

# Understanding drug repurposing from the perspective of biomedical entities and their evolution: a bibliographic research using aspirin

Xin Li, Justin F. Rousseau, Ying Ding, Min Song, Wei Lu

Submitted to: JMIR Medical Informatics  
on: October 19, 2019

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

## *Table of Contents*

---

<b>Original Manuscript</b> .....	<b>4</b>
<b>Supplementary Files</b> .....	<b>49</b>
Multimedia Appendixes .....	<b>50</b>
Multimedia Appendix 0.....	<b>50</b>
<b>Figures</b> .....	<b>51</b>
Figure 1.....	<b>52</b>
Figure 2.....	<b>53</b>
Figure 3.....	<b>54</b>
Figure 4.....	<b>55</b>
Figure 5.....	<b>56</b>
Figure 6.....	<b>57</b>
Figure 7.....	<b>58</b>
Figure 8.....	<b>59</b>
Figure 9.....	<b>60</b>
Figure 10.....	<b>61</b>
Figure 11.....	<b>62</b>

# Understanding drug repurposing from the perspective of biomedical entities and their evolution: a bibliographic research using aspirin

Xin LiPhD, ; Justin F. RousseauMD, MMSc, ; Ying DingPhD, ; Min SongPhD, ; Wei LuPhD,

## Corresponding Author:

Wei LuPhD,

Phone: +8602768752757

Email: weilu@whu.edu.cn

## Abstract

**Background:** Nowadays, 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.

**Objective:** The purpose of this study was to understand the drug repurposing from the perspective of bio-entities and their evolution.

**Methods:** In the work reported in this paper, we extended the bibliometric indicators of bio-entities mentioned in PubMed to detect potential patterns of bio-entities in various phases of drug research, and investigated the factors driving DR. We used aspirin (acetylsalicylic acid) as the subject of the study, since it can be repurposed for many applications. We applied four easy, transparent measures based on entitymetrics: Popularity Index (P1); Promising Index (P2); Prestige Index (P3); and Collaboration Index (CI), to investigate DR for aspirin.

**Results:** We found that the maxima of P1, P3 and CI are closely associated with the different repurposing phases of aspirin. These metrics enabled us to observe the way in which bio-entities interacted with the drug during the various phases of DR, and to analyze the potential driving factors of DR at the entity level. P1 and CI are indicative of the dynamic trends of a specific bio-entity over a long time period, while P2 is more sensitive to immediate changes. P3 reflects early signs of the practical value of bio-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 which can be valuable for enhancing decision making in the field of drug discovery and development.

(JMIR Preprints 19/10/2019:16739)

DOI: <https://doi.org/10.2196/preprints.16739>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

**Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

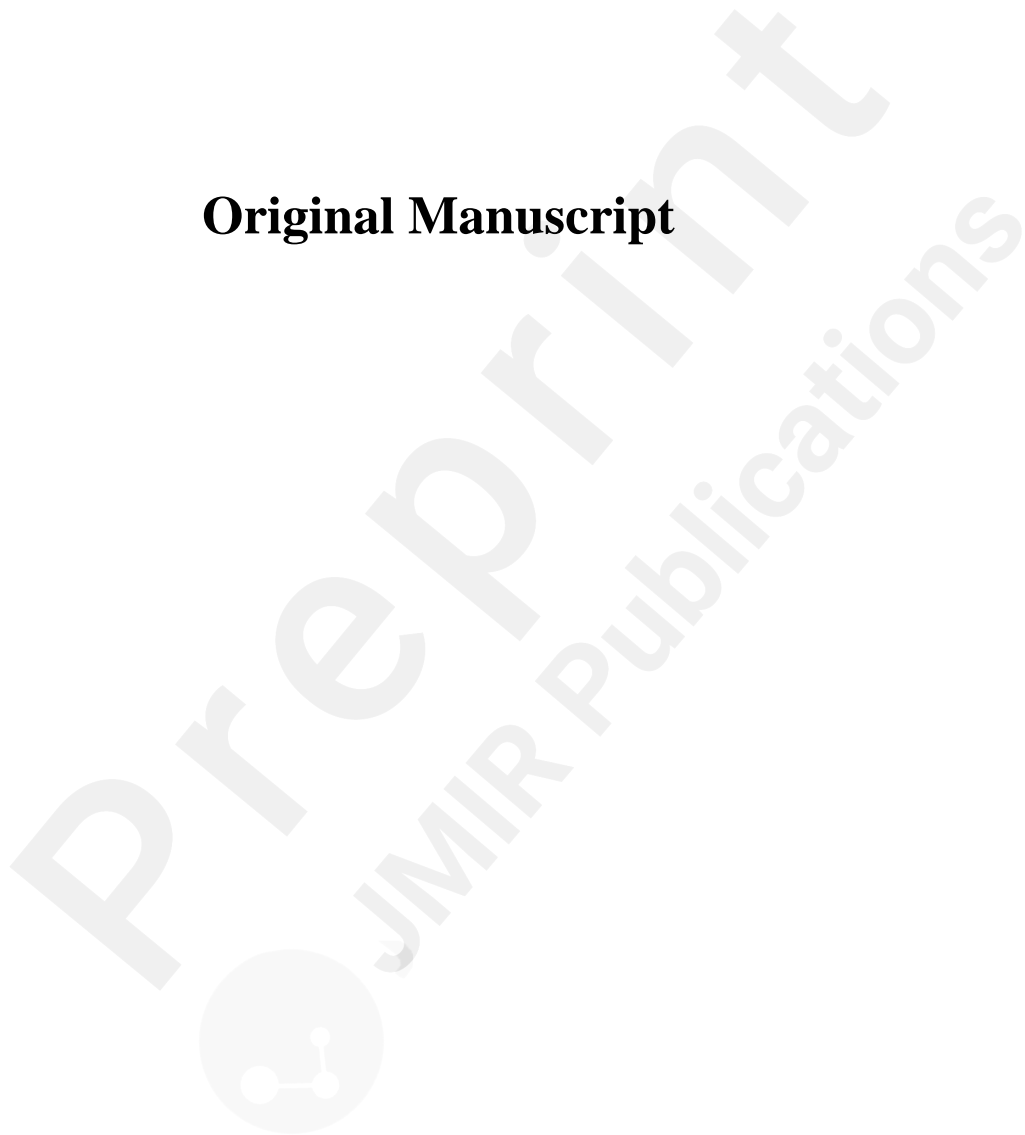
2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

**Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http](#)

**Original Manuscript**



# Understanding drug repurposing from the perspective of biomedical entities and their evolution: a bibliographic research using aspirin

## Abstract

**Background:** Nowadays, 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 (e.g., diseases, drugs and genes) studied in the biomedical texts, make it possible for researchers to measure the knowledge evolution and transfer of drug research.

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

**Methods:** In the work reported in this paper, we extend 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 use aspirin (acetylsalicylic acid) as the subject of the study, since it can be repurposed for many applications. We propose four easy, transparent measures based on entitymetrics: Popularity Index ( $P_1$ ); Promising Index ( $P_2$ ); Prestige Index ( $P_3$ ); and Collaboration Index (CI), to investigate DR for aspirin.

**Results:** We find that the maxima of  $P_1$ ,  $P_3$  and CI are closely associated with the different repurposing phases of aspirin. These metrics enable 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 of DR at the entity level.  $P_1$  and CI are indicative of the dynamic trends of a specific biomedical entity over a long time period, while  $P_2$  is more sensitive to immediate changes.  $P_3$  reflects 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 which can be valuable for enhancing decision making in the field of drug discovery and development.

**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 3,000, 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: the practice of identifying novel clinical indicators for existing marketed drugs [5–7].

The past decades have produced a few successful cases of drug repurposing. For examples, sildenafil, originally developed to treat cardiovascular disease, was unexpectedly discovered to be effective against erectile dysfunction (ED) [8]. Thalidomide, once used for morning sickness, has been repurposed for the treatment of multiple myeloma [9]; and metformin, originally a treatment for type II 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. Bin et al. [15] designed a system-based algorithm called Reverse Gene Expression Score (RGES) 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. Miao et al. [16] found that emricasan had an inhibitory effect on the Zika virus, by screening more than 6,000 compounds. With the rapid development of natural language processing (NLP) and deep learning techniques, robust solutions have been proposed recently 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 3,000 new articles currently published per day [23]. No researcher or clinician can read and comprehend all of 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 the PubMed literature, 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 relationship using the PubMed literature [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 extend bibliometric indicators for biomedical entities mentioned in the PubMed literature to investigate drug repurposing. We use 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 1,500 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 ten 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 aim to identify patterns in the different repurposing phases of aspirin by analyzing the diseases, drugs and genes related to aspirin. We propose four measures based on entitymetrics: Popularity Index ( $P_1$ ), Promising Index ( $P_2$ ), Prestige Index ( $P_3$ ) and Collaboration Index (CI), to identify the characteristics and patterns of DR for aspirin.

## 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 to multiple myeloma [9], and sildenafil for erectile dysfunction (ED) [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 two 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), a database for more than 7,000 gene-expression profiles of 1,309 compounds, has been widely used in this context in previous work. Liu et al. [40] found that the anti-cancer drugs KM-00927 and BR-K75081836 can be used to inhibit histone deacetylase, using a systematic analysis tool, L1000FWD, and CMap. Kidnapillai et al. [41] used gene expression signature data and CMap to identify 10 drugs, including camptothecin, nimesulide, and recinnamine, which could be effective against bipolar disorder (BD).

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) [42]. Using GWAS-driven methods, Sanseau et al. 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 [43]. 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 [43]. Enrico and Pankaj [44]

combined a CMap-based approach with perturbation of transcriptional profiles and disease data from GWASs for target prioritization and drug repurposing. These researchers pointed out that the 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 the traditional high-throughput screening. Dakshanamurthy et al. [45] proposed a proteochemometric method called TMFS to conduct molecular docking of over 3,000 FDA approved compounds across the crystal structures of more than 2,000 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. [46] designed a stand-alone 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 to obesity [47], both of which were “happy accidents”. Recently, Lun and Pankaj [48] generated human phenotypic profiles for drugs, based on over 3,000 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. [49] 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 five indications of 1,234 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 [50,51], link prediction [52,53], entity prediction [52], and path prediction [18,54]. Liang et al. [52] represented biomedical entities and their relationships in a heterogeneous network using graph2vec and knowledge2vec [55], 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. [54] treated drug repurposing as a binary classification problem, and combined the meta-path-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 was used to discover relationships between biomedical entities, such as Raynaud’s syndrome and fish oil [25], migraine headaches and magnesium [26], and 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 [56], network-based entity extraction [57] and semantic path-based storytelling [58]. 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. [59] 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 [60], Gene Ontology [61], and SIDER [62], 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. [63] designed a graph mining algorithm to predict potential relationships between different biomedical entities. The case they studied demonstrated that the anti-diabetic 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 academic texts 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. [64] 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. [65] 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 fine-grained level. Lv et al. [59] 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 in [17,23,24,27,28,54–56, 63]. In this research, we extend 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, 60]. The final search query is shown in Box 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.

Box 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]))

To better understand the drug repurposing process of aspirin, the relevant research was divided into four 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 to do the preprocessing (such as removing the punctuation and stop words) of titles and abstracts in the NLP pipeline. In addition, a novel and reliable method of author name disambiguation proposed by Lerchenmuller and Sorenson [66] was used to count distinct authors.

Table 1 Descriptive statistics of the four phases of aspirin research.

Phases	Time span	Number of publications	Number of authors	Avg. number of authors	Number of journals
1. original use	1951-1955	208	318	1.76	117
	1956-1960	299	498	1.88	159
	<b>1951-1960</b>	<b>507</b>	<b>794</b>	<b>1.83</b>	<b>218</b>
2. in-depth studies	1961-1965	748	1310	2.01	301

of pharmacological mechanisms and side effects	1966-1970	1268	2167	2.12	418
	1971-1975	2766	4880	2.40	696
	1976-1980	3797	7419	2.71	895
	1981-1985	4395	10011	3.16	1033
	1986-1990	4470	11600	3.50	1101
	<b>1961-1990</b>	<b>17444</b>	<b>31787</b>	<b>2.90</b>	<b>2153</b>
3. repurposing for cardiovascular diseases	1991-1995	5164	14044	3.69	1256
	1996-2000	6353	17694	4.10	1314
	<b>1991-2000</b>	<b>11517</b>	<b>28818</b>	<b>3.91</b>	<b>1798</b>
4. repurposing for other diseases	2001-2005	8099	27784	4.22	1719
	2006-2010	9366	35313	4.94	1974
	2011-2015	10436	44603	5.78	2410
	2016-2018	6018	30796	6.73	1881
	<b>2001-2018</b>	<b>33919</b>	<b>118857</b>	<b>5.33</b>	<b>3865</b>
	<b>Total</b>	<b>63387</b>	<b>171559</b>	<b>4.39</b>	<b>5443</b>

## Biomedical entity Extraction

The biomedical entity extraction module provided by the Biomedical Entity Search Tool (BEST) [67], 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 dictionary of BEST is built from 12 different public sources, including NCBI Entrez Gene, DrugBank, T3DB, Animal TFDB, Therapeutic Target DataBase, PubChem, and MeSH [67]. We obtained 1,472 unique disease names, 1,640 unique drug names and 3,184 unique gene names from the titles and abstracts. Table 2 shows the top 10 biomedical entities of three 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	Drugs	Frequency	Genes	Frequency
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, four transparent and easy entitymetric indexes (P3C) are developed: the Popularity Index, the Promising Index, the Prestige Index, and the Collaboration Index. These indicators can be considered as the extensions of the indicators proposed by Kissin for measuring the academic interest of a drug or a technique at the article level [33, 68]. In this paper, we adapt them from the perspective of biomedical entities with the goal to understand 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 (e.g., diseases, drugs, and genes) and non-biomedical entities (e.g., authors) that are related to the given drug. Detailed explanations of these measures are as follows:

- **Popularity Index ( $P_1$ )** 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 five years. The popularity of a biomedical entity  $i$ ,  $PopularityIndex(i)$ , is given by:

$$P_1(i) = \frac{N_i}{N_T} \times 100\% \quad (1)$$

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 ( $P_2$ )** 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:

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

where  $\frac{N_{pi}}{N_{pT}}$  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  $i$  increases, and *vice versa*.

- **Prestige Index ( $P_3$ )** 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:

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

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 (JIF) 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 six journals are highly associated with aspirin-related specialty areas, such as *Circulation*, *Blood*, and *The European Heart Journal*. The full list of the top 20 journals is shown in Table S1. The Prestige Index ( $P_3$ ) 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 [68]. We employed a threshold of 5% to indicate that  $P_3$  was of interest [68].

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

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

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 [68].

## Results

### Overview of aspirin-related studies



Figure 1 Number of aspirin-related studies in PubMed over time. The background colors indicate the four phases of aspirin research.

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 of articles in PubMed per year, respectively. The details of publications, authors, and journals are shown in Table 1. In the course of 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 Figure S1 in Supplementary Information). 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 anti-platelet effect [69], the mechanism of inhibition of prostaglandin synthesis [70], and the acetylation of the cyclo-oxygenase enzyme [71]. The percentage of aspirin-related articles in PubMed reached its peak in 1981, at about 0.32%, and then decreased. Kune et al. (1988) reported that aspirin could effectively reduce the incidence of colorectal cancer [72], 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 investigation of the in-depth studies of pharmacological mechanisms and side effects of aspirin.

Phase 3 (1991-2000) witnesses a steady and stable growth in the number and percentage of aspirin-related articles per year in PubMed (Figure 1). Compared to the first ten years (1951-1960), the number of articles increased by over 22 times, and the number of distinct authors increased by over 36 times. As shown in Figure S1, in both 1991-1995 and 1996-2000, four of the top five 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 (2,164) in 2015, but the percentage significantly reduced (Figure 1). From Table 1, we note that the numbers of articles, distinct authors and journals were all higher than those of the previous three periods. The average number of authors in this period had exceeded the total average (4.39). Journals covering other topics, for examples, *Cancer Management and Research*, *Drugs & Aging*, and *World Neurosurgery*, were increasingly represented (Figure S1), 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 four phases from the biomedical entity perspective, we first compute 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 reorganize the thirty biomedical entities (the top 10 disease, drugs, and genes) into the four phases of aspirin research, according to when each achieved its maximum Popularity Index, which indicates the focus of research in the field of aspirin. In each phase, we further analyzed the change patterns of P3C indicators for the most popular biomedical entities, to investigate the features of different phases of drug repurposing, the association between entities and P3C indicators, and the possible factors driving drug repurposing at the biomedical entity level.

## Before repurposing

Only “rheumatoid arthritis” (RA) reached its maximum Popularity Index ( $P_1$ ) in Phase1, at 9.36%, as shown in Figure 2 (1), and then exhibited a downhill trend for the rest of the three phases, and reaching a low of 0.63% in 2016-2018. As shown in Figure 2 (2), for the Promising Index ( $P_2$ ) of RA, there is only one significant rise of more than 0 in all four 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 of 1945-1950. It can also be observed from Figure 2 (3) that the Prestige Index ( $P_3$ ) 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 three 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 2 (4), the Collaboration Index (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.

