THE D3 PHARMACOVIGILANCE SYSTEM: CHARACTERIZATION OF DRUG-DRUG INTERACTIONS BY STATISTICAL INFERENCE BASED ON BIG DATA MINING AND SEMANTIC WEB TECHNOLOGIES

by

Adeeb Redha Noor

B.A., Taibah University, 2005

M.S., University of Colorado at Boulder, 2010

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Computer Science

2015

This thesis entitled:
The D3 Pharmacovigilance System: Characterization of Drug-Drug Interactions By Statistical Inference Based on Big Data Mining and Semantic Web Technologies
written by Adeeb Redha Noor
has been approved for the Department of Computer Science

_____
Prof. James H Martin

_____
Dr. Elizabeth White

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Noor, Adeeb Redha (Ph.D., Computer Science)

The D3 Pharmacovigilance System: Characterization of Drug-Drug Interactions by Statistical Inference Based on Big Data Mining and Semantic Web Technologies

Thesis directed by Prof. James H. Martin

Drug-drug interactions (DDIs) constitute a major cause of adverse drug events (ADEs), which may result in morbidity, mortality, and increased healthcare expenditures. As the role of drug therapy continues to expand and polypharmacy becomes more common, the prevalence of significant DDIs also increases, potentially limiting the therapeutic benefits of medication therapy. While existing DDI research largely focuses on single levels of DDI, a method that incorporates clinical, pharmacological, and physiological (genetic) factors can offer an improved approach to identifying and characterizing potential DDIs. Current limitations to efficient DDI characterization that can affect DDI prevention include: limited availability of clinical studies, shortcomings within study designs, limited accessibility of drug interaction information due to storage in disparate sources, and omissions in DDI reporting within information sources, especially their mechanisms of interaction and clinical significance.

This thesis presents the novel Drug-drug interaction Discovery and Demystification (D3) pharmacovigilance system, which employs Big Data mining and Semantic Web technologies to predict potential DDIs, assess their clinical significance, and describe the mechanism(s) responsible for predicted interactions.

By integrating drug information from a variety of trusted biomedical sources into a coherent, comprehensive DDI knowledge base, D3 leverages the power of big data to construct inference-based predictions to support probabilistic identification, validation, and mechanistic classification of potential DDIs. To qualify the effectiveness of D3, an unbiased benchmark was constructed to characterize the likelihood determinations of D3's predictions against both DDIs reported by its own knowledge sources as well as DDIs reported for five commonly prescribed medications with high propensities for interaction by Micromedex, a respected commercial knowledge source. D3 demonstrates a 93.4% recall rate against DDIs from its own knowledge sources and performs comparably within the margin of error for DDIs reported by Micromedex. The application of D3 to DDIs predicted by five publicly available pharmacovigilance systems indicates that these systems appear to be vastly over-stating the number of DDIs. These results indicate the potential of D3 as an investigative tool for clinicians and researchers to gain some foresight into the likelihoods and causes of potential interactions.

DEDICATION

To the most amazing woman on the earth Wafaa and to my sibling Redhwan

ACKNOWLEDGMENTS

First, I would like to address my sincerest thanks to both Prof. Martin and Dr. White, who enabled me to survive during my graduate life. Their helpful guidance, advice, care and comments allowed me to focus my research directions, and guided me through my PhD journey at the University of Colorado at Boulder. Moreover, their criticisms, constructive suggestions, patience, insights and research style during the preparation of this thesis, have been integral to the success of this work. Without their professional guidance and help, this work would not have been achieved. Also, I am so grateful to Dr. Michel Dumontier from Stanford University for his steadfast help and encouragement. Abdullah Assiri from Purdue University-School of Pharmacy showed the patience of Job in helping me understand the pharmacology science. I am also grateful to the King Abdulaziz University for offering me a scholarship, which helped me undertake my Masters and PhD degrees, and the Deanship of e-Learning and Distance Education for supporting my research. Last but not least, I would like to express my gratitude to my family, my parents (who passed away 11 years ago), my amazing and supportive wife Wafaa, my son Ray, and my coming daughter Aleen for their endless trust, support, encouragement and love, as well as critiques and advice whenever I was lost or helpless in this journey. This doctoral thesis is dedicated to them.

CONTENTS

APPENDICES

TABLES

Table

FIGURES

Figure

CHAPTER I

INTRODUCTION

I.1 Motivation and Problem Definition

One of the earliest scientists, Francis Bacon, once said of drug treatments, “the remedy is worse than the disease." This statement certainly rings true when considering adverse patient outcomes that occur due to drug-drug interactions (DDIs). DDIs constitute a major cause of adverse drug events (ADEs), and occur when multiple drugs interact within an individual’s body and produce unintended toxic effects not initially predicted by the administration of the individual drugs. ADEs represent a major health burden around the world, resulting in significant hospitalization, morbidity, mortality, and healthcare utilization costs (Bates et al., 1995; Classen, Pestotnik, Evans, Lloyd, & Burke, 1997). In 2013, ADEs accounted for 711,232 cases of serious illness in the US, with 117,752 of these cases resulting in death *(“FDA Adverse Events Reporting System (FAERS) > FAERS Reporting by Patient Outcomes by Year,” n.d.).* In addition, medical expenses attributed to ADEs are estimated at over US$1 billion per year to health care systems (Grymonpre, Mitenko, Sitar, Aoki, & Montgomery, 1988; Hamilton, Briceland, & Andritz, 1998;

Shad, Marsh, & Preskorn, 2001). Unfortunately, the impact of ADEs, and DDIs in particular, can be expected to increase exponentially due to the rise in the number of drugs being prescribed to each patient (Percha & Altman, 2013).

Several factors work against early identification of potential DDIs, one important factor being the unavailability of DDI information before drugs reach the market (LePendu et al., 2013; Reis & Cassiani, 2010). Clinical trials conducted prior to a new drug's approval are usually insufficient to test its interactions with other medications, since potential interacting drugs are rarely prescribed to the patients enrolled in the trial. In fact, since clinical trials are designed to study the safety and efficacy of a new medication, patients that are taking drugs that may interact with the trial medication are frequently explicitly excluded from the studies (van der Heijden, van Puijenbroek, van Buuren, & van der Hofstede, 2002). Other factors that may predispose individuals to DDIs include diet, dosage determination, and age-related changes in physiology (Patel, Rana, Suthar, Malhotra, & Patel, 2014). Because of these many complexities, a growing need exists to reduce ADEs through efficient characterization and prevention of potential DDIs.

Pharmacovigilance is a type of pharmacological science that aims to monitor, collect, and synthesize research information about ADEs from multiple resources in order to prevent drug-related problems. Existing pharmacovigilance research proposes to minimize the risk of ADEs by analyzing available published resources and investigating potential DDIs through the use of diverse informatics approaches. These pharmacovigilance methods offer a valuable opportunity to

identify potential DDIs prior to their occurrence. Such analyses investigate the potential for DDIs by leveraging extensive biomedical data from scientific articles, electronic health records (EHRs), the FDA's Adverse Event Reporting System (FAERS), and other drug information sources (Böttiger et al., 2009; Iyer, Harpaz, LePendu, Bauer-Mehren, & Shah, 2014; Tari, Anwar, Liang, Cai, & Baral, 2010; Tatonetti, Ye, Daneshjou, & Altman, 2012a*)*. Despite their comprehensive approach, pharmacovigilance studies to date have captured only a portion of the available potential DDI information and fail to accurately assess the clinical relevance of predicted interactions. Moreover, contemporary pharmacovigilance informatics studies are limited by their ability to analyze available information based on only single mechanisms of interaction. However, the significant numbers of confounding clinical, environmental, genetic and other factors pose formidable challenges in identifying, studying and predicting potential DDIs (Lewis, 2010; Percha & Altman, 2013; Tatonetti, Fernald, & Altman, 2012).

As highlighted above, one of the most complex challenges in identifying potential DDIs is the single-focus approach of current pharmacovigilance research. One example of this is a recent study by Preissner, *et al.*, who developed a complete database (SuperCYP) for cytochrome P450 enzymes and their rules for drug interactions from Medline abstracts (Preissner et al., 2010). The limited focus of studies like these leads to a neglect of other less common yet important pathways, and can also fail to detect multi-pathway DDIs, which occur as a result of two or more interactive mechanisms (Tari et al., 2010; R. Zhang et al., 2014a). For

example, the interaction between statins and cyclosporine occur through both the metabolism (CYP3A4) and transport (P-glycoprotein) pathways (Holtzman, Wiggins, & Spinler, 2006).

Another profound challenge is the overall lack of explanation regarding potential DDI mechanisms. There remains a pressing need to study potential DDI mechanisms beyond those that focus solely on drug metabolism (Hutzler, Cook, & Fleishaker, 2011). Accordingly, Boyce, *et al*. have voiced the need for pharmacovigiliance research focused on modeling both pharmacokinetic and pharmacodynamic DDI mechanisms (R. D. Boyce, Collins, Horn, & Kalet, 2007).

The disparate, disconnected sources of drug information that are currently available pose an even greater challenge to efficient pharmacovigilance evaluation. Integrating these sources into a complete resource that can aid in accurate and early interaction discovery continues to challenge researchers. For instance, many clinically used potential DDI sources vary widely in their reporting of interactions (Hazlet, Lee, Hansten, & Horn, 2001; Peters, Bahr, & Bodenreider, 2015, p. -). As a result, currently there is neither one comprehensive pharmacovigilance source nor a single successful technique that uses all the available data and integrates it in a way that predicts interactions before they are observed clinically. Indeed, a recent study recommends checking more than one resource to validate a DDI (Scheife et al., 2015).

Clearly, the limitations in existing pharmacovigilance research constrain the discovery and characterization of potential DDIs. Ultimately, the research to

date has failed to comprehensively analyze the large variety of potential DDI mechanisms and is inadequate to provide the information required to effectively prevent DDIs in an age of complex pharmacological treatments.

I.2 Research Questions and Objectives

The general purpose of this thesis is to build an integrated data model from existing biomedical knowledge sources, automate validation of clinical relevance, automate identification of interaction mechanism(s), contribute to developing a new and clinically relevant pharmacovigilance mechanism resource, and evaluate known pharmacovigilance systems.

Identifying the clinical relevance of predicted interactions before they reach the market has been extremely challenging due to the limited availability of clinical studies, shortcomings within study designs, limited accessibility of drug interaction information due to storage in disparate sources, and omissions in DDI reporting within information sources, especially regarding their mechanisms and clinical significance. These limitations raise the following questions:

1. Can an effective pharmacovigilance system be developed to provide probabilistic validation and classification of potential DDIs using existing knowledge sources?
2. What is the proper way to identify potential DDIs?
3. Is there a way to provide detailed, mechanistic explanations for potential DDIs?

4. What kind of clinical determinations can be made from the D3 pharmacovigilance system we create?
5. Is there a way to reliably assess the clinical relevance of potential DDIs we discover?
6. Ultimately, can we build a complete and intelligent pharmacovigilance database that identifies DDIs along with the mechanisms of interaction that can provide potential DDI information to both inform scientific evaluation and reduce the burden of DDIs in clinical medicine?

To address these questions, the proposed thesis has the following objectives:

- **Characterization**: To develop a pharmacovigilance system of knowledge integration and discovery with the goal of supporting the study of potential DDIs.
- **Integration**: To use scalable Semantic Web technologies to fill in knowledge gaps across biomedical domains.
- **Efficiency**: To design an efficient pharmacovigilance system and an intelligent database for potential DDI study.
- **Utility**: To improve patient care by providing a system for clinicians and researchers to identify and understand potential DDIs.

I.3 Applications

Drug-drug interaction **D**iscovery and **D**emystification (D3) is an integrated and semantic-based pharmacovigilance system intended to aid both researchers and clinicians. It is designed to incorporate all essential available data from existing trusted sources and identify new and hidden knowledge that each source individually cannot describe. It is important to clarify that the D3 pharmacovigilance system is not intended to replace clinical trials, but to: 1) evaluate theoretically potential DDIs; 2) suggest an explanation for unusual responses or symptoms; and 3) propose new hypotheses about the mechanisms of potential DDIs. Furthermore, the D3 system will open the horizons for researchers to incorporate more details and formulate better predictions that can be used efficiently by clinicians. All of these findings should be considered speculative until clinically verified.

I.3.a Usage: The D3 pharmacovigilance system here developed can be used in various clinical and scientific applications. One important application is in using the output of the D3 database as a gold-standard in pharmacovigilance research. Another central clinical application is using it as a decision-making system to check and understand potential DDIs, as well as a system for recommending alternative drugs with fewer interactions. The main application is that the D3 system can be used by clinicians and researchers to gain some foresight into the clinical relevance of DDIs and causes of potential interactions. Finally, the proposed system can help

in the development of safe drug combinations, which may be increasingly used to treat complex or chronic conditions.

I.3.b Limitations: The D3 pharmacovigilance system, on the other hand, has several limitations that need to be addressed in future work. One limitation of the system is that it will not be able to predict the severity of DDIs. In fact, determining the severity of any interaction based on mechanistic information is a challenging task due to the limited availability of experimental data determining how the identified interactions may change the concentrations of the interacting drugs. Another limitation is that the system does not consider a potential DDI resulting from inappropriate dosing. As a matter of fact, interactions are naturally dosage-dependent, but the dose a patient receives often depends on patient-related factors such as age, sex, demographics and medical history, all of which are outside the scope of this thesis (Rogatko, Babb, Tighiouart, Khuri, & Hudes, 2005).

I.3.c D3 for big data: When performing research and designing a data-driven system for proof of concept, the scale of training data potentially can be limited due to the financial and material resources available and the scope of the design. The D3 system proved to be a reversal of this concept. The quantity of data readily available in the commercial-DDI world is incredibly limited due to many factors, including lack of sharing, limited research, and private control. In fact, it is this very limitation of information that has been a driving force behind the creation of many pharmacovigilance systems. However, what appeared to be lacking in this field was a means to evaluate the effectiveness of such systems from a clinical

standpoint. This was the impetus behind the design of the D3 system. With the goal of D3 being to raise the standard of predicted DDIs to a clinically-relevant level, the necessary training of D3 needed to be clinically-researched and, hopefully, proven interaction information. Therefore, it was prudent for D3 to use and reuse whatever existing knowledge was available to meet this standard. As a result, we have integrated 29 different biomedical and drug interaction sources. While there are other sources available that could potentially be vetted, D3 was intended to be a proof of concept without limitations of scale that could only be improved with additional training information. As such, while the quantity of data currently and readily available to train D3 was somewhat limited by the typically conceived notions of "big data", this is also a limitation characteristic of the industry itself. If all possible DDI information were available, D3 would be unnecessary. Therefore, it is the very limitation of DDI information that creates the need for D3. As the availability of additional data increases, so will the performance of D3, and hopefully one day with complete DDI information at a clinical level, D3 will have outlived its usefulness. For now, enough data was utilized to prove the concept of D3 as a viable evaluation method for pharmacovigilance systems.

I.4 Contributions

The D3 system addresses the challenges outlined above by using the power of interconnected diverse biomedical data sources augmented with big data mining of the Semantic Web. Fundamentally, this thesis addresses the primary difficulties in aggregating DDI information sources and automated inferences for

pharmacovigilance discovery and explanation, with the ultimate goal of constructing a new and clinically relevant pharmacovigilance mechanism database. The core of the system is the use of Semantic Web technologies, community-curated ontologies and datasets to tackle pharmacovigilance research challenges. The proposed system will advance scientific knowledge by identifying significant new pathways involved in potential DDIs and meeting the following aims.

Aim One addresses the challenge of moving beyond a single-level drug interaction to focus on more complex interactions. By integrating DDI information from fifteen diverse sources into a coherent knowledge base, the D3 pharmacovigilance system can provide complex inferences, semantic similarities, and mechanistic information to identify and explain reported and potential DDIs at multiple biomedical levels. These levels include all factors that could contribute to DDIs: pharmacokinetic (interactions altering the absorption, distribution, metabolism, or excretion processes), pharmacodynamic (interactions affecting physiological systems within the body) and pharmacogenetic (when genetic factors alter drug exposure through pharmacokinetic or pharmacodynamic mechanisms), or multi-pathway (interaction occurring through interactive mechanisms) interactions.

Aim Two handles the real need for determining the clinical relevance of DDIs. By mining 10 clinically relevant DDI resources and 12 biomedical sources, the D3 pharmacovigilance system can assist in determining the clinical relevance of DDIs using a novel statistical inference model based on similarities within proven

DDI lists and biomedical features. Nine biomedical features are considered for similarity identifications: targets, side effects, enzymes, transporters, carriers, indications, mechanisms of action, genetic variations, and physiological features.

Aim Three is to leverage the coherent D3 knowledge base to identify most likely mechanisms for the interactions. These mechanisms include the following:

1. Pharmacokinetic interactions:
    a. Metabolism (inhibition and induction)
    b. Transporter (inhibition and induction)
    c. Protein binding
2. Pharmacodynamic interactions:
    a. Additive (both drugs share a mechanism of action)
    b. Competitive (both drugs act on the same pharmacologic target)
3. Pharmacogenetic interactions:
    a. Known single nucleotide polymorphisms (SNPs) in genes that may alter drug exposure
4. Multi-pathway interactions:
    a. Interactions occurring as a result of two or more interactive mechanisms

Aim Four tackles the need for annotating existing pharmacovigilance information resources with the mechanistic information. Fifteen different DDI sources, 14 that have been collected by Ayvaz *et al.* (including 5 clinical sources, 4 natural language corpora and 5 pharmacovigilance sources) and ClinicalTrials.gov,

will be annotated by D3 with mechanistic information (Ayvaz et al., 2015). The DDI pairs will be represented and stored as Unified Medical Language System (UMLS) concept identifiers and then the D3 resource will be constructed (Bodenreider, 2004). Below is the list of anticipated thesis contributions:

1. A new pharmacovigilance platform for knowledge discovery and integration in the biomedical field.
2. A mechanism for grouping semantic properties.
3. A pharmacovigilance system to support probabilistic validation and classification of potential DDIs utilizing existing knowledge sources.
4. A query model to support potential DDI queries utilizing 15 different DDI sources.
5. A method to support queries using different drug names (generic, brand and chemical) to find interactions.
6. A model to infer potential multi-pathway DDIs.
7. A novel mechanism to predict potential DDIs based on similarities in mechanisms of interaction.
8. A model to identify when potentially interacting drugs may also interact on the genetic level in patients possessing known polymorphisms (SNPs) in drug-metabolizing enzymes or transporters.
9. A similarity-based method to automate validation of clinical relevance.
10. A similarity-based method to automate identification of interaction mechanism(s).

11. A tool to evaluate known pharmacovigilance systems.
12. A tool to annotate known pharmacovigilance systems with most likely mechanism(s) of interactions.

I.5 Thesis Outline

This thesis is structured in 7 following chapters:

Chapter II introduces the dangers of DDIs in health and science. Then it discusses: (i) a traditional method of discovering potential DDIs and (ii) a pharmacovigilance method of discovering potential DDIs. In the related works section, we particularly examine, compare and contrast how these methods are designed and note current limitations of the state of the art. In the latter part, we introduce the Semantic Web: its concepts and architecture, its way of representing the knowledge and data integration, and finally, its impact on the clinical research area.

Chapter III discusses the process of developing a D3 knowledge base that integrates twelve biomedical data types and fifteen DDI resources considering all factors that could contribute to DDIs from pharmacokinetics (metabolism, transporter, and protein binding), pharmacodynamics (additive and competition), pharmacogenetics (SNPs), and multi-pathway (enzyme and transporter) levels. The construction of the D3 semantic infrastructure consists of three phases: (i) collecting, (ii) mapping, and (iii.) integrating and loading. Each phase is discussed

in more detail in a separate section. Then, we enrich the knowledge base by inserting the Minimum Information Required in the Annotation of Models (MIRIAM) registry property and demonstrate how easily D3 can be expanded by linking it to external sources (Novère et al., 2005). Next, we verify and validate the resulting semantic infrastructure by using reasoning techniques to check for consistency of the knowledge base. Finally, we demonstrate the usefulness of such integration by identifying embedded associations for a chemotherapy medication.

Chapter IV presents a pilot study of the thesis. In this chapter, we propose a novel method to identify multi-pathway potential DDIs using interconnected diverse biomedical data obtained by mining of the Semantic Web. First, we develop a semantic rule-based model to identify potential DDIs arising from metabolic and transporter interactions. We then demonstrate the utility and validity of our approach using irinotecan, a chemotherapy agent. Finally, the statistical significance of the result is analyzed along with the precision and recall of predicted interactions.

Chapter V describes in detail the design and development phases of the D3 pharmacovigilance system. In particular, Phase One discusses the identification and clinical examination of the most common mechanisms of DDIs; this yields nine mechanisms of interaction. Then we convert these mechanisms to semantic inferences that are proposed to validate and classify the proven and potential DDIs.

Phase Two illustrates the processes of building D3, which are divided into two main sub-phases: (i) building an inferential query model, and (ii) building an inferential probabilistic model. Each model is discussed in more detail in a separate section. In Phase Three, we exploit the D3 system by introducing a set of experiments conducted to test the two inferential models of the D3 system. Finally, we discuss the results and outcomes of these tests.

Chapter VI describes the process of annotating current pharmacovigilance resources with mechanistic information. First, we show the need for a complete and comprehensive pharmacovigilance database by analyzing seven DDI sources that report the interaction of two well-researched drugs. We also demonstrate the high level of diversity in reported interactions by pairing 15 DDI sources and then computing the average Jaccard Index between them. Further, we statistically analyze the fifteen DDI resources used in this study and prove the lack of their support with regards to DDI mechanisms of interaction. Then we propose two main contributions: (1) annotating the fifteen DDI resources with the mechanistic information and (2) building the D3 highly clinically pharmacovigilance database by extracting DDI pairs from 5 non-clinical resources.

Chapter VII defines the validation and benchmarking of the D3 pharmacovigilance system. First, we test the coverage of the D3 system's query-based model for the extraction of explicit DDIs from 15 DDI sources by computing

its recall against each source as well the whole group. Then the D3 probabilistic-based model was unbiasedly benchmarked to identify its likelihood of predictions against both DDIs reported by its own knowledge resources as well as DDIs reported by Micromedex, a respected commercial system, for five commonly prescribed medications with high propensities for interaction. Then, we use the D3 system model as a tool to evaluate five non-clinical resources in terms of reliability in reporting only clinically relevant DDIs. Finally, we present and discuss the evaluation of the results of this testing.

Chapter VIII gathers the conclusions of this thesis, together with its limitations and complementary future work to be addressed.

CHAPTER II

BACKGROUND AND SIGNIFICANCE

Polypharmacy, defined as concomitant prescriptions of five or more medications, has grown considerably in recent years, and unfortunately, patients taking multiple medications have an increased risk of experiencing major DDIs (Guthrie, Makubate, Hernandez-Santiago, & Dreischulte, 2015). Apart from the risk of death, DDIs are implicated in 0.054% of ER visits, 0.57% of hospital admissions and 0.12% of re-hospitalizations (Becker et al., 2007). Because of these increasing risks and costs, a wide range of techniques and approaches has been developed to study potential DDIs. This chapter critically examines and discusses these techniques and approaches in order to gain a more comprehensive view of how to manage potential DDIs. In this chapter, we also shed light on the Semantic Web technologies that are used in this thesis. This chapter is divided into three main sections. In the first, we discuss the concept of DDI. In the next section, we survey the current methods for studying potential DDIs by classifying them as traditional or novel (pharmacovigilance). The last section illustrates briefly the concept of Semantic Web technologies and how they are useful in this context. Finally, we

summarize the chapter with a focus on the limitations of current approaches to investigating potential DDIs.

II.1 Drug-Drug Interactions

DDIs constitute an emerging medical problem around the world, contributing significantly to morbidity and mortality. Generally, unintended interaction between two drugs causes either toxicity or inefficacy; neither of these is a desirable effect. This necessitates alterations in dosage or the pursuit of alternative treatments for therapeutic intervention to avoid the development of clinically significant ADEs. Mechanisms involved in DDIs can be pharmacokinetic (causing alterations in drug exposure), or pharmacodynamic (affecting physiological systems within the body) in nature, or both (Williams & Feely, 2012). Pharmacokinetic interactions have been known to affect drug absorption, distribution, metabolism (biotransformation) and elimination, while pharmacodynamic interactions change the actual effects of a drug (Pleuvry, 2005). Pharmacogenetic factors that contribute to drug interactions constitute a rapidly emerging field of study and involve specific genetic factors that predispose individuals to DDIs via pharmacokinetic or pharmacodynamic mechanisms. **Figure 2.1** shows the general concept of DDIs involving pharmacokinetic and pharmacodynamic mechanisms, as well as the contribution of pharmacogenetic factors.

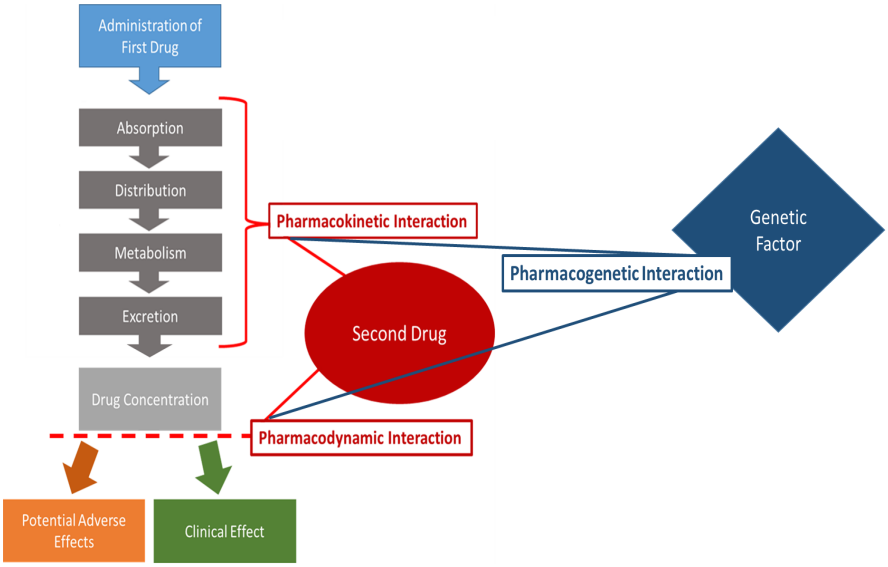


Figure 2.1: Mechanisms of DDI following multiple drug administration

Unfortunately, DDIs may result in discomfort, debilitating illness, and in extreme cases, death. For instance, when simvastatin (a CYP3A4 substrate) is administered with posaconazole (a CYP3A4 inhibitor), the statin accumulates in the body due to the inhibition of its metabolism by posaconazole, leading to risk of myopathy and rhabdomyolysis (Krishna et al., 2012). As illustrated in the previous example, DDIs can cause ADEs, which are associated with morbidity, mortality, and increased healthcare expenditure. As a result, development of methods to avoid these problems is critical since more accurate and comprehensive information about potential DDIs will increase patient safety as well as health quality. In practice, potential DDIs are challenging to study as they depend on many clinical, environmental, genetic and other factors. In a recent review, Percha & Altman argue that several of these factors work against early identification of potential

DDIs (Percha & Altman, 2013). They list the most prohibitive factors as being the lengthy time needed to perform clinical studies and the variance of genetic and demographic features in patient populations, which can produce or hide potential DDIs. In the next section, we will describe existing methods of potential DDI study.

II.2 The State of the Art

A wide range of clinical, computational and pharmacovigilance approaches have been developed to guide potential DDI research. Here, we discuss these techniques and approaches by categorizing them as either traditional or pharmacovigilance models for potential DDI discovery. First, we discuss the traditional potential DDI model. This includes *in vitro*, *in vivo*, and retrospective studies. Next, we discuss the pharmacovigilance DDI model. The pharmacovigilance discovery model includes mining clinical notes and published literature, as well as use of cheminformatics and the Semantic Web.

