. *Clin Res* 2019;2333. [doi: [10.1158/1538-7445.AM2019-2333](https://doi.org/10.1158/1538-7445.AM2019-2333)]
39. Baker NC, Ekins S, Williams AJ, Tropsha A. A bibliometric review of drug repurposing. *Drug Discov Today* 2018 Mar;23(3):661-672 [FREE Full text] [doi: [10.1016/j.drudis.2018.01.018](https://doi.org/10.1016/j.drudis.2018.01.018)] [Medline: [29330123](https://pubmed.ncbi.nlm.nih.gov/29330123/)]
40. Wang Z, Lachmann A, Keenan AB, Ma'ayan A. L1000FWD: fireworks visualization of drug-induced transcriptomic signatures. *Bioinformatics* 2018 Jun 15;34(12):2150-2152 [FREE Full text] [doi: [10.1093/bioinformatics/bty060](https://doi.org/10.1093/bioinformatics/bty060)] [Medline: [29420694](https://pubmed.ncbi.nlm.nih.gov/29420694/)]
41. Liu T, Hsieh Y, Chou C, Yang P. Systematic polypharmacology and drug repurposing via an integrated L1000-based Connectivity Map database mining. *R Soc Open Sci* 2018 Nov;5(11):181321 [FREE Full text] [doi: [10.1098/rsos.181321](https://doi.org/10.1098/rsos.181321)] [Medline: [30564416](https://pubmed.ncbi.nlm.nih.gov/30564416/)]
42. Kidnapillai S, Bortolasci CC, Udawela M, Panizzutti B, Spolding B, Connor T, et al. The use of a gene expression signature and connectivity map to repurpose drugs for bipolar disorder. *World J Biol Psychiatry* 2018 Aug 03;1-9. [doi: [10.1080/15622975.2018.1492734](https://doi.org/10.1080/15622975.2018.1492734)] [Medline: [29956574](https://pubmed.ncbi.nlm.nih.gov/29956574/)]
43. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007 Jun 07;447(7145):661-678 [FREE Full text] [doi: [10.1038/nature05911](https://doi.org/10.1038/nature05911)] [Medline: [17554300](https://pubmed.ncbi.nlm.nih.gov/17554300/)]
44. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, et al. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol* 2012 Apr 10;30(4):317-320. [doi: [10.1038/nbt.2151](https://doi.org/10.1038/nbt.2151)] [Medline: [22491277](https://pubmed.ncbi.nlm.nih.gov/22491277/)]
45. Ferrero E, Agarwal P. Connecting genetics and gene expression data for target prioritisation and drug repositioning. *BioData Min* 2018;11:7 [FREE Full text] [doi: [10.1186/s13040-018-0171-y](https://doi.org/10.1186/s13040-018-0171-y)] [Medline: [29881461](https://pubmed.ncbi.nlm.nih.gov/29881461/)]
46. Dakshanamurthy S, Issa NT, Assefnia S, Seshasayee A, Peters OJ, Madhavan S, et al. Predicting new indications for approved drugs using a proteochemometric method. *J Med Chem* 2012 Aug 09;55(15):6832-6848 [FREE Full text] [doi: [10.1021/jm300576q](https://doi.org/10.1021/jm300576q)] [Medline: [22780961](https://pubmed.ncbi.nlm.nih.gov/22780961/)]
47. Li YY, An J, Jones SJM. A computational approach to finding novel targets for existing drugs. *PLoS Comput Biol* 2011 Sep;7(9):e1002139 [FREE Full text] [doi: [10.1371/journal.pcbi.1002139](https://doi.org/10.1371/journal.pcbi.1002139)] [Medline: [21909252](https://pubmed.ncbi.nlm.nih.gov/21909252/)]
48. Leinung MC, Grasso P. [D-Leu-4]-OB3, a synthetic peptide amide with leptin-like activity, augments the effects of orally delivered exenatide and pramlintide acetate on energy balance and glycemic control in insulin-resistant male C57BLK/6-m db/db mice. *Regul Pept* 2012 Nov 10;179(1-3):33-38. [doi: [10.1016/j.regpep.2012.08.006](https://doi.org/10.1016/j.regpep.2012.08.006)] [Medline: [22960403](https://pubmed.ncbi.nlm.nih.gov/22960403/)]
49. Yang L, Agarwal P. Systematic drug repositioning based on clinical side-effects. *PLoS One* 2011;6(12):e28025 [FREE Full text] [doi: [10.1371/journal.pone.0028025](https://doi.org/10.1371/journal.pone.0028025)] [Medline: [22205936](https://pubmed.ncbi.nlm.nih.gov/22205936/)]