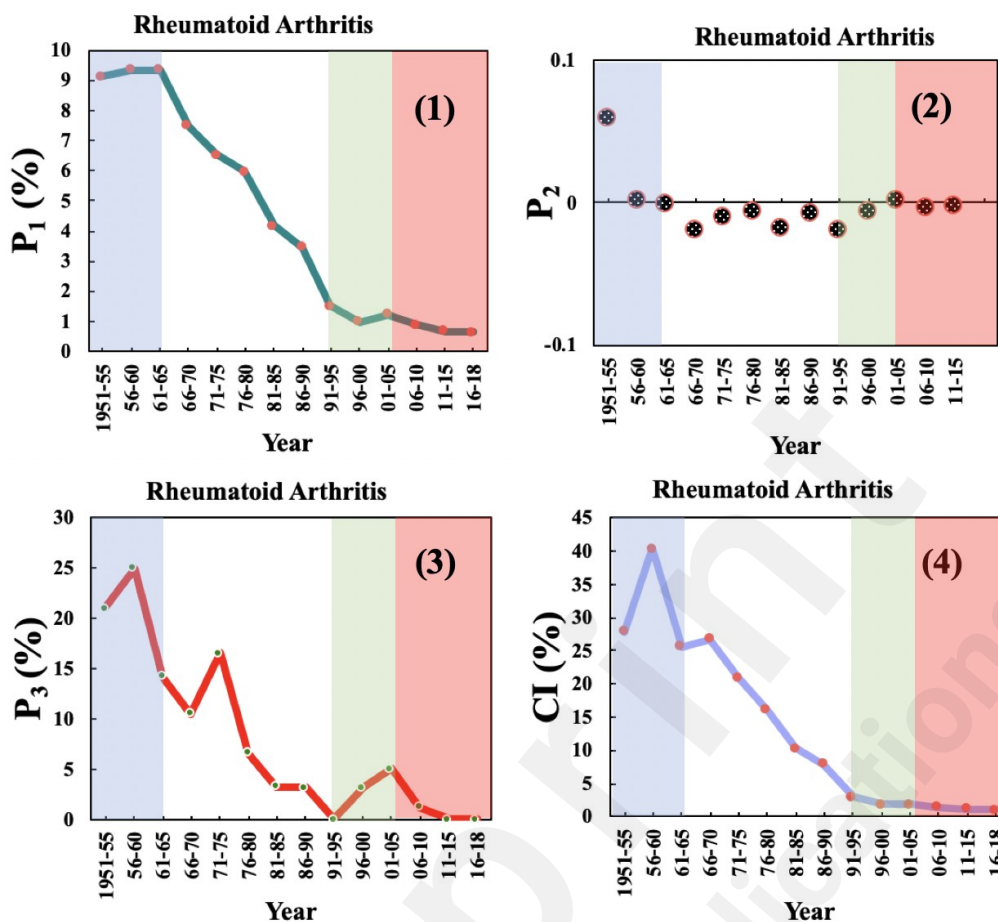


Figure 2 The P3C of biomedical entity “Rheumatoid Arthritis” over time. The background colors indicate the four phases of aspirin research

In summary, in Phase 1, the P<sub>1</sub>, P<sub>2</sub>, P<sub>3</sub> and CI of RA reached their maxima, or showed a significant rise, 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 three phases, which means that aspirin was studied in relation to other diseases, and is thus an ideal example of drug repurposing.

### The scientific basis for repurposing

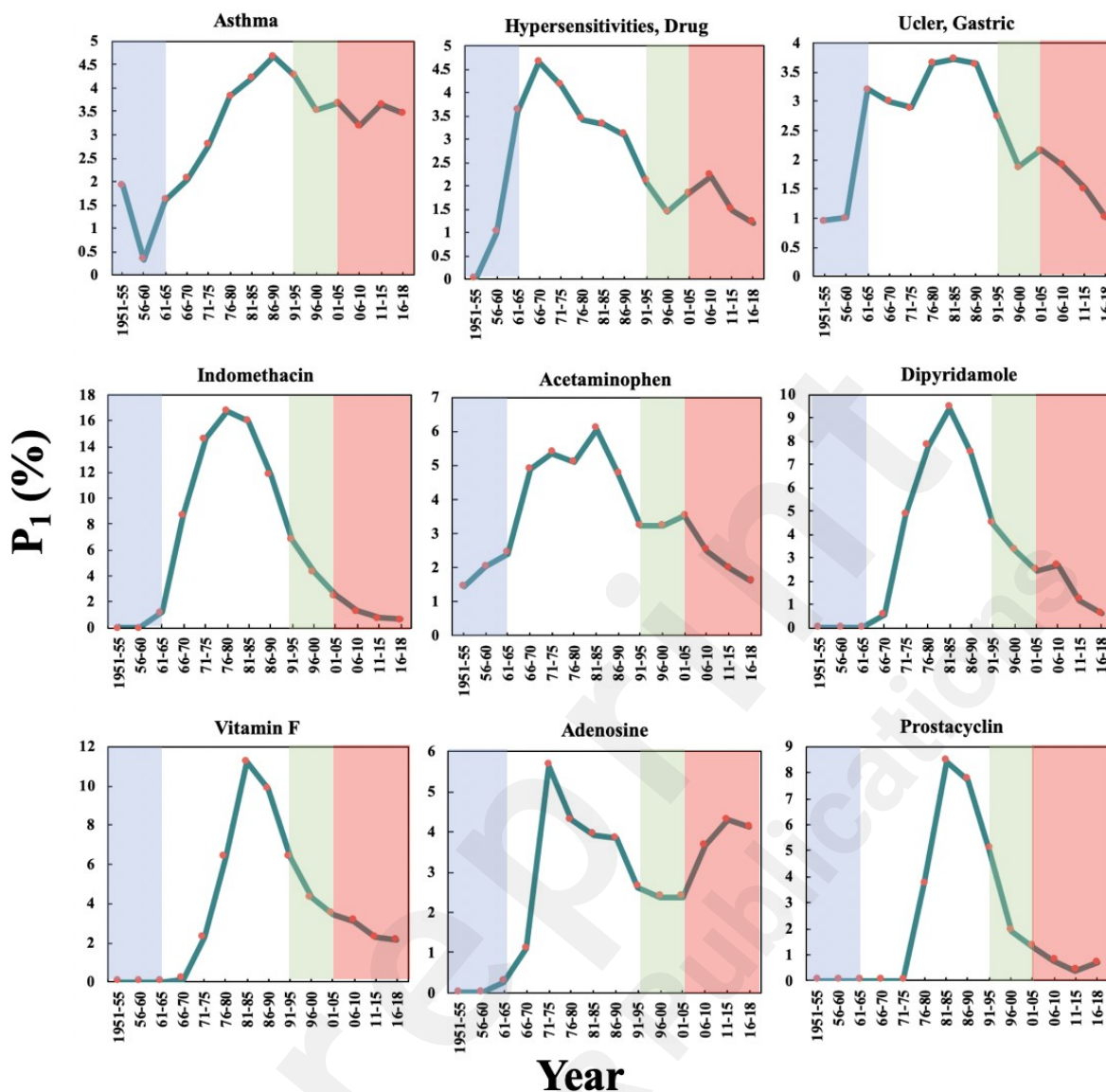


Figure 3 The Popularity Index ( $P_1$ ) of the biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the four phases of aspirin research.

As shown in Figure 3, there are nine top biomedical entities in the aspirin domain which reached their maximum  $P_1$  in Phase 2, including three diseases (“asthma, “hypersensitivities, drug” and “ulcer, gastric”) and six drugs (indomethacin, acetaminophen, dipyridamole, vitamin F, adenosine and prostacyclin”. The three diseases can all be side effects of aspirin, while the six drugs can be divided into three 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 the mechanisms of aspirin were studied in depth in Phase 2. The  $P_1$  of indomethacin in 1976-1980 (16.75) was the highest among these nine entities in Phase 2, and vitamin F in 1981-1985 (11.19) ranked second.

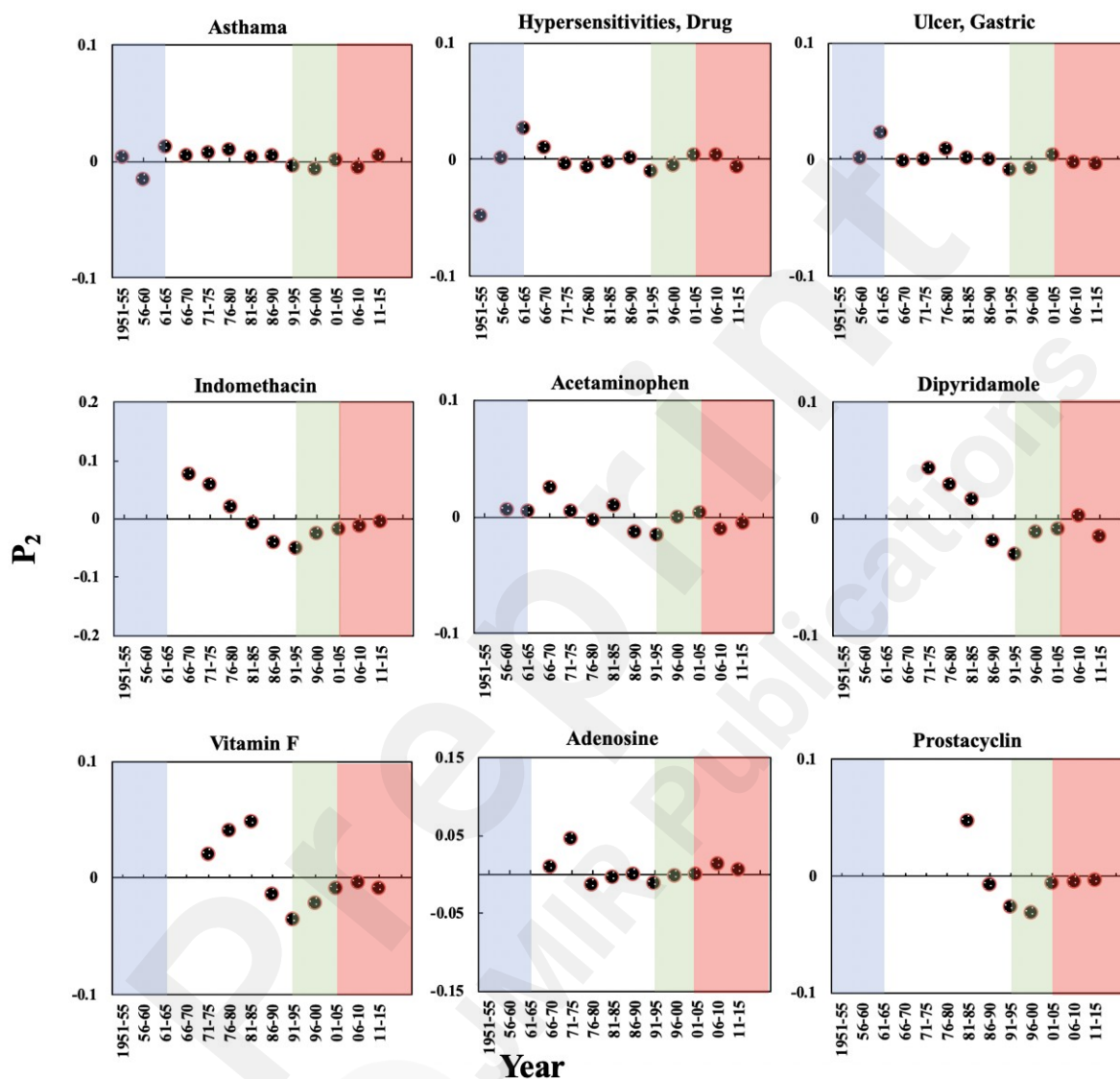


Figure 4. Promising Index ( $P_2$ ) of biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the four phases of aspirin research.

Figure 4 shows the  $P_2$  of these nine biomedical entities in the aspirin domain over time. The  $P_2$  of three side effects had a significant rise of more than zero in Phase 2, 1961-1965 and 1976-1980 for “asthma”, 1961-1965 for “hypersensitivities, drug”, and 1961-1965 for “ulcer, gastric”, indicating that interest in the side effects of aspirin increased sharply. The time periods in which the  $P_2$  of the six drugs showed significant rises are generally later than those of the side effects, such as 1971-1975 for indomethacin and 1981-1985 for prostacyclin. This observation indicates that the discovery and

in-depth studies of side effects may have positive effects on the discovery of the mechanism of action of aspirin, as well as the development of its alternatives with fewer side effects.

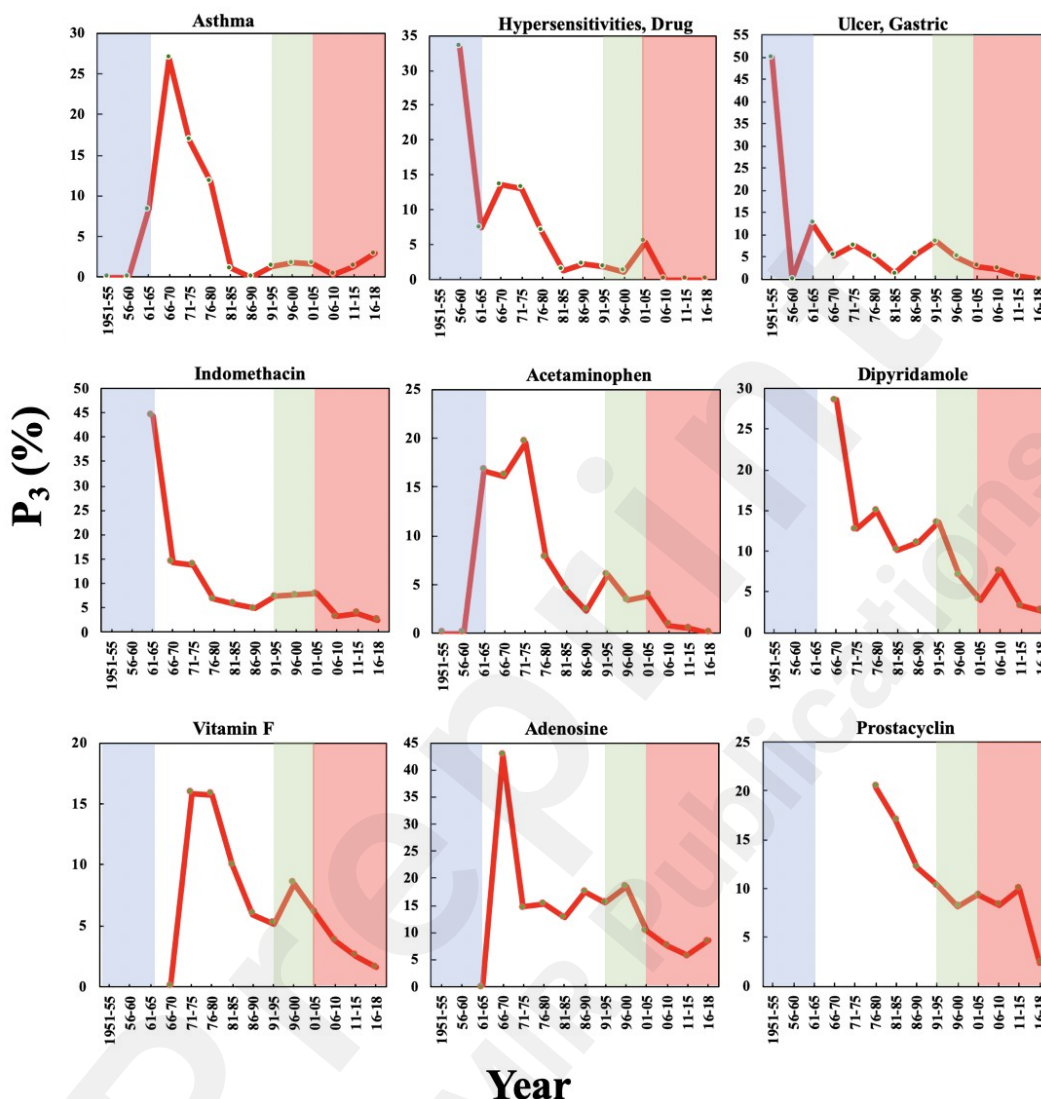


Figure 5 Prestige Index ( $P_3$ ) of biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the four phases of aspirin research.

Figure 5 shows the  $P_3$  of these nine biomedical entities in the aspirin domain, demonstrating a feature common to all nine entities: a gradual decline with fluctuation in  $P_3$  after reaching a maximum in Phase 1 or Phase 2. The  $P_3$  of “hypersensitivities, drug” and “ulcer, gastric” had a highest initial value 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 two entities are over the 5% threshold: 5.82% for adenosine and 10% for prostacyclin. In the aspirin domain, papers studying these two 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 two entities were still important foci of research in the aspirin domain.

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
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

It can be observed from Figure 3, Figure 5 and Table 3 that  $P_3$  on average achieved their 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 the Prestige Index can reflect an early sign of academic interest into biomedical entities, a phenomenon which could be potentially valuable for tracking the research frontiers of a drug.