II.2.a The Traditional DDI Discovery Model

Historically, DDIs have been proven through laboratory or clinical studies of drug interactions. The traditional discovery model is used to identify the effect of one drug on a biological target or system in relation to a second drug *(*"Pharmacology Condensed, 2nd Edition | Maureen Dale, Dennis Haylett | ISBN 9780443067730," n.d.*)*. The interaction often results in a change in concentration of the second drug. Cytochrome p450 metabolism interaction is the most common

example of this type of DDI ("*Flockhart DA. Drug Interactions: Cytochrome P450 Drug Interaction Table," n.d.; Lynch & Price, 2007*).

In vitro and *in vivo* studies mandated during the drug development process assist in discovering some interactions before a drug reaches the market. *In vitro* studies are particularly valuable to characterize a new drug's route(s) of biotransformation and to examine the potential of a new drug to alter enzymatic activity. Moreover, they are designed to determine the relationship between a drug's concentration and its pharmacologic effects. After *in vitro* studies, *in vivo* experiments leverage *in vitro* data to identify the volume of distribution, clearance, bioavailability, and other pharmacokinetic parameters of the drug (Thummel & Wilkinson, 1998). The US Food and Drug Administration (FDA) recently issued guidance to inform *in vitro* and *in vivo* drug interaction studies in order to make more detailed investigation of potential DDIs part of the approval process for new drugs (S.-M. Huang, Temple, Throckmorton, & Lesko, 2007).

One advantage of traditional *in vitro* and *in vivo* methods for discovering potential DDIs is their ability to characterize the pharmacokinetics of a drug. In other words, the processes of absorption, distribution, metabolism and excretion can be accurately tested inside the living organism. Another important advantage of these models is the minimization of human risk while studying potential DDIs (Wienkers & Heath, 2005). Although *in vitro* and *in vivo* studies warn the medical community about interactions, these investigative processes are slow and often test only small numbers of drugs and targets (Hutzler et al., 2011; Venkatakrishnan,

von Moltke, Obach, & Greenblatt, 2003). As a result, these processes do not allow researchers to study the interactions of new drugs as quickly as these new drugs are being added to the market. Another limitation of *in vitro* and *in vivo* studies is that clinical trials that are conducted prior to a new drug's approval largely focus on that drug alone. This is obviously not sufficient to test a drug's interactions with other medications, which are usually not prescribed to the patients enrolled in the studies. Finally, FDA instruction to the pharmaceutical industry to guide DDI studies focuses most commonly on the cytochrome enzyme (CYPs) and transporter interactions, while important potential DDI factors such as pharmacogenetics are not discussed. Thus, these guidelines do not provide comprehensive instruction for adequate study and understanding of potential DDIs.

Retrospective studies are another method of traditional DDI discovery. Retrospective studies look backward in time using patients' health records to identify potential DDIs. Many retrospective studies have produced promising results, providing useful and easy-to-implement information to clinicians for a variety of applications, including dose spacing (Liu & Unni, 2014; Peng et al., 2003; "Reference Guide For Foreign Pharmacy Licensing Exam Pharmacy Management & Pharmacoeconomics Question And Answers," n.d.). However, retrospective studies are often subject to bias and contain limited sample sizes that reduce both their statistical power to assess research questions and their external validity (generalizability). In other words, because these studies are limited to existing records, they do not allow researchers to manipulate trial design to most efficiently

examine clinical questions, and they are subject to error in the collection of data within the original medical record (Rosholm, Bjerrum, Hallas, Worm, & Gram, 1998).

An additional limiting characteristic of traditional discovery models for investigation of potential DDIs is the focus on either one particular enzyme or a single drug class interaction. For instance, Williams & Feely have extensively reviewed pharmacokinetic and pharmacodynamic drug interactions for the cholesterol-lowering statins (D. D. Williams & Feely, 2012). The same group has studied potential DDIs related to UDP-glucuronosyltransferase substrates (J. A. Williams et al., 2004). Even though these studies have shown promising results, they are still considering only single pathway interactions. These analyses do not address the fact that DDIs can be complex processes with interactions extending beyond the single pathway level.

The medical burden associated with DDIs will undoubtedly continue due to the long timeframe required to discover potential DDIs using traditional approaches. As a result, recent research has concluded that new methods are needed to aid in early DDI detection (Hennessy & Flockhart, 2012; Percha & Altman, 2013). In the next subsection, we will discuss these novel (pharmacovigilance) methods.

II.2.b The Pharmacovigilance DDI Discovery Model

The pharmacovigilance approach leverages different computational methods to predict potential DDIs. This novel model for potential DDI discovery accounts for important DDI information, which may be embedded in text forms such as scientific articles, electronic health records (EHRs), the FDA's Adverse Event Reporting System (FAERS), or drug information sources (Iyer et al., 2014; Tari et al., 2010; Tatonetti, Ye, et al., 2012a; Wishart et al., 2008). It also includes cheminformatic methodologies, which identify potential DDIs by using 2D/3D QSAR (quantitative structure-activity relationships) to find similarities among sets of drugs (Korhonen et al., 2005; Yap & Chen, 2005). Therefore, researchers have recently begun experimenting with diverse pharmacovigilance approaches that employ available published resources to investigate and minimize the potential risk of DDIs.

Mining medical texts to manually represent DDI data in a formal representation has been widely used and has shown promising results in potential DDI discovery. Medical text-mining includes extracting potential DDIs from scientific articles such as Medline, FAERS and EHRs *(*"FDA Adverse Events Reporting System (FAERS) > FAERS Reporting by Patient Outcomes by Year," n.d.*)*. An example of utilizing EHRs for pharmacovigilance study is the informatics for integrating biology & the bedside (i2b2) system (Murphy et al., 2010). Other works have proposed a classification system to not only identify potential DDIs but to also explain existing ones (Coulet, Shah, Garten, Musen, & Altman, 2010; Percha, Garten, & Altman, 2012). The method of mining texts has produced

promising results, but the extracted medical literature data is often error-prone because it requires substantial manual revision before use.

Other pharmacovigilance models have incorporated text mining with Semantic Web technologies specifically to utilize the Semantic Web's reasoning technique for potential DDI discovery. For instance, Tari *et.al* mixed text-mining methods and reasoning capability and found interactions based on transcription factors (Tari et al., 2010). More recent work has found potential DDIs based on the physiological effect of drugs at the level of biological functions. This study showed that though two drugs are not linked together by targets or proteins, they still could produce interactions when the two drugs are targeting the same biological process (R. Zhang et al., 2014a).

The cheminformatic method of potential DDI discovery depends on the idea that if two drugs are similar in chemical and pharmacological features, they could interact. For example, protein structure, molecular structure and target-based interactions are applied to identify potential DDIs (Vilar, Lorberbaum, Hripcsak, & Tatonetti, 2015). The cheminformatic methods account for a similarity between drug classes as well. A significant advantage of this method, in addition to its capacity to identify potential DDIs, is the ability to provide extensive information about mechanisms of drug interaction.

The manual curation of DDI data into a formal representation has been applied to find potential DDIs as well. This includes building a potential DDI knowledge base, combining multiple biomedical resources related to DDIs, and

utilizing the Semantic Web's capabilities in reasoning. For instance, one contributing work on potential DDIs used only Semantic Web converted patient EHRs in a Resource Description Framework (RDF)*("Resource Description Framework (RDF) Model and Syntax Specification," n.d.)*, which is a standard way of representing data on the Semantic Web. After conversion, the researchers queried the possible interactions between cardiovascular and gastroenterology drugs from the RDF to identify whether clinically reported DDIs in publically available resources were observed in real patient data.

Another recent contribution to pharmacovigilance models for potential DDI discovery has been the development of drug recommendation frameworks using a business rule engine (Doulaverakis, Nikolaidis, Kleontas, & Kompatsiaris, 2014). This system was used to find both potential DDIs and drug-disease interactions, and was employed by Takarabe *et al.* to develop a DDI knowledge base containing both pharmacokinetic and pharmacodynamic information (Takarabe, Shigemizu, Kotera, Goto, & Kanehisa, 2011). A chief advantage of pharmacovigilance models of potential DDI discovery relative to traditional models is their ability to combine different scientific disciplines implicated in DDI study. However, while recent efforts in pharmacovigilance discovery have offered valuable advancements, these methods still possess substantial shortcomings, including a reliance on small data sets and an inability to assess the clinical relevance of predicted DDIs. This thesis, on the other hand, will focus on more complex methods of potential DDI discovery,

provide possible mechanisms of interaction, and assess the clinical relevance of discovered potential DDIs using heterogeneous data.

Overall, based on the available literature discussed in the state of the art section, we identify two common weaknesses of current DDI study. First, most of these studies focus largely on pharmacological aspects of the problem and provide only a single-pathway of discovery. However, in reality, an interaction between two drugs may occur due to many factors. Therefore, these methodologies are greatly limited in discovering potential DDIs caused by other mechanisms, such as by genetic, transporter, or physiological factors. The second weakness is that most studies lack an explanation regarding the mechanisms of potential DDIs; an extensive literature review revealed only one study that addresses mechanistic details, and these were limited solely to pharmacokinetic interactions. Consequently, since designing clinical trials to comprehensively characterize potential DDIs is heavily reliant on preliminary information from pre-clinical studies describing the nature of potential interactions, providing more detailed information about the mechanisms of proposed interactions is essential to aiding researchers in efficient clinical trial design.

II.3 The Semantic Web in Life Science

The Semantic Web is a new technology that aims to enhance the current World Wide Web (WWW) to make it understandable to both humans and machines. Thus, information presented on the WWW can easily be connected and interpreted,

leading to accurate knowledge retrieval and integration. The World Wide Web Consortium (W3C) supports the Semantic Web ("World Wide Web Consortium (W3C)," n.d.). The main goal of the W3C's work is to allow data to be shared and reused across multiple applications, users, and organizations by providing an interoperability framework. Data from various resources are explicitly described in the Semantic Web, leading to easy integration, understanding and automated computer processing.

With the rapid increase and nature of numerous biomedical data sets, knowledge in information systems is often represented in different formats. The need addressed by the Semantic Web to accomplish standard representation and integration is essential to the life sciences industry to achieve better biomedical studies (Tim Berners-Lee & Hendler, 2001; Hendler, 2003). In this section, we will describe the Semantic Web concept architecture, its applications for data integration, its methods of knowledge representation, and its important role within the biomedical domain.

II.3.a Concepts and Architecture

The Semantic Web architecture includes multiple layers (**Figure 2.2**). The first layer is a Semantic Web language layer, termed the Universal Resource Identifier (URI) (T. Berners-Lee, Fielding, & Masinter, 2005). It is used to standardize the language of the resources in the Semantic Web. The actual syntax

used in the Semantic Web is the Extensible Markup Language (XML) schema ("Extensible Markup Language (XML) 1.0," n.d.).

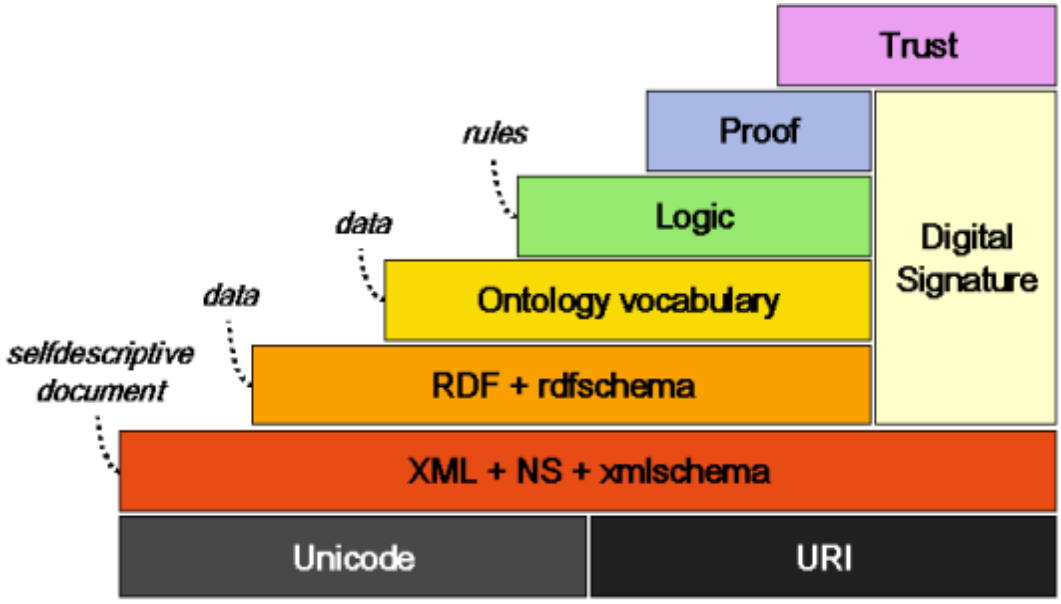


Figure 2.2: Diagram representation of Semantic Web layers. Adopted from W3C

Another important syntax in the Semantic Web is the RDF. An RDF is a graphical representation of resources, written in the form of XML. Resource Description Framework Schema (RDFS) extends the RDF to allow classes, properties and types of resources to be explicitly declared ("RDF Schema 1.1," n.d.). Also, Web Ontology Language (OWL) adds more vocabulary for describing classes and properties ("OWL Web Ontology Language Overview," n.d.). Inferences and reasoning within the Semantic Web data can be done via RDFS and OWL. Another important component in the Semantic Web architecture is the SPARQL Protocol

and RDF Query Language (SPARQL) (Prud'hommeaux & Seaborne, n.d.). SPARQL is a query language for the Semantic Web.

II.3.b Knowledge Representation

Knowledge in biomedical systems often comes in different and complex formats, leading to the risk of unsuccessful discovery of a desirable goal and processing with regards to biomedical study (Ruttenberg, Rees, Samwald, & Marshall, 2009). In fact, any knowledge is useless if it is not well represented and accessible. The use of ontologies as a knowledge representation tool on the Semantic Web helps to avoid such problems. An ontology is a formal representation or conceptualization that aims to make knowledge understandable and reusable across different platforms (Gruber, 1993). The ontology also offers sophisticated reasoning and inter-operational abilities relevant to a semantic context. The main components of an ontology as described above are: classes, relationships, instances, and, finally, axioms. Biomedical knowledge provides an example: classes (genes, diseases, etc.), relationships between different classes (e.g., a protein is encoded by a gene), sets of instances mapped to classes (e.g., mouse p53 is a gene), and finally axioms (e.g., if drugs from different classes share a common property, such as a side effect, they may interact). Representation of all these components is easily done using OWL in the Semantic Web. Despite the advantages of the ontology in making the information more suitable for inference and processing, its development is time-consuming and complex (e.g., it requires ontology engineers and domain experts).

Another problem with building any new ontology in the medical field is that biomedical data continues to grow rapidly, and its ontology needs to change accordingly. As a result, it has been recommended to reuse existing ontologies whenever it is possible instead of developing new ones ("Wiley," n.d.). Likewise, to encourage the reuse of existing ontologies, particularly within complex ontologies (i.e., from different knowledge domains), it is important to provide mapping between different ontologies to describe merging and constraining processes.

II.3.b.1 Resource Description Framework (RDF)

The RDF is a standard way of representing data on the Semantic Web in graph-or-tree form. <Subject, Predicate, Object> are the components of the RDF statement known as triples. RDF statements are modeled as a graph with <Subject, Object> as its vertices connected by directed arcs <Predicate>. In addition, each RDF statement creates its own unique meaning. However, a single RDF statement can be linked to other graphs. Eventually, the result is a well-defined web forming a structure like a graph. Every resource in the RDF graph model has a unique URI reference. This URI guarantees global integration and avoids redundancy. Finally, having our data represented in a logical graph associated with relationships and constraints provides an easy way to manage and analyze it.

II.3.b.2 Web Ontology Language (OWL)

The OWL is used to capture knowledge of interest about specific domains by adding more vocabulary to Semantic Web data for describing properties and classes. An important use of the OWL is to define concepts and add relationships among them (Antoniou & van Harmelen, 2004). OWL also can be used to define constraints. In addition, there are many open source ontology editors such as *Protégé* that are used to develop an ontology with all its components (classes, properties, instances, and axioms) (Horridge, Knublauch, Rector, Stevens, & Wroe, 2004, p. -). The structure of classes within the ontology is a taxonomy with super- and sub-classes. In the OWL, properties are usually used to provide the semantic link between concepts using binary relations. Instances represent individuals, while the axioms are logic-based in the form of rules. A significant strength of OWL is its ability to provide not only inheritance from classes of properties, but also subsumptive reasoning that allows interoperability between different systems.

II.3.c Data Integration

The Semantic Web, as a new version of the current web, is known as web2; data from various resources are explicitly described leading to easy integration, understanding and automated computer processing ("The Semantic Web," n.d.). The Semantic Web and its technologies (RDF, RDFS, and OWL) represent concepts in graphical format using a common language. As is stated above, the RDF is the foundation of Semantic Web technology. The need for both RDFS and OWL with

RDF is to embody concepts of specific-domain knowledge and add relationships. Once the knowledge is represented in Semantic Web format, accurate reasoning from this knowledge can be achieved. The use of URI in the Semantic Web to provide a physical location of the information helps to avoid data redundancy. If two URIs point to the same concept, the RDF will join them together with one unique URI. This approach allows seamless integration of data from diverse sources in addition to providing the interoperable representation of it; this is in contrast to traditional data integration. Another important feature of the Semantic Web is that the RDF model is easily changeable; i.e., there is no need to rebuild the model from the beginning. Thus, the Semantic Web is dynamic, and the result is an ability to add new information quickly.

II.3.d Toward Translational Research

Recently, there have been tremendous advances in designing and implementing biomedical ontologies. The Gene Ontology (GO), Chemical Entities of Biological Interest (ChEBI), and many other projects have adopted Semantic Web knowledge representations (Ashburner et al., 2000; Degtyarenko et al., 2008). Moreover, many consortium efforts such as Health Care and Life Sciences (HCLS) by W3C and Open Biomedical Ontology (OBO) have been proposed in order to develop standard biomedical ontologies (Cheung, Yip, Townsend, & Scotch, 2008; Smith et al., 2007). Hence, the Semantic Web has been used effectively in the life sciences, and there has been a great deal of positive progress in this area.

II.4 Conclusion

In this chapter, we first demonstrated the concept of DDI along with its clinical and economic impact in health and life science. We then discussed, examined and analyzed different studies from two main potential DDI discovery models: traditional and pharmacovigilance methods. Overall, the available literature from both models pose significant limitations to the process of discovering and understanding potential DDIs. These limitations are summarized as follows:

- None of the previously examined studies have proposed a comprehensive approach for discovering potential DDIs at different and multiple mechanistic levels. All of the studies focus either on a single interaction type or study a single interaction mechanism. As a result, there is a need for an approach that is able to study potential DDIs at multiple mechanistic levels in order to provide not just potential interactions, but also to identify new mechanisms of interaction.
- Most potential DDI discovery approaches are developed for specific domains and are difficult to re-apply in other settings. Therefore, a more generic approach that is able to model potential DDIs is needed.
- Most of the studies that utilize potential DDI information for their research are limited in that they only consider narrow types of DDI discovery, which are mostly pharmacokinetic. Therefore, there is a

need for a broader approach that is able to model, adopt and utilize diverse DDI dimensions.

- Potential DDI information sources consist of rich information, which can provide a great benefit to a pharmacovigilance customization system. Most of the studies discussed in this chapter are incapable of gathering and interpreting hidden semantic knowledge from these sources. We argue that a good pharmacovigilance system is one that can infer and exploit hidden semantic information from multiple sources and that this pharmacovigilance system can be used to provide better results.
- None of the previously examined pharmacovigilance studies have proposed an effective approach for knowing the clinical relevance of predicted DDIs. All of the studies seem to produce more and more DDIs, which may not necessarily be important. Consequently, there is a need for a method to evaluate and identify the most clinical relevance of existing DDIs.

Finally, we gave a brief introduction to the Semantic Web and its related technologies, and we concluded by showing some well-known successful applications of it.

CHAPTER III

THE DRUG-DRUG INTERACTION DISCOVERY AND DEMYSTIFICATION KNOWLEDGE BASE

As discussed in the previous chapter, existing pharmacovigilance systems suffer from a number of limitations that restrain potential DDI discovery. A critical shortcoming of these systems is their limited focus on single pharmacological aspects of potential DDIs. The main problem with such a focus is that it ignores other important interaction factors that may not be easily recognized. This is generally because the sources currently available for DDI discovery are disparate and disconnected. Therefore, most of the DDI detection methods (both traditional and pharmacovigilance) that we examined earlier, which rely on single levels of interaction, cannot accurately forecast potential DDIs in an age of complex pharmacological treatments. One way to address this problem is by developing a complete and comprehensive knowledge base that integrates multiple existing DDI information sources.

In clinical practice, healthcare practitioners often have diverse backgrounds, training, expertise and clinical preferences, creating an essential need for resources that guide clinician judgment toward optimal patient care. For the

treatment of disease, clinical practice guidelines are created that synthesize the best medical evidence available to provide evidence-based recommendations that optimize patient outcomes (Woolf, Grol, Hutchinson, Eccles, & Grimshaw, 1999). For the prevention and management of DDIs, however, no single comprehensive resource exists to guide clinicians. By providing more detailed explanations and integrating a greater amount of information (including mechanistic information) to assess potential DDIs, the **drug-drug interaction discovery and demystification (D3)** system we are developing seeks to aid clinicians in making appropriate decisions regarding DDIs. As a result, the D3 system aims to provide clinicians with an invaluable decision-making support tool by creating a comprehensive DDI knowledge base that is able to model and answer complex DDI questions and keep pace with an increasingly complex clinical picture.

In this chapter, we propose the D3 semantic infrastructure that integrates existing biomedical data from pharmacokinetic (interactions involving metabolism, transporters, and protein binding), pharmacodynamic (additive, synergistic, and competitive interactions), pharmacogenetic (single nucleotide polymorphisms that influence drug exposure), and multi-pathway (interactions involving more than one of the aforementioned pathways) interaction levels. The innovation of this D3 semantic infrastructure is both the integration of diverse knowledge sources for drug entities and the inference of new relationships using semantic synergism. The D3 semantic infrastructure is, therefore, able to identify and explain reported and

potential DDIs. In order to achieve this complex task, a number of requirements must be taken into consideration:

- The D3 knowledge base should contain the necessary information for DDI discovery at multiple interaction levels.
- Data in the D3 knowledge base should be extracted, transformed, and stored in a formal and semantic representation.
- The D3 knowledge base should be able to provide precise answers to complex DDI questions.
- The D3 knowledge base should be able to easily adapt to changes.
- The D3 knowledge base should be generic enough to be integrated and deployed to provide a wide range of DDI discoveries.

Building upon these requirements, in the next section we propose building a novel D3 knowledge base.

III.1 Building the D3 Knowledge Base

For the accurate discovery and explanation of drug interactions occurring through multiple pathways as planned in this thesis, a knowledge base needs to include information from different biomedical levels related to drug mechanisms. The development of this framework is necessary since there is, unfortunately, no single knowledge base that provides the requisite information. In fact, modeling DDI mechanisms is not a trivial task because the mechanisms are often missing

from sources or not up-to-date (R. D. Boyce et al., 2007). For example, Boyce *et al.* have developed a drug interaction knowledge base, but it is restricted to only metabolic-based interactions (R. Boyce, Collins, Horn, & Kalet, 2009a, 2009b). Other contributions such as the Quantitative Drug Interactions Prediction System (Q-DIPS) , the Pharmacogenomics Knowledge Base (PharmGKB), and SuperCyp have shown important results for modeling DDI mechanisms (Bonnabry, Sievering, Leemann, & Dayer, 1999; Klein et al., 2001; Preissner et al., 2010); however, they mostly focus on single mechanisms or certain drug classes. Due to the limited coverage in existing work with regard to modeling DDI mechanisms, this thesis proposes the building of a knowledge base called D3. The D3 knowledge base will model drug interaction mechanisms at pharmacokinetic, pharmacodynamic, pharmacogenetic, and multi-pathway levels. Moreover, D3 will be used as a model basis for data integration and a schema for knowledge discovery. Data will be semantically integrated into the D3 RDF network.

III.1.a Data Sources

The scientific goal of the D3 knowledge base is to support mechanistic DDI research. Toward such a goal, we attempt to provide background knowledge that serves our research purpose and to choose a number of sources that can aid in achieving complex research. It is important to mention that we will build the D3 knowledge base so as to make it not only scalable for reasoning but also accurate at the conceptualization level. Moreover, while designing the D3 knowledge base, we

will consider high-level granularity by reusing as much information from existing knowledge bases as we can. The final goal while constructing the D3 knowledge base is to provide it with extensibility and flexibility by offering a well-structured schema for the purpose of future integration of new information sources. The D3 knowledge base will be developed initially using fifteen different biomedical sources. These biomedical sources are listed below:

1. The Unified Medical Language System (UMLS) (a terminology integration system produced by the National Library of Medicine (NLM)) (Bodenreider, 2004).
2. The Open Data Drug & Drug Target Database (DrugBank) (a database containing detailed information about drugs and their targets) (Wishart et al., 2008).
3. The Pharmacogenomics Knowledge Base (PharmGKB) (a database identifying significant relationships that show the effect of genetic variation (SNPs) on drug disposition) (Klein et al., 2001).
4. The National Drug File – Reference Terminology (NDF-RT) (an ontology that describes and models drugs within pharmacokinetic, pharmacodynamic, physiological, and related disease domains)(Brown et al., 2004).
5. The National Cancer Institute thesaurus (NCIt) (an ontology that covers vocabulary for medical and translational research with a focus on cancer) (de Coronado, Haber, Sioutos, Tuttle, & Wright, 2004).

6. Gene Ontology (GO) (an ontology that aims to describe genes and gene products at the molecular function, biological process, and cellular component levels) (Ashburner et al., 2000).
7. Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway (a comprehensive knowledge base for pathway information found in genes, gene products, drugs and diseases) (Kanehisa & Goto, 2000).
8. BioCarta Pathways (an online service to provide pathway information at molecular levels, for the understanding of gene interactions) ("BioCarta_Pathways," n.d.).
9. UniProtKB/Swiss-Prot (an integrated knowledge base that provides information on protein functions and sequences) (Boutet, Lieberherr, Tognolli, Schneider, & Bairoch, 2007).
10. Entrez Gene (a well-known, comprehensive, gene-specific database with a focus on completed sequenced genomes from National Center for Biotechnology Information (NCBI) (Maglott, Ostell, Pruitt, & Tatusova, 2005)).
11. Drug-Drug Adverse Drug Event Associations database (TWOSIDES) (a drug-drug interaction resource derived from the analysis of spontaneously reported ADEs) (Tatonetti, Ye, et al., 2012a).
12. Gene-GO file from NCBI (a text file that includes gene-gene ontology relationship information)("Home - Gene - NCBI," n.d.).

13. Side Effect Resource (SIDER) (a free source describing drugs with their side effects, targets and indications) (Kuhn, Campillos, Letunic, Jensen, & Bork, 2010).
14. ClincalTrials.gov (an international source for information from clinical trials)("Home - ClinicalTrials.gov," n.d.).
15. DDI source by Ayvaz *et al* that combines 14 different drug interaction sources including 5 clinical sources, 4 natural language corpora and 5 pharmacovigilance sources (Ayvaz et al., 2015).

These data sources and ontologies have been chosen to be semantically represented, integrated and stored in the D3 knowledge base in order to fill in knowledge gaps in drug information addressed by the thesis, as well as to achieve better knowledge discovery by other researchers. In the next section, we will describe in detail the process of building the D3 knowledge base. This includes the identification of biomedical entities, normalization and insertion of semantic relationships, and the process of transformation, normalizing, and storing D3 RDF data, which includes direct insertions from UMLS and mapping external sources.