50. Ye H, Liu Q, Wei J. Construction of drug network based on side effects and its application for drug repositioning. *PLoS One* 2014;9(2):e87864 [FREE Full text] [doi: [10.1371/journal.pone.0087864](https://doi.org/10.1371/journal.pone.0087864)] [Medline: [24505324](https://pubmed.ncbi.nlm.nih.gov/24505324/)]
51. Olayan RS, Ashoor H, Bajic VB. DDR: efficient computational method to predict drug-target interactions using graph mining and machine learning approaches. *Bioinformatics* 2018 Apr 01;34(7):1164-1173 [FREE Full text] [doi: [10.1093/bioinformatics/btx731](https://doi.org/10.1093/bioinformatics/btx731)] [Medline: [29186331](https://pubmed.ncbi.nlm.nih.gov/29186331/)]
52. Ding H, Takigawa I, Mamitsuka H, Zhu S. Similarity-based machine learning methods for predicting drug-target interactions: a brief review. *Brief Bioinform* 2014 Sep;15(5):734-747. [doi: [10.1093/bib/bbt056](https://doi.org/10.1093/bib/bbt056)] [Medline: [23933754](https://pubmed.ncbi.nlm.nih.gov/23933754/)]
53. Liang X, Li D, Song M, Madden A, Ding Y, Bu Y. Predicting biomedical relationships using the knowledge and graph embedding cascade model. *PLoS One* 2019;14(6):e0218264 [FREE Full text] [doi: [10.1371/journal.pone.0218264](https://doi.org/10.1371/journal.pone.0218264)] [Medline: [31194807](https://pubmed.ncbi.nlm.nih.gov/31194807/)]
54. Chen B, Ding Y, Wild DJ. Assessing drug target association using semantic linked data. *PLoS Comput Biol* 2012;8(7):e1002574 [FREE Full text] [doi: [10.1371/journal.pcbi.1002574](https://doi.org/10.1371/journal.pcbi.1002574)] [Medline: [22859915](https://pubmed.ncbi.nlm.nih.gov/22859915/)]
55. Fu G, Ding Y, Seal A, Chen B, Sun Y, Bolton E. Predicting drug target interactions using meta-path-based semantic network analysis. *BMC Bioinformatics* 2016 Apr 12;17:160 [FREE Full text] [doi: [10.1186/s12859-016-1005-x](https://doi.org/10.1186/s12859-016-1005-x)] [Medline: [27071755](https://pubmed.ncbi.nlm.nih.gov/27071755/)]
56. Wang Q, Mao Z, Wang B, Guo L. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Trans. Knowl. Data Eng* 2017 Dec 1;29(12):2724-2743. [doi: [10.1109/tkde.2017.2754499](https://doi.org/10.1109/tkde.2017.2754499)]
57. Mukhopadhyay S, Palakal M, Maddu K. Multi-way association extraction and visualization from biological text documents using hyper-graphs: applications to genetic association studies for diseases. *Artif Intell Med* 2010 Jul;49(3):145-154. [doi: [10.1016/j.artmed.2010.03.002](https://doi.org/10.1016/j.artmed.2010.03.002)] [Medline: [20382004](https://pubmed.ncbi.nlm.nih.gov/20382004/)]
58. Cameron D, Bodenreider O, Yalamanchili H, Danh T, Vallabhaneni S, Thirunarayan K, et al. A graph-based recovery and decomposition of Swanson's hypothesis using semantic predications. *J Biomed Inform* 2013 Apr;46(2):238-251 [FREE Full text] [doi: [10.1016/j.jbi.2012.09.004](https://doi.org/10.1016/j.jbi.2012.09.004)] [Medline: [23026233](https://pubmed.ncbi.nlm.nih.gov/23026233/)]
59. Song M, Heo GE, Ding Y. SemPathFinder: Semantic path analysis for discovering publicly unknown knowledge. *Journal of Informetrics* 2015 Oct;9(4):686-703. [doi: [10.1016/j.joi.2015.06.004](https://doi.org/10.1016/j.joi.2015.06.004)]
60. Lv Y, Ding Y, Song M, Duan Z. Topology-driven trend analysis for drug discovery. *Journal of Informetrics* 2018 Aug;12(3):893-905. [doi: [10.1016/j.joi.2018.07.007](https://doi.org/10.1016/j.joi.2018.07.007)]

61. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014 Jan;42(Database issue):D1091-D1097 [FREE Full text] [doi: [10.1093/nar/gkt1068](https://doi.org/10.1093/nar/gkt1068)] [Medline: [24203711](https://pubmed.ncbi.nlm.nih.gov/24203711/)]
62. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004 Jan 01;32(Database issue):D258-D261 [FREE Full text] [doi: [10.1093/nar/gkh036](https://doi.org/10.1093/nar/gkh036)] [Medline: [14681407](https://pubmed.ncbi.nlm.nih.gov/14681407/)]
63. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res* 2016 Jan 04;44(D1):D1075-D1079 [FREE Full text] [doi: [10.1093/nar/gkv1075](https://doi.org/10.1093/nar/gkv1075)] [Medline: [26481350](https://pubmed.ncbi.nlm.nih.gov/26481350/)]
64. He B, Tang J, Ding Y, Wang H, Sun Y, Shin JH, et al. Mining relational paths in integrated biomedical data. *PLoS One* 2011;6(12):e27506 [FREE Full text] [doi: [10.1371/journal.pone.0027506](https://doi.org/10.1371/journal.pone.0027506)] [Medline: [22162991](https://pubmed.ncbi.nlm.nih.gov/22162991/)]
65. Williams RS, Lotia S, Holloway AK, Pico AR. From scientific discovery to cures: bright stars within a galaxy. *Cell* 2015 Sep 24;163(1):21-23 [FREE Full text] [doi: [10.1016/j.cell.2015.09.007](https://doi.org/10.1016/j.cell.2015.09.007)] [Medline: [26406364](https://pubmed.ncbi.nlm.nih.gov/26406364/)]
66. Zhu Y, Song M, Yan E. Identifying Liver Cancer and Its Relations with Diseases, Drugs, and Genes: A Literature-Based Approach. *PLoS One* 2016;11(5):e0156091 [FREE Full text] [doi: [10.1371/journal.pone.0156091](https://doi.org/10.1371/journal.pone.0156091)] [Medline: [27195695](https://pubmed.ncbi.nlm.nih.gov/27195695/)]
67. Lerchenmueller MJ, Sorenson O. Author Disambiguation in PubMed: Evidence on the Precision and Recall of Authority among NIH-Funded Scientists. *PLoS One* 2016;11(7):e0158731 [FREE Full text] [doi: [10.1371/journal.pone.0158731](https://doi.org/10.1371/journal.pone.0158731)] [Medline: [27367860](https://pubmed.ncbi.nlm.nih.gov/27367860/)]
68. Lee S, Kim D, Lee K, Choi J, Kim S, Jeon M, et al. BEST: Next-Generation Biomedical Entity Search Tool for Knowledge Discovery from Biomedical Literature. *PLoS One* 2016;11(10):e0164680 [FREE Full text] [doi: [10.1371/journal.pone.0164680](https://doi.org/10.1371/journal.pone.0164680)] [Medline: [27760149](https://pubmed.ncbi.nlm.nih.gov/27760149/)]
69. Kissin I. What Can Big Data on Academic Interest Reveal about a Drug? Reflections in Three Major US Databases. *Trends Pharmacol Sci* 2018 Mar;39(3):248-257. [doi: [10.1016/j.tips.2017.12.005](https://doi.org/10.1016/j.tips.2017.12.005)] [Medline: [29358009](https://pubmed.ncbi.nlm.nih.gov/29358009/)]
70. O'Brien J. EFFECTS OF SALICYLATES ON HUMAN PLATELETS. *The Lancet* 1968 Apr;291(7546):779-783. [doi: [10.1016/s0140-6736\(68\)92228-9](https://doi.org/10.1016/s0140-6736(68)92228-9)]
71. Vane JR. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nat New Biol* 1971 Jun 23;231(25):232-235. [doi: [10.1038/newbio231232a0](https://doi.org/10.1038/newbio231232a0)] [Medline: [5284360](https://pubmed.ncbi.nlm.nih.gov/5284360/)]
72. Roth GJ, Stanford N, Majerus PW. Acetylation of prostaglandin synthase by aspirin. *Proc Natl Acad Sci U S A* 1975 Aug;72(8):3073-3076 [FREE Full text] [doi: [10.1073/pnas.72.8.3073](https://doi.org/10.1073/pnas.72.8.3073)] [Medline: [810797](https://pubmed.ncbi.nlm.nih.gov/810797/)]
73. Kune GA, Kune S, Watson LF. Colorectal cancer risk, chronic illnesses, operations, and medications: case control results from the Melbourne Colorectal Cancer Study. *Cancer Res* 1988 Aug 01;48(15):4399-4404 [FREE Full text] [Medline: [3390835](https://pubmed.ncbi.nlm.nih.gov/3390835/)]