The results of the CI of these nine biomedical entities in the aspirin domain are presented in Figure 6, which shows that the CIs for these biomedical entities have similar trends to  $P_1$  over time. Indomethacin achieved the highest maximum of CI in 1976-1980 (19.79%) among all nine biomedical entities during 1951-2018, indicating that it became a strong competitor to aspirin as an analgesic agent in Phase 2. This result also demonstrates that during the last five-year period (2011-2015), only two of the nine entities' CI were over 5%, meaning that the two entities were still the subject of research of a considerable number of scientists (more than 2,230) in the aspirin research community in 2011-2015. The two 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 side effects and the 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.

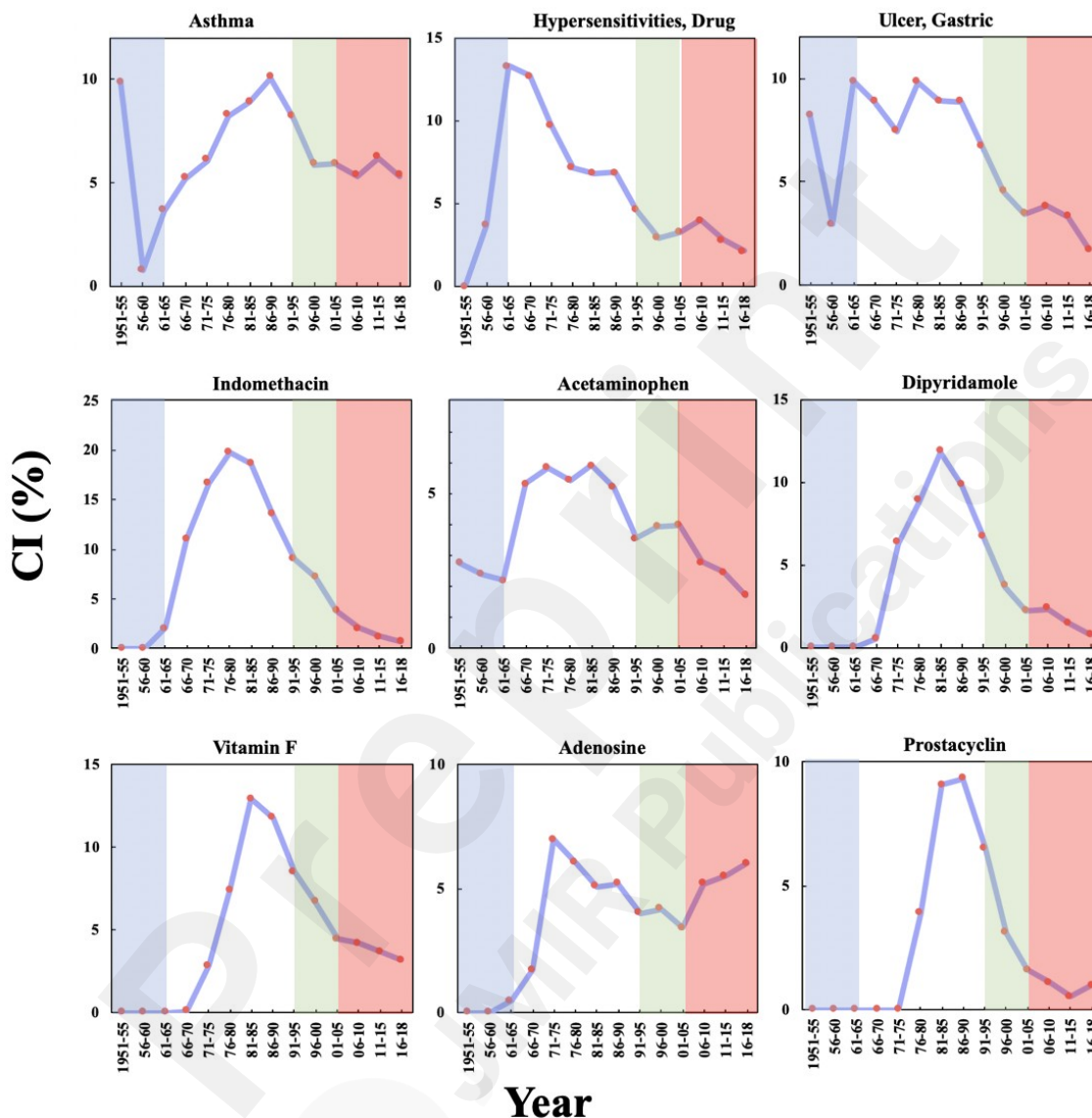


Figure 6 Collaboration Index (CI) of biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the four phases of aspirin research.

### Repurposing aspirin for cardiovascular-related diseases

In Phase 3, five top biomedical entities comprising four diseases and one drug reached their maximum Popularity Index, as shown in the first row of Figure 7. The four diseases were all cardiovascular related, including “coronary disease (CD)” ( $P_1 = 18.88\%$  in 1996-2000), “cerebral ischemia” ( $P_1 = 2.57\%$  in 1996-2000), “intracranial vascular disorder” ( $P_1 = 5.73\%$  in 1991-1995)

and “ischemic heart disease (IHD)” ( $P_1 = 3.01\%$  in 1996-2000). Compared with Figure 2 and Figure 3, the  $P_1$  of the previous ten biomedical entities that peaked in the first or second phases were considerably lower than that of CD, indicating that the cardiovascular-related disease was the focus of the aspirin domain in that time. CD is often referred as IHD, 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 reported to be the first and second most common causes of death worldwide in the early 21st century [73]. The demand for the prevention and treatment of such fatal diseases could be one of the factors that drive the repurposing of aspirin for cardiovascular-related diseases.

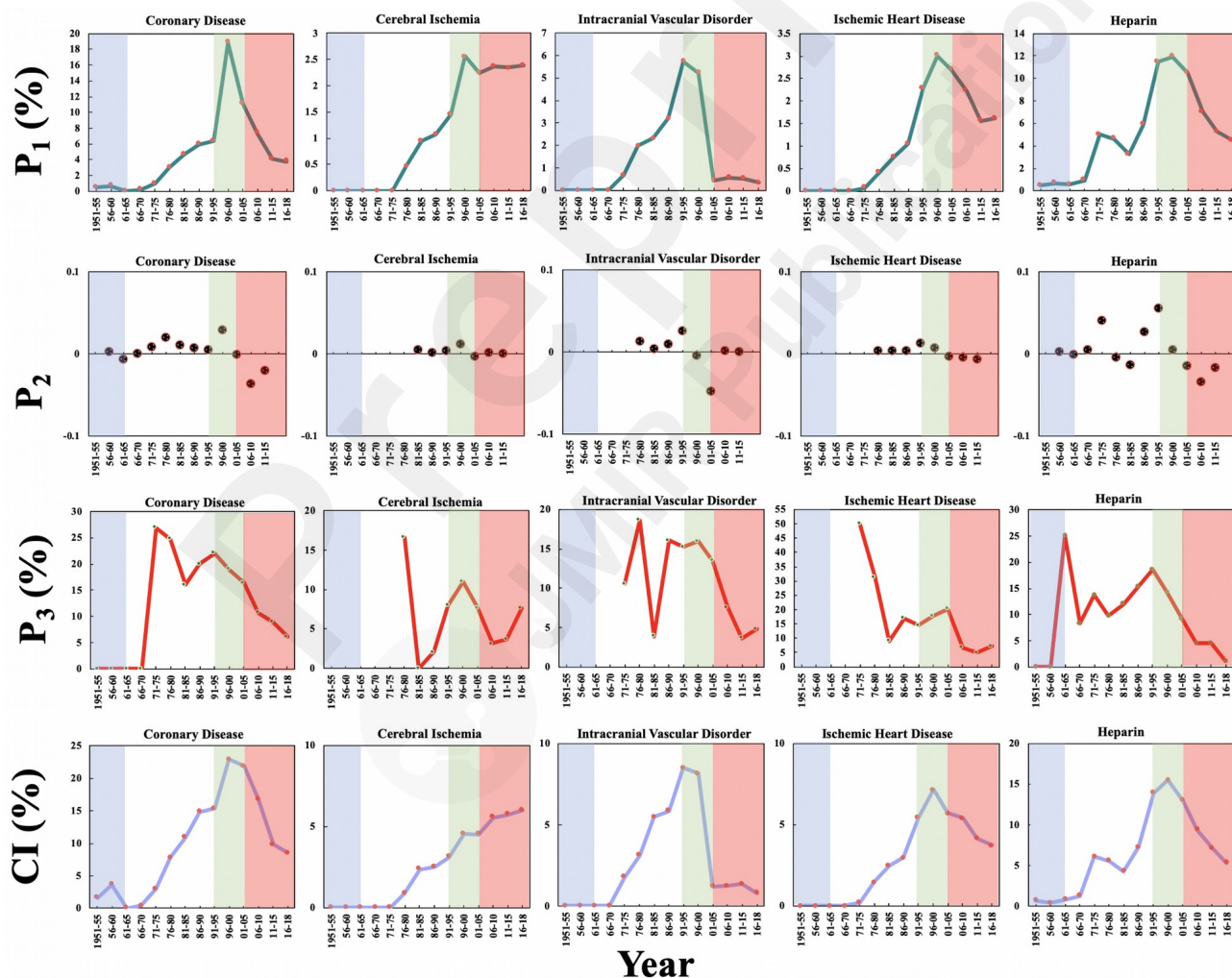


Figure 7 The P3C of biomedical entities on the cardiovascular diseases in aspirin domain over time. The background colors show the four phases of aspirin research

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 anti-platelet effect of aspirin in Phase 2.

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

The pattern of  $P_3$  for these five entities over time is displayed in the third row of Figure 7. All five 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 in 1961-1965. The difference from the previous phases is that the  $P_3$  of these five biomedical entities peaked again in Phase 3. For instance, “coronary disease” peaked in 1991-1995, and heparin in 1991-1995, indicating that these biomedical entities were important topics of research in both Phase 1 and Phase 3.

The last row of Figure 7 shows the CI of five biomedical entities during 1951-2018, in which the CI demonstrated a dynamic trajectory very similar to that of the Popularity Index. 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 Phase 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 four biomedical entities, which increased in the 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 its popularity and collaboration so there is still more increases expected to come.

## Repurposing aspirin for other diseases

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

Figure 8 also shows that the  $P_1$  of two 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 Muller discovered that aspirin had the effect of lowering blood sugar. Inspired by this observation, scientists have since been trying to use aspirin for the treatment of diabetes [74]. 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 [5,75]; 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 three drugs also peaked in Phase 4, including the anti-

platelet drugs clopidogrel and ticlopidine, which are competitors of aspirin as anti-platelet 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 [76,77].

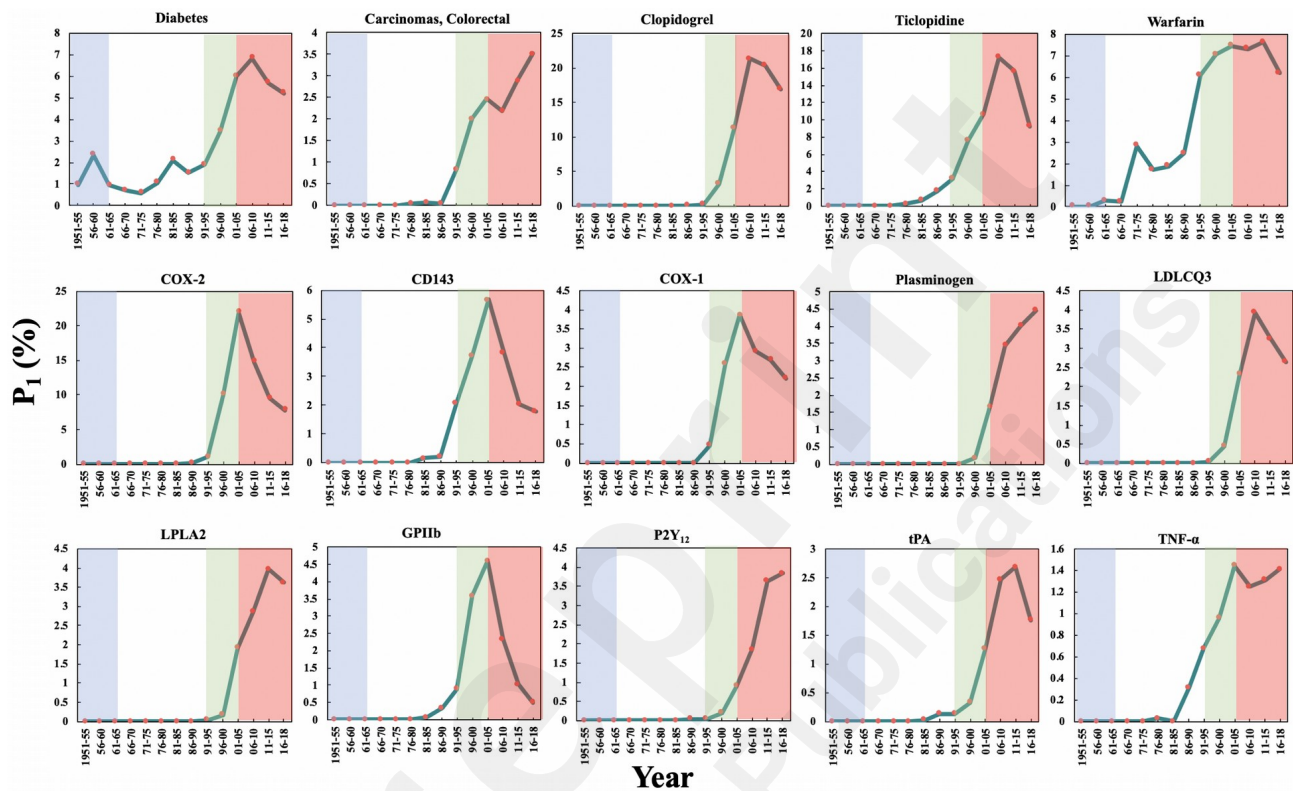


Figure 8 Popularity Index ( $P_1$ ) of biomedical entities on repurposing aspirin for other diseases over time. The background colors show the four phases of aspirin research.

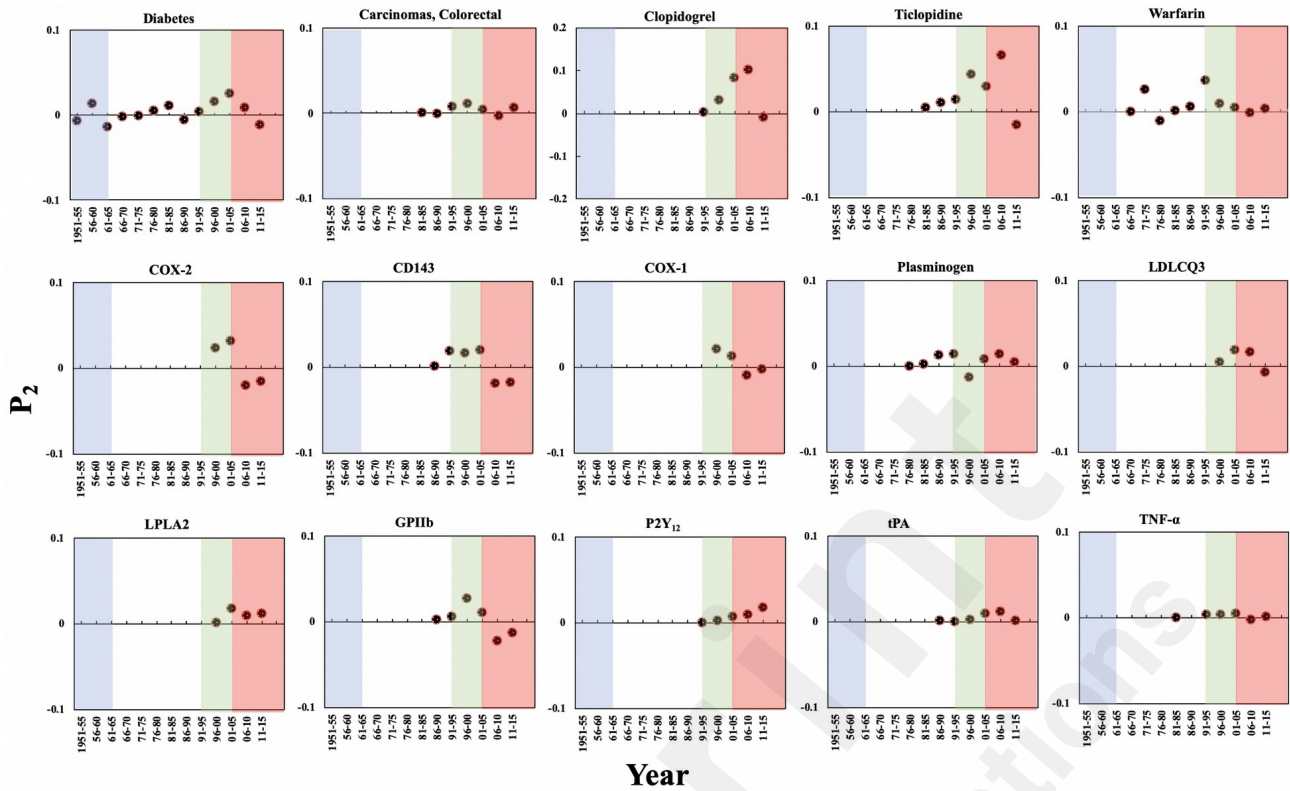


Figure 9 Promising Index ( $P_2$ ) of biomedical entities on repurposing aspirin for other diseases over time. The background colors show the four phases of aspirin research.