III.1.b Constructing the D3 OWL (Schema) and Inserting Semantic Relationships

Different biomedical sources record similar information, but vary in the level of detail and ways they represent that information. For example, the DrugBank knowledge base records information about warfarin and its related enzymes,

transporters, etc. The side effect resource SIDER also holds warfarin side effect information. In order to incorporate these sources, their entities (known as resources in the Semantic Web) need to be identified consistently. These entities are identified and named by URIs. The URIs in the Semantic Web serve two functions: (1) as identifiers, and (2) as the locations of network endpoints. There have been many efforts to normalize URIs, such as the Life Science Record Name (LSRN) project , OKKAM , the Minimum Information Required in the MIRIAM registry, and Bio2rdf ("Life Science Resource Name (LSRN)," n.d., "Okkam - Thinking identifiers!," n.d.; Nolin et al., n.d.; Novère et al., 2005). Though these efforts have offered different URIs that work toward seamless integration, there is still no complete schema for identifying biomedical entities (Sahoo, Bodenreider, Zeng, & Sheth, 2007). As a result, we chose to define our own schema as the Drug-Disease ontology. For representing and storing D3 knowledge base concepts, we chose to rely on the UMLS. More specifically, the UMLS Concept Unique Identifiers (CUIs) will be converted into the D3 knowledge base URIs, and these form the reference set of concepts by which the other data sources are to be mapped. That is to say, the UMLS produced by NLM (version 2015 AA) was used as the backbone of the D3 knowledge base since UMLS has been a stable integrated biomedical information source for more than 20 years ("UMLS - Release Notes," n.d.; Vizenor, Bodenreider, Peters, & McCray, 2006).

Apart from its longevity, the UMLS has also been chosen for the following reasons: (1) UMLS by nature is an integration system (**Figure 3.1**), thus providing an integrated knowledge base for the proposed system;

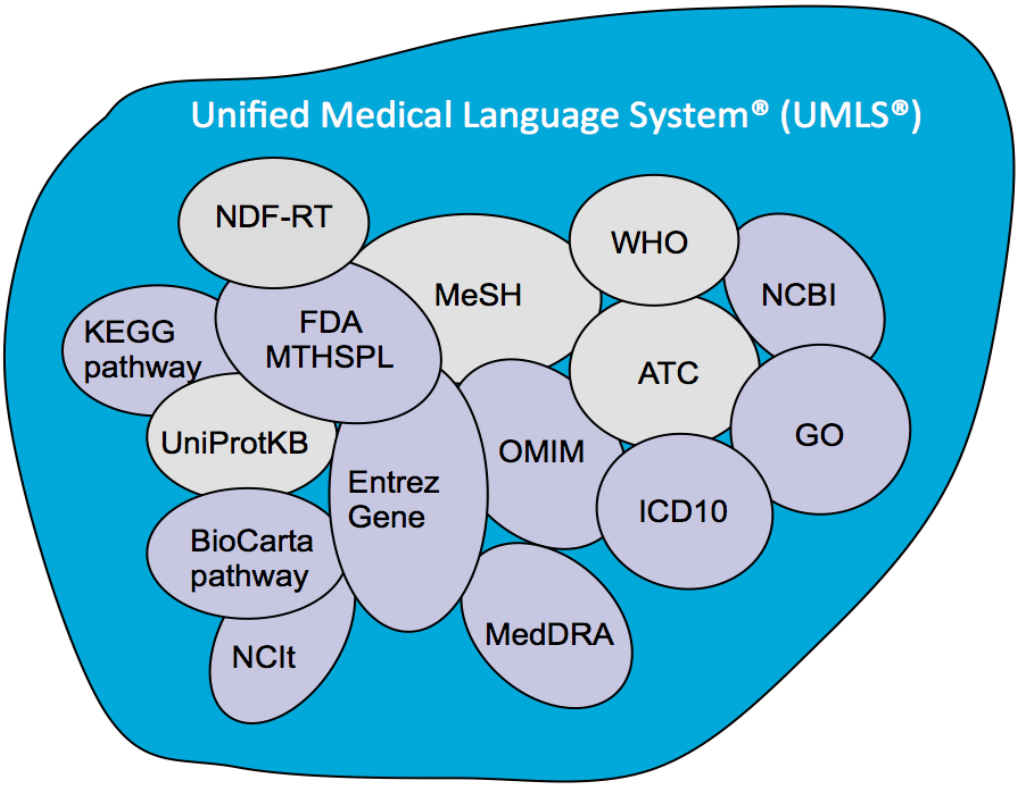


Figure 3.1: Different biomedical sources already integrated into UMLS

(2) UMLS is an enormous repository, currently containing 170 source vocabularies and biomedical ontologies, thus providing system extensibility; (3) UMLS is updated periodically, and using it as a basis makes the D3 system updatable to extend its utility lifetime; and (4) UMLS covers a comprehensive list of biomedical sources and provides semantic relationships so that D3 can be used as a model for data integration and a schema for knowledge discovery. In the next section, we will

discuss the process of building the D3 knowledge base. It contains two construction steps: (1) creation of the D3 OWL profile, and (2) establishment of the D3 RDF network.

III.1.b.1 D3 OWL Construction

As discussed before, OWL extensions to the RDF help capture additional knowledge. For example, OWL supports many useful features such as inheritances between classes, restrictions, inferences and consistency checking. However, most biomedical sources are not presented in OWL format. As a result, converting them into OWL is labor intensive and time consuming. In addition, there is no single comprehensive OWL for all types and relationships to describe objects within the biomedical area. The Semanticscience Integrated Ontology (SIO), the Semantic Network of the UMLS, and several studies reviewed by Smith and others are all examples of OWL schemas for linking entities in the Semantic Web (At, 1989; Dumontier et al., 2014; Smith et al., 2005). We examined all of these and did not find one that could cover all the relationships and objects in the D3 knowledge base. Thus, for this thesis, we use a self-created OWL for storing the D3 knowledge base schema (classes and relationships) and for consistency checking.

Protégé version 4.3 was chosen for building the OWL representation. The OWL has been used to create classes and subclasses in the D3 knowledge base to hold drug profiles including: pharmacological profiles, with subclasses (involving drugs, pharmacokinetics, and proteins); bimolecular profiles, with subclasses

(involving genes, pathways, cellular components, biological processes and molecular functions); physiological profiles, with subclasses (involving mechanisms of action and physiology); genetic profiles, with subclasses (involving alleles, SNPs and haplotypes); and finally, phenomenological profiles, with subclasses (involving diseases and side effects)(Knublauch, Fergerson, Noy, & Musen, 2004). In preparation for the remaining thesis steps, semantic relationships that link instances in the D3 knowledge base's RDF graph have been inserted from UMLS and numerous external resources.

III.1.b.2 Normalization and Insertion of D3 Semantic Relationships

Meaningful relationships between entities are key not only for reasoning but also for integration of biomedical entities into the D3 knowledge base. However, as we stated above, there is unfortunately no one formal model for defining semantic relationships. Hence, in D3, semantic relationships between instances have been inserted and normalized from the UMLS Metathesaurus and numerous external resources. From UMLS, the needed semantic relationships were retrieved from the MRREL, which is a table in UMLS that stores semantic relationships, using a MySQL query. Other semantic relationships have been inserted from the external ontologies such as DrugBank, PharmGKB, and others listed above. Semantic relations such as 'induces', 'inhibits' and 'substrate' were stored as text in the DrugBank knowledge base; thus, it was necessary to build a Java regular expression function to extract and normalize them and then to map them to the D3

knowledge base. Moreover, all semantic relationships between entities have been studied and grouped as much as possible. For instance, it was recommended by NDF-RT to group the four different relationships *may_treat, may_prevent, may_diagnose*, and *induces* into only one causal relationship, *has_indication*, between drug-disease pairs ("National Drug File – Reference Terminology (NDF-RT™) Documentation," n.d.). To provide another example, NCIt linked irinotecan to the CYP3A4 enzyme using the *chemical_or_drug_is_metabolized_by_enzyme* relationship, while DrugBank linked them by *substrate_of*. Since both relationships connect the same resources (irinotecan to CYP3A4), the D3 knowledge base grouped and stored them as *drug_is_metabolized_by_enzyme*. This grouping simplifies the relations in the knowledge base and removes redundancies. In its current phase of development, the D3 knowledge base has a total of 116 semantic relationships that link instances in its knowledge base, providing it with a rich set of relationships for inferencing purposes.

III.1.c Creation and Loading of RDF into the D3 Knowledge Base

The challenge of data integration is well documented in functional areas such as life science and translational research ("Health Care and Life Sciences (Semantic Web) Current Status - W3C," n.d.). The principles of the Semantic Web specify solutions for efficient integration, reuse and discovery of a knowledge base. As stated above, an RDF represents data in the form of a graph. <Subject, Predicate, Object> are the components of the RDF statement known as triples. These triples are represented as one database object. Moreover, these triples form a complex

graph like the World Wide Web and can be linked together in enormous and intricate graphs. RDF by nature has one universe for all RDF data stored. Therefore, when two entities share the same URI, they will be linked. In our study, we are using RDF to incorporate the data that we have identified as significant to achieve our goal. In this section, we will illustrate the process of creating D3's RDF network. The process includes extracting data from UMLS and mapping external sources.

III.1.c.1 Extracting Data from UMLS into the D3 RDF Network

As mentioned earlier, we are building our D3 knowledge base based on the UMLS Metathesaurus, and all other biomedical domains are integrated, or mapped, to it. The UMLS Metathesaurus contains multiple terms and vocabulary from different resources with their relationships. When terms or vocabulary mean the same thing, UMLS groups them and assigns a CUI (**Figure 3.2**).

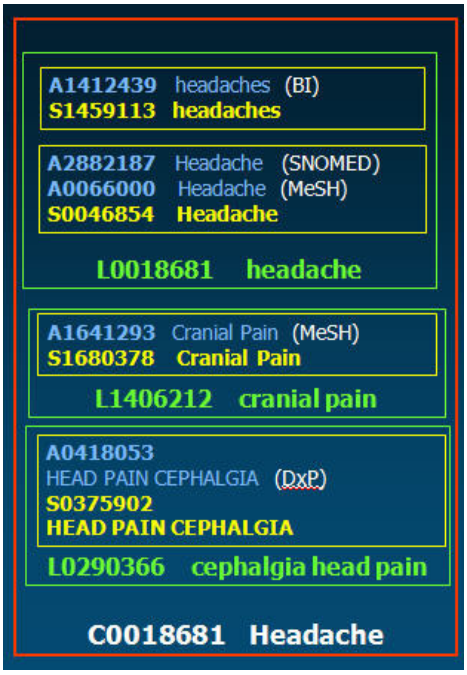


Figure 3.2: UMLS integration methodology. Adapted from http://www.nettab.org/2007/slides/Tutorial_Bodenreider.pdf

In short, UMLS is semantically oriented. Additionally, when UMLS integrates sources, it retains those sources' original identifiers. For instance, in **Figure 3.2**, UMLS grouped multiple terms (synonyms) related to the concept of headache and assigned C0018681 as a unique ID. In doing this, it also stored A0066000, which is the concept ID for headache in the Medical Subject Headings (MeSH) knowledge base (Lipscomb, 2000). We are using the appropriate external IDs stored by UMLS for mapping purposes.

The UMLS Metathesaurus dataset version 2015AA was downloaded from the UMLS knowledge server with a MySQL loading script. Metathesaurus has 48 files, of which we are using four: MRCONSO to standardize drug names, MRREL to add semantic relationships, MRSAT to map external sources, and MRSTY to check correct semantic groups and assure the quality of the inter-source mapping process (Bodenreider & McCray, 2003). Jena application-programming interface (API) is used to create and store the D3 knowledge base (Carroll et al., 2004; Owens, Seaborne, Gibbins, & Schraefel, 2008). For example, to add pathway instances using KEGG from UMLS, a MySQL query was issued to retrieve all CUIs (instances) under specific conditions. For all vocabulary and ontologies inside UMLS, the same MySQL technique has been applied to generate the D3 RDF with small modifications to reflect the source of interest in MRSAT fields. In preliminary work, an RDF network of GO, NDF-RT, KEGG, NCIt, and NCBI genes from UMLS has been successfully generated and stored in a local D3 RDF triple store database. Moreover, for each entity in the D3 knowledge base, we have added a label, a

semantic type and its external references. UMLS has 133 semantic types, which provides for terminology categorization in multiple biomedical domains (At, 1989). We have inserted semantic types as properties for our RDF for the purpose of checking each entity type.

III.1.c.2 Mapping External Resources into D3 Knowledge Base

Despite the fact that RDF triples have been successfully created and stored in a local D3 triple store file, in order to achieve better knowledge discovery it is also necessary to add other external biomedical resources to fill in gaps between multiple biomedical domains described by the source database. Therefore, PharmGKB, DrugBank, ClinicalTrials.gov, SIDER, DDI source by Ayvaz, and NCBI (gene-GO associations) have also been added to the D3 knowledge base. This mapping is based on a cross-referencing technique. That is, if two different resources in two different databases refer to each other by ID, they can be linked together through what is known in the Semantic Web as an x-reference. As stated above, when UMLS integrates sources, it retains the original identifiers, which can be used to establish the D3 mapping process. To map the DrugBank dataset, for instance, a Java program has been written to query the dataset and look for cross-referencing identifiers in drugs, enzymes, transporters, drug carriers and targets. After that, cross-references were checked one by one and a MySQL query (in a specific field that matches the cross-references) was issued to link the cross-references to its UMLS CUI. If a link was found, the DrugBank triple was

integrated into UMLS. For example, irinotecan was added to the D3 knowledge base from DrugBank by cross-referencing its entry in the Anatomical Therapeutic Chemical (ATC) classification system, which codes drugs based on therapeutic, pharmacological and chemical properties, active ingredients, and systems or organs affected ("WHOCC - ATC/DDD Index," n.d.). The ATC system uniquely identifies irinotecan with the value L01XX19, and since the ATC data source is already included in the UMLS Metathesaurus, UMLS was queried to retrieve the CUI associated with the L01XX19 value. The result was the CUI of irinotecan, C0123931. Then, the SPARQL/update technique was applied to modify the DrugBank resource by replacing the DrugBank identifier of DB00762 with C0123931 (Seaborne & Manjunath, 2007). In this way, the D3 knowledge base was enriched by inheriting central properties and objects from DrugBank. In this study, all external sources were stored as CUIs to avoid redundant linking between D3 URIs and to keep inference syntax simple. The same methodologies are being applied to integrate the other external biomedical resources in a consistent way. For other data sources where there was neither a direct link nor an intermediate source to provide a link, we mapped them by an exact case-insensitive match of a semantic string. For example, ClinicalTrials.gov was mapped by string-matching a drug name in the ClinicalTrials.gov database with UMLS, and the consistency of mapping was checked by identifying the correct semantic type of CUI in the UMLS. Data sources covering information such as DDI sources and gene-GO relationships that come in text form were converted to RDF format and added to the D3

knowledge base as well. All biomedical resources listed in the data source section have been successfully integrated into the D3 in its current form, leading to a solid knowledge base at different biomedical levels. After integrating all significant biomedical resources into the D3, its knowledge base size is 2.01 GB. Further, the number of relationship triples is more than 30 million, with 32 classes (genes, drugs, diseases, etc), 116 properties (*has_indication, has_SNPs*, etc), 35,158 drugs, and more than 50,000 DDI pairs. **Figure 3.3** shows the aspirin resource inside the D3 knowledge base.

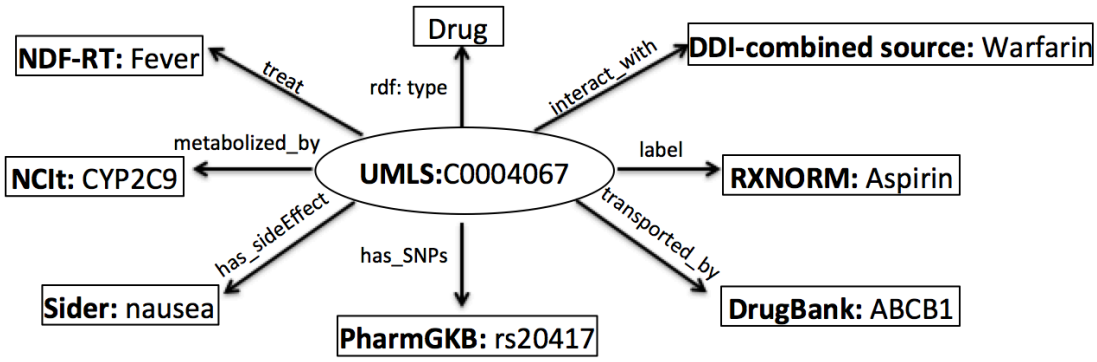


Figure 3.3: Aspirin resource in D3 knowledge base

III.1.d Consistency Checking and Validation of the D3 Knowledge Base

This thesis proposes an integrative identification, validation, and classification system for proven and potential DDIs using existing knowledge sources and the Semantic Web. An innovative aspect of this system is its intention

to both integrate diverse knowledge sources for drug-centric entities and show the benefit of semantic synergism.

The D3 knowledge base was built as discussed above by integrating various biomedical sources, which could introduce inconsistencies, a common problem when integrating different sources in the Semantic Web. Thus, it is important to check for these inconsistencies to avoid incorrect inference. Pellet, which is an OWL-DL reasoner, has been chosen to check for inconsistencies (Sirin, Parsia, Grau, Kalyanpur, & Katz, 2007). Pellet is integrated with Jena and supports extensive reasoning functionalities such as consistency checking, classification and realization. Moreover, Pellet's resulting inferred model of reasoning can be efficiently queried using SPARQL. First, the Pellet library has been added and Java code has been written to validate the D3 knowledge base.

The D3 knowledge base has been found to be consistent, and the classification and realization processes were completed successfully by adding main classes for each instance in the knowledge base (**Figure 3.4**).

```
Adeebs-MacBook-Pro-2:DDID adeebnoor$ java -cp ddidRun.jar -Xmx8G -XX:MaxPermSize=8g ddid.TestCaseReasoner
 ... Test Case 1: Peleet Reasoner E   ...
 ... Connecting to DDID tdb ...
 ... Running Pellet Reasoner and Storing results in new TDB ...
creting infeer dataset
creting OntModel
adding schema (OWL) to OntModel
Classifying 39 elements
Classifying:  100% complete in 00:00
Classifying finished in 00:00
Realizing 39 elements
Realizing:  100% complete in 00:00
Realizing finished in 00:00
adding data (RDF) to OntModel
data is consistencies
saving infead model
closing infeed dataset
 ... Connecting to Inferred tdb ...
Done
Adeebs-MacBook-Pro-2:DDID adeebnoor$
```

Figure 3.4: Results of D3 knowledge base consistency checking and validation

III.2 Adding MIRIAM Registry for Linking Life Science Data

The power of the Semantic Web is its usefulness in linking separate biomedical sources. Data from different sources and formats can be easily integrated after modeling in the RDF. Every resource in the RDF graph model has a unique URI reference, and this URI is the key to linking biomedical sources. When two different sources use the same URIs, they can be easily linked. Even better, related concepts can also be linked via URIs, as when one database, **source1**, uses a URI from another database, **source2**, as a cross-reference. For example, a PharmGKB source can be seamlessly linked to a UMLS concept, because the PharmGKB source refers to the UMLS concept; hence it is easy to connect the two sources (**Figure 3.5**).

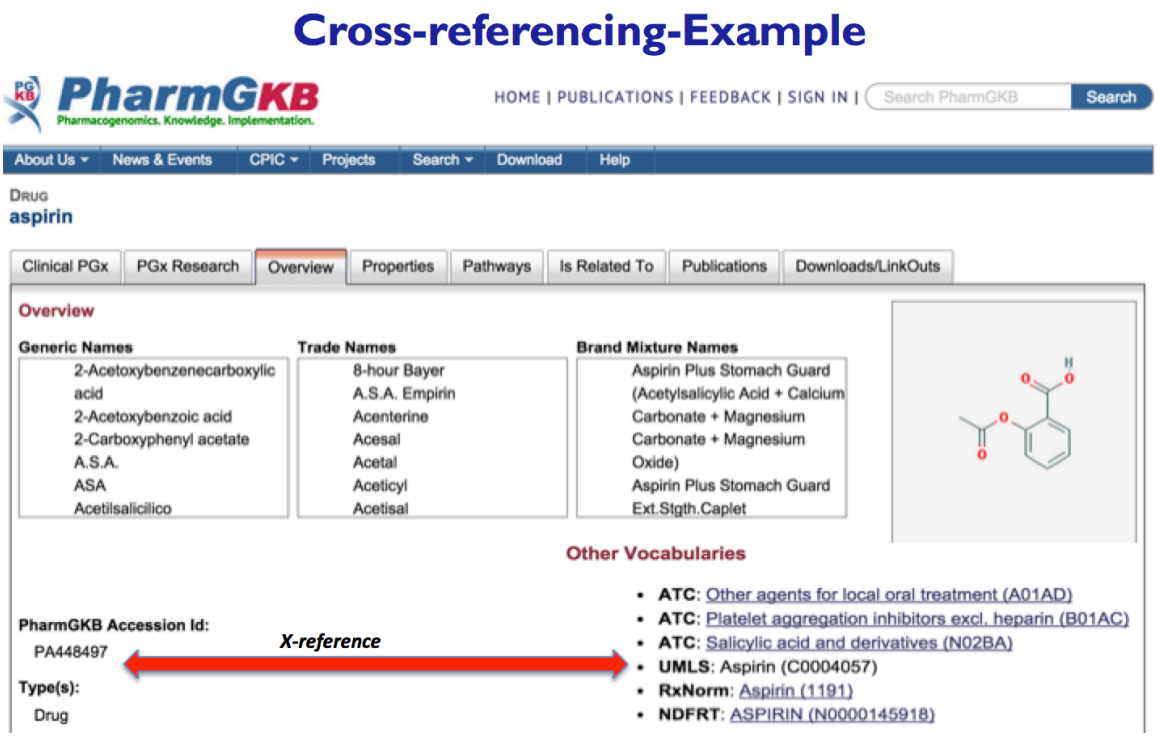


Figure 3:5: PharmGKB x-reference to UMLS

As a result, it is important that scientists use the same URI when referring to the same thing. Unfortunately, with the explosive growth of biomedical sources in recent years, many URIs are redundant or inconsistent. In response to these issues, scientists and researchers have produced different schemas to normalize URIs, such as OKKAM, MIRIAM, and Bio2rdf.

The MIRIAM registry maintained by the European Bioinformatics Institute provides a unique URI to normalize biomedical concepts. The core of life science data in MIRIAM's registry uses *Identifers.org* to provide online resources for seamless integration with external sources (Juty, Novère, & Laibe, 2012). MIRIAM recently has been adopted by the Semantic Web community in frameworks such as Bio2RDF and EMBL-EBI for normalizing URIs and tracing links mediated by cross-referencing (Jupp et al., 2014). The MIRIAM registry has published 557 life science data sources such as NCBI Gene, GO, and DrugBank, and has provided 734 URIs ("MIRIAM Registry," n.d.). In this section, we will show how we have added cross-referencing URIs to the D3 knowledge base for easy linking, and we will demonstrate the effectiveness of adding the MIRIAM registry URIs by linking the Online Mendelian Inheritance in Man (OMIM) source into the D3 knowledge base (Hamosh, Scott, Amberger, Bocchini, & McKusick, 2005).

While constructing the D3 knowledge base, our main goal was to make it applicable for a wide range of potential DDI discoveries. One way to achieve such a goal is by linking D3's knowledge to external knowledge bases. We chose the MIRIAM registry to provide the cross-referencing link for the D3 knowledge base.

The process of adding the registry for cross-referencing URIs into D3 included the following steps:

1. When adding a resource from UMLS to the D3 knowledge base, store its original ID and original database in a list along with its CUI.
2. Check whether the MIRIAM registry website has provided an identification scheme for the original source.
3. If there is identification, retrieve both the namespace as well as the URI; otherwise; the link cannot be established.
4. Finally, add a new property to the D3 resource with the namespace as name of the relationship and URI as object + the original ID.

ABCB1 gene provides an example to illustrate this process. The ABCB1 gene is stored in the D3 knowledge base as D3:C037662.

As stated above, one advantage of UMLS is keeping the original ID when it integrates and assigns the CUI. The C0376622 has its original OMIM id = 171050. To add the MIRIAM registry URI, we checked its website and found that MIRIAM provided a unique URI for each OMIM source. As a result, we created a new property for the C0376622 concept in the D3 knowledge base to store the x-OMIM reference (**Figure 3.6**).

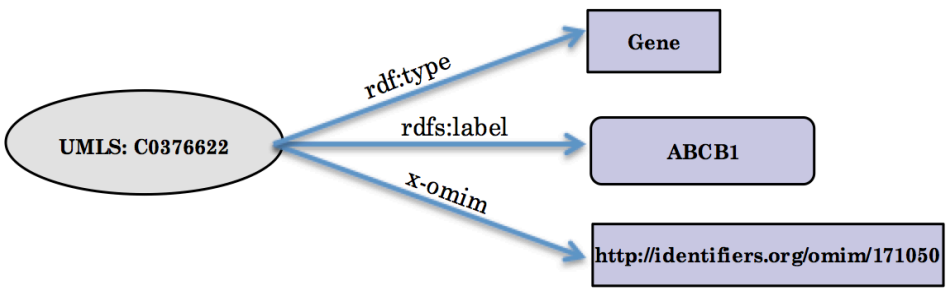


Figure 3.6: ABCB1 resources with Miriam registry URIs

In the D3 knowledge base, we have successfully added the external MIRIAM registry URIs for these sources: DrugBank (drugs and targets), PharmGKB (drugs, genes, and diseases), Entrez Gene, Gene Ontology, OMIM, KEGG, NDF-RT, SIDER, and NCIt. These URIs will provide our D3 knowledge base with implicit linkability. Next, we illustrate how such a link can be easily achieved using the MIRIAM registry URIs. Specifically, we would like to be able to easily query the OMIM source using the D3 knowledge base.

OMIM is a human genotype and phenotype knowledge base that supports human clinical genetics research. It has been widely used to identify the relationships between genes and diseases (Feramisco, Sadreyev, Murray, Grishin, & Tsao, 2009; J. Wang et al., 2008). The OMIM source was transformed and stored as an RDF file by the Bio2RDF team, and they have made the source available for query or download ("Statistics for Online Mendelian Inheritance in Man," n.d.). We downloaded the RDF version from the Bio2rdf server and stored it using a Jena triple store database. The OMIM RDF version has a property *Bio2RDF: x-identifiers.org* that links a resource to its URIs in the MIRIAM registry. We used this property when querying to link between our D3 resources and the Bio2RDF one. **Figure 3.7** shows an example of a SPARQL query where we linked our D3 with the OMIM source and asked about the ABCB1 gene ID in the Bio2rdf source:

```
PREFIX ddids:<https://csel.cs.colorado.edu/~noor/Drug_Disease_ontology/DDID.owl#>
PREFIX ddidd:<https://csel.cs.colorado.edu/~noor/Drug_Disease_ontology/DDID.rdf#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX bio2rdf: <http://bio2rdf.org/omim_vocabulary:>

SELECT DISTINCT  *
WHERE {
?umlsCUI ddids:x-omim <http://identifiers.org/omim/171050> .
?subject bio2rdf:x-identifiers.org <http://identifiers.org/omim/171050>
}

```

Figure 3.7: Link D3 to Bio2RDF via Miriam registry

The result of the query can be found at Bio2rdf website. We used the Bio2rdf server for illustration purposes here. Using the power of the Semantic Web and the utility of x-identifiers.org, we show how seamless the integration between biomedical sources can be. This will greatly aid in aggregating information to answer complex biomedical questions. We have also showed how D3 can be easily expanded and linked to important sources.

III.3 Benefits of Semantic Mining in Identifying Embedded Associations for a Chemotherapy Medication

While more complex cases demonstrating the utility of the D3 system in modeling and answering a specific DDI question will be discussed in detail in **Chapter V** of this thesis, the current case serves to illustrate the D3 system's ability to uncover new relationships between information stored in distinct sources. As seen with irinotecan, effectively integrating information from multiple sources may contribute to a more detailed understanding of the subject being queried. Furthermore, this example shows the potential utility of the D3 knowledge base for additional clinical applications.