74. World Health Organization. The top 10 causes of death Internet URL: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> [accessed 2019-07-16]
75. Ebstein W, Müller J. Brenzkatechin in dem Urin eines Kindes. *Archiv f. pathol. Anat* 1875 Feb;62(4):554-560. [doi: [10.1007/bf01928660](https://doi.org/10.1007/bf01928660)]
76. Pignone M, Alberts MJ, Colwell JA, Cushman M, Inzucchi SE, Mukherjee D, American Diabetes Association, American Heart Association, American College of Cardiology Foundation. Aspirin for primary prevention of cardiovascular events in people with diabetes. *J Am Coll Cardiol* 2010 Jun 22;55(25):2878-2886 [FREE Full text] [doi: [10.1016/j.jacc.2010.04.003](https://doi.org/10.1016/j.jacc.2010.04.003)] [Medline: [20579547](https://pubmed.ncbi.nlm.nih.gov/20579547/)]
77. van Walraven C, Hart RG, Singer DE, Laupacis A, Connolly S, Petersen P, et al. Oral anticoagulants vs aspirin in nonvalvular atrial fibrillation: an individual patient meta-analysis. *JAMA* 2002 Nov 20;288(19):2441-2448. [doi: [10.1001/jama.288.19.2441](https://doi.org/10.1001/jama.288.19.2441)] [Medline: [12435257](https://pubmed.ncbi.nlm.nih.gov/12435257/)]
78. Mohr J, Thompson J, Lazar R, Levin B, Sacco R, Furie K, et al. A Comparison of Warfarin and Aspirin for the Prevention of Recurrent Ischemic Stroke. *N Engl J Med* 2001 Nov 15;345(20):1444-1451. [doi: [10.1056/nejmoa011258](https://doi.org/10.1056/nejmoa011258)]
79. Couzin J. Drug safety. Withdrawal of Vioxx casts a shadow over COX-2 inhibitors. *Science* 2004 Oct 15;306(5695):384-385. [doi: [10.1126/science.306.5695.384](https://doi.org/10.1126/science.306.5695.384)] [Medline: [15486258](https://pubmed.ncbi.nlm.nih.gov/15486258/)]
80. Sanmuganathan PS, Ghahramani P, Jackson PR, Wallis EJ, Ramsay LE. Aspirin for primary prevention of coronary heart disease: safety and absolute benefit related to coronary risk derived from meta-analysis of randomised trials. *Heart* 2001 Mar;85(3):265-271 [FREE Full text] [doi: [10.1136/heart.85.3.265](https://doi.org/10.1136/heart.85.3.265)] [Medline: [11179262](https://pubmed.ncbi.nlm.nih.gov/11179262/)]
81. Kissin I. Can a bibliometric indicator predict the success of an analgesic? *Scientometrics* 2010 Nov 30;86(3):785-795. [doi: [10.1007/s11192-010-0320-7](https://doi.org/10.1007/s11192-010-0320-7)]
82. Koenig MED. Determinants of expert judgement of research performance. *Scientometrics* 1982 Sep;4(5):361-378. [doi: [10.1007/bf02135122](https://doi.org/10.1007/bf02135122)]

## Abbreviations

**CI:** Collaboration Index

**DR:** drug repurposing

**GWAS:** genome-wide association study

**P<sub>1</sub>**: Popularity Index  
**P<sub>2</sub>**: Promising Index  
**P<sub>3</sub>**: Prestige Index  
**P3C**: the 4 entymetric indicators for biomedical entities  
**RA**: rheumatoid arthritis

*Edited by G Eysenbach; submitted 19.10.19; peer-reviewed by J Yang, W Griffin, Y Li; comments to author 06.12.19; revised version received 08.01.20; accepted 31.03.20; published 16.06.20*

*Please cite as:*

*Li X, Rousseau JF, Ding Y, Song M, Lu W*

*Understanding Drug Repurposing From the Perspective of Biomedical Entities and Their Evolution: Bibliographic Research Using Aspirin*

*JMIR Med Inform 2020;8(6):e16739*

*URL: <https://medinform.jmir.org/2020/6/e16739>*

*doi: [10.2196/16739](https://doi.org/10.2196/16739)*

*PMID: [32543442](https://pubmed.ncbi.nlm.nih.gov/32543442/)*

©Xin Li, Justin F. Rousseau, Ying Ding, Min Song, Wei Lu. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 16.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.