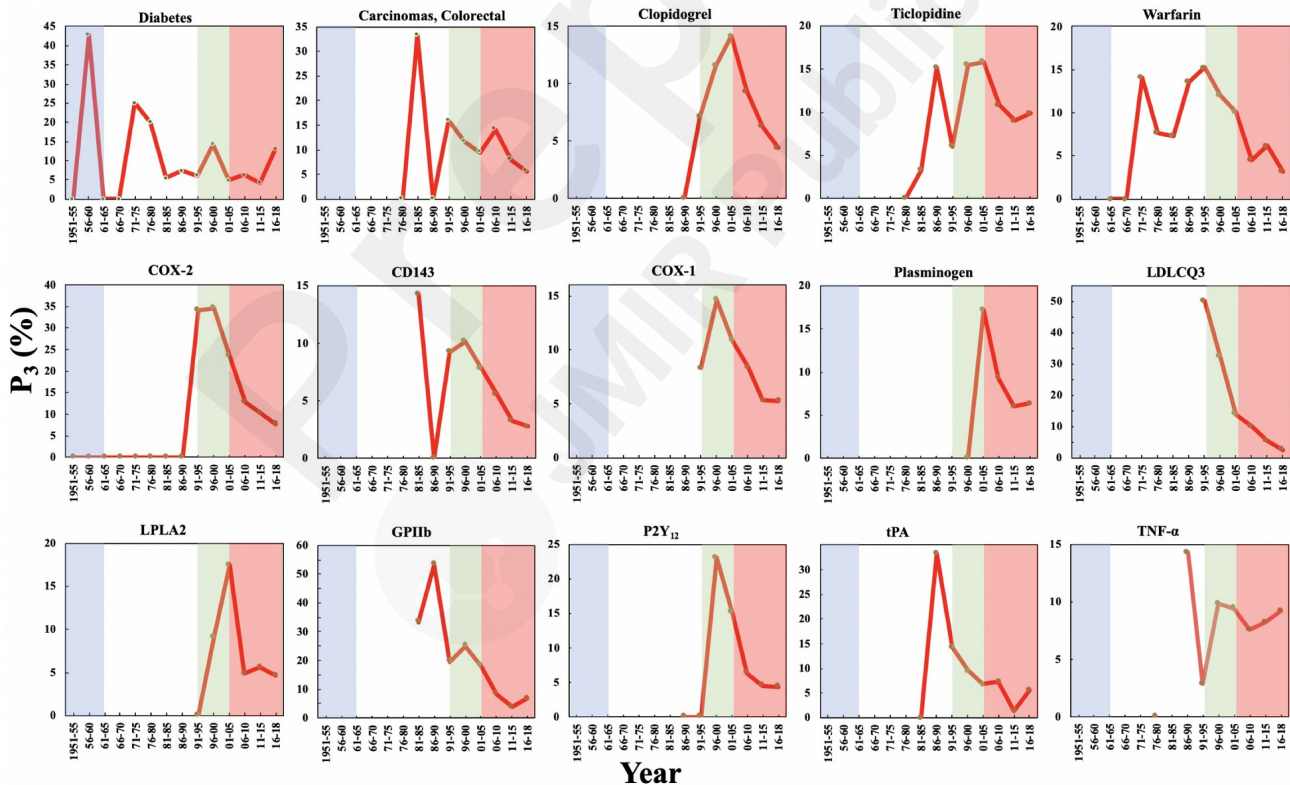


Figure 10 Prestige Index ( $P_3$ ) of biomedical entities on repurposing aspirin for other diseases over time. The background colors show the four phases of aspirin research.

Figure 9 presents the changes in  $P_2$  of these fifteen biomedical entities over time. All of the genes demonstrate a rise of more than 0 in Phase 4. Unlike these genes, the diseases and drugs

showed several significant rises of over 0 in different phases, which reflects a longer history of research in the aspirin domain. For example, 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 fifteen 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 two or more 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 the third or fourth phases, illustrating that these genes are relatively new topics in the aspirin domain.

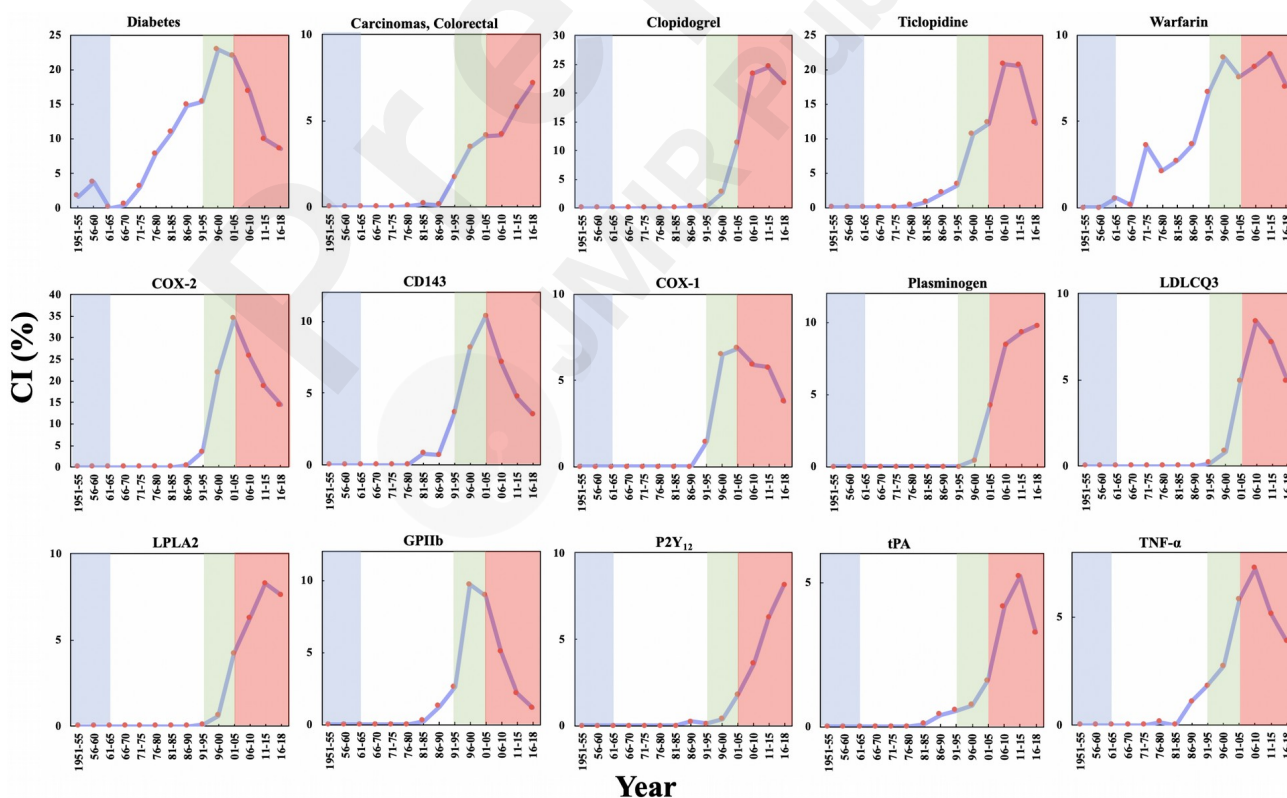


Figure 11 Collaboration Index (CI) of biomedical entities on repurposing aspirin for other diseases over time. The background colors show the four phases of aspirin research.

The CI data for these fifteen biomedical entities are presented in Figure 11, which shows that



the maximum of COX-2's CI is the highest, at 34.37%, in 2001-2015 for COX-2, denoting that COX-2 was the focus of aspirin research in Phase 4; the research and development of Vioxx, a selective COX-2 inhibitors with less side effects, may be one of the reasons [78]. The CI of two drugs, clopidogrel (25.54% in 2001-2015) and ticlopidine (20.74 in 2006-2010), reveals fierce competition between aspirin and these alternative anti-platelet drugs. This competition could have driven the repurposing of aspirin for other diseases, especially cancers, which have an urgent demand for effective treatment.

## 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 to identify potential patterns among biomedical entities in various phases of the drug research timeline. The main contribution of this paper is twofold. First, this paper proposes four entitymetric indices (P3C) to quantify changes in academic interest in biomedical entities and to reveal the fine-grained process of drug repurposing. Second, we divide aspirin research into four phases, including the 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 three fine-grained perspectives—disease, drug, and gene—which 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 rheumatoid arthritis [34], with a P3C which 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 rise of  $P_2$  for the relevant biomedical entities in 1961-1990. The anti-platelet effect [69], inhibition of prostaglandin synthesis [70] and the acetylation effect on the enzyme cyclo-oxygenase [71] were uncovered. These discoveries had laid a solid knowledge foundation for the successful repurposing of aspirin. The highest  $P_1$  in 1961-1990 was for indomethacin (16.75), denoting a fierce competition of aspirin for its original use. This could be one of the factors contribute to the repurposing of aspirin.

In Phase 3, aspirin was successfully used for several cardiovascular-related diseases because of its anti-platelet effect [79]. The related diseases and drugs achieved their highest values of  $P_1$  and CI, as well as significant rises in  $P_2$ , in 1991-2000. As these diseases are the most common diseases worldwide, according to WHO statistics [73], the demand for the prevention and treatment of the fatal diseases is also probably another factor driving drug repurposing. In the last phase, there were 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 level of genes. Ten genes had a maximum  $P_1$  and CI, as well as an apparent rise 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, two fatal diseases—diabetes and colorectal carcinoma—as well as three competitive drugs of aspirin as an anti-platelet agent—clopidogrel, ticlopidine and warfarin (an anticoagulant and competed with aspirin for stroke prevention)—also had peak  $P_1$  and CI values, as well as a great rise in  $P_2$ .

Methodologically, in this study we develop four entitymetrics, and demonstrate how to use them to investigate the process of drug repurposing. The results demonstrate that the maximums of  $P_1$ ,  $P_3$  and CI are closely associated with the different phases of research into 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 rises 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, the  $P_3$  can reflect a research focus far earlier than the other three indices, which means that a continuing high  $P_3$  may be valuable as an early signal of the emergence and transfer of research topics in drug research. If the Prestige Index 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 [80,81]. 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 EHR, 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 and pharmaceutical could affect the drug repurposing process. Second, there are several ways of measuring the impact of a journal, such as Impact Factor (IF) and Relative Citation Ratio (RCR). Third, this paper mainly focuses on investigating the repurposing journey of aspirin, but we didn't test whether it can be used for predicting future drug repurposing. In the future studies, we will evaluate different impact measures of a journal and choose a proper one fitting better to the chosen drug. Furthermore, we also aim to test the proposed metrics on the other drugs to understanding their repurposing journeys (e.g., metformin) to see whether there exist generalized patterns in different repurposing processes.

## Acknowledgements

This study was supported by the Major Project of the National Social Science Foundation of China (17&ZDA292). The support provided by China Scholarship Council (CSC) during a visit of 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.

## References

1. Schneider G. Automating drug discovery. *Nat Rev Drug Discov* 2018 Feb;17(2):97–113. [doi: 10.1038/nrd.2017.232]
2. Parrish MC, Tan YJ, Grimes K V, Mochly-Rosen D. Surviving in the Valley of Death: Opportunities and Challenges in Translating Academic Drug Discoveries. *Annu Rev Pharmacol Toxicol Annu Rev Pharmacol Toxicol* 2019;59:405–421. [doi: 10.1146/annurev-pharmtox]
3. Deftereos SN, Andronis C, Friedla EJ, Persidis A, Persidis A. Drug repurposing and adverse event prediction using high-throughput literature analysis. *Wiley Interdiscip Rev Syst Biol Med*. John Wiley & Sons, Ltd; 2011 Aug; 323–334. [doi: 10.1002/wsbm.147]
4. Daphna Laifenfeld C, Research G, Cha Y, Erez T, Reynolds IJ, Kumar D, Ross J, Koytiger G, Kusko R, Zeskind B, Risso S, Kagan E, Papapetropoulos S, Grossman I, Laifenfeld D. Drug repurposing from the perspective of pharmaceutical companies. *Br J Pharmacol* 2018;175:168–180. [doi: 10.1111/bph.v175.2/issuetoc]
5. Strittmatter SM. Overcoming Drug Development Bottlenecks With Repurposing: Old drugs learn new tricks. *Nat Med* Nature Publishing Group; 2014 Jun 5;20(6):590–591. [doi: 10.1038/nm.3595]
6. Corsello SM, Bittker JA, Liu Z, Gould J, McCarren P, Hirschman JE, Johnston SE, Vrcic A, Wong B, Khan M,

- Asiedu J, Narayan R, Mader CC, Subramanian A, Golub TR. The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat Med* 2017 Apr 1;23(4):405–408. [doi: 10.1038/nm.4306]
7. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Guillems T, Latimer J, McNamee C, Norris A, Sanseau P, Cavalla D, Pirmohamed M. Drug repurposing: Progress, challenges and recommendations. *Nat Rev Drug Discov* 2018; 41–58. [doi: 10.1038/nrd.2018.168]
  8. Fenig DM, McCullough A. Sildenafil in the treatment of erectile dysfunction. *Aging health* 2007 May 14;3(3):295–303. [doi: 10.2217/1745509X.3.3.295]
  9. Singhal S, Mehta J, Desikan R, Ayers D, Roberson P, Eddlemon P, Munshi N, Anaissie E, Wilson C, Dhodapkar M, Zeldis J, Siegel D, Crowley J, Barlogie B. Antitumor Activity of Thalidomide in Refractory Multiple Myeloma. *N Engl J Med* 2002 Nov 18;341(21):1565–1571. [doi: 10.1056/nejm199911183412102]
  10. Lee A, Morley JE. Metformin decreases food consumption and induces weight loss in subjects with obesity with type II non-insulin-dependent diabetes. *Obes Res* 1998 Jan 1;6(1):47–53. [doi: 10.1002/j.1550-8528.1998.tb00314.x]
  11. Berstein LM. Metformin in obesity, cancer and aging: Addressing controversies [Internet]. *Aging* 2012; 320–329. PMID:22589237
  12. Pascual J, Rivas MT, Leira R. Testing the combination beta-blocker plus topiramate in refractory migraine. *Acta Neurol Scand* 2007 Feb 1;115(2):81–83. [doi: 10.1111/j.1600-0404.2006.00772.x]
  13. Diener H-C, Tfelt-Hansen P, Dahlf C, Linez MJA, Sandrini G, Wang S-J, Neto W, Vijapurkar U, Doyle A, Jacobs D, StudyGroup on behalf of the M-003. Topiramate in migraine prophylaxis. *J Neurol* 2004 Aug;251(8):943–950. [doi: 10.1007/s00415-004-0464-6]
  14. Simsek M, Meijer B, van Bodegraven AA, de Boer NKH, Mulder CJJ. Finding hidden treasures in old drugs: the challenges and importance of licensing generics. *Drug Discov Today* 2018 Jan 1;23(1):17–21. [doi: 10.1016/J.DRUDIS.2017.08.008]
  15. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua M-S, So S, Butte AJ. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun* 2017 Dec 12;8(1):16022. [doi: 10.1038/ncomms16022]
  16. Xu M, Lee EM, Wen Z, Cheng Y, Huang W-K, Qian X, TCW J, Kouznetsova J, Ogden SC, Hammack C, Jacob F, Nguyen HN, Itkin M, Hanna C, Shinn P, Allen C, Michael SG, Simeonov A, Huang W, Christian KM, Goate A, Brennand KJ, Huang R, Xia M, Ming G, Zheng W, Song H, Tang H. 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. [doi: 10.1038/nm.4184]
17. Mons B, van Haagen H, Chichester C, Hoen P-B 't, den Dunnen JT, van Ommen G, van Mulligen E, Singh B, Hooft R, Roos M, Hammond J, Kiesel B, Giardine B, Velterop J, Groth P, Schultes E. The value of data. *Nat Genet* 2011 Apr 29;43(4):281–283. [doi: 10.1038/ng0411-281]
  18. Gao Z, Fu G, Ouyang C, Tsutsui S, Liu X, Yang J, Gessner C, Foote B, Wild D, Ding Y, Yu Q. edge2vec: Representation learning using edge semantics for biomedical knowledge discovery. *BMC Bioinformatics* 2019 Dec 10;20(1):306. [doi: 10.1186/s12859-019-2914-2]
  19. Chen B, Dong X, Jiao D, Wang H, Zhu Q, Ding Y, Wild DJ. Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data. *BMC Bioinformatics* 2010 May 17;11(1):255. [doi: 10.1186/1471-2105-11-255]
  20. Hamilton WL, Bajaj P, Zitnik M, Jurafsky D, Leskovec J. Embedding Logical Queries on Knowledge Graphs. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18)*, Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, and Nicolò Cesa-Bianchi (Eds.). Curran Associates Inc., USA, 2030-2041.
  21. Chang Y, Park H, Yang H-J, Lee S, Lee K-Y, Kim TS, Jung J, Shin J-M. Cancer Drug Response Profile scan (CDRscan): A Deep Learning Model That Predicts Drug Effectiveness from Cancer Genomic Signature. *Sci Rep* 2018 Dec 11;8(1):8857. [doi: 10.1038/s41598-018-27214-6]
  22. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* 2018 Sep 1;34(17):i821–i829. [doi: 10.1093/bioinformatics/bty593]
  23. Jinha AE. Article 50 million: an estimate of the number of scholarly articles in existence. *Learn Publ* 2010 Jul 1;23(3):258–263. [doi: 10.1087/20100308]
  24. Ding Y, Stirling K. Data-driven Discovery: A new era of exploiting the literature and data. *J Data Inf Sci* 2016;1(4):1–9. [doi: 10.20309/jdis.201622]
  25. Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect Biol Med* 1986;30(1):7–18. PMID:3797213
  26. Swanson DR. Undiscovered Public Knowledge. *Libr Q [Internet]* University of Chicago Press ; 1986 Apr 29;56(2):103–118. [doi: 10.1086/601720]
  27. Swanson DR. Two medical literatures that are logically but not bibliographically connected. *J Am Soc Inf Sci* 1987 Jul 1;38(4):228–233. [doi: 10.1002/(SICI)1097-4571(198707)38:4<228::AID-ASI2>3.0.CO;2-G]
  28. Cory KA. Discovering Hidden Analogies in an Online Humanities Database. *Comput Hum* 1997;31(1):1–12.