To demonstrate the usefulness of our semantically integrated knowledge base, we will use a case study to find drug-pathway associations. It is true that biochemical pathways can be straightforwardly retrieved from different knowledge bases, such as KEGG, but the implicit biochemical pathways are hard to find. Using our D3 knowledge base, we are going to find the implicit pathway for irinotecan, a chemotherapy drug for colon cancer treatment. As has been done in related work

(Qu, Gudivada, Jegga, Neumann, & Aronow, 2007), we propose 3 different SPARQL queries to identify the implicit pathways for irinotecan:

1. **Through irinotecan directly:** The query retrieves only one pathway, Caspase Cascade in Apoptosis (**Figure 3.8**).
2. **Through irinotecan interacting with genes:** The query retrieves 103 pathways (**Figure 3.9**).
3. **Through irinotecan's clinical features:** The query retrieves 21 pathways (**Figure 3.10**).

The apoptosis pathway was found through the second and third queries, while the first query yielded the Caspase Cascade, a system of aspartic acid-specific proteases responsible for the apoptosis process, with the result shown here ("Caspases: Their Role in Cell Death and Cell Survival | Sigma-Aldrich," n.d.).

```
 9 SELECT DISTINCT ?pathwaysRetrieved
10 WHERE {
11 ddidd:C0123931 ddids:has_target ?gene .
12 ?gene ddids:is_element_in_pathway ?pathways.
13
14 ?pathways ddids:label ?pathwaysRetrieved .}
15
16 ----------------------------
17 | pathwaysRetrieved         |
18 ============================
19 | "Caspase Cascade Pathway" |
20 ----------------------------
```

Figure 3.8: Mining irinotecan pathways via target

```
SELECT DISTINCT ?pathwaysRetrieved
WHERE {
ddidd:C0123931 ddids:has_target ?gene .
?gene ddids:has_interaction_gene ?gene2 .
?gene2 ddids:is_element_in_pathway ?pathways.

?pathways ddids:label ?pathwaysRetrieved .}

-----------------------------------------------------------
| pathwaysRetrieved                                       |
===========================================================
| "Apoptosis"                                             |
| "PPAR Signaling Pathway"                                |
| "Cell Cycle"                                            |
| "CDK Regulation Pathway"                                |
| "Cyclin E Degradation Pathway"                          |
| "Cyclin Regulation of Cell Cycle Pathway"               |
| "E2F1 Degradation Pathway"                              |
| "RB Checkpoint Signaling Pathway"                       |
| "p27 Regulation Pathway"                                |
| "p53 Signaling Pathway BioCarta"                        |
| "G1/S Checkpoint Pathway"                               |
| "G1 to S Transition Pathway"                            |
| "CTCF Pathway"                                          |
```

Figure 3.9: Mining irinotecan pathways via interacting genes

```

SELECT DISTINCT ?pathwaysRetrieved
WHERE {
ddidd:C0123931 ddids:has_indiction ?disease.
?disease ddids:disease_has_associated_gene ?gene .
?gene ddids:is_element_in_pathway ?pathways.

?pathways ddids:label ?pathwaysRetrieved .}

-------------------------------------------------------
| pathwaysRetrieved                                   |
=======================================================
| "Apoptosis"                                         |
| "Cell Cycle"                                        |
| "p53 Signaling Pathway KEGG"                        |
| "ATM Signaling Pathway"                             |
| "CDC25 and CHK1 Regulatory Pathway"                 |
| "Hypoxia Pathway"                                   |
| "RB Checkpoint Signaling Pathway"                   |
| "p53 Signaling Pathway BioCarta"                    |
| "Apoptosis DNA Damage Pathway"                      |
| "Cancer Susceptibility Pathway"                     |
| "Cell Cycle Progression Pathway"                    |
| "G1/S Checkpoint Pathway"                           |
| "G2-M Cell Cycle Checkpoint"                        |
```

Figure 3.10: Mining irinotecan pathways via clinical features

In fact, the mechanism of action of irinotecan is known to lead to apoptosis (Pommier, 2006). Irinotecan works by inhibiting the topoisomerase enzyme, which is responsible for DNA copying and transcribing. The inhibition happens when irinotecan binds to the topoisomerase I-DNA complex that leads to the destruction

of that complex. As a result, the cell's DNA no longer copies and transcribes, and this leads to apoptosis. Other retrieval pathways identified by D3, such as the p53-signaling pathway, are shown to be related to cancer (Sherr & McCormick, 2002). These preliminary results could be used for repurposing existing drugs for new indications. Studying these inferred pathways within a disease and the role that irinotecan plays in these pathways could help to find a new possible indication for other types of cancer. Further studies would need to be conducted to determine these pathways and to clarify their role in the cell cycle in relation to irinotecan. In summary, the Semantic Web provides an excellent way to mine biomedical resources to explore clinical applications.

III.4 Conclusion

In this chapter, we have introduced a complete and comprehensive DDI knowledge base using Semantic Web technologies. This knowledge base integrates multiple biomedical sources by implicitly transforming and mapping their instances into one complete resource. In order to accurately map and represent the knowledge discovered by D3, we introduced two novel methods. One exploits the UMLS semantic types network to accurately identify a concept; while the second groups similar semantic relationships into one.

The result of this work is a complete D3 knowledge base that improves pharmacovigilance study. Unlike other approaches in the literature such as (Bonnabry et al., 1999; R. Boyce et al., 2009a, 2009b), or (Preissner et al., 2010) D3

covers the DDI mechanisms at multiple biomedical levels. In particular, our proposed knowledge base is capable of modeling, discovering and explaining the different biomedical mechanisms that could contribute to DDIs. This D3 knowledge base is also able to expand, adapt and update information easily. Another main focus highlighted in this chapter is that the proposed D3 knowledge base is generic and flexible enough to be integrated with diverse biomedical sources. In this chapter, we have illustrated this innovation by integrating OMIM sources.

In order to show the usefulness of our D3 knowledge base, we also identified a new potential application for a chemotherapy medication, showing that D3 could be useful for drug repurposing and discovering ways that existing medications could be used to effectively treat new diseases (Boguski, Mandl, & Sukhatme, 2009).

Notably, a fundamental issue that is not addressed in this chapter is the critically important process of discovering potential DDIs utilizing the D3 knowledge base. In general, our approach builds a complete DDI knowledge base that models different mechanisms of interaction, allowing us to identify, validate, and classify reported and potential DDIs in a more effective way. In the next chapter, we investigate this potential by generating different inference tasks for the D3 system.

CHAPTER IV

DRUG-DRUG INTERACTION KNOWLEDGE DISCOVERY USING PROTOTYPE SEMANTIC MINING INFRASTRUCTURE - A METABOLIC AND TRANSPORTER INTERACTION CASE STUDY

In the previous chapter, we established a comprehensive knowledge base called D3 that models DDI mechanisms at multiple biomedical levels. So far, this D3 knowledge base has been dealt with solely as a collection of individual sources, ignoring the potential for valuable semantic relations that might exist between them. In this chapter, we examine our fundamental hypothesis that by employing the semantic relationships contained within the D3 knowledge base, we provide the means to better identify and characterize both proven and potential DDIs and to offer a better pharmacovigilance system.

This chapter focuses on utilizing the semantically integrated D3 knowledge base to identify complex interactions as well as to reveal previously unsuspected mechanistic interactions. The novelty of the interaction model proposed here is its ability to identify potential DDIs across multiple biomedical levels. Such a model can then be used to discover hidden DDI information in existing sources, improving

understanding of DDIs at the mechanistic level. However, in order to develop an effective model, we need to carefully address three questions:

- What factors should be considered when designing a model to utilize the rich semantic relationships of the D3 knowledge base?
- How do we exploit the vast integrated relationships within the D3 knowledge base to comprehensively assess the potential for DDIs?
- How do we develop a model to find potential DDI information that is not explicitly evident within knowledge sources?

In this chapter, we address these questions by introducing the importance of multi-pathway DDIs, a complex type of interaction, which occurs due to two or more interactive mechanisms. For instance, cyclosporine interacts with many cholesterol-lowering HMG-CoA reductase inhibitors (statins) by both inhibiting CYP3A4-mediated drug metabolism and organic anion transport polypeptide- (OATP-) mediated transport into hepatocytes, the site of the statins' therapeutic activity (Asberg, 2003). The result of this interaction is not only a decrease in the statins' therapeutic efficacy, but also an increase in their toxicity due to higher drug plasma concentrations because of reduced metabolism and drug elimination.

To study multi-pathway DDIs, we present a novel method using the Semantic Web's inference capabilities to examine drug information across multiple mechanistic levels. Specifically, we propose a rule-based model to discover potential DDIs arising from metabolic and transporter interactions. We then demonstrate the

validity of this approach using irinotecan, a chemotherapy agent. By linking mechanistic data from available DDI sources, we demonstrate the capacity of the D3 system to identify potential interactions with irinotecan, including the discovery of new potential interactions across previously unidentified mechanisms.

IV.1 Method

To demonstrate the usefulness of the semantically integrated D3 knowledge base in identifying complex interactions, a preliminary case study is designed to identify multi-pathway interactions. This includes choosing a real world test case drug and applying inference.

IV.1.a A Motivating Scenario: The Colon Cancer Medication Irinotecan as a Case Study

We choose to explore new interactions for irinotecan, a chemotherapy drug. Irinotecan is a nuclear DNA topoisomerase I (Top1) inhibitor used to treat metastatic colorectal cancer. Its mechanism of action involves inhibition of the Top1 enzyme, a critical mediator of DNA transcription and replication. As a result, irinotecan promotes cancer cell death, leading to its therapeutic efficacy. Despite the effectiveness of irinotecan, myelosuppression (the slowing of bone marrow activities related to blood and immune cell production) and diarrhea, which results from the direct action of irinotecan on the intestinal mucosa, are established toxic effects of irinotecan (Xu & Villalona-Calero, 2002).

Irinotecan is eliminated from the body in the bile via the efflux transporter P-glycoprotein (P-gp). As a result, any induction or inhibition of the P-gp transporter will effect irinotecan concentrations. For instance, cyclosporine (which was one of the results our model infers) has been shown to decrease the elimination of irinotecan by 39% to 64% due to inhibition of P-gp (Innocenti et al., 2004). Additionally, the enzyme CYP3A4 plays an important role in irinotecan metabolism (Haaz, Rivory, Riché, Vernillet, & Robert, 1998, p. -450). The impact of CYP3A4 on irinotecan metabolism is seen when voriconazole, a CYP3A4 inhibitor, is co-administered with irinotecan, resulting in increased irinotecan plasma concentrations and subsequently increased cellular toxicity. Therefore, DDIs involving irinotecan can occur at either the metabolism or transporter levels. For example, nefazodone (which was another of the results our model infers), an antidepressant that inhibits CYP3A4 and induces P-gp, influences the disposition of irinotecan via two distinct mechanisms when the medications are co-administrated (Ma & McLeod, 2003). To comprehensively characterize the potential for drug interaction with irinotecan, then, a DDI knowledge base must be able to integrate drug information across a variety of mechanisms in order to accurately model drug disposition.

Finally, irinotecan has been the subject of extensive DDI study due to its very complicated pharmacokinetic profile that makes its interactions with other drugs more likely (Mathijssen et al., 2001). More specifically, it has been shown that most of the toxic effects of irinotecan result from variations of the UGT phase II enzymes

(Hahn, Wolff, & Kolesar, 2006). Therefore, due to its complex pharmacokinetic profile and high potential for serious ADRs, such as myelosuppression and diarrhea, we selected irinotecan as a robust case to illustrate the utility of the D3 system in identifying potential drug interactions.

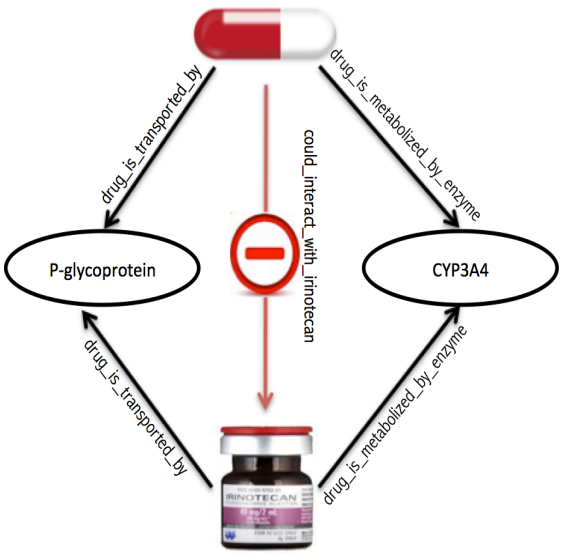
A semantic rule-based model has been designed to infer drugs that could potentially interact with the same metabolic pathways and transporters that mediate irinotecan disposition. Following this hypothesis, the PharmGKB knowledge base will be used as a source of evidence to adjust the semantic rule to account only for drugs that share metabolic and transporter pathways with irinotecan.

IV.1.b Identifying Inference Rules

Inference—the process of reasoning from evidence and conditions to produce conclusions *("Apache Jena - Reasoners and rule engines: Jena inference support," n.d.)*—is one of the most powerful tools of the Semantic Web. This process can be done using mainly OWL, RDFS (subClassOf) or RDF, using different inferring techniques such as description logic (DL) and rule engines (RE) ("Apache Jena - Reasoners and rule engines: Jena inference support," n.d., "The Description Logic Handbook," n.d.). Jena offers inferences from a knowledge base through reasoners and rule engines. For this study, we use Jena's generic rule reasoner for inferring purposes. The engines produce valid statements within a system based on rules. Moreover, Jena's generic rule reasoner includes three models: forward chaining,

backward chaining and a hybrid model. The forward chaining works by looking at the facts to derive a conclusion; a typical example of this model occurs in medical diagnosis. For instance, if a patient has a positive ELISA test for anti-HIV antibodies and a detectable HIV viral load in his blood sample, we can diagnose him with HIV. The backward chaining, on the other hand, works in the opposite way; i.e., it starts with a conclusion (goal) and tries to satisfy it. Considering the same HIV example, a physician (using backward chaining) assumes a patient has HIV and then tries to find all information (facts) to validate his assumption. The last model is hybrid chaining, in which both forward and backward models can be combined.

For this study, a semantic rule-based model is predefined for reasoning throughout the D3 knowledge base to identify potential interactions involving the chemotherapy medication irinotecan. Basically, we reason through the D3 system using the forward chaining model. Potential interaction candidates for irinotecan are identified on the basis of pharmacological similarities as stated before. That is, drugs with the same metabolic pathways (e.g., being processed by the CYP3A4 enzyme) and transporter (e.g., being transported by P-gp) as irinotecan are assumed to potentially interact with this drug. **Figure 4.1** shows a graphical representation of the inference along with its forward chaining rule that is defined to identify the drugs that potentially would interact with irinotecan at the metabolic and transporter levels.




```
if      (?potenialDDIs D3:drug_is_metabolized_by_enzyme ?enzyme)
        (?potenialDDIs D3:drug_is_transported_by ?transporter)
        (?potenialDDIs D3:has_pharmGKB_gene_association ?GKBgene)
        (?targetDrug D3:label "irinotecan")
        (?irinotecanEnzyme D3:label "CYP3A4")
        (?irinotecanTransporter D3:label "P-gp")
        notEqual(?drugs,?targetDrug)
        Equal(?enzyme,?irinotecanEnzyme)
        Equal(?transporter,? irinotecanTransporter)

then    ?potenialDDIs rdf:type D3:could_interact_with_irinotecan)
```

Figure 4.1: Multi-pathway inference discovery model

IV.2 Results

After the semantic rule-based model was applied to infer irinotecan drug interactions, 215 FDA-approved drugs were identified that satisfied the requisite conditions for an interaction: they were metabolized by the CYP3A4 enzyme and transported by P-gp. To validate the outcomes, PharmGKB was used as a source of evidence to check associations between the candidate drugs and the CYP3A4 enzyme/P-gp transporter. Only 116 of the 215 drugs identified by D3 had evidence for an association listed in PharmGKB (**Table 4.1**).

Table 4.1: Model outcome validation of the irinotecan 116 predicted drugs

Validation Source	Predicted Interactions
Curated DDI sources	80 (69%)
Literature	12 (10%)
Never studied	24 (21%)

This table shows the overall validation results after running the possible drug interactions of irinotecan. 80 possible interactions have been found in the selected curated DDI sources. Among those not found in the selected sources, 12 interactions were reported in medical literature and other sources that were not used in the first analysis. This validation process yielded 24 potential interactions, which have never been studied or discussed in either the medical literature or commercial and free DDI sources.

Among the 116 drugs identified by PharmGKB, 28 were possible candidates for metabolism-based interactions (either inhibition or induction), 28 were indicated for possible transporter-based interactions (either inhibition or induction), and 8 drugs, despite not inhibiting or inducing CYP3A4 or P-gp, were co-substrates for both pathways with irinotecan (similar disposition). Additionally, 52 drugs demonstrated the potential to interact with irinotecan through both transporter and metabolic mechanisms (**Figure 4.2**).

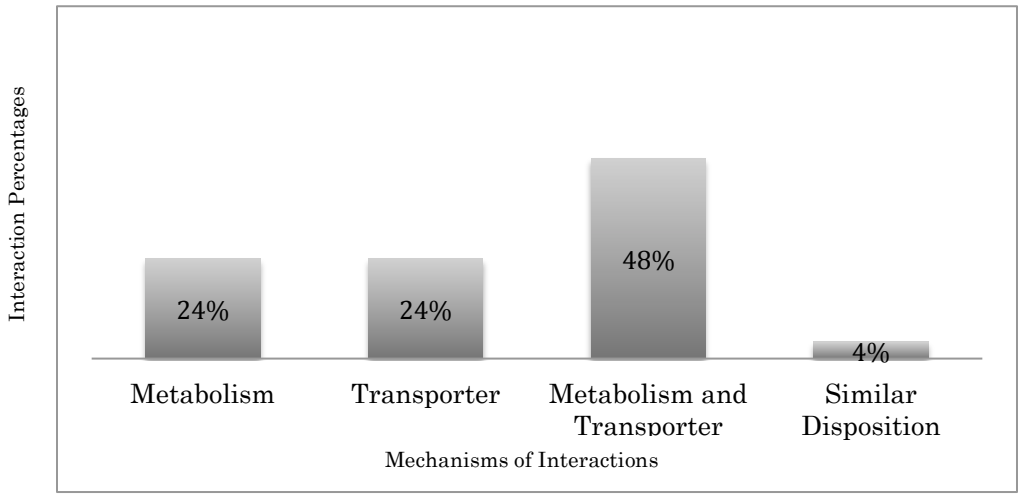


Figure 4.2: 116 Interactive drugs predicted by the model were classified into four different categories (metabolism, transporter, metabolism and transporter, and similar disposition)

Based on the observed propensity for drugs to interact with irinotecan simultaneously across multiple mechanisms, the importance of developing a knowledge base that accurately integrates information detailing all potential mechanisms of interaction is essential. Notably, the D3 system successfully detailed interaction between nefazodone and irinotecan (an example previously discussed) at both the transporter and metabolism mechanistic levels. Conversely, decision support tools such as Micromedex, Lexicomp, and Facts and Comparisons all identified only interactions between the two medications at the metabolism level through CYP3A4 inhibition ("Facts & Comparisons® eAnswers | Clinical Drug Information," n.d., "Lexicomp® Online | Clinical Drug Information," n.d., Micromedex® Healthcare Series., 2015). Comprehensively identifying all mechanisms of drug interaction is critical since interaction through one mechanism may potentiate or mitigate interaction at another, and thereby potentially contribute to clinically meaningful DDIs (Hinton, Galetin, & Houston, 2008). Therefore, the complex case of irinotecan highlights the insufficiencies of conventional DDI identification systems, namely their ability to identify DDIs solely related to one mechanism of interaction, and demonstrates the value of using integrated Semantic Web technology systems such as D3 to more accurately model potential DDIs and inform clinical decision support tools.

IV.2.a Validation

The 116 semantic rule-based DDI predictions were compared against five different commercial and free license DDI information sources, specifically PDDIs, Medscape ("Latest Medical News, Clinical Trials, Guidelines – Today on Medscape," n.d.), Lexi-Comp, Drugs.com ("Drugs.com | Prescription Drug Information, Interactions & Side Effects," n.d.), and Micromedex Solutions. These five sources were used because no single comprehensive DDI source exists. Zhang et al. suggest that using the curated data sources (such as the ones mentioned above) for validation could lead to many false positives due to the incompleteness of available curated data sources, and thus, precision could suffer (R. Zhang et al., 2014b). Tari et al. also report the limitations of using DrugBank as a gold standard for validation; they found only 11% of their results reported while 77% were found by searching the medical literature (Tari et al., 2010). These studies demonstrate that weak recall is a commonly encountered limitation when using commonly available sources. When we consider only those 116 irinotecan-interactive drugs found also in PharmGKB, the D3 validation process identified 80 DDIs collectively reported among the sources, while 36 of the potential DDIs identified in our study had no interaction reported in any of the sources investigated. Twelve of the 36 remaining drugs had evidence associated from the literature and other clinical websites (**Appendix A**).

IV.2.b Statistical Significance of Predicted Interactions

The DDIs source by Ayvaz et al. was selected as the gold standard for this study, since it integrated 14 publically available sources of DDI information from natural language corpora and from clinical and pharmacovigilance sources. Fisher's exact test was used to find the significance of the overlap between the results generated by the proposed model and DDIs. First, a two-by-two contingency table for irinotecan was generated. Then, a significant overlap between the drugs predicted by the model and the drugs that were listed to have interactions with it in the DDIs source was measured (**Table 4.2**).

Table 4.2: Fisher's exact test two-tailed p-value

	True	**False**	**P-value**
Predicted	55	37	<0.05
Unpredicted	166	1356	

The table illustrates the significant overlap between the drugs predicted by our model (116 drugs) to have interaction with irinotecan and the drugs that are listed to have interactions with irinotecan in DDIs (two-tailed Fisher's exact test, p-value = <0.05).

IV.2.c Precision and Recall of Predicted Interactions

For further evaluation of the potential interaction candidates retrieved for irinotecan, this study calculated the recall and precision of those candidates. As mentioned before, the results of initial semantic predictions were 116 drugs, of which 80 were proven (True Positive-TP) and 36 were potential (False Positive-FP), during the validation process. To identify the quality of search results, the total number of irinotecan interactions was first retrieved individually from the

sources DDIs, Lexi-Comp, Drugs.com, Micromedex Solutions and Med-scape. Then, the resulting lists of irinotecan interactions were combined, and overlapping information was removed, leading to 547 drugs being reported to interact with irinotecan among these seven sources. The recall then was computed as: $(\frac{Tp\ (80)}{TP\ (80)+\ FN\ (467)} \times 100\% = 14\%)$, and the precision was calculated as: $(\frac{Tp\ (80)}{TP\ (80)+\ FP\ (37)} \times 100\% = 68\%)$. Poor recall was due to the restriction of the inference model (i.e., a drug being metabolized by the CYP3A4 enzyme and transported by the P-gp transporter). In fact, in the application of such a DDI inquiry, relevant and precise information is more important than improper information; that is to say, the precision is more important than recall in this case study. However, in order to improve the recall, the model could be modified to account for any enzyme-transporter co-interactions.

IV.3 Limitations

Though the proof-of-concept rule-based model was able to identify possible interactions for irinotecan with 79% correct detection (69% from DDI sources + 10% from literature and clinical websites), this test study has several limitations that should be addressed in future work. First, the rule-based model checked only for potential interactions, but it did not determine the clinical significance of those interactions. In fact, determining the severity of any interaction based on mechanistic information is complex without pursuing clinical pharmacokinetic/ pharmacodynamic studies to (1) determine how the identified interactions may change the interacting drugs' concentrations, and (2) assess the

therapeutic index of the interacting medications, that is, how changes in drug concentrations will impact the drugs' therapeutic efficacy and the potential for adverse effects. This model, however, provides potential interaction candidates for pharmacologists to pursue in clinical studies to ascertain the true significance of these interactions. Another limitation is that this study examines interactions for only one drug, irinotecan. Even though irinotecan's complex pharmacokinetic profile allowed us to test the full range of our model, the value of the DDI knowledge base would be proven more definitively if more examples were provided.

IV.4 Conclusion

In this project, the D3 knowledge base is employed to identify potential DDIs and to model their mechanisms of interaction. We focus on our semantically integrated knowledge base as the main knowledge resource in this case study, due to its ability to represent more complex methods of interaction than other DDI identification systems.

In this chapter, the fundamental hypothesis is that by discovering the semantic knowledge contained within the D3 knowledge base, we obtain the means for providing a better way to understand drug interaction mechanisms and hence to provide a better pharmacovigilance system. In order to access the semantic knowledge from the D3 knowledge base, we first need a way to identify how drugs are related to each other. Therefore, in this chapter we proposed a

novel semantic method that aims to identify interaction at two mechanistic levels. This method inferred the interactions by using the information that is associated with a drug in the D3 knowledge base. Moreover, in this chapter, we also identify potential interactions that have not yet been studied clinically and characterize potential and proven DDIs based upon their mechanisms of interaction. We also conduct different evaluations to assess different aspects of our proposed approach. In the evaluation processes, seventy-nine percent of possible irinotecan-based DDIs were corroborated by 8 well-known DDI sources, while 21% were not (67% precision, 15% recall). Also, significant differences (based on a two-tailed Fisher's exact test, p-value < 0.05) were detected when a rule-based model was compared to the DDI source by Ayvaz et al. The obtained results demonstrate the effectiveness of using our method to identify new and potential drug interactions. Finally, we summarize the significant contributions in this chapter as follows:

- We developed a method to identify multi-pathway drug interactions based on metabolic and transporter similarities.
- We identified a number of potential interactions that have not been clinically studied.
- We categorized and explained the mechanisms of DDIs.
- We realized when evaluating our outcome that the establishment of a comprehensive, accurate, and evidence-based resource for DDIs would

offer significant value for both drug interaction research and clinical practice.

CHAPTER V

DATA-DRIVEN SYSTEM FOR VALIDATION AND CLASSIFICATION OF DRUG-DRUG INTERACTIONS

In **chapter III**, we established a comprehensive DDI knowledge base by semantically integrating and modeling drug information at the pharmacokinetic, pharmacodynamic, pharmacogenetic, and multi-pathway interaction levels. This D3 knowledge base was developed to be used for automated reasoning about potential and proven DDIs and their mechanisms and has an effective means of adapting to changes in drug profiles, that is, the drug information present within the knowledge bases within D3 upon which inferences are generated. We showed the ability of the semantically integrated D3 knowledge base to identify hidden pathways of drug action, which was effectively utilized in our example to propose a potential drug repositioning application for a chemotherapy medication. In **Chapter IV**, we showed that the D3 knowledge base can be used to identify, validate and classify multi-pathways DDIs. By proposing a novel semantic rule-based model that is able to exploit hidden knowledge from the relationships among D3 integrated sources, we demonstrated the capacity to identify multi-pathway potential DDIs that arise from metabolic and transporter

interactions. Up to this point, however, we have only tested the ability of our system to identify potential interactions for one mechanism type, the multi-pathway interaction. Having demonstrated the full range of our model by testing these complex multi-pathway interactions, we will proceed to assess the ability of the D3 knowledge base to characterize DDIs via each of the previously proposed mechanisms of interaction. One current limitation of our use of the D3 system is that we have not yet considered the clinical relevance of predicted DDIs. Therefore, considering both the clinical relevance of predicted DDIs and their precise mechanisms of interactions are two critical factors that we will address in this chapter.

Unfortunately, existing pharmacovigilance systems typically either identify just one type of interaction or analyze drug interaction using a single mechanistic pathway. In the former systems, the main focus is usually on pharmacokinetic interactions, such as those that occur with CYP enzymes, which are well described in clinical literature. In the latter systems, the multi-pathway potential DDI is completely ignored as these systems only possess the capacity to model potential DDIs via a single pathway of interaction. Both approaches are unlikely, then, to discover potential DDIs that have not already been identified through clinical study. In addition, these approaches are unlikely to detect DDIs through less common mechanisms of interaction. However, it is evident that multiple mechanisms of DDIs must be taken into account when providing pharmacovigilance services. This is because in real life, a potential

DDI may happen due to the interplay between many different and even simultaneously active mechanisms; hence, effective pharmacovigilance systems should be able to account for complex drug mechanisms. We also acknowledge that existing sources of DDIs contain a high proportion of non-clinically relevant DDIs.

Our aim is to design a pharmacovigilance system that can provide effective DDI discovery services. However, providing an effective pharmacovigilance system is a very complex task. This is because first, existing pharmacovigilance systems cannot support the complex and diverse mechanisms of interaction that need to be added. Second, most of the available DDI resources that could be used to assess and identify the clinical relevance of predicted DDIs vary widely in how comprehensively or consistently they report already proven DDIs. Our paradigm of providing effective DDI discovery services includes the validation of the clinically relevant DDIs and characterization of their specific mechanisms of interaction. This requires us to address the following challenging questions:

- How do we collect and integrate existing DDI information?
- Which DDI sources do we trust to forecast the clinical relevance of predicted DDIs?
- When identifying similarities between two drugs, what are the biomedical features to be considered that may constitute an interaction?
- How do we collect and organize mechanistic interaction information?

- What types of mechanisms should be considered for interaction?
- How do we represent and model mechanisms in drug profiles?
- How do we examine interactions between competing mechanisms in an effective way?
- In conclusion, how do we provide a pharmacovigilance system that is able to assess and integrate DDI information across multiple mechanisms of interaction?

In this chapter, our objective is to develop a pharmacovigilance system, called **D**rug-drug interactions **D**iscovery and **D**emystification (D3) that aims to use Semantic Web technologies to address the limitations of existing pharmacovigilance research. The D3 system, herein proposed, holds a proven DDI list collected from fifteen trusted sources and examines different mechanisms of interaction. The D3 system then employs these sources about drug interactions, and their mechanisms, as part of a pharmacovigilance recommender system. Ideally, D3 will contain two novel inferential models to study proven and potential DDIs: (1) a query-based model and (2) a probabilistic-based model.

This chapter is divided into three main sections that can be seen as the phases of development of the D3 system. In the first phase (Phase I), we discuss the identification and clinical examination of the most common mechanisms of an interaction. Then we convert those mechanisms into a set of semantic inferences

that will be used by D3 for automated discovery to both identify the mechanisms of potential DDIs and clarify the proven DDIs. The second section (Phase II) illustrates the process of constructing the D3 system. This second section is divided into two main sub-phases: (1) building an inferential query-based model, and (2) building an inferential probabilistic-based model. In last section (Phase III), we exploit both D3 models (query and probabilistic) for DDI study.

V.1 Phase 1: Declaration of D3 Semantic Inferences

Fundamentally, there are multiple avenues to pursue when characterizing DDIs, but one of the most proven methods is to examine DDIs by first determining the mechanisms of interaction between drugs and then comparing those mechanisms to known mechanisms of DDI. Understanding mechanisms of interaction is essential to informing the strategies that will enable clinicians and scientists to avoid the interaction altogether or to employ strategies that will mitigate its severity, thereby providing the means of reducing harmful ADEs. Thus, there is a pressing need to study potential mechanisms that go beyond a single level. Boyce, *et al*. have voiced the need for pharmacovigiliance research focused on modeling both pharmacokinetic and pharmacodynamic DDI mechanisms, for example (R. D. Boyce et al., 2007). In the first development phase of D3, we aim to define the most common mechanisms of DDIs and convert them into semantic inferences. In this process, D3 comprises and amalgamates facts to be used for inferences from nine different well-known biomedical sources. After that, we build

nine different inferences to extract hidden information about mechanisms of interaction from these sources. Specifically, we build nine inferences that study the following DDI mechanisms: (1) protein binding; (2) metabolism inhibition; (3) metabolism induction; (4) transporter inhibition; (5) transporter induction; (6) additive pharmacodynamic effect; (7) competitive pharmacological effect; (8) pharmacogenetics; and finally, (9) multi-pathway interactions. Notably, two described mechanisms of DDI are not included as inferences in D3 for distinct reasons. Synergistic drug effects, occurring when the pharmacological or side effects of one drug increase the pharmacological efficacy of another, were not included as they are accounted for in clinical decision-making and do not typically lead to ADEs or negative patient outcomes. Antagonistic pharmacodynamic drug effects, which occur when a pharmacological or other side effect of one drug reduces the pharmacological effect of another, are not included in our inferences since D3's similarity-based method for integrating DDI information is not able to effectively capture this interaction type.

We are using these nine semantic inferences for validation and classification of proven and potential DDIs and their mechanisms. Below, we give the clinical definition of each semantic inference, a real clinical example, and categorization for each semantic inference based upon its interaction type. After that, we give the fact source that is being used by the D3 pharmacovigilance system for inference and specify the symbolic description for each inference.

1. **Protein binding-based interaction**

Clinical definition and example: Protein binding increases the unbound drug concentration of one of a pair of interacting drugs**.** Therefore, this can lead to intensified pharmacological effects and possible toxicity, especially with drugs that have a narrow therapeutic window. For example, nitazoxanide when coadministered with warfarin results in the displacement of warfarin from its plasma protein binding sites, increasing unbound warfarin concentrations and, accordingly, increasing the chance of warfarin toxicity (Mullokandov E, 2014).

Interaction type: pharmacokinetic.

Fact source: DrugBank.

D3 inference definition: *DrugBank stores the protein binding values as a human readable text, so we develop a Java pattern recognition method to extract numbers from the text. Then we only consider a protein-binding interaction if both drugs have a high affinity (> 70% of the drug is bound) to bind to the same protein.*

2. **Metabolism induction-based interaction**

Clinical definition and example: Enzyme induction refers to the process by which a drug increases the activity of a drug-metabolizing enzyme, thereby increasing the rate at which the enzyme metabolizes drugs (either other drugs or the drug itself as in the case of auto-induction) in the body (Spina & de Leon, 2007). Induction of drug-metabolizing enzymes is a common mechanism for potential DDIs and is well described in the CYP family of enzymes. Enzyme induction may lead to

either toxicity (particularly through formation of a toxic metabolite) or therapeutic inefficacy. For example, the combination of rivaroxaban and rifampin results in reduced rivaroxaban concentrations and is associated with an increased risk of stroke due to a loss of rivaroxaban's therapeutic effect (Baciewicz, Chrisman, Finch, & Self, 2013).

Interaction type: pharmacokinetic.

Fact source: NCIt NDF-RT, and UniProtKB/Swiss-Prot.

D3 inference definition: *When one drug induces the metabolic processing of the other, D3 infers an interaction and the inducer drug that causes the interaction.*

3. **Metabolism inhibition-based interaction**

Clinical definition and example: In opposition to enzyme induction, enzyme inhibition refers to the process by which a drug reduces the function of a drug-metabolizing enzyme. Enzyme inhibition increases the plasma concentrations of drugs metabolized by the enzyme, which may lead to either toxicity or therapeutic inefficacy (as in the case of a prodrug). For example, the antimicrobial agent erythromycin inhibits the metabolism of warfarin when the two are co-administered, resulting in an increase in warfarin serum concentrations and a subsequent increased risk of bleeding (Rice, Perry, Afzal, & Stockley, 2003).

Interaction type: pharmacokinetic.

Fact source: NCIt NDF-RT, and UniProtKB/Swiss-Prot.

D3 inference definition: *When one drug prevents the metabolic processing of the other, D3 infers an interaction and the inhibitor drug that causes the interaction.*

4. **Transporter induction-based interaction**

Clinical definition and example: Generally, transporter interaction plays an important role in drug absorption and disposition. Induction of a transporter leads to increases in drug relocation within the body via the transporter, which, depending on the location of the transporter and whether the transporter promotes drug influx or efflux, may either increase or decrease the amount of drug absorbed into the body or available at the site of drug action. As a result, induction of drug transporters may result in serious toxicities or therapeutic inefficacy. For instance, carbamazepine induces the P-gp mediated efflux of paliperidone and thus reduces paliperidone's therapeutic effect when the two are co-administered (Yasui-Furukori, Kubo, Ishioka, Tsuchimine, & Inoue, 2013).

Interaction type: pharmacokinetic.

Fact source: DrugBank.

D3 inference definition: *When one drug induces transport of another, resulting in alterations in the elimination of the second drug, D3 infers the interaction and inducer drug that causes the interaction.*

5. **Transporter inhibition-based interaction**

Clinical definition and example: Transporter inhibition results in decreased transporter function and occurs more commonly than transporter induction. As with

induction, inhibition of drug transporters can increase or decrease drug exposure systemically or at the site of action depending on the location and function of the transporter being inhibited; the resulting changes in drug concentration can lead to ADEs or therapeutic inefficacy. For instance, quinidine significantly increases serum concentrations of digoxin when the two are co-administered by inhibiting P-gp-mediated digoxin efflux in the intestinal wall *(König, Müller, & Fromm, 2013)*.

Interaction type: pharmacokinetic.

Fact source: DrugBank.

D3 inference definition: *When one drug inhibits the transport of another, resulting in alterations in the elimination of the second drug, D3 infers the interaction and inhibitor drug that causes the interaction.*

6. **Multi-pathway-based interaction**

Clinical definition and example: A multi-pathway interaction refers to a type of complex DDI wherein the interaction occurs as a result of two or more interactive mechanisms. This type of interaction is difficult to accurately describe as the net effect of drug action across multiple simultaneous pathways of interaction must be characterized. For example, DDIs between some statins and cyclosporine occur through the metabolism (CYP3A4) and transport (P-gp) pathways (Holtzman et al., 2006). This leads to complicated interactions that are hard to resolve.

Interaction type: pharmacokinetic.

Fact source: NCIt NDF-RT, and UniProtKB/Swiss-Prot.

D3 inference definition: *When both drugs share at least one enzyme as well as one transporter, D3 infers the interaction and the pharmacological actions of both enzyme and transporter.*

7. **Competitive pharmacological effect-based interaction**

Clinical definition and example: This interaction occurs when two or more drugs bind competitively to the same molecular target, with the result that the binding of one drug reduces the affinity of another drug or both drugs for the target. While the significance of the interaction depends on the pharmacological effects of the competing medications, the relative affinities of the medications for the receptors, and the therapeutic index of the interacting drugs, this interaction type results in a reduction of drug action (which may potentially decrease therapeutic efficacy) and the potential for increased drug adverse effects. For example, albuterol-mediated bronchodilation, achieved through agonism at the beta-2 adrenergic receptors, is competitively disrupted when co-administered with the non-selective beta-adrenergic receptor antagonist, propranolol (Johnsson, Svedmyr, & Thiringer, 1975).

Interaction type: pharmacodynamic.

Fact source: Entrez Gene, Gene-GO, NCIt, and GO.

D3 inference definition: *When both drugs act on the same pharmacologic target, D3 infers the interaction and the affected target.*

8. **Additive pharmacodynamic effect-based interaction**

Clinical definition and example: When two drugs with similar pharmacodynamic effects are administered, summation of the effects of both drugs occurs, potentially resulting in excessive toxicity and pharmacodynamic response. For instance, the combination of glyburide and metformin could result in hypoglycemia since both agents reduce blood glucose levels through different mechanisms (Q. Wang, Cai, Van de Casteele, Pipeleers, & Ling, 2011).

Interaction type: pharmacodynamic.

Fact source: NCIt, NDF-RT and UniProtKB/Swiss-Prot.

D3 inference definition: *When both drugs share the same pharmacological effect, D3 infers the interaction.*

9. **Pharmacogenetic-based interaction**

Clinical definition and example: Pharmacogenetic-based interactions occur when two drugs that share a common metabolic pathway or a transporter are given to a patient with a polymorphism in the drug metabolizing gene or transporter known to alter drug exposure. For example, individuals with a genetic alteration (SNP) associated with the poor activity of a drug-metabolizing enzyme may experience drug interactions during co-administration of drugs that are co-substrates for the affected enzyme despite the fact that an appreciable DDI is not seen when the same two drugs are co-administered in individuals with wild-type enzyme function. Increases in the hypotensive effects of metoprolol (a CYP2D6 substrate) have been described upon co-administration with paroxetine, also a

substrate for CYP2D6, in patients with known alleles for reduced CYP2D6 function (Lynch & Price, 2007).

Interaction type: pharmacogenetic.

Fact source: Pharmgkb.

D3 inference definition: *When potentially interacting drugs may also interact on the genetic level in patients possessing known polymorphisms (SNPs) in drug-metabolizing enzymes or transporters, D3 infers the responsible SNPs in the interaction.*

These nine predefined semantic predictions for each of these levels that the D3 system utilizes for DDI study are summarized in **Table 5.1.**

Table 5.1: Nine different inferences are provided by the D3 system for potential and proven DDI discovery and explanation, where x is a drug (object), z is a drug (perpetrator) and y represents a mechanism of interaction, respectively.

Interaction inference	Clinical definition	Conditions for interaction	Fact sources	Clinical example
Protein binding-based interactions	When two drugs have high affinity (> 70% of the drug is bound) to bind to the same protein.	***x,z*** *bind_to* ***y*** ***x,z*** *> 70%*	DrugBank	nitazoxanide /warfarin
Metabolism induction-based interactions	When one drug induces the metabolic processing of the other.	***x*** *metabolized_by* ***y*** ***z*** *induces* ***y***	NCIt NDF-RT, and UniProtKB/Swiss-Prot	rivaroxaban / rifampin
Metabolism inhibition-based interactions	When one drug prevents the metabolic processing of the other.	***x*** *metabolized_by* ***y*** ***z*** *inhibits* ***y***	NCIt and NDF-RT, and UniProtKB/Swiss-Prot	warfarin/ erythromycin

Transporter induction-based interactions	When one drug induces the transport of another, altering its elimination.	***x*** *transported_by* ***y*** ***z*** *induces* ***y***	DrugBank	carbamazepine/ talinolol
Transporter inhibition-based interactions	When inhibits transport of another, altering its elimination.	***x*** *transported_by* ***y*** ***z*** *inhibits* ***y***	DrugBank	quinidine/ digoxin
Multi-pathway-based interactions	When both drugs share at least one enzyme and one transporter.	***x*** *metabolized_by* ***y*** ***x*** *transported_by* ***y2*** ***z*** *metabolized_by* ***y*** ***z*** *transported_by* ***y2***	NCIt NDF-RT, and UniProtKB/Swiss-Prot	cyclosporine/ statins
Competitive pharmacological effect-based interaction	When both drugs act at same pharmacologic target.	***x*** *targets* ***y*** ***z*** *targets* ***y***	Entrez Gene, Gene-GO, and NCIt, GO	propranolol/ albuterol
Additive pharmacodynamic effect-based interaction	When both drugs share mechanism of action.	***x*** *share_moa_with* ***y*** ***z*** *share_moa_with* ***y***	NCIt, NDF-RT and UniProtKB/Swiss-Prot	glyburide/ metformin
Pharmacogenetic-based interaction	When two drugs that share a common metabolic pathway or a transporter are given to a patient with a polymorphism in the drug metabolizing gene or transporter known to alter drug exposure	***x*** *associated with _SNPs* ***y*** ***z*** *associated with _SNPs* ***y***	Pharma GKB	metoprolol/ paroxetine

V.2 Phase 2: Construction of the D3 System

The aim of this phase is to address the question of how an effective pharmacovigilance system can be developed to provide validation and classification

of proven and potential DDIs using existing knowledge sources. This system was motivated by the following requirements:

- The D3 should be able to not only predict an interaction, but also should provide the likely mechanisms causing the interaction.
- The D3 should be able to support the identification of proven DDIs by amalgamating 15 different DDI knowledge sources.
- The D3 should be able to support DDI queries using different drug names (generic, brand and chemical).
- Many pharmacovigilance systems are limited to specific types of interaction. Our system should be comprehensive enough to consider diverse interaction types.
- Many pharmacovigilance systems do not take into account rich mechanistic information with regards to DDI. Our system should be capable of discovering and exploiting hidden mechanistic information from multiple diverse text formats, data schemas, and controlled vocabularies, and it should interlink these with diverse datasets extracted from the biomedical domain to improve the quality of pharmacovigilance recommendations.
- Not all pharmacovigilance systems take into account genetic variation (SNPs) when studying DDIs. Our system should be capable of providing an alert when interacting drugs may also interact on the genetic level in

patients possessing known polymorphisms (SNPs) in drug-metabolizing enzymes or transporters.

- Many pharmacovigilance systems do not take into account all of the details present within drug information from existing sources for proven DDIs. Our system should be capable of discovering similarities of interaction between proven DDIs and then exploiting them to provide the likelihood of a potential DDI.
- No existing pharmacovigilance systems assess the clinical relevance of their predicted DDIs. Our system should be able to consider the clinical relevance of potential DDIs and report only those DDIs likely to impact clinical decision-making.
- The D3 should be able to provide only the likely mechanisms of interaction.

To address these challenges and requirements, in this section we integrate our proposed techniques from **Chapters II-IV**, as well as those presented in the first section of **Chapter V**, to model potential DDIs and their mechanisms in order to provide a comprehensive pharmacovigilance system.

In this section, the main goal is to show that D3 is capable of identifying proven DDIs from different sources along with their mechanisms of interaction and to demonstrate D3's ability to infer potential DDIs and their mechanisms accurately to provide a valuable decision tool for informing clinical practice. To achieve such a

goal, two types of inferential models are being developed and added to the D3 system: (1) query-based and (2) probabilistic-based. With regards to the former type, it should be noted that the system determines only whether there is a reported interaction or not, along with the common pathways it finds to account for the interactions. If the D3 query-based model does not find a reported interaction, it will infer potential DDIs based on nine different inferences that were described in the **V.1 Phase 1: Declaration of D3 Semantic Inferences** section and that could cause the interaction. The query-based model does not rank the mechanisms of DDIs according to importance when there is more than one, and it does not provide the clinical relevance of predicted interactions. The probabilistic-based model, on the other hand, is designed as an improvement, which will provide the likely mechanisms for proven DDIs to help avoid the need for manual filtering of results to recognize the strongest cause of the interactions. It will also use a novel and complex biomedical similarity-based method to identify the clinical relevance of potential DDIs.

V.2.a Phase 2: D3 Inferential Query Model

The main aim of this section is to provide evidence for the usefulness of the Semantic Web approaches to identify not only potential DDIs, but also to explain the cause of the interactions. As a result, the semantic inferential query-based model we propose, in conjunction with a probabilistic-based model to be described later, will help in overcoming the current challenges in pharmacovigilance systems.

Specifically, these proposed models are designed to move beyond a single-level of drug interaction to focus on more complex interactions and to identify all possible mechanisms that contribute to the interaction between two drugs. The D3 inferential query-based model, as the name suggests, is based on the query language of the Semantic Web, SPARQL. SPARQL is the query language for RDF graphs that produces conclusions only if requisite conditions are satisfied. Further, the conditions are based on the D3 knowledge model; hence SPARQL supports indirect inference by retrieving information that is not explicitly stated in the knowledge base. SPARQL query syntax can be understood as finding a path in the RDF graphs by utilizing a set of semantic relationships; **Figure 5.1** shows a D3 SPARQL query to identify an inhibition of metabolism-based interaction that can be read as follows: *Find a drug that is a substrate of an enzyme and another drug that is an inhibitor of this enzyme.*


```
SELECT DISTINCT *
WHERE
{

?drug1 substrate_of  ?enzyme .
?drug2 inhibits ? enzyme .

?enzyme label ?enzymeName .
?enzyme rdf:type Enzyme.
}
```

Figure 5.1: Metabolism inhibition inference

At the macroscopic level, the D3 inferential query-based model works by querying the D3 knowledge base in search of an interaction between two drugs. The

model then returns a determination of whether the two drugs are reported for potential interaction. If a DDI is reported, the model then runs all semantic inferences to provide the common pathways it finds that could account for the interaction. If the DDI has a potential significance that has not been reported for an interaction in the integrated DDI sources, it will infer a potential DDI based on all semantic predictions that could cause the interaction. **Figure 5.2** shows a flowchart of the D3 inferential query-based model.

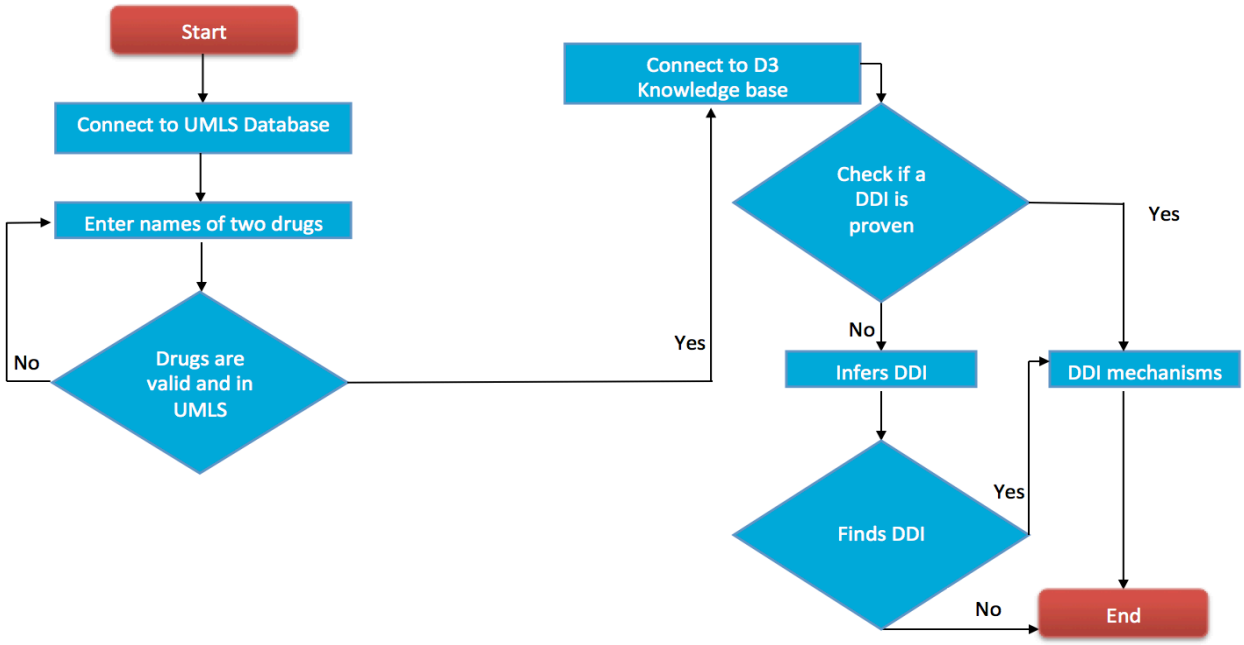


Figure 5.2: D3 inferential query-based model flowchart

Practically, the D3 inferential query-based model contains the following important functions:

1. **Normalized Naming of Drugs Using RxNorm:** A drug by nature has several names: brand, generic, chemical, etc. Aspirin, for example, has 35 different drug names ("Bayer (Aspirin) Patient Information: Side Effects and Drug Images at RxList," n.d.). In our system, we normalize all drug names using RxNorm, which is a tool provided by the National Library of Medicine to assist with standardization of drug names. RxNorm currently holds drug names and synonyms from 15 drug information sources ("RxNorm Overview," n.d.). This normalizing step is necessary because users might have different preferences when searching for an interaction. **Algorithm 1** shows the D3 normalized naming using the RxNorm function that takes users' input and converts it to the RxNorm identifier.

Algorithm 1: A method to return a RxNorm ID from user input (drug)

```
Input  : input drug.
Output: RxNorm Id.
begin
 RxNormId = (inputDrug, umlsServer);
if RxNormId is not empty then
    return RxNormId;
else
    break;
```

2. **Proven DDI Identification:** once the drug name is standardized using the RxNorm function, the model will start investigating the requested DDI. The model checks the enormous volume of information available

within DDI sources to identify the interaction. The novelty of the D3 inferential query-based model is its ability to identify reported interactions by checking DDI information collected and integrated from the 15 DDI resources that are stored in the D3 knowledge base. **Table 5.2** shows the DDI resources considered for use in the proven DDI identification.

Table 5.2: Fifteen DDI resources are used by the D3 query-based model

DDI source	DDI Numbers	Definition	Source type
Crediblemeds.org(CredibleMeds, n.d.)	82	List of important DDIs.	Clinical
PK-Corpus(R. Boyce, Gardner, & Harkema, 2012)	165	DDI resource derived from all drug product labels.	Texts
DrugBank(Wishart et al., 2008)	11840	List of DDIs along with their explanations.	Bioinformatics
The Drug Interaction Knowledge Base(R. Boyce et al., 2009a)	560	List of pharmacokinetic DDIs along with their evidence.	Bioinformatics
NDF-RT("National Drug File – Reference Terminology (NDF-RT™) Documentation," n.d.)	9883	DDI resource derived by Veteran's Administration health care system.	Clinical
ClinicalTrials.gov("Home - ClinicalTrials.gov," n.d.)	3645	List of sever DDIs that result of ADEs.	Clinical
KEGG(Takarabe et al., 2011)	26328	DDI resource derived from Japanese product labels.	Bioinformatics
DDI-Corpus-2011(Segura Bedmar, Martínez, & Sánchez Cisneros, 2011)	569	DDI resource derived from biomedical texts.	Texts
DDI-Corpus-2013(Segura Bedmar et al., 2011)	1282	Updated list from DDI-Corpus-2011.	Texts
National Library of Medicine cardiovascular Corpus(Stan, 2014)	246	DDI resource derived from cardiovascular drug product labels.	Biomedical texts
ONC-High Priority(Phansalkar et	1150	List of important DDIs by the Office of the National Coordinator for	Clinical

al., 2012)		Health Information Technology.	
ONC-Non-interruptive(Phansalkar et al., 2013)	2079	List of non-interruptive DDIs by the Office of the National Coordinator for Health Information Technology.	Clinical
OSCAR EMR(OSCAR Electronic Medical Record, n.d.)	7753	DDI resource derived from EHRs.	Clinical
SemMedDB from PubMed(Kilicoglu, Shin, Fiszman, Rosemblat, & Rindflesch, 2012)	3536	DDI resource derived from PubMed abstracts.	Bioinformatics
Twosides(Tatonetti, Ye, Daneshjou, & Altman, 2012b)	63333	DDI resource derived from the analysis of spontaneously reported adverse events	Bioinformatics

Using a complete list of available DDI resources is essential due to the high level of diversity in reported interactions among DDI information sources. Once an interaction is found, the D3 inferential query-based model will run all semantic inferences to provide the common pathways it finds between two drugs to account for the discovered interaction. **Algorithm 2** shows the D3 proven DDI identification function that takes users' input and returns a report of DDI information from knowledge sources along with the possible mechanisms of the interaction

Algorithm 2: A method to return sources of DDI with its mechanisms

```
Input  : DDI.
Output: DDI sources + mechanisms of interaction.
begin
 isProvenDDI = (drug1,drug2, D3knowledgeBase);
if isProvenDDI is true then
    return
    (1) DDI sources report the interaction;
    (2) DDI mechanisms of the interaction;
else
    false;
```

3. **Potential DDI Inference:** This function will perform and play a role only if the proven DDI identification method returns null. In other words, this function is only run when the requested interaction is not reported in any of the 15 DDI sources. The determination of the interaction using the potential DDI inference method then depends on the occurrence of at least one of the D3 semantic inferences to count for a potential DDI. **Algorithm 3** shows the D3 potential DDI inference function that takes users' input and determines if a DDI is potential or not based upon the D3 semantic inferences.

Algorithm 3: A method to return mechanisms of potential DDI

```
Input  : DDI.
Output: mechanisms of potential interaction.
begin
 isProvenDDI = (drug1,drug2, D3knowledgeBase);
if isProvenDDI is ture then
    (1) DDI sources report the interaction;
    (2) DDI mechanisms of the interaction;
else
    comment: This interaction is potential;
    semanticPredictions.size();
    for (i:=0  to semanticPredictions.size() i++);
     inferDDI = (drug1,drug2, D3knowledgeBase, semanticPredictions[i]);
    if inferDDI is null then
        This DDI may be a negative;
    else
     └ return inferDDI
```

Finally, **Algorithm 4** shows how the D3 inferential query-based model works.