[doi: 10.1023/A:1000422220677]

29. Ding Y, Song M, Han J, Yu Q, Yan E, Lin L, Chambers T. Entitymetrics: Measuring the Impact of Entities. Bar-Ilan J, editor. PLoS One 2013 Aug 29;8(8):e71416. [doi: 10.1371/journal.pone.0071416]
30. Wang H, Ding Y, Tang J, Dong X, He B, Qiu J, Wild DJ. Finding Complex Biological Relationships in Recent PubMed Articles Using Bio-LDA. Langowski J, editor. PLoS One 2011 Mar 23;6(3):e17243. [doi: 10.1371/journal.pone.0017243]
31. Song M, Han N-G, Kim Y-H, Ding Y, Chambers T. Discovering Implicit Entity Relation with the Gene-Citation-Gene Network. Hayasaka S, editor. PLoS One 2013 Dec 17;8(12):e84639. [doi: 10.1371/journal.pone.0084639]
32. Ding Y, Zhang G, Chambers T, Song M, Wang X, Zhai C. Content-based citation analysis: The next generation of citation analysis. J Assoc Inf Sci Technol 2014 Sep 1;65(9):1820–1833. [doi: 10.1002/asi.23256]
33. Kissin I, Edwin L B. Top journals selectivity index: is it acceptable for drugs beyond the field of analgesia? Scientometrics 2011 August 1; 88(2): 589–597. [doi: 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 1;113:1–8. [doi: 10.1016/J.VPH.2018.10.008]
35. Bordons M, Bravo C, Barrigó S, Barrigón S. Time-Tracking of the Research Profile of a Drug Using Bibliometric Tools. J Am Soc Inf Sci Technol 2004;55(5):445–461. [doi: 10.1002/asi.10397]
36. Gilligan MM, Gartung A, Sulciner ML, Norris PC, Sukhatme VP, Bielenberg DR, Huang S, Kieran MW, Serhan CN, Panigrahy D. Aspirin-triggered proresolving mediators stimulate resolution in cancer. Proc Natl Acad Sci U S A 2019 Mar 26;116(13):6292–6297. PMID:30862734
37. Volz J, Mammadova-Bach E, Gil-Pulido J, Nandigama R, Remer K, Sorokin L, Zernecke A, Abrams SI, Ergün S, Henke E, Nieswandt B. Inhibition of platelet GPVI induces intratumor hemorrhage and increases efficacy of chemotherapy in mice. Blood 2019 Jun 20;133(25):2696–2706. PMID:30952674
38. Jackson SS, Zhu B, Pfeiffer R, Liu Z, Gadalla S, Koshiol J. Abstract 2333: Aspirin may extend biliary tract cancer survival: Results from population-based cohort. Clin Res 2019. p. 2333–2333. [doi: 10.1158/1538-7445.SABCS18-2333]
39. Baker NC, Ekins S, Williams AJ, Tropsha A. A bibliometric review of drug repurposing. Drug Discov Today 2018;23(3):661–672. PMID:29330123
40. Liu T-P, Hsieh Y-Y, Chou C-J, Yang P-M. Systematic polypharmacology and drug repurposing via an integrated L1000-based Connectivity Map database mining. R Soc open Sci 2018 Nov;5(11):181321. PMID:30564416
41. Kidnapillai S, Bortolasci CC, Udawela M, Panizzutti B, Spolding B, Connor T, Sanigorski A, Dean OM, Crowley

- T, Jamain S, Gray L, Scarr E, Leboyer M, Dean B, Berk M, Walder K. The use of a gene expression signature and connectivity map to repurpose drugs for bipolar disorder. *World J Biol Psychiatry* 2018 Aug 3;1–9. [doi: 10.1080/15622975.2018.1492734]
42. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007 Jun;447(7145):661–678. [doi: 10.1038/nature05911]
  43. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, Mooser V. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol* 2012 Apr 10;30(4):317–320. [doi: 10.1038/nbt.2151]
  44. Ferrero E, Agarwal P. Connecting genetics and gene expression data for target prioritisation and drug repositioning. *BioData Min* 2018 Dec 31;11(1):7. PMID:29881461
  45. Dakshanamurthy S, Issa NT, Assefnia S, Seshasayee A, Peters OJ, Madhavan S, Uren A, Brown ML, Byers SW. Predicting New Indications for Approved Drugs Using a Proteochemometric Method. *J Med Chem* 2012 Aug 9;55(15):6832–6848. PMID:22780961
  46. Li YY, An J, Jones SJM. A Computational Approach to Finding Novel Targets for Existing Drugs. Bourne PE, editor. *PLoS Comput Biol* 2011 Sep 1;7(9):e1002139. [doi: 10.1371/journal.pcbi.1002139]
  47. 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. PMID:22960403
  48. Yang L, Agarwal P. Systematic Drug Repositioning Based on Clinical Side-Effects. Csermely P, editor. *PLoS One* 2011 Dec 21;6(12):e28025. [doi: 10.1371/journal.pone.0028025]
  49. Ye H, Liu Q, Wei J. Construction of Drug Network Based on Side Effects and Its Application for Drug Repositioning. *PLoS One* 2014 Feb 4;9(2):e87864. [doi: 10.1371/journal.pone.0087864]
  50. 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 1;34(7):1164–1173. PMID:29186331
  51. 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. PMID:23933754
  52. 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 Jun 13;14(6):e0218264. [doi: 10.1371/journal.pone.0218264]
  53. Chen B, Ding Y, Wild DJ. Assessing Drug Target Association Using Semantic Linked Data. *PLoS Comput Biol* 2012;8(7):e1002574. [doi: 10.1371/journal.pcbi.1002574]
  54. 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 Dec 12;17(1):160. [doi: 10.1186/s12859-016-1005-x]
55. 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]
56. 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 Elsevier*; 2010 Jul 1;49(3):145–154. [doi: 10.1016/J.ARTMED.2010.03.002]
57. Cameron D, Bodenreider O, Yalamanchili H, Danh T, Vallabhaneni S, Thirunarayan K, Sheth AP, Rindfleisch TC. A graph-based recovery and decomposition of Swanson’s hypothesis using semantic predications. *J Biomed Inform* 2013 Apr 1;46(2):238–251. [doi: 10.1016/J.JBI.2012.09.004]
58. Song M, Heo GE, Ding Y. SemPathFinder: Semantic path analysis for discovering publicly unknown knowledge. *J Informetr* 2015 Oct 1;9(4):686–703. [doi: 10.1016/J.JOI.2015.06.004]
59. Lv Y, Ding Y, Song M, Duan Z. Topology-driven trend analysis for drug discovery. *J Informetr* 2018;12(3):893–905. [doi: 10.1016/j.joi.2018.07.007]
60. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, Adamjee S, Dame ZT, Han B, Zhou Y, Wishart DS. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 2014 Jan 1;42(D1):D1091–D1097. [doi: 10.1093/nar/gkt1068]
61. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* 2004 Jan 1;32(90001):258D – 261. [doi: 10.1093/nar/gkh036]
62. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res* 2016 Jan 4;44(D1):D1075–D1079. [doi: 10.1093/nar/gkv1075]
63. He B, Tang J, Ding Y, Wang H, Sun Y, Shin JH, Chen B, Moorthy G, Qiu J, Desai P, Wild DJ. Mining relational paths in integrated biomedical data. *PLoS One* 2011;6(12):e0027506. [doi: 10.1371/journal.pone.0027506]
64. Williams RS, Lotia S, Holloway AK, Pico AR. From Scientific Discovery to Cures: Bright Stars within a Galaxy. *Cell* 2015;163(1):21–23. [doi: 10.1016/j.cell.2015.09.007]
65. 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. [doi: 10.1371/journal.pone.0156091]
66. Lerchenmueller MJ, Sorenson O. Author disambiguation in PubMed: Evidence on the precision and recall of author-ity among NIH-funded scientists. *PLoS One* 2016 Jul 1; 11(7):e0158731. [doi: 10.1371/journal.pone.0158731]
67. Lee S, Kim L, Lee K, Choi J, Kim S, Jeon M, Lim S, Choi D, Kim S, Tan A-C, Kang J. BEST: Next-Generation

- Biomedical Entity Search Tool for Knowledge Discovery from Biomedical Literature. 2016; 11(10): e0164680. [doi: 10.1371/journal.pone.0164680]
68. Kissin I. What Can Big Data on Academic Interest Reveal about a Drug? Reflections in Three Major US Databases. *Trends Pharmacol Sci*. 2018; 39(3): 248–257. PMID:29358009
  69. O'Brien JR. Effects of Salicylates on Human Platelets. *Lancet* 1968 Apr 13;291(7546):779–783. [doi: 10.1016/S0140-6736(68)92228-9]
  70. VANE JR. Inhibition of Prostaglandin Synthesis as a Mechanism of Action for Aspirin-like Drugs. *Nat New Biol* 1971 Jun;231(25):232–235. [doi: 10.1038/newbio231232a0]
  71. Roth GJ, Stanford N, Majerus PW. Acetylation of prostaglandin synthase by aspirin. *Proc Natl Acad Sci U S A* 1975 Aug 1;72(8):3073–6. PMID:810797
  72. 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 1;48(15):4399–404. PMID:3390835
  73. World Health Organization. The top 10 causes of death [Internet]. [cited 2019 Jul 16]. p. 2018-05–24. Available from: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
  74. Ebstein W, Müller J. Brenzkatechin in dem Urin eines Kindes. *Arch für Pathol Anat und Physiol und für Klin Med* [Internet] Springer-Verlag; 1875 Feb [cited 2019 Aug 31];62(4):554–560. [doi: 10.1007/BF01928660]
  75. Pignone M, Alberts MJ, Colwell JA, Cushman M, Inzucchi SE, Mukherjee D, Rosenson RS, Williams CD, Wilson PW, Kirkman MS, Williams CD, Wilson PW, Kirkman MS. Aspirin for Primary Prevention of Cardiovascular Events in People With Diabetes. *J Am Coll Cardiol* 2010 Jun 22;55(25):2878–2886. [doi: 10.1016/j.jacc.2010.04.003]
  76. van Walraven C, Hart RG, Singer DE, Laupacis A, Connolly S, Petersen P, Koudstaal PJ, Chang Y, Hellemons B. Oral anticoagulants vs aspirin in nonvalvular atrial fibrillation: an individual patient meta-analysis. *JAMA* 2002 Nov;288(19):2441–8. PMID:12435257
  77. Mohr JP, Thompson JLP, Lazar RM, Levin B, Sacco RL, Furie KL, Kistler JP, Albers GW, Pettigrew LC, Adams HP, Jackson CM, Pullicino P. 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]
  78. Couzin J. Drug safety. Withdrawal of Vioxx casts a shadow over COX-2 inhibitors. *Science* 2004 Oct 15;306(5695):384–5. PMID:15486258
  79. 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 1;85(3):265–71. PMID:11179262

80. Kissin I. Can a bibliometric indicator predict the success of an analgesic? *Scientometrics* 2011;86(3):785–795. [doi: 10.1007/s11192-010-0320-7]
81. Koenig MED. Dterminants of Expert Judgement of Research Performance. *Scientometrics*. 1982;4(5):361-378. [doi: 10.1007/BF02135122]



# Supplementary Information

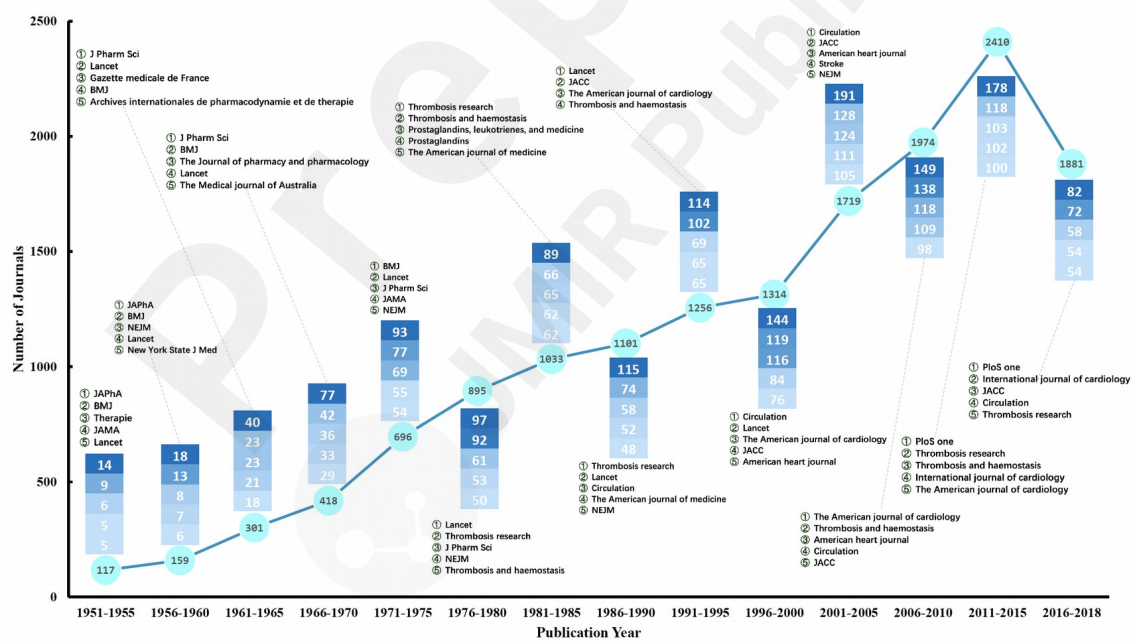


Figure S1 Changes in the number of journals with aspirin-related publications during 1951-2018. The top five journals and their frequencies are indicated using heatmaps for every five-year period.

Table S1 Top 20 journals related to aspirin research.