Algorithm 4: D3 inferential query-based model

```
Input  : DDI.
Output: identification, validation, and classification of DDI
begin
reasonsOfDDI = (ProteinBinding, MetabolismInduction,
MetabolismInhibition TransporterInduction, TransporterInhibition,
MultiPathways, Synergistic, Competition, Additive, Pharmacogenetic) ;
sourcesOfDDI = (Crediblemeds,PK-Corpus,DrugBank DIKB, NDF-RT,
ClinicalTrials.gov,ONC-High Priority,
ONC-Non-interruptive,OSCAR,SemMedDB,Twosides KEGG,
DDI-Corpus-2011, DDI-Corpus-2013, NLM-Corpus) ;
 RxNormId = (DDI, umlsServer);
if RxNormId is not empty then
    return RxNormId;
else
    break;
 isInD3knowledgeBase = (DDI, D3knowledgeBase);
if D3knowledgeBase is not empty then
     isProvenDDI = (drug1,drug2, D3knowledgeBase);
    if isProvenDDI is ture then
        comment: This interaction is reported;
         sourcesOfDDI = (drug1,drug2, D3knowledgeBase, DDIsources);
         reasonsOfDDI = (drug1,drug2, D3knowledgeBase,
        semanticPredictions);
        return sourcesOfDDI + reasonsOfDDI
    else
        comment: This interaction is potential;
         inferDDI = (drug1,drug2, D3knowledgeBase,
        semanticPredictions);
        if inferDDI is null then
            This DDI may be a negative;
        else
            return inferDDI
else
    one or two drugs are not stored in the D3 knowledgeBase;
    exist the system;
```

At the end of this phase, two types of DDI studies were proposed. These types are: reported and potential studies. In each type, the D3 inferential query-based model can provide complex inferences to identify reported and potential interactions along with their possible mechanisms of interactions. An example of each semantic inference is given in **Appendix B**.

V.2.b Phase 2: D3 Inferential Probabilistic Model

In the previous section we introduced the inferential query-based model that can be seen as a proof-of-concept for identifying and explaining complex DDIs and hidden mechanisms of interaction. By integrating extensive DDI information from diverse sources into a coherent knowledge base and using SPARQL as an inference technique, the query model was able to provide useful inferences to identify and explain reported and potential DDIs. This typical way of inferencing has been widely used in many studies within the biomedical field such as (Ruttenberg et al., 2009; Sahoo et al., 2007). However, this inference technique has one notable limitation: researchers are required to filter the query results. This limitation affects our D3 inferential query-based model in two important ways. First, it will not allow the model to automate the ranking of the mechanisms of interactions according to importance when there is more than one. Thus, the possible mechanisms of interactions would need to be manually filtered to recognize the strongest cause of the interactions. Second, it also will not provide the probability of the potential interaction. As a result, it is almost impossible to assess the clinical

relevance of discovered potential interaction using solely the D3 inferential query-based model.

The main aim of this section is to overcome these two limitations by introducing the second part of the D3 pharmacovigilance system, the inferential probabilistic model. The inferential probabilistic model of D3 is proposed as a perfection of the inferential query-based model described earlier. The probabilistic model, unlike the query-based one, eventually should return the probability of the potential interactions, assist in determining the clinical relevance of the interactions, and place the possible mechanisms of interaction in order according to importance when there is more than one. Thus, it will provide a means of identifying novel and clinically relevant potential DDIs to minimize the risk of ADEs.

In this D3 inferential probabilistic model, the hidden information and similarity between DDIs, their interaction lists, and their mechanisms of interaction are discovered using a set of complex probabilistic algorithms. The discovered knowledge is then exploited to infer and gain a better understanding of DDI mechanisms of interaction, and this in turn results in providing better pharmacovigilance services. In order to provide such services, we require answers to these challenging questions:

- Is it possible to automate prediction of potential DDIs to some degree?
- Is there a way to reliably assess the clinical relevance of discovered potential DDIs?

- How can we probabilistically identify and provide the most likely mechanisms of interaction according to importance?

Based on these questions, we divide this phase into two important development stages: (1) **the learning process** and (2) **statistical inferences**. The learning stage includes three main steps: (1) filtering existing and available DDI resources to create relevant DDI lists; (2) constructing a drug's biomedical similarity matrix, which allows the comparison of similarities between two drugs in terms of their pharmacokinetic, pharmacodynamic, and pharmacogenetic properties; and (3) extracting all possible mechanisms of interaction for DDIs. The statistical inferences stage is then designed to utilize the learning stage functionalities to perform calculations that provide the probability of potential DDIs, to assess the clinical relevance of predicted DDIs, and to order the mechanisms of interactions according to importance. In the next section, we shall discuss the learning process of the D3 inferential probabilistic model.

Stage 1: Learning Process

In this stage, we propose developing a pharmacovigilance profile from the existing DDI resources and biomedical sources that are scattered throughout the biomedical domain. This process includes building lists of important DDIs, constructing the biomedical features similarity matrix, and identifying mechanisms of interaction. We consider several existing DDI resources for building the list of

drug interactions, as they have rich information that can be exploited to identify the clinical relevance of discovered potential interactions. However, a drug might have different lists of interactions in different resources. For instance, while National Drug File - Reference Terminology (NDF-RT) listed 7 drugs that interact with irinotecan, Drugs.com reported 329 DDIs involving irinotecan. Such information cannot be used unless we have a sufficient DDI list that records all the interactions of irinotecan in each resource. If this information is available, then we need a way to learn how important a DDI is in a particular clinical assessment (e.g., how important is it for a physician to watch for a specific DDI that might cause ADEs?). According to Horn, important DDI identification requires both the object and perpetrator drugs to be characterized along with the mechanism(s) of their interaction (Horn, Hansten, & Chan, 2007). Following this hypothesis, he implemented the Drug Interaction Probability Scale (DIPS), a tool that includes thoughtful questions to evaluate the probability of the clinical relevance of discovered potential interactions. In our learning process, we have taken into consideration the DIPS recommendations for discovering novel and potential DDIs. In particular, we consider (1) examining both drugs in order to identify the interaction and (2) investigating their mechanisms of interactions. First addressing the former, we primarily pre-filter DDI lists (i.e., those stored in the existing, publically available DDI resources) and consider just the clinically relevant drug interactions. Then, we study all drugs' profiles utilizing 9 different biomedical features in order to assure the identification of clinically relevant DDIs. For the

latter, we propose to examine 9 different well-known mechanisms of interaction to identify the most likely cause of the interaction. The D3 learning process consists of three main steps: filtering DDI resources to build a list of clinically relevant interactions, constructing the drug's biomedical features similarity matrix, and gathering all possible pathways of interaction. Before presenting each step, here we define some concepts that are essential in this learning process:

- ***Clinical relevance likelihood:*** *This is the likelihood that a clinical interaction will occur. I.e a DDI that impacts the clinical judgments.*
- ***Clinically relevant DDIs:*** *These represent DDI pairs that are reported in either a clinically oriented resource or extracted from biomedical literature. This type of DDI is essential for two main reasons. The first is that it would be associated only with clinically relevance interactions. The second reason is that it could be used to discriminate between important and less significant DDIs. That is, the D3 pharmacovigilance system should provide relevant DDIs for the current drug and exclude other DDIs that are clinically irrelevant to the current drug interaction.*
- ***DDI filtering:*** *A drug and its list of interactions in the D3 inferential probabilistic model are filtered based on their clinical impact. That is, all irrelevant DDIs would not be considered in the learning process.*
- ***DDI biomedical features similarity profile:*** *This profile contains all the modeled DDI information associated with relevant biomedical features. Each biomedical feature models and contains information that is relevant in just*

one specific biomedical entity. As a result, a DDI might have more than one biomedical feature based on the number of drug dimensions that are taken into account in the D3 pharmacovigilance system. In this, D3 is unlike other systems that list each biomedical feature in a separate profile.

Stage 1 Step 1: Filtering Existing DDI Resources

When providing pharmacovigilance services, only the most essential reported DDIs that are relevant to clinical practice should be selected. Therefore, in this step, we first need to evaluate the reputable, available DDI resources that hold different lists of DDIs to consider the most comprehensive and clinically relevant DDIs. This can be done using two different modeling techniques. One possible modeling technique is to treat each source separately. However, such a mechanism has two limitations: (1) a DDI list in one resource might overlap with other resources; thus, this can lead to extensive duplication; and (2) other important DDIs may be overlooked when a single resource does not report them. Instead, in this step we suggest an integrated learning technique to quantify and model those DDIs that are clinically relevant. For instance, if a particular DDI is found in two DDI resources, then we add one to the D3 probabilistic-model list. The advantage of this technique is that there is no need to be concerned about the potential overlap within DDI sources, which can lead to inaccurate prediction. In building the integrated and clinically relevant DDI list, we consider 10 different publically available DDI resources. Specifically, in the **V.2.a Phase 2: D3 Inferential Query Model**

section, we showed how the query-model can effectively search for an interaction exploiting 15 different DDI resources (6 clinical resources, 4 natural language corpora and 5 pharmacovigilance resources). However, we mentioned that such a model is unable to assist in knowing the clinical relevance of this interaction. This is because of two important reasons. The first is that the pharmacovigilance resources, unlike the clinical ones, though they have some measure of success, are not derived from clinical studies. Rather, they result from an inference process and often are not clinically relevant. For example, Twosides (a pharmacovigilance resource derived from the analysis of spontaneously reported adverse events) is an example of a DDI resource excluded through our learning process because as stated above their DDIs were not clinically relevant. Another limitation of the pharmacovigilance resources is that they list only DDIs that are already reported in clinical resources. Thus, considering them will add more duplication to our D3 lists. Therefore, the step of filtering existing DDI resources is essential, as we aim only to provide the most clinically relevant DDIs. Therefore, from the 15 resources used in the D3 query-based model, we consider 10 DDI resources, six of which are clinically oriented and four of which are derived from analyzing biomedical literature (*in vitro -in vivo* studies, pharmacology case reports, lab reports, medical literature, and drug product labels). As illustrated in **Table 5.3**, the D3 inferential probabilistic model comprises the most comprehensive collection of DDI information from publicly available resources to be considered for use in the learning process by which D3 will learn to identify potential clinically relevant DDIs.

Table 5.3: Ten DDI resources are used by D3 probabilistic -based model

DDI resources	DDI Numbers	Definition	Source type
Crediblemeds.org	82	List of important DDIs	Clinical
PK-Corpus	165	DDI resource derived from all drug product labels	Biomedical Texts
NDF-RT	9883	DDI resource derived by Veteran's Administration health care system	Clinical
ClinicalTrials.gov	3645	List of severe DDIs that result in ADEs	Clinical
DDI-Corpus-2011	569	DDI resource derived from biomedical texts	Biomedical Texts
DDI-Corpus-2013	1282	Updated list from DDI-Corpus-2011	Biomedical Texts
National Library of Medicine cardiovascular Corpus	246	DDI resource derived from cardiovascular drug product labels	Biomedical texts
ONC-High Priority	1150	List of important DDIs by the Office of the National Coordinator for Health Information Technology	Clinical
ONC-Non-interruptive	2079	List of non-interruptive DDIs by the Office of the National Coordinator for Health Information Technology	Clinical
OSCAR EMR	7753	DDI resource derived from EHRs	Clinical

Once all the DDIs are extracted from our base of DDI sources, a list of clinically relevant DDIs can be generated. **Algorithm 5** shows the process of integrating all resources and removing the overlap between them to generate a unique clinically relevant DDI list.

Algorithm 5: Create clinical relevance DDIs list

Input : 10 reported DDI sources

Output: an integrated and non-duplicated DDIs list

begin

if *DDI source is not empty* **then**

```
List CredibleMeds = getDDIlistFromSource("CredibleMeds");
List ClinicalTrials.gov = getDDIlistFromSource("ClinicalTrials.gov");
List PK-Corpus = getDDIlistFromSource("PK-Corpus");
List NDF-RT = getDDIlistFromSource("NDF-RT");
List DDI-Corpus-2011 = getDDIlistFromSource("DDI-Corpus-2011");
List DDI-Corpus-2013 = getDDIlistFromSource("DDI-Corpus-2013");
List NLM-CV = getDDIlistFromSource("NLM-CV ");
List ONC-Priority = getDDIlistFromSource("ONC-High-Priority");
List ONC-interruptive = getDDIlistFromSource("ONC-interruptive");
List OSCAR = getDDIlistFromSource("OSCAR");
D3-inferential-probabilistic-model-DDIList= union AllLists;
return (D3-inferential-probabilistic-model-DDIList)
```

else

this source has no DDI;

After integrating and removing overlapping processes, we generate a list of 21,897 clinically important and non-duplicated DDIs. This list, as will be discussed in detail in the inference stage, will be used for determining the clinical relevance of predicted interactions as well as for ordering the mechanisms of interaction.

Stage 1 Step 2: Building a Biomedical Features Similarity Matrix for Drug Comparison

Once all clinically relevant DDIs from D3's DDI sources are filtered and stored in a list, the list will be used in the next step of learning and modeling DDIs and their mechanism profiles. For example, we can compare a DDI by using a list of

interactions: by creating an interaction profile for each drug within the DDI, D3 can mine the knowledge base to find similarities between two interactions' profiles. In this step, we aim to compute the similarities of each drug in terms of multiple biomedical features. Indeed, the similarity-based method has already been widely used in the field of DDI study. This includes, but is not limited to, studying similarities between two drugs in terms of chemical and molecular structure, ADE profile, drug indication class, and interactions profile. For example, studies showed that similarity between the molecular structures of drugs from AERS was useful in predicting drugs that might potentially interact to produce adverse events such as rhabdomyolysis and inflammation of the pancreas (Vilar et al., 2011; Vizenor et al., 2006). Similarly, a recent contribution has combined 6 different similarity-models to enhance the detection of DDIs that can cause arrhythmia (Vilar et al., 2015). These existing studies provide the motivation for our work. However, the content of drug information in knowledge sources is vastly different in the D3 system than in previous research.

Fundamentally, we are applying the concept of the Jaccard metric (**Equation 5.1**) to establish our Jaccard similarity equation for computing the pair-wise biomedical similarity of interacting drugs.

$$J(A,B) = \frac{A \cap B}{A \cup B}$$

Equation 5.1: Jaccard similarity equation

Indeed, there are a larger number of similarity measures that can be applied to replace Jaccard. Those measures differ in three important properties: (1) counting for negative matches (lack of similarity across a pair resulting in stronger likelihood of interaction), (2) correcting matches (alternative similarity of mined information), and (3) classifying similarities. Choi et al (Choi & Cha, 2010), have classified 76 similarity measures based on their relationships and convergences that could be considered. These could be evaluated based on regression of various samples, constituting DDI classes, to determine the measures' effectiveness and applicability. At the current stage of the work, D3 uses the Jaccard similarity measure to identify clinically relevant DDIs based on computing the similarity between drugs within their biomedical features.

As drugs may have more than one common feature (e.g., side effects, molecular target, etc.), a drug profile in D3 is modeled for each entity independently. Formally, let F = {f1, f2,...,fn} be a set of all the biomedical features, let D = {d1,d2,...,dn} be a set of all drugs mapped to F (i.e., based on the D3 knowledge base), and let FX = {fx1, fx2, ..., fxn} be a drug dimension that contains different biomedical features fx, where $dn \times fx \rightarrow [0, 1]$. Next, we propose an averaged Jaccard similarity metric to compute the pair-wise DDI biomedical similarity, which consists of two sub-phases:

Step 2 Sub-Phase 1: In this sub-phase, the goal is to define the biomedical features that we are considering for creating a drug's biomedical features similarity matrix. As the goal of the D3 inferential probabilistic model is to identify the

clinical relevance of predicted DDIs, we apply nine different biomedical features, unlike the existing research, which typically considers only one or two features, to evaluate similarities between two drugs. These nine biomedical features are: (1) drug targets, (2) drug indications, (3) drug enzymes, (4) drug transporters, (5) drug side-effects, (6) drug carriers, (7) drug mechanisms of action, (8) genetic variations, and (9) physiological effects. **Table 5.4** summarizes all biomedical features along with their sources of information and shows the nine biomedical features that are used to create the D3 similarity matrix.

Table 5.4: Nine biomedical features are used to create similarity matrix.

Biomedical Features	**Fact sources**
Drug targets	DrugBank
Drug indications	NDF-RT
Drug enzymes	DrugBank, NCIt NDF-RT, and UniProtKB/Swiss-Prot
Drug transporters	DrugBank
Drug side effects	SIDER
Drug carriers	DrugBank
Drug mechanisms of action	NCIt and NDF-RT
Genetic variations	PharmaGKB
Physiological effects	NDF-RT

Step 2 Sub-Phase 2: Here we describe our rules for computing the biomedical similarity between two drugs; the rules are:

1. If both drugs only have exact elements (match) within a biomedical feature, they are similar.
2. If both drugs lack the same feature, they are similar.
3. All biomedical features are equally weighted.
4. The similarity score is a result of summing each feature's individual Jaccard score divided by the total number of biomedical features.

Figure 5.3 shows the process of computing the similarity of the nine-biomedical features as well as the equation of the final similarity score.

$$Enzyme - Jaccard(drug1, drug2) = \frac{drug1Enzyme \cap drug2Enzyme}{drug1Enzyme \cup drug2Enzyme}$$

$$Target - Jaccard(drug1, drug2) = \frac{drug1Target \cap drug2Target}{drug1Target \cup drug2Target}$$

$$Indications - Jaccard(drug1, drug2) = \frac{drug1Indications \cap drug2Indications}{drug1Indications \cup drug2Indications}$$

$$Transporters - Jaccard(drug1, drug2) = \frac{drug1Transporters \cap drug2Transporters}{drug1Transporters \cup drug2Transporters}$$

$$SideEffects - Jaccard(drug1, drug2) = \frac{drug1sideEffects \cap drug2sideEffects}{drug1sideEffects \cup drug2sideEffects}$$

$$Carriers - Jaccard(drug1, drug2) = \frac{drug1Carriers \cap drug2Carriers}{drug1Carriers \cup drug1Carriers}$$

$$SNPs - Jaccard(drug1, drug2) = \frac{drug1SNPs \cap drug2SNPs}{drug1SNPs \cup drug2SNPs}$$

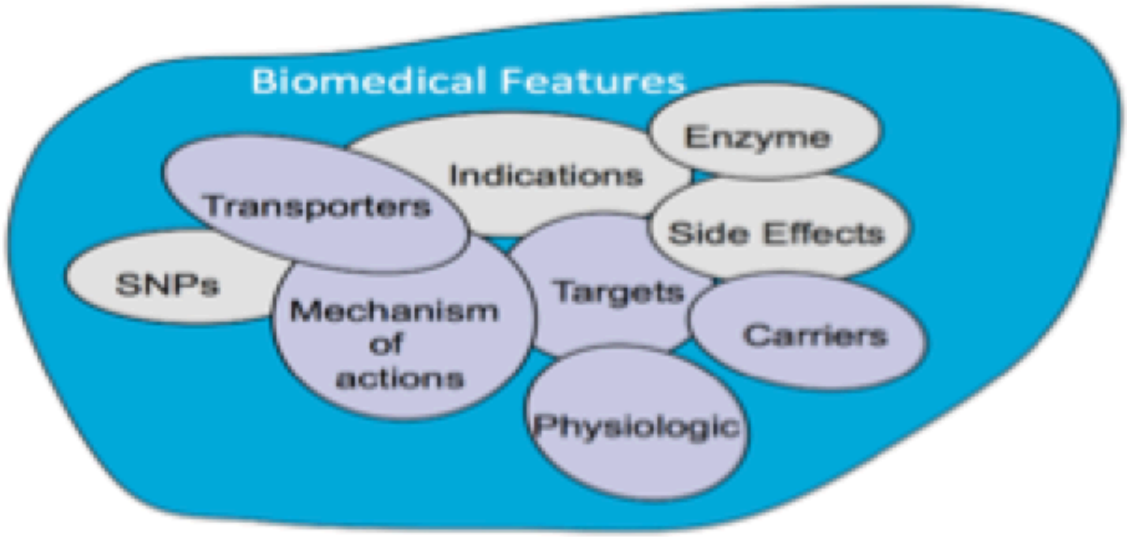
$$MoA - Jaccard(drug1, drug2) = \frac{drug1MoA \cap drug2MoA}{drug1MoA \cup drug2MoA}$$

$$Physiologic - Jaccard(drug1, drug2) = \frac{drug1Physiologic \cap drug2Physiologic}{drug1Physiologic \cup drug2Physiologic}$$

$$Allfeatures Jaccard(drug1, drug2) = \frac{All biomedical features Jaccard scores}{Total Number 0f Biomedical features}$$

Figure 5.3: D3 similarity score calculation

The final similarity score will be used to generate the biomedical similarities matrix for ranking the candidate DDIs. **Figure 5.4** shows the flowchart of building the drug's biomedical features similarity matrix.



$$biomedicalSimilarityCalculation(drug1, drug2) = \frac{drug1 \cap drug2}{drug1 \cup drug2}$$

DDI	Similarity								
Drug1, drug2	0	0.98	0	0.5	1	0	0	0	1

$$AverageJaccard(drug1, drug2) = 0.39$$

Figure 5.4: Overall process of building the D3 matrix

Stage 1 Step 3: Extracting All Possible Mechanisms of Interaction

After creating a list of clinically relevant DDIs and constructing the D3 biomedical features similarity matrix, we move to the last step in our learning process: retrieving all potential mechanisms of interaction.

The aim of this step is to capture all possible mechanisms of interaction that can be discovered from the D3 knowledge base. Understanding such DDI mechanisms can allow a pharmacovigilance system to uncover hidden pathways of interaction. In order to capture and discover DDI mechanisms of interaction in this step, we utilize the nine proposed mechanisms of DDI at the pharmacokinetic, pharmacodynamic, pharmacogenetic, and multi-pathway levels described in the **V.1 Phase 1: Declaration of D3 Semantic Inferences** section. We then design a function that takes a potential DDI and returns any findings from the D3 semantic inferences that account for its potential mechanisms of interaction. **Algorithm 6** shows the extraction of potential mechanisms of interaction for a given potential DDI.

Algorithm 6: Extracting All possible mechanisms of interaction

Input : DDI

Output: All possible mechanisms of interaction

begin

D3 mechanisms of interaction = (Protein binding, Metabolism induction, Metabolism inhibition, Transporter induction, Transporter inhibition, Multi-pathways, Competitive, Additive, Pharmacogenetic)

hasMechanisms = (DDI, D3 knowledgebase)

if *hasMechanisms is not empty* **then**
  return all possible mechanisms of interaction;
else
  no mechanisms of interaction found

At the end of this stage, the three steps of the learning process are initialized. To review, these steps are: creation of a clinically relevant DDI list, construction of a biomedical features similarity matrix, and extraction of potential mechanisms of interaction. In each step, a set of results will be available for inference purposes. In the next stage, we shall show how these results will be exploited in order to infer a likelihood score for a potential interaction, assist in determining its clinical relevance, and suggest the most likely mechanisms of interaction.

Stage 2: Statistical Inferences

One important question might be raised regarding how to use existing biomedical knowledge about a drug, its interactions list, and its mechanistic information to discover novel DDIs that are both highly probable and clinically relevant. Because drug information, while of essential use in creating a DDI characterization tool, is both heterogeneous in its presentation and ever-evolving, we argue that an effective pharmacovigilance system needs a means of coherently organizing current drug information scattered across different sources in order to provide highly relevant and likely DDIs with their mechanisms. The main target of this stage is to answer the following questions:

- When selecting a proven and clinically relevant DDI (based on clinical literature) from our created list of clinically relevant DDIs, how similar is this proven DDI to the potential one being investigated?

• Is there a way to provide detailed, mechanistic explanations for potential and proven DDIs?

The main assumption is that the proven DDIs can provide us with a direct bridge to understand accurate criteria for proposing potential DDIs based on similarities within biomedical features. **Figure 5.5** illustrates the general hypothesis of the probabilistic inferences.

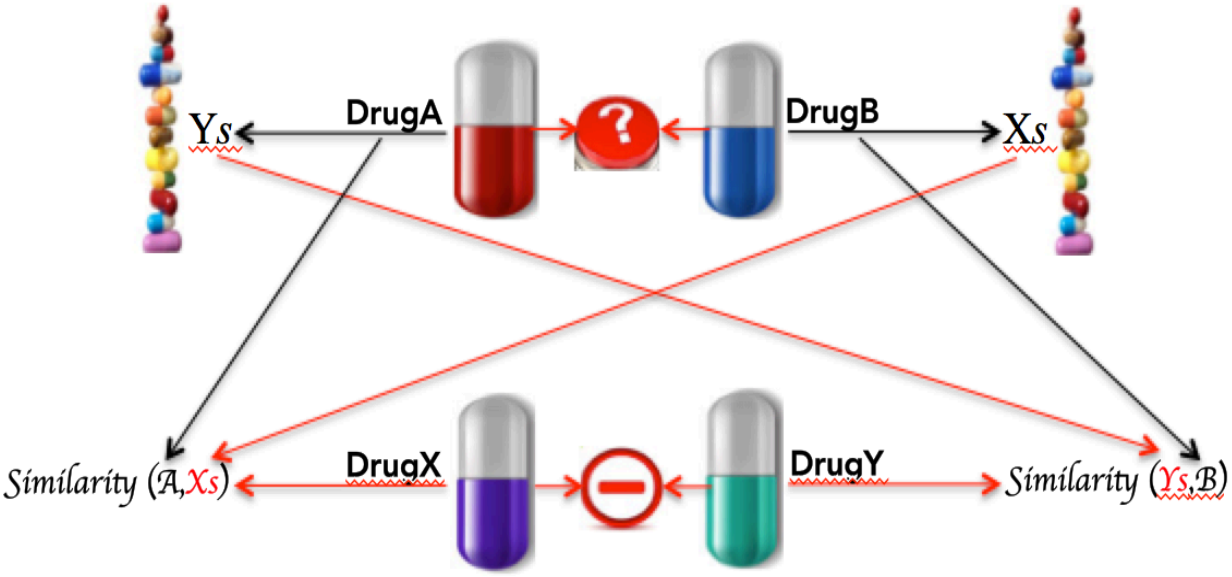


Figure 5.5: D3 probabilistic inferences hypothesis

More specifically, in this stage, we aim to carefully address three questions:

- How likely is a DDI to occur?
- How clinically relevant is a proposed potential DDI?
- Why, mechanistically, does a DDI occur?

Our strategies for answering these questions are based on the learning process stage. Specifically, we base the inferences on three principles: (1) the construction of interaction profiles for a DDI (i.e., an interaction profile for each drug in the DDI) by exploiting the filtered DDI list, (2) the computation of a biomedical similarity score for the DDI, and (3) the extraction of mutual mechanisms between the interacting drugs. The main inference strategy in this stage is based on our novel probabilistic inferences that were introduced in the section **V.2.a Phase 2: D3 Inferential Probabilistic Model**. Moreover, this statistical inferences stage has three main methods: (1) identifying the likelihood of DDIs, (2) inferring the clinical relevance of potential DDIs, and (3) sorting the mechanisms of DDIs according to importance. The D3 statistical inferences model is distinctive as it uses six different steps to probabilistically infer and recommend novel and potential DDIs along with their mechanisms of interaction. The three steps outlined above are considered when determining the likelihood and the clinical impact of a potential DDI, while another three steps are employed when determining and ordering the DDI mechanisms of interactions: (1) identifying the mutual mechanisms of DDIs, (2) filtering these mutual mechanisms based on prevalence and biomedical similarity calculation, (3) and, finally, ordering the mechanisms of interaction. Before presenting each step, we define some concepts that are essential in this statistical inferences stage:

- ***Existing drug:*** *a drug that is found in both the D3 knowledge base and the clinically filtered DDI list.*
- ***New drug:*** *a drug that is found in the D3 knowledge base but not in the clinically filtered DDI list.*

These definitions for distinguishing between drugs are critical because they determine whether D3 should look in clinical trials for a known DDI or calculate a statistical inference on a potential one.