NO	Name
1	THE NEW ENGLAND JOURNAL OF MEDICINE
2	LANCET
3	JAMA
4	BRITISH MEDICINE JOURNAL
5	NATURE
6	NATURE REVIEWS. DRUG DISCOVERY
7	NATURE MEDICINE
8	SCIENCE
9	PHARMACOLOGICAL REVIEWS
10	ANNALS OF INTERNAL MEDICINE
11	THE JOURNAL OF CLINICAL INVESTIGATION
12	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCE OF THE UNITED STATES OF AMERICA
13	TRENDS IN THE PHARMACOLOGICAL SCIENCES
14	THE JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS.
15	CIRCULATION
16	JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
17	BRITISH JOURNAL OF HAEMATOLOGY
18	BLOOD
19	EUROPEAN HEART JOURNAL
20	JOURNAL OF THROMBOSIS AND HAEMOSTASIS

Table S2 The entitymetric results for the top 30 biomedical entities

Time Period	Phase 1 (1951-1960)				Phase 2 (1961-1990)				Phase 3 (1991-2000)				Phase 4 (2001-2018)			
	$P_1$ (%)	$P_2 > 0$	$P_3$ (%)	$CI$ (%)	$P_1$ (%)	$P_2 > 0$	$P_3$ (%)	$CI$ (%)	$P_1$ (%)	$P_2 > 0$	$P_3$ (%)	$CI$ (%)	$P_1$ (%)	$P_2 > 0$	$P_3$ (%)	$CI$ (%)
coronary disease						(71-90)	26.92 (71-75)		18.88 (96-00)	(96-00)	22.02 (91-95)	22.91 (96-00)				
asthma				9.84 (51-55)	4.68 (86-90)	(61-65) (71-80)	26.92 (66-70)	10.08 (86-90)			1.79 (96-00)					
diabetes	2.34 (56-60)	(56-60)	42.86 (56-60)	3.68 (56-60)	2.12 (81-85)	(81-85)	25.00 (71-75)			(96-00)	14.22 (96-00)	22.91 (96-00)	6.83 (06-10)	(01-10)		
hypersensitiviti es, drug			33.33 (56-60)		4.65 (66-70)	(61-70)	13.56 (66-70)	13.28 (61-65)							5.37 (01-05)	
ulcer, gastric			50 (51-55)	9.85 (61-65)	3.73 (81-85)	(61-65) (76-80)	12.5 (61-65)				8.45 (91-95)					
cerebral ischemia							16.67 (76-80)		2.57 (96-00)	(96-00)	11.04 (96-00)					5.99 (16-18)
intracranial vascular disorder							18.67 (76-80)		5.73 (91-95)	(91-95)	15.96 (96-00)	8.50 (91-95)				
ischemic heart disease							50 (71-75)		3.01 (96-00)	(91-00)		7.14 (96-00)			20.18 (01-05)	
carcinomas, colorectal							33.33 (81-85)			(91-00)	15.91 (91-95)		3.52 (16-18)		14.15 (06-10)	7.15 (16-18)
rheumatoid arthritis	9.36 (56-60)	(51-55)	25 (56-60)	40.44 (56-60)			16.57 (71-75)								5.05 (01-05)	
clopidogrel										(96-00)			21.32 (06-10)	(01-10)	14.13 (01-05)	24.54 (11-15)
ticlopidine						(86-90)	15.19 (86-90)			(91-00)			17.24 (06-10)	(01-10)	15.83 (01-05)	20.74 (06-10)
heparin			25 (61-65)			(71-75) (86-90)	13.67 (71-75)	6.08 (71-75)	11.92 (96-00)	(91-95)	18.58 (91-95)	15.50 (96-00)				
indomethacin			44.44 (61-65)		16.75 (76-80)	(66-80)		19.79 (76-80)								
warfarin						(71-75) (86-90)	14.10 (71-75)	3.59 (71-75)		(91-00)	15.23 (91-95)	8.69 (96-00)	7.62 (11-15)		6.16 (11-15)	8.89 (11-15)
vitamin f			15.87 (71-75)		11.19 (81-85)	(76-85)		12.90 (81-85)			8.46 (96-00)					
dipyridamole					9.42 (81-85)	(76-85)	28.57 (66-70)	11.80 (81-85)			13.54 (91-95)				7.6 (06-10)	
adenosine					5.64 (71-75)	(71-75)	42.85 (66-70)	6.95 (71-75)			18.54 (96-00)			(06-15)		
acetaminophen					6.07 (81-85)	(61-70)	19.59 (71-75)	5.90 (81-85)			6.02 (91-95)		3.51 (01-05)		3.87 (01-05)	

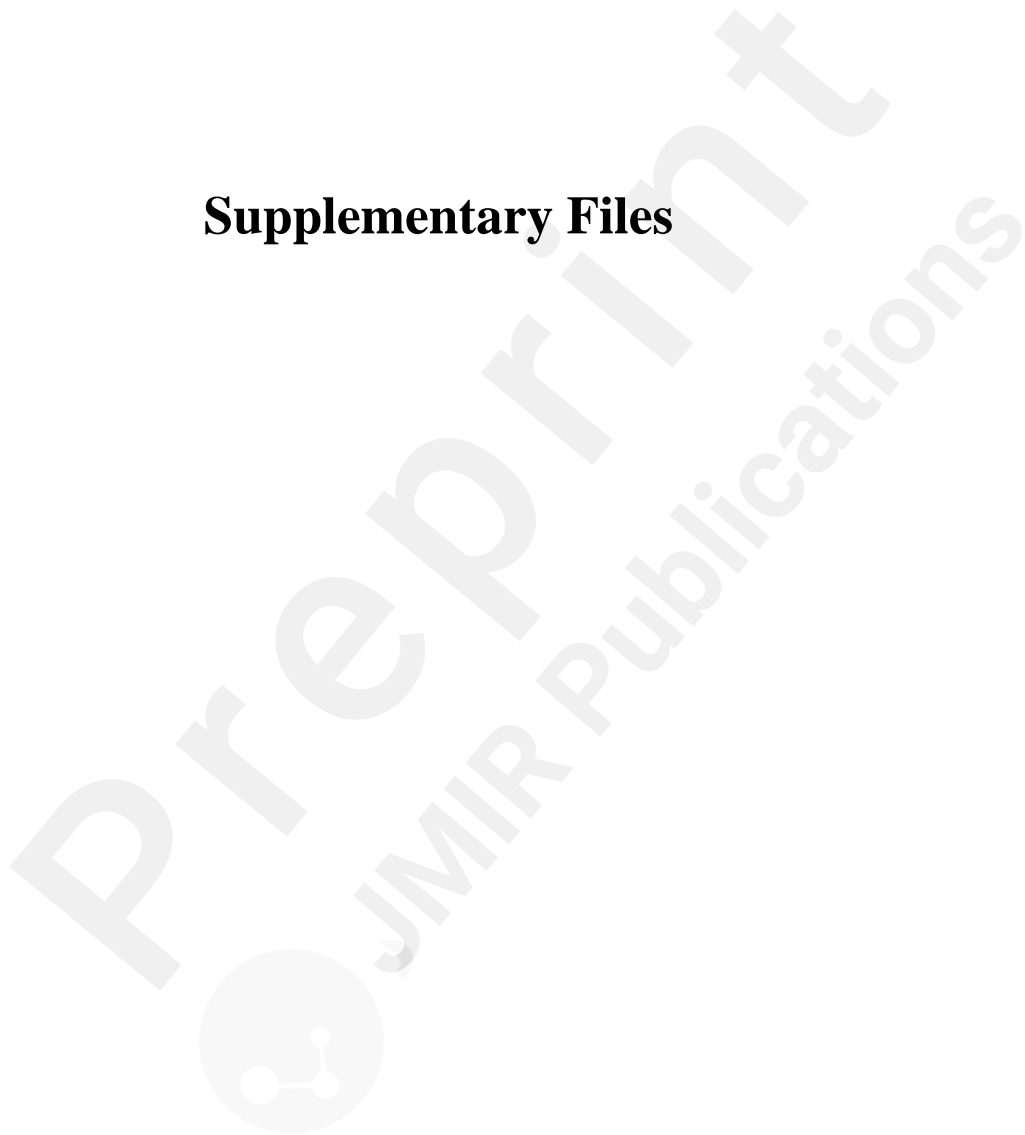
prostacyclin					8.42 (81-85)	(81-85)	20.42 (76-80)	9.32 (86-90)							10 (11-15)	
COX-2									(96-00)	34.58 (96-00)		21.97 (01-05)	(01-05)			34.37 (01-05)
CD143									(91-00)	10.17 (96-00)		5.68 (01-05)	(01-05)			10.36 (01-05)
COX-1									(96-00)	14.63 (96-00)		3.86 (01-05)	(01-05)			6.88 (01-05)
Plasminogen						(86-90)			(91-95)			4.45 (16-18)	(01-10)	17.16 (01-05)		9.79 (16-18)
LDLCQ3										50 (91-95)		3.93 (06-10)	(01-10)			8.38 (06-10)
LPLA2												3.97 (11-15)	(01-15)	17.41 (01-05)		8.27 (11-15)
GPIIb							53.33 (86-90)		(91-00)	25 (96-00)	9.72 (96-00)	4.59 (01-05)	(01-05)			
P2Y <sub>12</sub>										23.07 (96-00)		3.84 (16-18)	(01-15)			8.17 (16-18)
tPA							33.33 (86-90)					2.68 (11-15)	(01-10)	7.36 (06-10)		5.22 (11-15)
TNF- $\alpha$							14.28 (86-90)		1.44 (01-05)	9.84 (96-00)			(01-05)			7.27 (06-10)

Table S2 summarized the peaks of the P1, P3 and CI as well as the rise of P2 for all the top 30 biomedical entities. The details can be found in the supplementary information section, including the phase 2 (1961-1990, the scientific basis for repurposing), the phase 3 (1991-2000, repurposing aspirin for cardiovascular-related diseases) and phase 4 (2001-2018, repurposing aspirin for other diseases).

Preprint  
JMIR Publications



## Supplementary Files

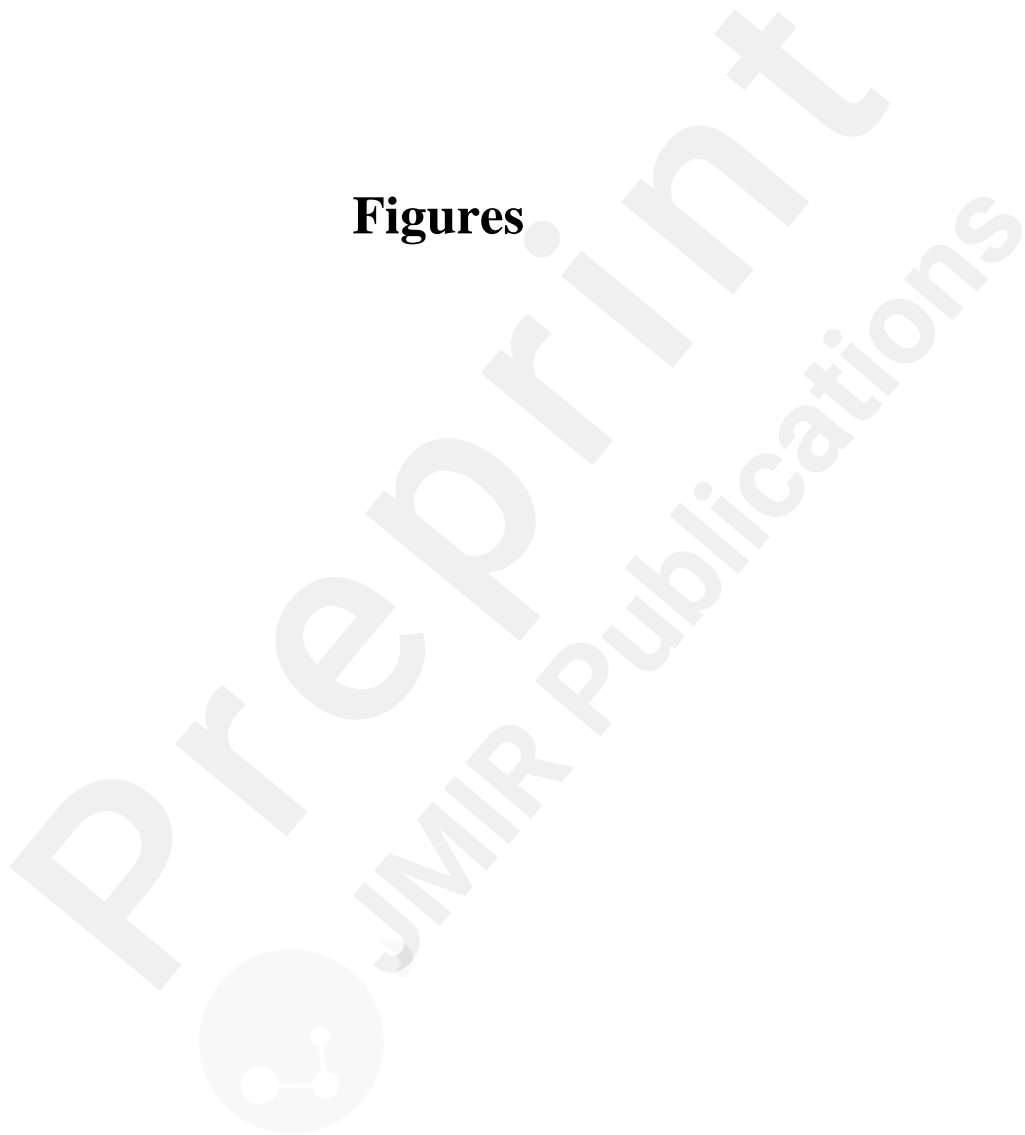


## Multimedia Appendixes

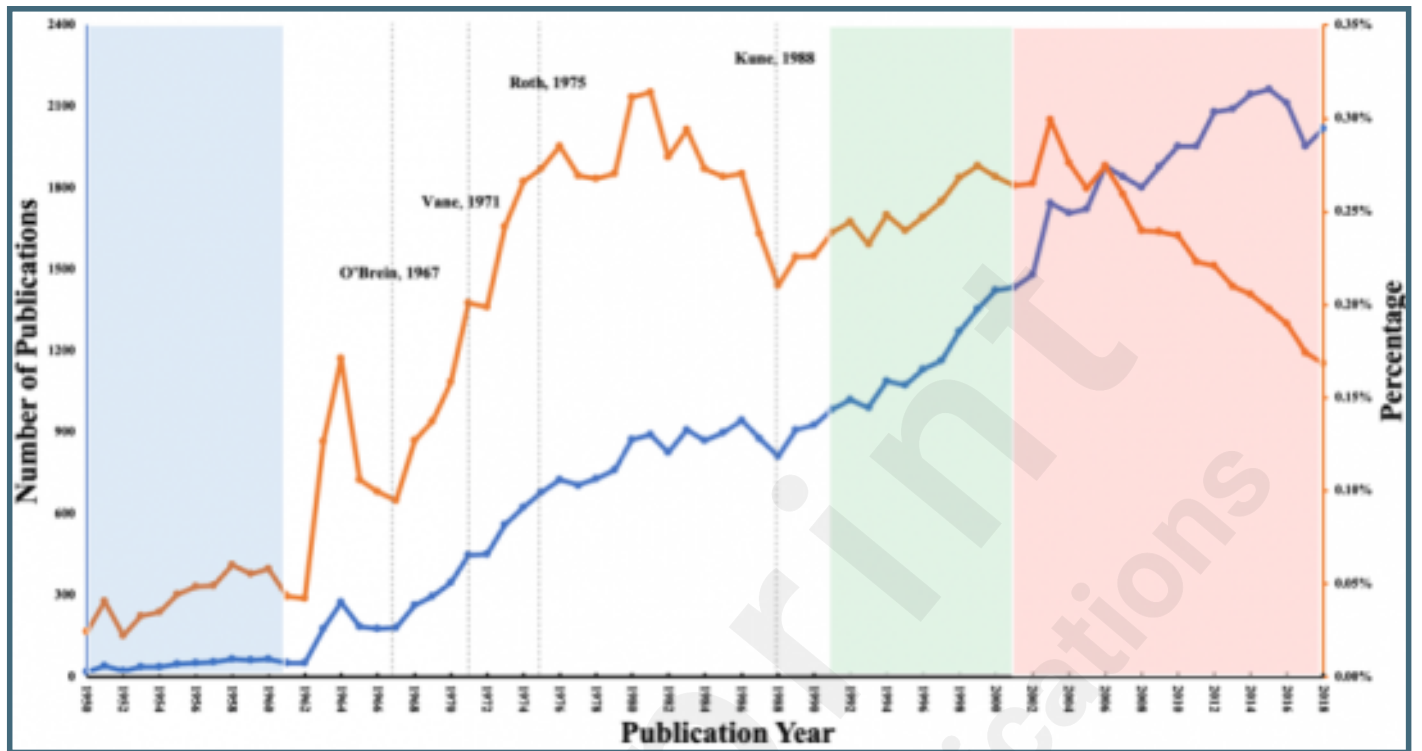
Table S2 summarizes the peaks of P1, P3, and CI as well as the increase in P2 for all the top 30 bioentities. The details can be found in the supplementary information section, including Phase 2 (1961-1990, the scientific basis for repurposing), Phase 3 (1991-2000, repurposing aspirin for cardiovascular-related diseases), and Phase 4 (2001-2018, repurposing aspirin for other diseases).

URL: <https://asset.jmir.pub/assets/2abde2ebb66857c7e0a34afb81ea1a0f.docx>

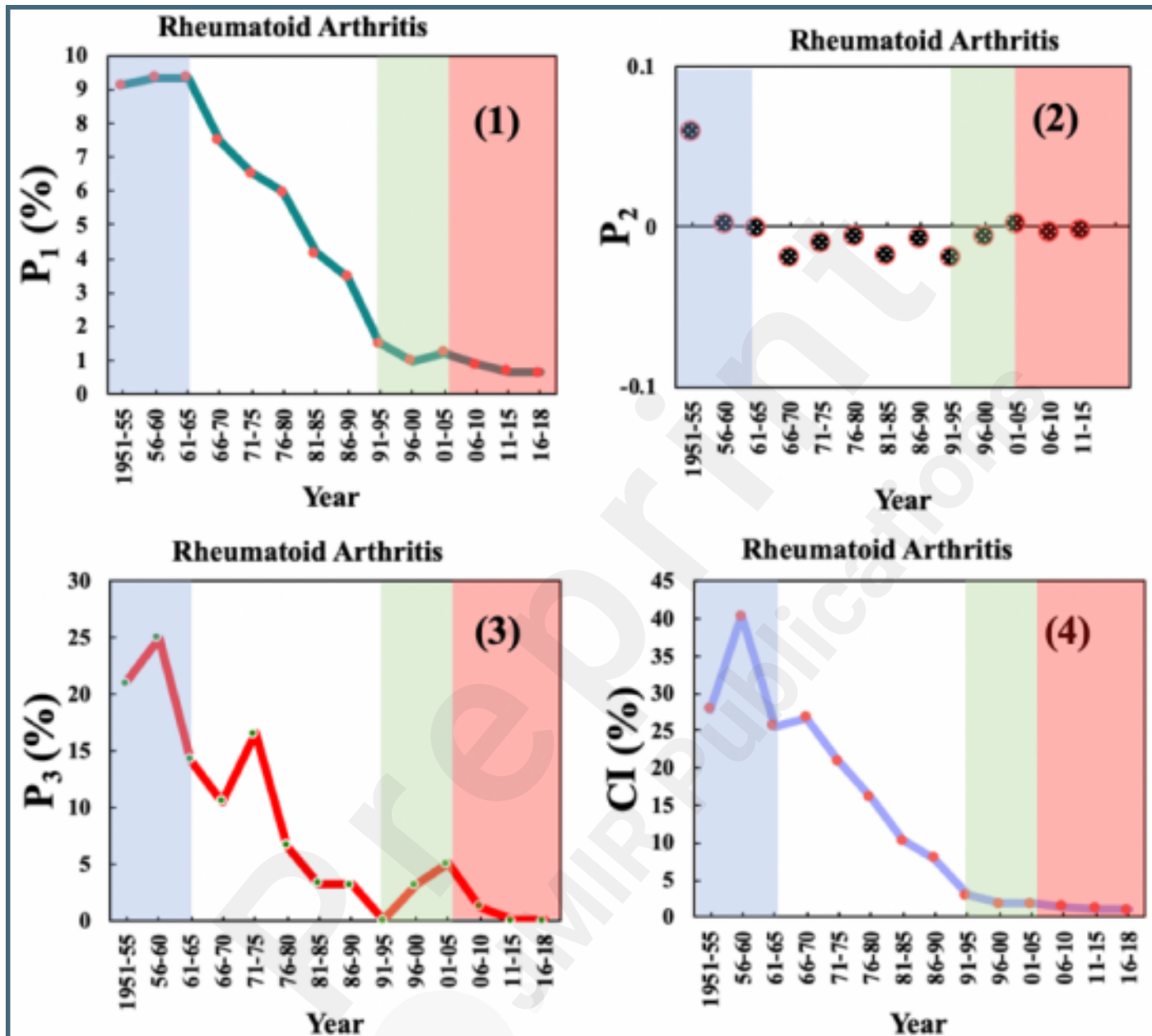
## Figures



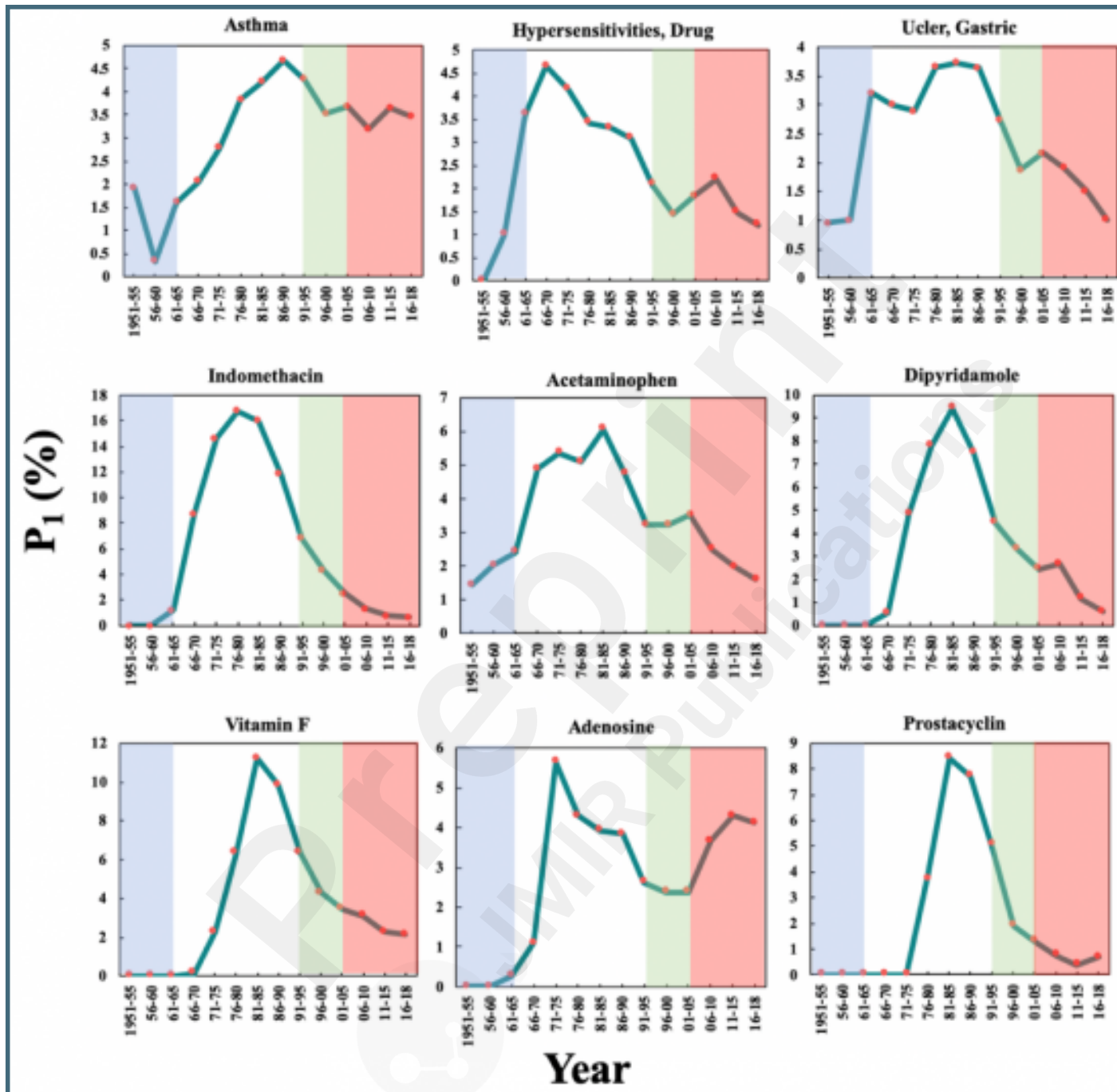
Number of aspirin-related studies in PubMed over time. The background colors indicate the 4 phases of aspirin research.



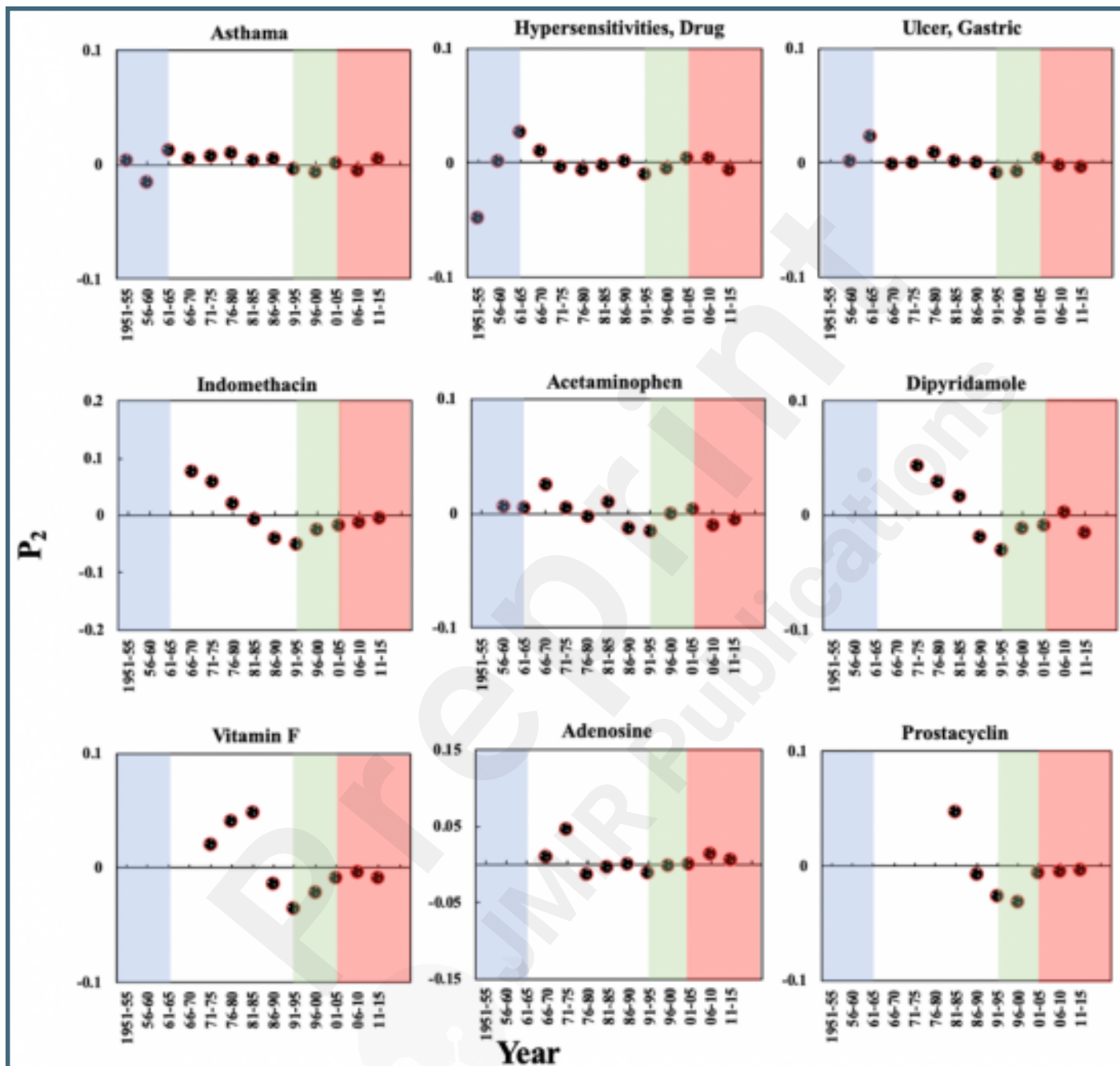
The 4 entitymetric indexes of the biomedical entity ‘‘Rheumatoid Arthritis’’ over time. The background colors indicate the 4 phases of aspirin research.



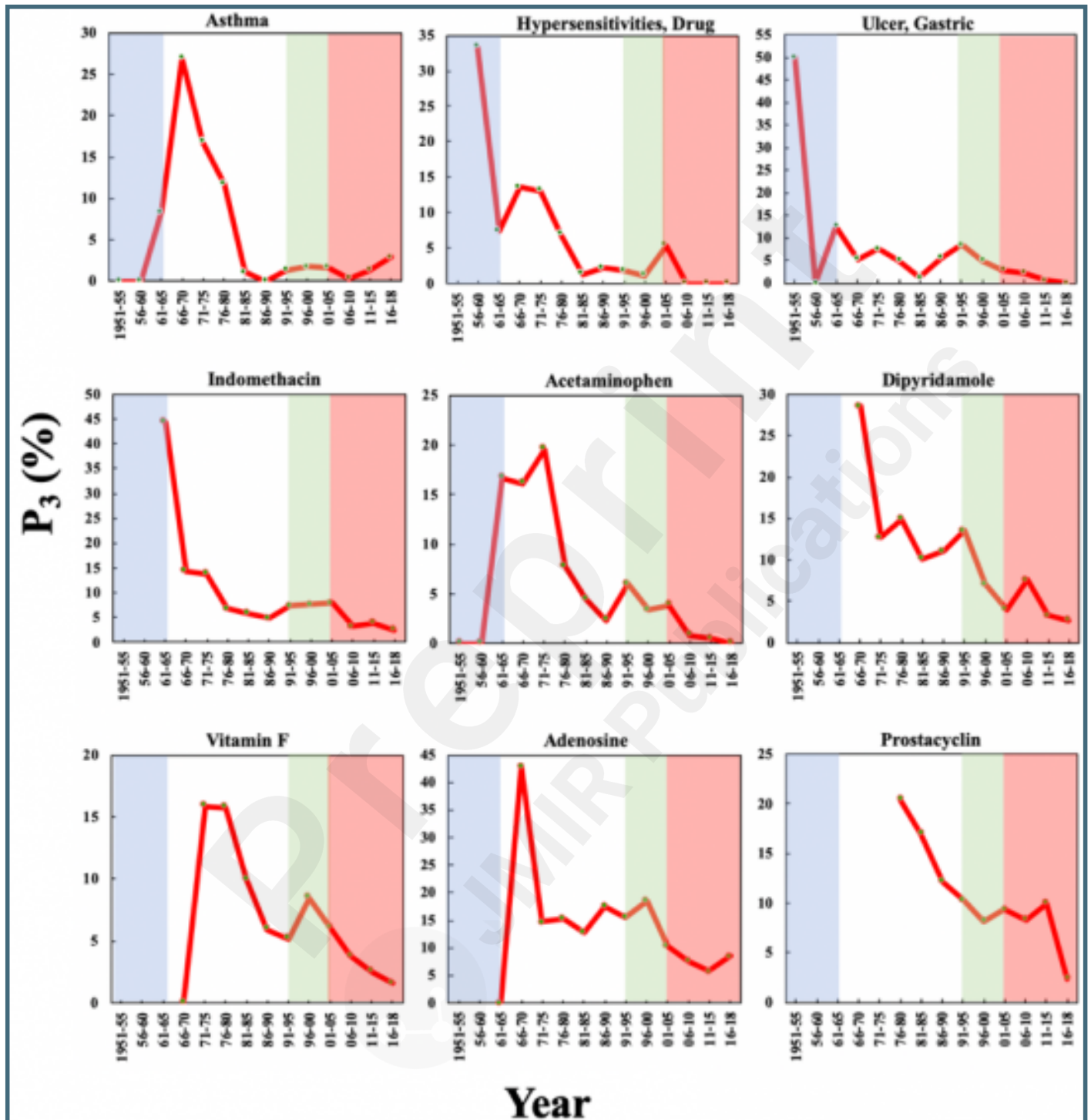
The Popularity Index ( $P_1$ ) of the biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the 4 phases of aspirin research.



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.

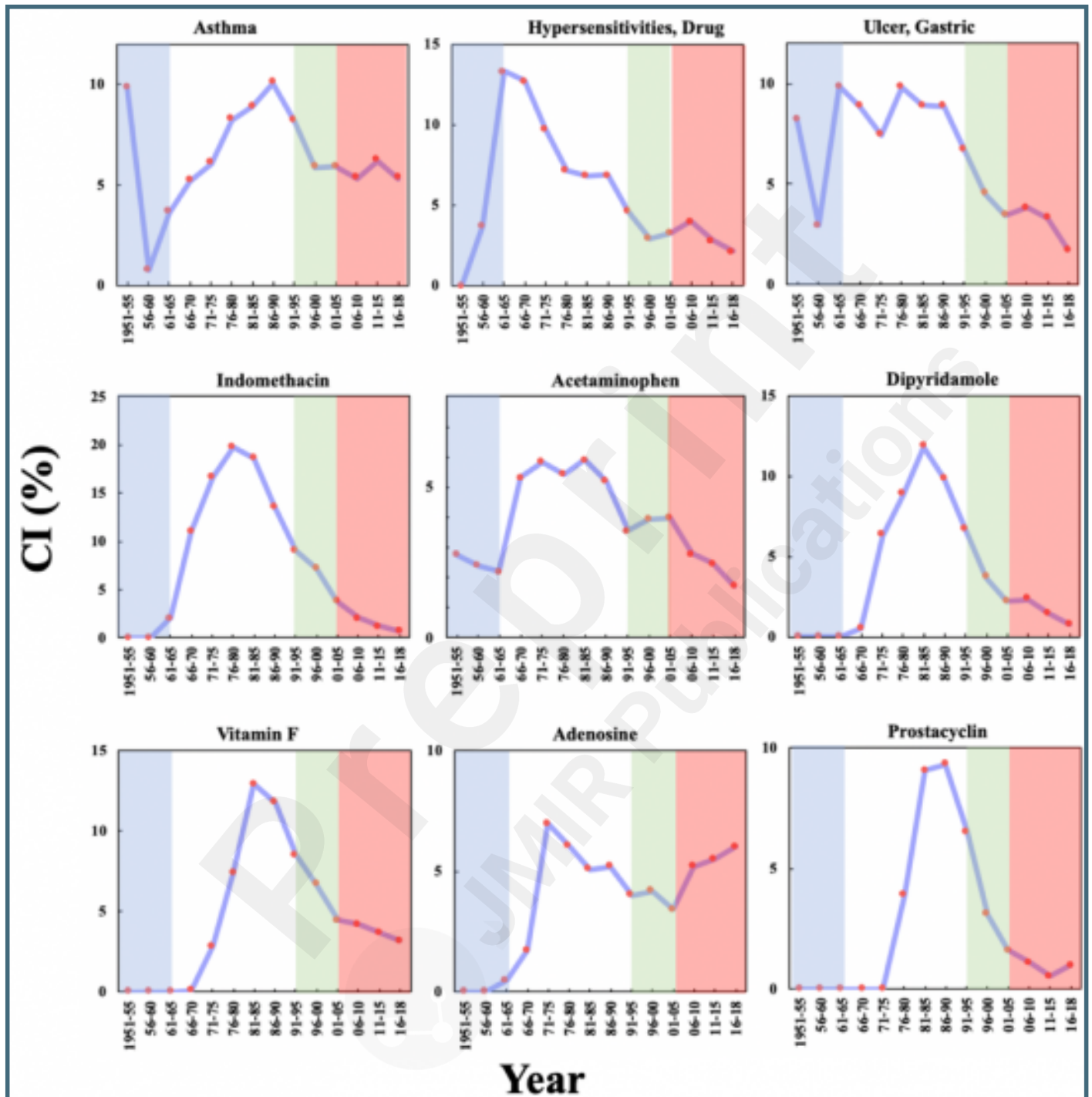


The Prestige Index ( $P_3$ ) of the biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the 4 phases of aspirin research.