Stage 2 Step 1: Identifying the Likelihood and Clinical Relevance of a DDI

In this method, the goal is to use existing knowledge about a drug, including its clinically relevant list of interactions and biomedical features, to provide the likelihood of important potential DDIs. This step requires four main methods: (1) generating interaction profiles for a DDI, (2) constructing a biomedical similarity matrix for the interaction profiles of a DDI, (3) computing the likelihood of a DDI, and finally, (4) determining the clinical relevance of a DDI.

Step 1 Method 1: Generating Interaction Profiles for DDIs

In this step, the goal is to quantify a DDI between two drugs. That is, we want to identify the level of reported interactions for each drug in the DDI. Formally, let DDI = {drugX and drugY} where drugX ≠ drugY, let drugX = {drugX1, drugX2,..., drugXn} be a set of all reported interactions for drugX, and let drugY =

{drugY1, drugY2, ..., drugYn} be a set of all reported interactions for drugY. Next, we propose a method to retrieve two lists of interactions for both drugX and drugY from the learning process stage. **Algorithm 7** shows the process of retrieving a drug interaction list to create the interaction profile for a DDI.

Algorithm 7: Generating an interaction profile for drug

```
Input  : drug
Output: list of its interaction drugs
begin
 listOfDDI = (drug, clincalDDIlist);
if listOfDDI is not empty then
    return list of drug interaction
else
    drug is not found to have any clinically important interaction;
    exit;
```

Step 1 Method 2: Constructing a Biomedical Similarity Matrix for the Interaction Profiles with DDIs

Once profiles of interactions have been generated for a DDI, all drugs in both interaction profiles are used in the next step of identifying how similar the interaction profiles are for the drugs proposed in the DDI. In this step, we aim to compute the biomedical similarity for each drug within the proposed DDI based on their interaction profiles. Considering a first interaction profile for a drugX = {drugX1, drugX2,..., drugXn} and the second interaction profile for a drugY = {drugY1, drugY2,..., drugYn}, we want to identify how similar the {drugX1, drugX2,..., drugXn} list is to the drugY list and, vice versa, how similar the drugX list is to the {drugY1, drugY2,..., drugYn} list. Therefore, we create a

biomedical similarity matrix for each drug in the DDI. We then apply the methodology described in **Stage 1 Step 2: Building a Biomedical Features Similarity Matrix for Drug Comparison** to identify the similarity score for each drug profile with other drugs' profiles. **Figure 5.6** illustrates the construction of the biomedical similarity matrix to compute the similarity of drugX's interaction profile to drugY's.

	Enzyme	Target	Transporter	SideEffect	SNPs	MoA	Indication	Physiologic
*drugX1,**drugY***	0	1	0	1	0	1	0	0
*drugX2,**drugY***	1	1	0	0	0	0	1	0
*drugXn,**drugY***	0	0	0	0	0	0	0	0

Figure 5.6: Biomedical similarity matrix

The result of this step is two lists: (1) all drugs that interact with drugX = {drugX1, drugX2,..., drugXn} and are similar to drugY, and (2) all drugs that interact with drugY = {drugY1, drugY2,..., drugYn} and are similar to drugX. **Algorithm 8** shows the process of scoring each drug in the interaction profiles.

Algorithm 8: Similarity score for drug and interaction profile

```
Input  : interactionProfile, drug
Output: similarity score for all drugs in the interaction profile
begin
for (i=0;i¡interactionProfile.size;++);
 D3biomedicalSimilarityScore = (interactionProfile (i), drug);
if D3biomedicalSimilarityScore == 0 then
    go back to the loop
else
    return score;
```

Step 1 Method 3: Computing the Likelihood of DDIs

The main goal of this step is to use all calculations of **Step 1 Methods 1** and **2** to provide the likelihood of a DDI. This requires checking for an existing interaction between two lists (from Step 1 Method 2) and averaging the similarity scores. Once we compute the similarity of the two interaction profiles within a DDI, the next step is to check for an interaction between drugs in both lists. More clearly, we want only to consider the clinically proven interactions from both lists below a predefined threshold. Formally, let drugXddiSimilartoY = {drugX1ddiSimilartoY, drugX2ddiSimilartoY,..., drugX(n)ddiSimilartoY } be a list of drugs that are proven to interact with drugX and that are shown to have similar biomedical features to drugY, and let drugYddiSimileatoX ={drugY1ddiSimilartoX, drugY2ddiSimilartoX,..., drugY(n)ddiSimilartoX } be a list of drugs that are proven to interact with drugY and that are shown to have similar biomedical features to drugX. Following this, we want to check for an interaction between all drugs from both lists; i..e., (drugX(n)ddiSimilartoY vs. drugY(n)ddiSimilartoX). Assuming drugX**10**ddiSimilartoY was proven to interact with drugY**3**ddiSimilartoX, we then retrieve their biomedical similarity scores, which were computed in Step 1 Method 2. We average these similarity scores for both drugs to compute the likelihood of a DDI, using **Equation 6.2:**

$$LikelihoodOfDDI = \frac{drugX(n)SimilartoY + drugY(n)SimilartoX}{2}$$

We follow the same process for all drugs in both lists until there are no drugs left. The result is a new list of proven DDIs along with their similarity scores. We then take the DDI list and sort it in descending order. The highest value is considered to have the highest likelihood of a potential DDI.

Step 1 Method 4: Determining the Clinical Relevance of DDIs

DDIs remain a major challenge to effective administration of medications around the world. Researchers, drug companies and clinicians spend considerable effort attempting to mitigate DDIs and their potential to cause ADEs. However, the challenge of preventing DDIs is still unresolved. The most important reason for this is that there is still as yet no proper way to adequately identify significant DDIs. In other words, there is still no clear method for identifying potential clinically relevant DDIs. The medical need to have a clear method for recognizing clinically relevant DDIs is increasing as new drug therapies are introduced to the market and treatment strategies become increasingly complex, resulting in a theoretical increase in the number of potential DDIs.

In this section, we provide a novel computational method of recognizing clinically relevant potential DDIs based on the similarities not only within the interaction profiles for the DDIs, but also in their biomedical features. The main assumption is that the list of integrated and proven clinically relevant DDIs from Stage 1 Step 1 of our learning process (**Filtering Existing DDI Resources)** and the similarity score from the statistical inferences from Step 1 Method 2

(**Constructing a Biomedical Similarity Matrix for the Interaction Profiles with DDIs)** can provide us with a tool to identify clinically relevant potential DDIs. Formally, let DDI = {drugX and drugY} where drugX ≠ drugY. As we apply Step 1 Methods 1, 2 and 3 of the statistical inferencing process, we obtain a likelihood score for a potential DDI. Once we get the score, we could classify it based on five well-known DDI severity categories:

No interaction represents a score of [0 - 0.1].

(1) Minor interaction represents a score of [0.2 - 0.3].

(2) Moderate interaction represents a score of [0.3 - 0.4].

(3) Major interaction represents a score of [0.4 - 0.7].

(4) Avoid combination (contraindication) represents a score of [0.7 - 1.0].

After we classify the likelihood score, we only consider a DDI to be clinically relevant if the score is within categories 4 or 5. In other words, if the score is equal to or above 0.4, we consider it a clinically relevant DDI. To provide evidence for the usefulness of our method to recognize clinically relevant potential DDIs or not, in **V.1 Phase 3: Exploiting D3 Models** below we shall show a real world test case using a DDI between irinotecan and levofloxacin. .

Stage 2 Step 2: Ordering the Mechanisms of DDIs

The last method of statistical inferences is the determination of the most likely mechanism of a DDI. In the **V.1 Phase 1: Declaration of D3 Semantic Inferences** section, we created 9 different inferences to deduce possible mechanisms of DDI at the pharmacokinetic, pharmacodynamic, pharmacogenetic, and multi-pathway interaction levels. Later, in the **V.2.a Phase 2: D3 Inferential Query Model** section**,** we used the 9 inferences in the D3 query-based model to identify novel and hidden pathways of interaction. We showed how successfully we were able to infer the well-known mechanisms of a DDI as well as to suggest novel ones (**Appendix B)**. However, we have yet to demonstrate the capacity of the D3 query-based model for determining the main mechanisms of DDIs. In this method, the main goal is to identify the most likely mechanisms of interaction. In other words, we aim to suggest the mechanisms with regard to their roles when there are more than one. We also want to be able to eliminate any irrelevant mechanisms of DDIs. To achieve such a goal, we first need to identify the overall level of supporting evidence within existing DDI resources for the D3 inferences. **Algorithm 9** shows the process of computing the overall level of completeness of the knowledge-sources for D3 inferences among the 15 DDI resources, and **Figure 5.7** shows the results of this analysis.

Algorithm 9: Overall level support for D3 inferences

Input : Aggregated and non-duplicated DDIs list from 15 resources

Output: Supporting level for each D3 interaction mechanism inferences

begin

D3InferenceMechanisms = (ProteinBinding, MetabolismInduction, MetabolismInhibition TransporterInduction, TransporterInhibition, MultiPathways, Synergistic, Competition, Additive, Pharmacogenetic) ;

For each DDI resource;

1. Identify set of recognized drugs from present DDI pairs;
2. Create a list of all possible pairs of drugs that could interact as a result of any inferential rule;

Then;

3. Union the 15 lists;
4. From union, identify the level of completeness for each rule; (proven for rule /identified by rule);

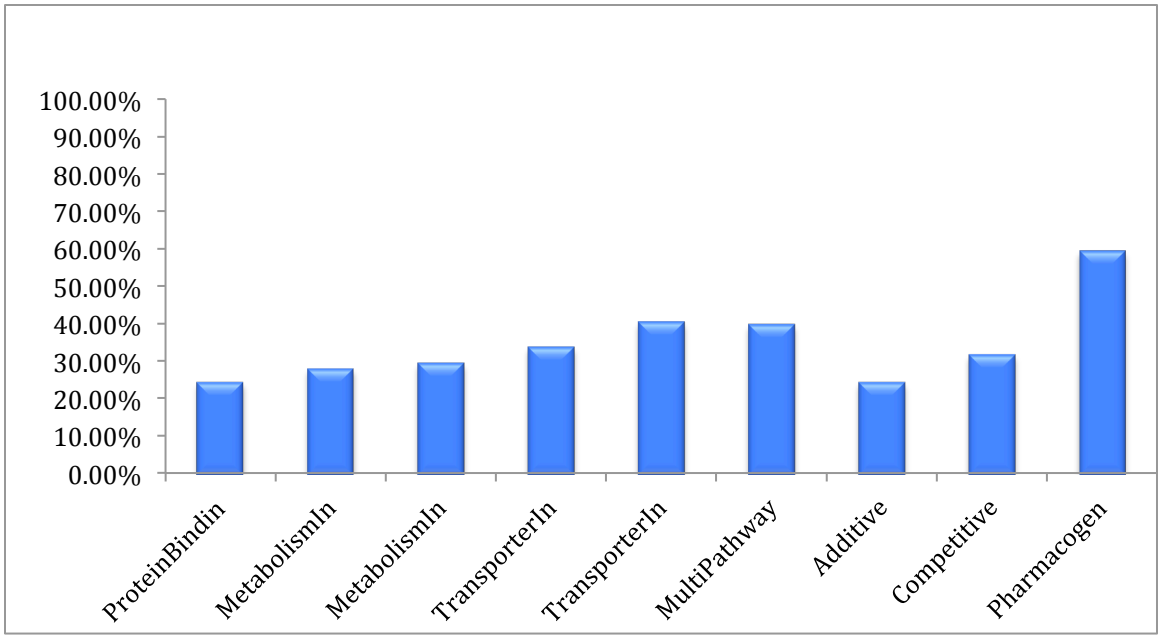


Figure 5.7: Bar chart showing the overall support that existing DDI resources provide for D3 inferences

This chart (**Figure 5.8**) shows how existing DDI resources support D3 inferences for the interaction mechanisms. The primary reason for this analysis is to identify why DDIs occur, but applying our D3 inferences across DDI resources does not show a clear independence. That is, the chart shows that DDIs are very complicated since there is not one dominant mechanism across all resources. Moreover, the chart clearly indicates the lack of supporting evidence for the D3 inferences across all interaction mechanisms. The mechanism with the largest support is pharmacogenetic, with over sixty percent, whereas the additive mechanism has the smallest support with just over twenty percent. Apart from the pharmacogenetic mechanism, reported evidence for the next highest supported mechanism provides support just below forty percent. This analysis indeed confirms findings from current research about the lack of studies supporting DDI mechanisms of interactions. Many studies indicate not only the lack of supporting DDI mechanism information but also the huge disparities in the information that is reported. One way to cope with this problem is by supporting decision-making based on mechanism information with available and accurate information about drug biomedical features. In this section, we present a novel method to probabilistically order the mechanisms of DDIs. The proposed method is divided into five main methods: (1) retrieving the mechanisms of DDI, (2) identifying similar proven interactions, (3) eliminating additional mechanisms of DDI, (4) computing the prevalence of mechanisms of interaction within DDI profiles, and, finally, (5) identifying the most likely mechanisms for DDIs. Before presenting each method,

we here define some concepts that are essential to our methods of ordering DDI mechanisms:

- ***Potential DDI:*** *a DDI in question.*
- ***Proven DDI:*** *two that are clinically proven to interact; ideally, proven DDIs will possess highly similar biomedical features to the potential DDIs in question.*

Step 2 Method 1: Retrieving the Mechanisms of a DDI

The main goal of this process is to query the D3 knowledge base to retrieve all possible mechanistic pathways for the potential DDI. We use the learning process **Stage 1 Step 3: Extracting All Possible Mechanisms of Interaction** to extract and create a list of all possible mechanisms of interaction.

Step 2 Method 2: Identifying a Proven Interaction Similar to a DDI

As we stated above, existing DDI resources are limited in their ability to accurately characterize mechanisms of interaction. To overcome such limitations, in this step, we first need to identify a proven DDI that is similar to a potential one. Therefore, we utilize the learning process **Stage 1 Step 2: Building a Biomedical Features Similarity Matrix for Drug Comparison** to detect a similar DDI. Once the similar DDI is found, then we retrieve all of its possible mechanisms of interaction.

Step 2 Method 3: Eliminating Additional Mechanisms of DDIs

The goal of this process is to filter the mechanisms of interaction. By filtering here we mean removing unimportant mechanisms from our list. In **Step 2 Method 2: Identifying a Similar Proven Interaction to a DDI,** we identified a DDI that was similar to the potential DDI, and we retrieved the mechanism(s) for the similar DDI. We have also (in **Step 2 Method 1: Retrieving Mechanisms of a DDI**) stored all possible mechanisms of a potential DDI. Therefore, we have two lists describing mechanisms of interaction. Our next step is then to identify the common mechanisms between our potential DDI and the proven similar DDI. Formally, let potentialDDI = {MoA1, MoA2} be a set of all possible mechanisms for potential DDIs, and let provenDDI = {MoA2, MoA3, MoA4} be a set of all possible mechanisms for proven DDIs; then potentialDDI ∩ provenDDI = {MoA2} is the set of remaining possible mechanisms for the proven and potential DDIs. The results of the intersection process are considered to be the most probable mechanisms of interaction that will be ordered in the next process.

Step 2 Method 4: Computing Prevalent Mechanisms of Interaction within the DDI Profile

Methods 1, 2, and 3 can be considered input processing stages. That is, all calculations of these methods will be used to provide inputs to Method 4. In this method, the aim is to support mechanisms of interaction based on a drug interaction profile as well as on biomedical similarity score. In other words, we

want to identify how common a mechanism is for a drug through its list of interactions. To answer this question we need to have two inputs: (1) a proven DDI that is highly similar to the potential DDI and (2) a non-sorted mechanisms-of-interaction list. **Step 2 Method 2: Identifying a Similar Proven Interaction to a DDI** provides us with the first input whereas **Step 2 Method 3: Eliminating Additional Mechanisms of a DDI** provides the second. Once all the required inputs are provided, two sub-phases are proposed to discover the commonality of a mechanism: (1) creating interaction profiles for the proven DDI and (2) identifying mechanisms common to the potential and proven DDI's interaction profiles.

Method 4 Sub-Phase 1: Creating Interaction Profiles for the Proven DDI

The main goal of Sub-phase 1 is to create interaction profiles that contain a list of drugs proven to interact with the examined DDI. From **Stage 2: Statistical Inferences,** we call **on Step 1 Method 1: Generating Interaction Profiles for DDIs** to generate two interaction profiles for the two interacting drugs in the DDI.

Method 4 Sub-Phase 2: Identifying Prevalent Mechanisms for Interaction Profiles

Once we generate the two interaction profiles of a DDI, Sub-phase 2 is designed to compute the commonality for each mechanism. Formally, let DDI =

{drugX and drugY} where drugX ≠ drugY. Let drugXinteractionProfile = {drugX1, drugX2,..., drugXn} be a set of all reported interactions for drugX, and drugYinteractionProfile = {drugY1, drugY2,..., drugYn} be a set of all reported interactions for drugY. Let DDImechanism = {MoA1, MoA2,..., MoA(n)} be a set of all possible mechanisms after eliminating all additional mechanisms that are not shared by the two lists. Next, we propose a novel commonality method to compute the importance of each mechanism within the interaction profile and the biomedical similarities. **Algorithm 10** shows how we compute the commonality score for each mechanism in the two profiles.

Algorithm 10: Mechanisms commonality for DDI

Input : drugX, drugY, DDImechanism

Output: commonality score for DDImechanism

begin

drugXinteractionProfile = drugX1, drugX2,..., drugX(n)
drugYinteractionProfile = drugY1, drugY2,..., drugY(n)

if *DDImechanism is empty* **then**
    There is no DDImechanism to test;
    break;
else
    for (MoA = 0; MOA *lessThan* DDImechanism; MoA++)

    comment: **score each drug in interaction profile with 1 if the mechanism applies**
    drugXmechanism = (drugXinteractionProfile, DDImechanism);
    drugYmechanism = (drugYinteractionProfile, DDImechanism);

    comment: **compute the similarity for drug with other drug profile of interaction**
    XsimilarityScore = (drugXmechanism, drugY);
    YsimilarityScore = (drugYmechanism, drugX);

    comment: **find commonality for each profile within a mechanism**
    XcommonalityScore = (drugXmechanism, drugY);
    YcommonalityScore = (drugYmechanism, drugX);

    comment: **sum all similarity scores for drugX and drugY**;
    for (x= 0; x *lessThan* XsimilarityScore; x++);
    XsumofSimilarityScore = x++;
    for (y= 0; x *lessThan* YsimilarityScore; y++);
    YsumofSimilarityScore = y++;

    comment: **Find commonality for drugX and drugY**;

$$XresultCommonalityScore(drugX, MoA1) = \frac{XsumofSimilarityScore}{drugXmechanism}$$

$$YresultCommonalityScore(drugY, MoA1) = \frac{YsumofSimilarityScore}{drugYmechanism}$$

    return to loop to test the next mechanism ;

Step 2 Method 5: Identifying the Most Likely Mechanisms of Interaction for a DDI

Finally, the last method of **Stage 2 Step 2: Ordering the Mechanisms of DDIs** is to order the mechanisms of interaction for DDIs, taking into account all the **Stage 2: Statistical Inferences** methods including: retrieving the mechanisms of a DDI, identifying a proven interaction similar to the DDI, eliminating extraneous mechanisms of the DDI, and, finally, computing the common mechanisms of interaction within the DDI profile in the previous methods, using **Equation 6.3:**

$$mechanismScore(drugX, drugY) = \frac{XresultCommonalityScore + YresultCommonalityScore}{2}$$

Now, at the end of Stage 2, three novel methods of inference have been initialized. These methods are: identifying the likelihood of DDIs, inferring the clinical relevance of potential DDIs, and sorting the mechanisms of DDIs according to importance. In each method, a set of complex statistical functions and calculations have been used to infer results. In Phase 3, we shall show how these inferences will be exploited in order to better understand and examine DDIs.

V.3 Phase 3: Exploiting the D3 System

Once both D3 inferential models are constructed, we can then study DDIs by discovering more relevant mechanisms of interaction that a might be involved in a DDI. In this phase we aim to show the usefulness of our D3 pharmacovigilance system for studying DDIs. We argue that the inferences in the D3 system are rich

with semantic information and computations, and by exploiting such computations we can infer novel semantic knowledge that can be used to provide a better understanding of DDIs and hence provide more relevant pharmacovigilance services. In **Chapter IV**, we have already proposed a model to infer a novel multi-pathway DDI from our D3 knowledge base. In that model, we proposed a rule-based method that allows our pharmacovigilance system to discover and obtain extra information about a chemotherapy medication's mechanisms of interaction, which, in turn, led to an enhanced understanding of the drug's potential therapeutic mechanism. In this phase, we argue that applying the same method to identify and explain mechanisms for all DDIs allows us to discover, infer and gain a better understanding of these DDIs and their mechanisms of interaction.

As explained above, the D3 pharmacovigilance system contains two main inferential models: (1) a query-based model and (2) a probabilistic-based model. The D3 query based-model has been designed to check for interactions using 15 DDI resources, as well as inferring 9 different mechanisms of interaction using SPARQL as an inference technique. The following scenarios have been proposed with the query-based model:

1. If both drugs exist in the D3 knowledge base and are listed in the 15 DDI resources as a reported interaction, the interaction is proven. In this case, D3 will infer all possible mechanisms of interaction.
2. If both drugs exist in the D3 knowledge base and are NOT listed in the 15 DDI resources as an interaction, this interaction is defined as potential.

D3 will infer all possible mechanisms of interaction to support the interaction identification.

3. If one of the drugs does NOT exist in either source, a clinical trial is recommended.
4. If neither drug exists in either source, a clinical trial is recommended.

The probabilistic-based model has been proposed to identify the likelihood of DDIs, assess their clinical relevance, and order the mechanisms of interaction. The following scenarios have been offered by the probabilistic-based model:

1. If both drugs exist in the D3 knowledge base and are listed in a D3 DDI list from Stage 1 Step 1 of our learning process (**Filtering Existing DDI Resources**) as a reported interaction, this DDI is listed as proven with high clinical relevance. D3 will then order all possible mechanisms of interaction according to importance.
2. If both drugs exist in the D3 knowledge base and exist in the clinically relevant DDI list but are not reported as an interaction, D3 will assess how likely this DDI is and what its clinical relevance is. Based on this assessment, D3 will then order all possible mechanisms of interaction according to importance.
3. If neither drug exists in either source, a clinical trial is recommended.
4. If neither drug exists in the D3 knowledge base BUT is listed in a clinically relevant DDI list as a reported interaction, the DDI will be listed

as proven with clinical relevance. However, D3 will not be able to provide mechanisms of interaction for this DDI.

5. If one of the drugs does NOT exist in either source, a one-sided calculation will assess how likely this DDI is and what its clinical relevance is. D3 will then order all possible mechanisms of interaction according to importance.

In the next phase, we propose two case studies: (1) a case study for the D3 inferential query-based model and (2) another case study for the D3 inferential probabilistic-based model. In each case, we will test both proven and potential DDIs.

Case Study 1: D3 Query-Based Model Inferences

In this case study, the main goal is to demonstrate the usefulness of the D3 query-based model by testing two real world DDIs. Specifically, we propose a DDI between aspirin and ibuprofen as a proven DDI and investigate interactions between irinotecan and levofloxacin as a potential DDI.

1. <u>Aspirin and Ibuprofen as a Proven DDI</u>

Aspirin and ibuprofen have been used for decades to relieve pain and other symptoms. A study was published by the MacDonald group that showed that co-administration of both drugs in cardiovascular patients led to increased all-cause mortality and cardiovascular mortality compared to patients who had received

aspirin only (MacDonald & Wei, 2003). In 2006, the US Food and Drug Administration (FDA) warned about interactions between these two agents(FDA, n.d.). Moreover, aspirin and ibuprofen's mechanisms of interaction are very complex and happen at multiple pharmacokinetic and pharmacodynamic levels (Awa, Satoh, Hori, & Sawada, 2012). Therefore, due to their complex interaction mechanisms and since this interaction has the potential to lead to serious ADEs, we tested our model's usefulness in studying interactions between the two drugs. Both drugs were entered into the model and 5 different DDI resources reported an interaction: NDF-RT, Twosides, Drugbank, Kegg, and OSCAR. Second, seven mechanisms of interactions were inferred (**Table 5.5**).

Table 5.5. Seven inferred mechanisms of interaction between aspirin and ibuprofen found by the D3 query-based model

Mechanism of interaction	Explanation
Pharmacokinetic (protein binding)	High protein binding affinity of aspirin(Wishart et al., 2008)
Pharmacodynamic (additive)	Cyclooxygenase Inhibitors*(*Micromedex® Healthcare Series.*, 2015)*
Pharmacokinetic (metabolism inhibition)	CYP2C9, PTGS2, PTGS1 enzymes (Samowitz et al., 2006)
Pharmacokinetic (metabolism induction)	CYP2C19 enzyme(Chen et al., 2003)
Pharmacokinetic (transporter inhibition)	SLC22A6 gene("Ibuprofen Pathway, Pharmacokinetics," n.d.)
Multi-pathway (metabolic- transporter)	CYP2C9 enzyme, ABCB1 gene, SLC22A6 gene, CYP2C19 enzyme, CYP2C8 enzyme(Samowitz et al., 2006) ("Ibuprofen Pathway, Pharmacokinetics," n.d.)
Pharmacogenetic	Rs20417(Lee, Kim, Wu, Wang, & Dionne, 2006)

The results in **Table 5.5** clearly show that our D3 query-based model can help in capturing hidden mechanistic information with regards to DDIs and provide different sources for reporting the interaction. Although these drugs have been thoroughly studied, knowing that D3 can infer potential mechanisms for proven DDIs at so many levels with well-known drugs increases our confidence that it can also identify potential interactions among less well-studied drugs.

2. Irinotecan and Levofloxacin as a Potential DDI

An effective way to infer potential DDIs is through identifying and understanding the mechanisms of interaction. We proposed to identify a possible interaction between irinotecan and levofloxacin and to understand the mechanisms of interaction. Irinotecan, which has been an FDA-approved drug since 1996, is an effective chemotherapeutic medication for treating colon cancer (Douillard et al., 2000). Irinotecan induces apoptosis by inhibiting the topoisomerase I enzyme, and consequently also inhibiting DNA replication and transcription (Pommier, Leo, Zhang, & Marchand, 2010). Unfortunately, patients treated with chemotherapy medications are at high risk of many different kinds of infections. According to cancer clinical practice guidelines, levofloxacin has good evidence supporting its use for primary infection prevention in patients undergoing chemotherapy (Freifeld et al., 2011). Therefore, irinotecan and levofloxacin were fed into the D3 model and the model returned 4 mechanisms that could cause interactions: metabolism inhibition (irinotecan and levofloxacin both inhibit the CYP3A4 enzyme), transporter

inhibition (levofloxacin inhibits the P-gp transporter), multi-pathway interaction (both drugs share the CYP3A4 enzyme and the P-gp transporter), and pharmacogenetic interaction (a specific SNP, rs1045642, is associated with altered P-gp activity for both drugs). To provide precise evidence, we queried the D3 knowledge base aiming to find a drug that was already known to interact with irinotecan and which had a similar pharmacokinetics profile to levofloxacin (an inhibitor of both the CYP3A4 enzyme and the P-gp transporter). Indinavir was retrieved. We could then hypothesize that there is a possible interaction between irinotecan and levofloxacin for two important pharmacological reasons. The first is based on the P-gp transporter. Indinavir has been shown to interact with irinotecan because it increases the level or effect of irinotecan (L. Zhang, 2010). Since levofloxacin and indinavir share the same transporter action (both are P-gp inhibitors), we can say that levofloxacin and irinotecan could interact via P-gp. The second reason is based on the CYP3A4 enzyme. Levofloxacin is a CYP3A4 inhibitor, and CYP3A4 plays an important role in irinotecan metabolism (Santos et al., 2000, p. 4). Therefore, when CYP3A4 inducers or inhibitors are administered, irinotecan concentration may change significantly. The ability to generate these two hypotheses was based on the capability of the D3 query-based model. Based on our hypothesis, we recommend one of two possible actions be taken as a follow-up:

1. Medical records either forward or retrospective for patients who have been prescribed both irinotecan and levofloxacin should be checked; or

2. Clinical trials should be conducted to prove the interaction between irinotecan and levofloxacin.

Case Study 2: D3 Probabilistic-Based Model Inferences

The D3 probabilistic-based model was designed to address three important limitations of the query-based model. The first limitation is that even though the query-based model infers potential interactions, it is not able to provide the likelihood of those interactions. The second limitation is that the query-based model will not be able to determine the clinical relevance of predicted interactions. The last limitation is that, though the query-based model infers potential mechanisms of interaction, those mechanisms require manual filtering to determine the most likely cause of the interactions. In this case study, we show how the D3 probabilistic-based model will overcome those limitations. We propose the same test cases we used in the query model to study the proven and potential DDIs.