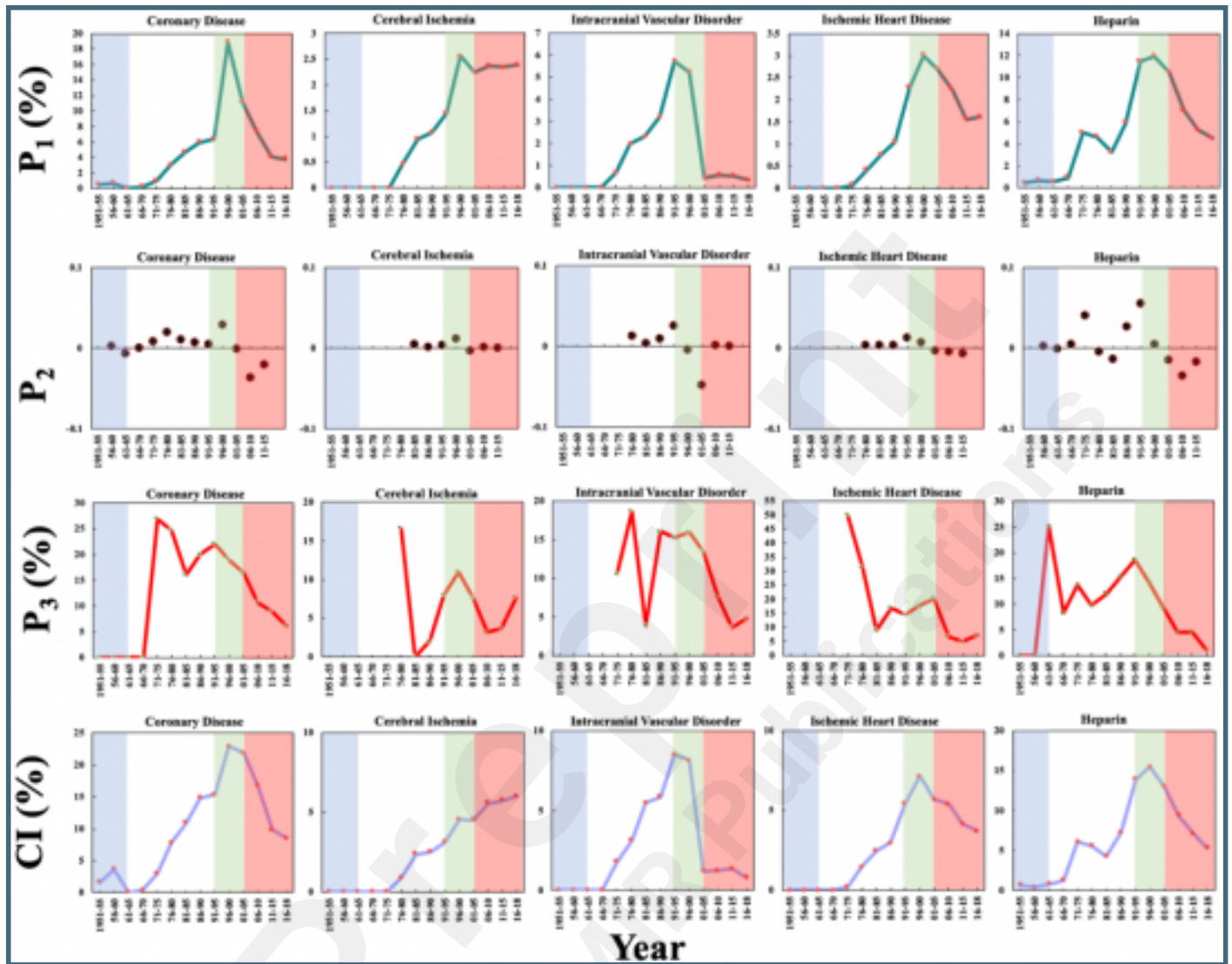




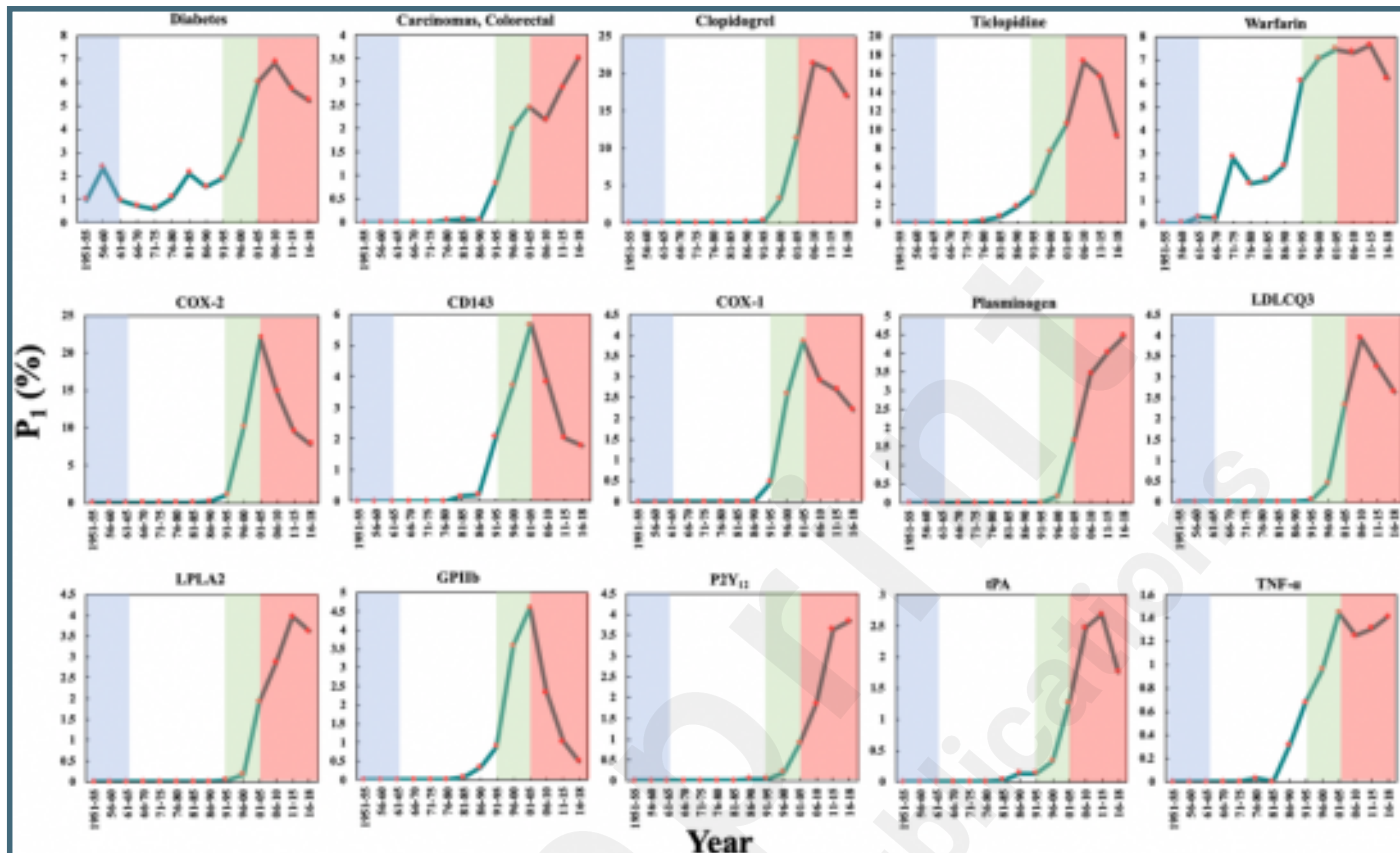
Collaboration Index (CI) of the biomedical entities on the pharmacological mechanisms and side effects of aspirin over time. The background colors show the 4 phases of aspirin research.



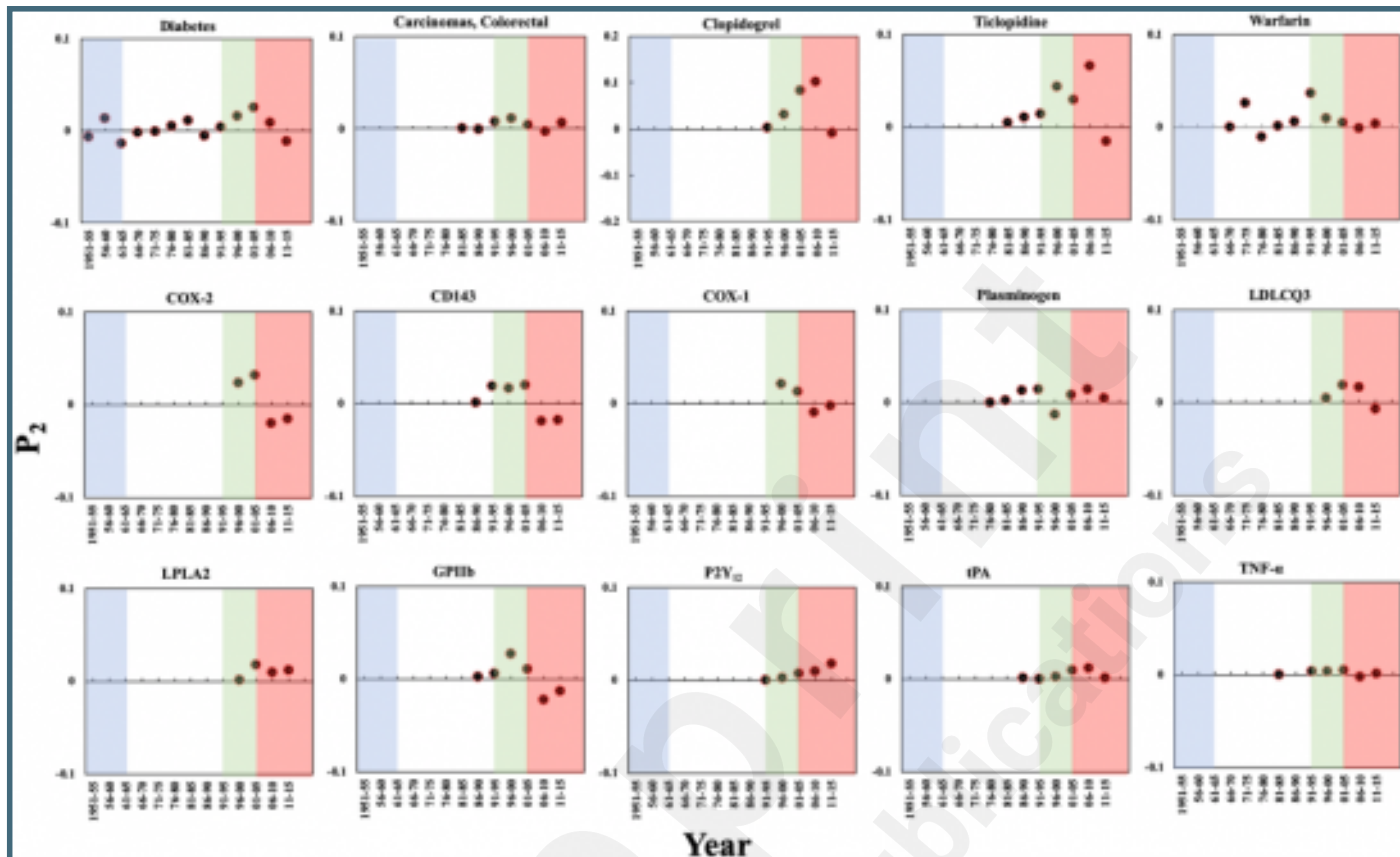
The 4 entymetric indexes of the biomedical entities on cardiovascular diseases in the aspirin domain over time. The background colors show the 4 phases of aspirin research.



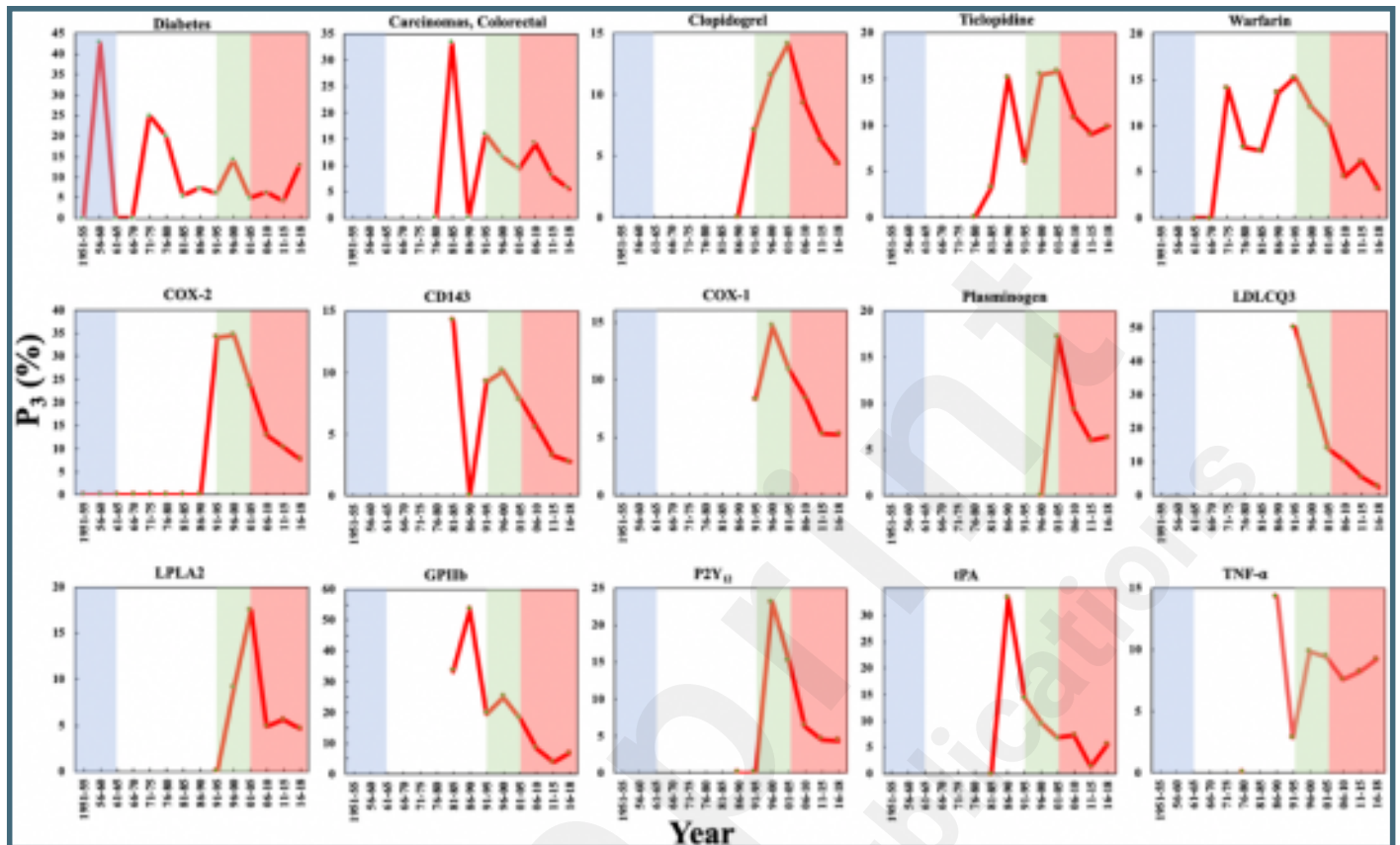
The Popularity Index ( $P_1$ ) of the biomedical entities on repurposing aspirin for other diseases over time. The background colors show the 4 phases of aspirin research.



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.



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.



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.