1. <u>Aspirin and Ibuprofen as a Proven DDI</u>

In the query-based model, we showed how we were able to identify the proven interaction between aspirin and ibuprofen, list all DDI resources that report the interaction, and suggest 7 different mechanisms of interaction. Using the probabilistic-based model, on the other hand, we want to demonstrate that we can capture the interaction with its likelihood, provide its clinical relevance, and, finally, order the mechanisms of interaction according to importance. When

we enter aspirin and ibuprofen into the D3 model, it returns 1 (i.e., 100%) as the likelihood of interaction. In **Stage 2: Statistical Inferences, Step 1 Method 4: Determining the Clinical Relevance of DDIs,** we classify our likelihood score of interaction to determine the clinical relevance. Based on our classification schema, any likelihood score above 0.4 is considered clinically relevant. Thus, since the score of aspirin and ibuprofen is 1, it is then identified by our model as a clinically relevant DDI. Finally, the model orders the 7 mechanisms of interaction found by the query-based model as follows:

A. 0.046651974 Metabolism Inhibition
B. 0.037774377 Transporter Inhibition
C. 0.03353279 Additive
D. 0.030037235 MultiPathways
E. 0.02711666 Protein Binding
F. 0.017725537 Metabolism Induction
G. 0.016276017 Pharmacogenetic

Based on the results, metabolism inhibition is shown to be the strongest mechanism of interaction. In accordance, one study suggested that aspirin and ibuprofen might interact due to the metabolism inhibition mechanism of interaction (Saxena, Balaramnavar, Hohlfeld, & Saxena, 2013). Our results clearly show how the D3 probabilistic-based model is able to give a probability value for a DDI, assist in determining its clinical importance, and order the mechanisms of interaction based on a common mechanism within drugs' interaction profiles.

2. Irinotecan and Levofloxacin as a Potential DDI

The potential interaction between irinotecan and levofloxacin was suggested via the D3 query-based model. Moreover, the potential interaction was suggested because of four different mechanisms of interaction. Here, we are using the D3 probabilistic-based model to evaluate this interaction. Therefore, the two drugs are given to the model and the model returns a likelihood value of 0.23119652. Using our classification method, an interaction could be possible but the model would consider it a minor interaction. The model also provides prednisone and etoposide as similar proven DDIs (in terms of biomedical features) to the potential irinotecan and levofloxacin. Finally, we order the four mechanisms of interaction that were found to contribute to the DDI as follows:

A. 0.07724251 Metabolism Inhibition
B. 0.06572684 Transporter Inhibition
C. 0.060484476 Multi-Pathways
D. 0.050410457 Pharmacogenetic

The results of both test cases demonstrate how the D3 probabilistic-based model is capable of identifying the likelihood of proven and potential DDIs, assessing the clinical relevance, and ordering the mechanisms of interaction based upon importance when there are more than one. A corollary to this model is that it can be used to provide putative explanations for reported DDIs, the majority of which are reported without an underlying mechanism, and it can evaluate existing DDI resources to report highly clinical relevant DDIs.

V.4 Conclusion

In this chapter, we presented a distinctive system that is able to use existing knowledge about drugs, their biomedical features, and their interaction lists to provide pharmacovigilance services. This system is based on the integration of two types of inferences: query and probabilistic. The D3 system is a generic system that consists of different techniques and algorithms to study, understand and exploit users' drug knowledge and DDI resources to provide pharmacovigilance services. Finally, using different biomedical features of drugs, we showed that considering different drug information in pharmacovigilance systems can provide more effective DDI discovery and explanation. We also show that our system's probabilistic method can assist in determining the clinical relevance of existing DDIs and provide putative explanations for reported DDIs that have no mechanistic information. The main contributions of this chapter are as follows:

- We presented two inferential models and studied some of their components and capabilities.
- We proposed a generic system that could be used by diverse applications to provide an explanation for DDIs. This system is also generic so it can assist in determining the clinical relevence of DDIs.
- We proposed a novel method to identify the likelihood of a DDI based on the proven interaction information that is collected from existing DDI resources, as well as a process to compute the clinical importance of the DDI.

- We extended the method of inferring and exploiting new hidden knowledge that was presented in **Chapter IV** to infer new mechanisms for DDIs.
- We introduced a novel system for ordering the mechanisms of DDIs to provide the most likely causes of the DDIs.
- We demonstrate the essential nature of DDI resources and other biomedical information in providing an effective pharmacovigilance service, and illustrate the capacity of the D3 system to more effectively integrate drug information from a wide array of existing, reputable drug sources to provide superior pharmacovigilance services.

CHAPTER VI

DEVELOPMENT OF A COMPREHENSIVE, MECHANISM-BASED DRUG-DRUG INTERACTION RESOURCE AND ITS UTILIZATION IN CHARACTERIZING THE MECHANISMS FOR DRUG-DRUG INTERACTIONS DESCRIBED IN EXISTING KNOWLEDGE RESOURCES

In the previous chapter, we established a distinctive system that is able to use publically available drug information to provide unique pharmacovigilance services. This includes, but is not limited to: (1) mining 15 different DDI resources to identify potential DDIs as well as determine their likelihood, (2) assisting in determining the clinical relevance of discovered interactions, and (3) providing mechanistic explanations to characterize interactions across nine mechanistic levels: protein binding, induction and inhibition of drug metabolism, induction and inhibition of drug transport, competitive pharmacological effect, additive pharmacodynamic effects, pharmacogenetic considerations, and multi-pathway interactions.

In this chapter, we focus on exploiting this system to address three well-known yet unresolved limitations of current pharmacovigilance systems. These are: (1) omissions in the comprehensive reporting of DDIs, (2) a lack of mechanistic information to describe interactions, and (3) the inability to determine the clinical

relevance of DDIs. These issues make it difficult for both clinicians and researchers to make sound, evidence-based decisions regarding the potential for drug interactions in the setting of clinical medicine. Indeed, providing a comprehensive DDI resource is a very complex task. This is because, first, existing publically available and commercial DDI reporting resources vary widely in their reporting of interactions. Frequently, one resource will report a particular DDI while another resource will not. For instance, the interaction of irinotecan with beta-blocker drugs is reported only in Lexi-Comp, but not in any other resource. Second, the mechanistic information describing DDIs is rarely available and often not well-organized. For example, CredibleMeds, a clinically-oriented resource, provides only unstructured text explanations about DDI mechanisms. Third, the mechanistic information of DDIs, if found, is very frequently limited to metabolic interactions. For example, the Drug Interaction Knowledge Base (DIKB), a pharmacovigilance DDI resource, is built solely for examining metabolic DDIs. Finally, and most challenging, is the assessment of the clinical relevance of interactions identified in DDI resources. A recent study confirms that there is not yet a systematic way to determine the clinical relevance of DDIs (Scheife et al., 2015). The same study also highlights the significant need to distinguish the clinical impact of identified DDIs due to the increasing prevalence of interactions as new drugs are developed and commercially marketed. Our general aim, then, is to build a pharmacovigilance resource called D3 that can provide comprehensive DDI information. Our paradigm of providing comprehensive DDI information includes specifically enhancing

existing resources (clinical and text) with mechanistic information using the nine different interaction mechanisms listed in **Chapter V** and evaluating existing pharmacovigilance resources to extract the most clinically relevant potential and unique DDI pairs that have not yet been clinically well-described. In order to achieve these complex tasks, a number of requirements must be taken into consideration:

- When enhancing existing DDI resources with mechanistic information, only the most likely mechanisms should be added.
- The D3 pharmacovigilance resource should contain only the most clinically relevant DDI information that is extracted from publicly available DDI resources.
- The D3 pharmacovigilance resource should transform and store DDIs in a formal and semantic representation.
- The D3 pharmacovigilance resource should be able to provide precise mechanistic information.
- The D3 pharmacovigilance resource should contain the original source of the DDI information.
- The D3 pharmacovigilance resource should be universal enough to be integrated into and to provide inspiration for a wide range of studies.

Building upon these requirements, we divide this chapter into four main sections: (1) illustrating the enormous variation among DDI resources, (2) highlighting the lack of mechanistic information within DDI resources, (3) demonstrating the accuracy of identified mechanistic information within D3 by utilizing proven, clinically relevant DDIs as examples, and (4) building the D3 pharmacovigilance resource.

VI.1 Searching for Consensus among DDI Resources

DDI information can be found in literature, on webpages, in medical reports, and in databases. Manipulating and representing the extensive knowledge from these disparate resources is a common problem in the clinical domain. Both researchers and clinicians are challenged in their efforts to find reliable DDI information because it must be identified and collected from so many different formats and places, and in the end, the various resources may not agree. The purpose of this section is to prove the presence of variance in the DDIs reported by leading information sources along with the degree of behavioral independence among them. This is a study of the collective reporting behaviors of seven leading commercial and free resources based on comparative sampling of select DDIs reported (or not reported). These sources are: TWOSIDES, DrugBank, Lexi-comp, NDF-RT, Drugs.com, Micromedex, and Medscape.

As demonstrated in **Chapter IV**- IV.2.a, divergent reporting behaviors surfaced during the validation process. Some sources frequently reported many

DDIs while others reported far fewer. For example, it was observed that NDF-RT only reported 7 drug interactions with irinotecan, although it did report more than 9000 DDIs for other drugs. In contrast, Drugs.com reported 329 DDIs involving irinotecan, and Drugbank reported 11. Of perhaps greater concern was the inconsistency found among the resources regarding the reporting of particular DDIs with irinotecan. Frequently, one source would report a particular interaction while another source would not. For instance, the interaction of irinotecan with beta-blocker drugs was reported in Lexi-Comp, but not in the other resources. Simple observation of these inconsistencies led to an investigation into the level of agreement or disagreement among the resources. A preliminary null hypothesis was made that there would be neither a consistent agreement nor a consistent disagreement among the resources. An agreement/disagreement scale was constructed ranging from zero to one with zero and one representing absolute disagreement and absolute agreement, respectively. The null hypothesis was placed in the middle of the scale ($H0 = 0.5$). The primary focus in choosing this scale and structure was to explore various forms of hypothesis testing to analyze the overall relative reporting behaviors of the sources collectively. Such hypothesis testing would also allow approximate determination of the level of independence in the reporting behavior of the resources. The sample was based on reporting behaviors of the 80 reported DDIs of irinotecan among the 116 drugs originally considered in the verification process. Thus, each DDI report sampled from the resources involved the interaction of irinotecan with one of the 80 drugs reported by our DDI knowledge

base to potentially interact with it. From the reports sampled, a conformity sample was constructed by finding the mean reporting behaviors of all pairs of the seven sources and the overall collection of sources. After computing confidence intervals for the collective sample, the lower confidence bound equaled 0.49 and the upper equaled 0.58. Thus, there was a 95% chance that the population's mean was between 0.49 and 0.58 (**Table 6.1**).

Table 6.1: 95% Confidence interval to show the variations when reporting irinotecan interaction among resources

Sample Number (True positive)	80	95.00% confidence interval	
Sample mean	0.53		
Sample standard deviation	0.20	Lower bound	0.49
95% Confidence T-Value for 80 degree of freedom	2	Upper bound	0.58

This table shows the 95% confidence interval of the 7 DDI resources compared to each other. 0 value means absolute disagreement, 0.5 shows variations (agreement/disagreement), and 1 indicates absolute agreement. After computing, there is a 95% chance that the population's mean is between 0.49 and 0.58, which means there are variations in reporting the interactions among 7 resources.

Comparing this to the constructed conformity scale led to the strong conclusion that there was no level of statistical agreement in the reporting behavior between the DDI resources. Additional confidence intervals were investigated on samples of individual pairings of sources to aid in identifying the dependence and independence of the sources. The analyses of these confidence intervals are provided in **Figure 6.1**.

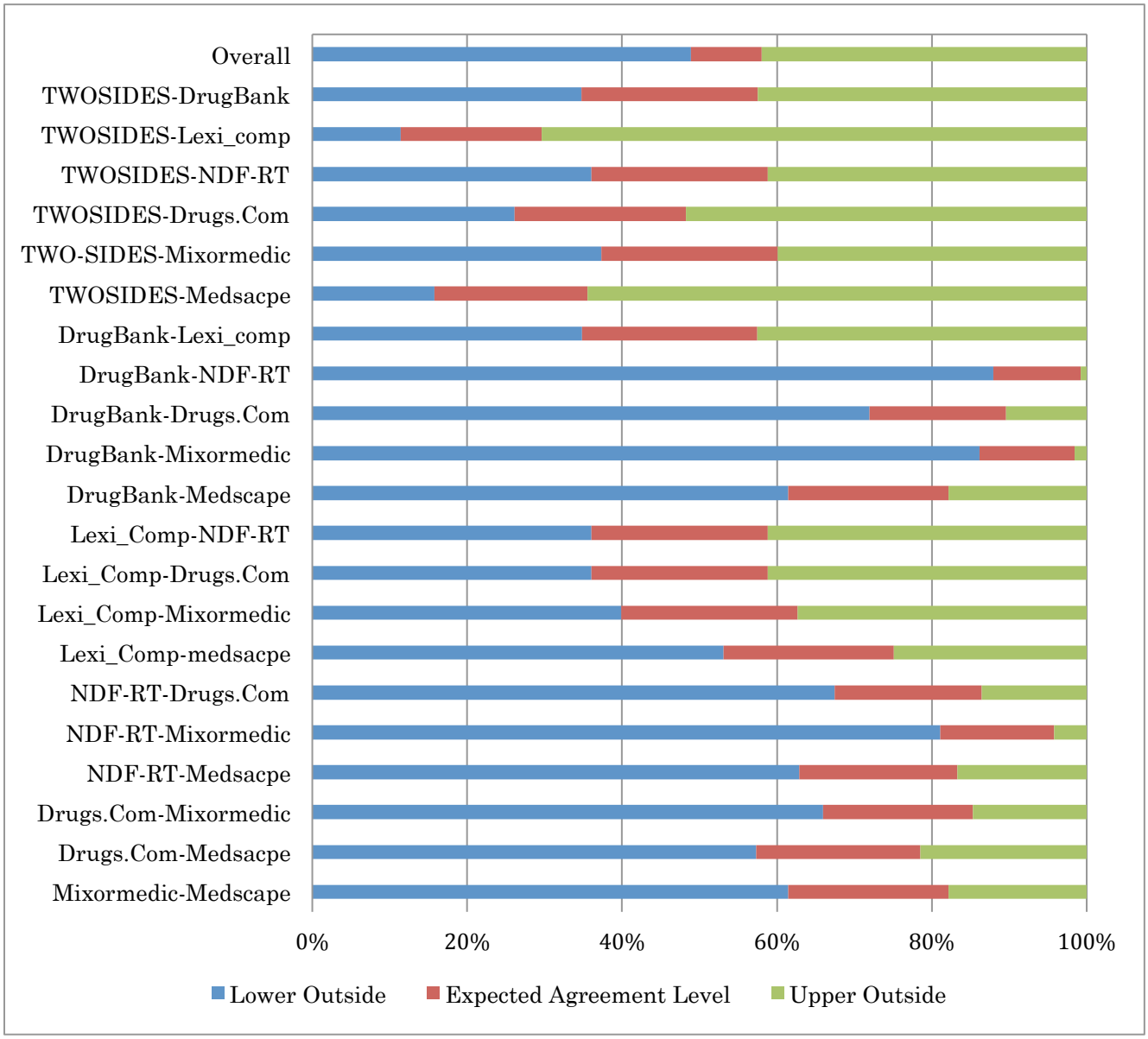


Figure 6.1: Confidence interval to show the variations when reporting irinotecan interaction among DDI resources

To put this result into perspective, we apply the same test for carbamazepine, a well-known drug used to treat seizures. Our carbamazepine result corroborates the enormous variation in DDI reporting among knowledge sources. **Figure 6.2** shows the expected agreement level for carbamazepine.

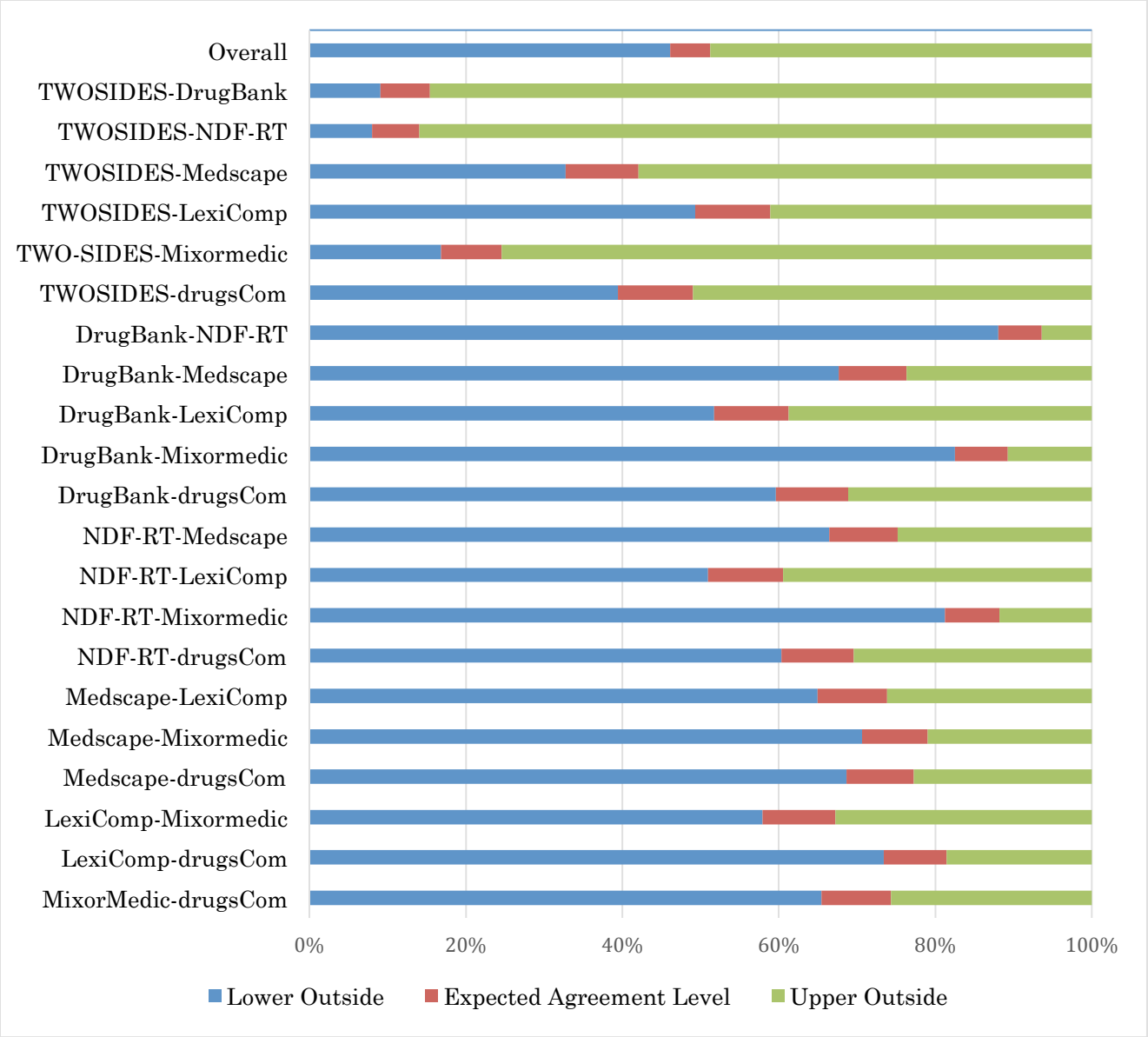


Figure 6.2: Confidence interval to show the variations when reporting carbamazepine interaction among DDI resources

The strength of agreement presumed from the testing intuitively infers a level of dependence among sources. Likewise, the strength of disagreement also seems to imply some level of dependence.

VI.2 The Need for Mechanistic Information among Data Resources

In **Chapter V, V.2.2, Method 3,** we computed the overall level of completeness of knowledge sources in supporting mechanistic information among 15 DDI resources. Our analysis showed that the existing DDI resources do not sufficiently assess the mechanisms of drug interaction. In this section, we would like to investigate this within an individual resource. That is, if a resource supports mechanistic information, how complete is the information in the resource and which types of mechanisms does it support? In order to answer these complex questions, an inference process is conducted to test the nine mechanisms resource by resource. We consider the 15 DDI resources used by the D3 query-based model in **Chapter V, V.2.1**. These 15 DDI resources are classified in 3 main categorizes: (1) clinical, (2) text, and (3) pharmacovigilance. After that, we test the nine mechanisms of interaction resource by resource and identify the number of occurrences for each mechanism. Eventually, this analysis will identify the popularity of a mechanism within multiple DDI resources. The consensus DDI mechanism is the one mentioned most consistently among the resources. **Algorithm 11** shows how we compute the popularity score for each mechanism within the DDI resource. The three F**igures 6.3, 6.4,** and **6.5** represent the results for each resource category.

Algorithm 11: Popularity of mechanism within DDI resource

Input : DDIresourceS, DDImechanismS

Output: score of each mechanism within DDI resource

begin

foreach DDIresource

1. Retrieve all DDI pairs
2. Create sets of pairs for each DDImechanism
3. Pair each set
4. Determine the size of the intersection of the pair divided by the size of the smaller set in the pair
5. **if** *result* == *1* **then**
   replace the larger set with the complement of the pair
6. **foreach** set

$$overallReportingRate(DDImechanism) = \frac{setSize}{numberOfDDIsInResource}$$

7. Sort the D3 mechanisms by reporting rate to identify the strengths of each DDI resource in each of the DDImechanisms

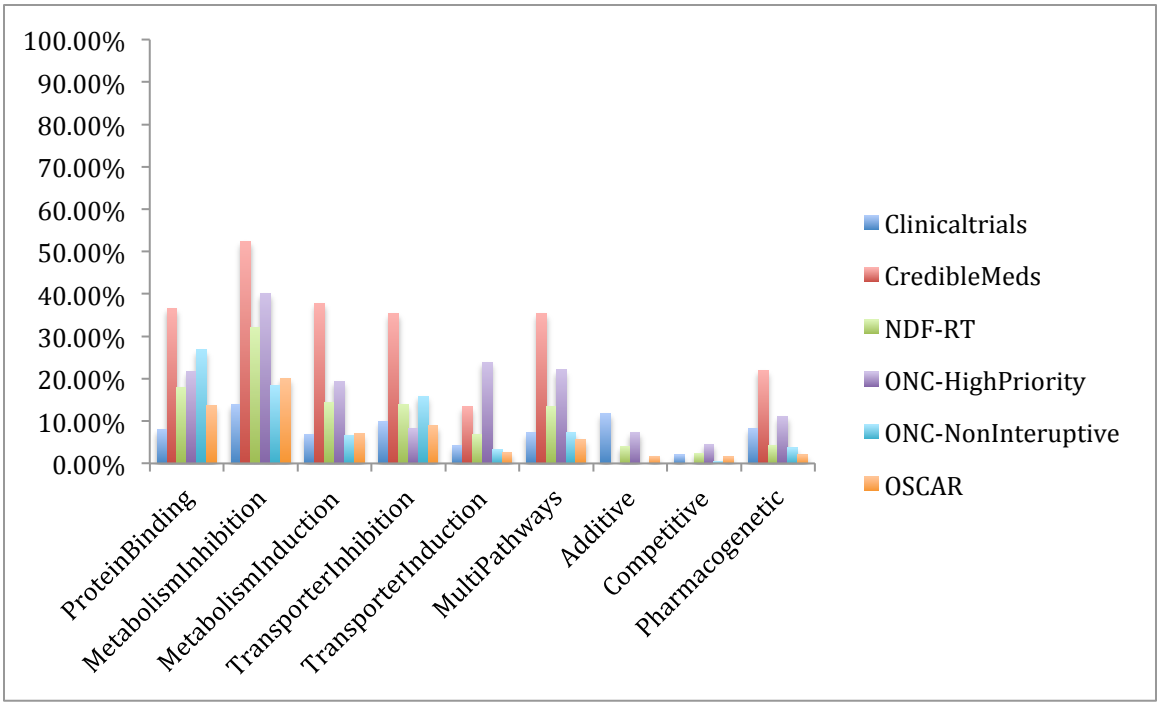


Figure 6.3: Mechanisms of interaction supporting levels by clinical resources

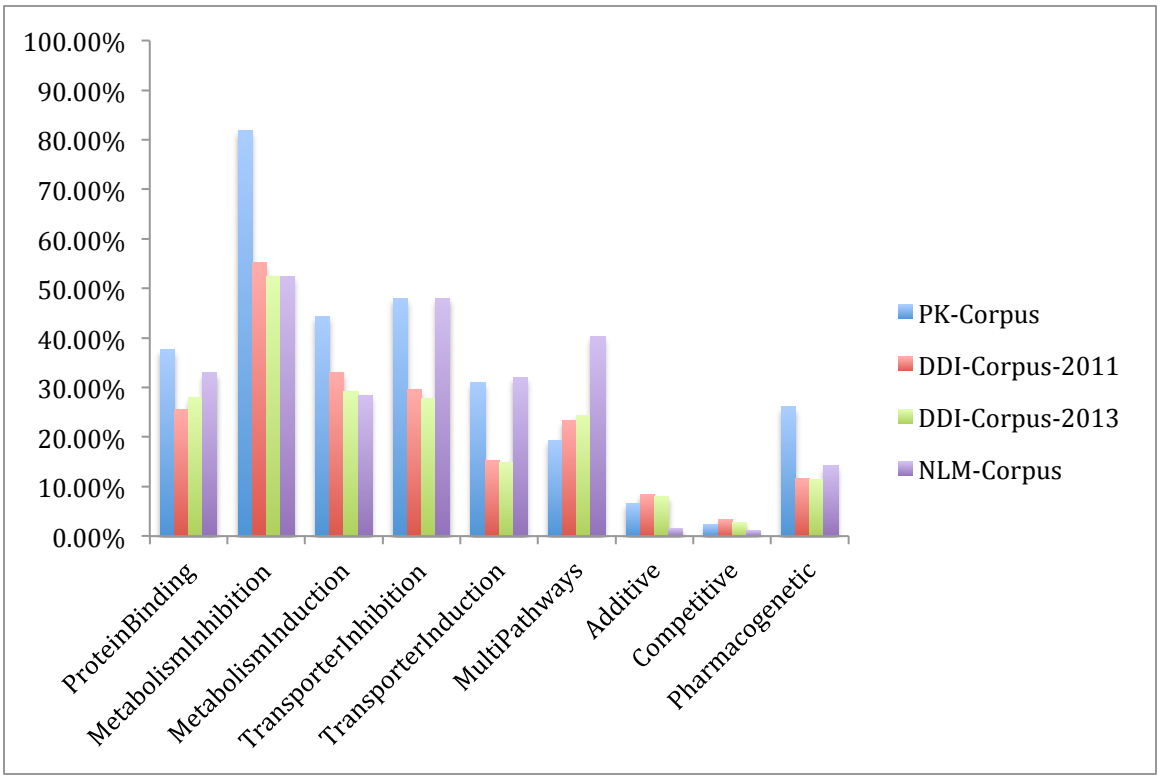


Figure 6.4: Mechanisms of interaction supporting levels by literature

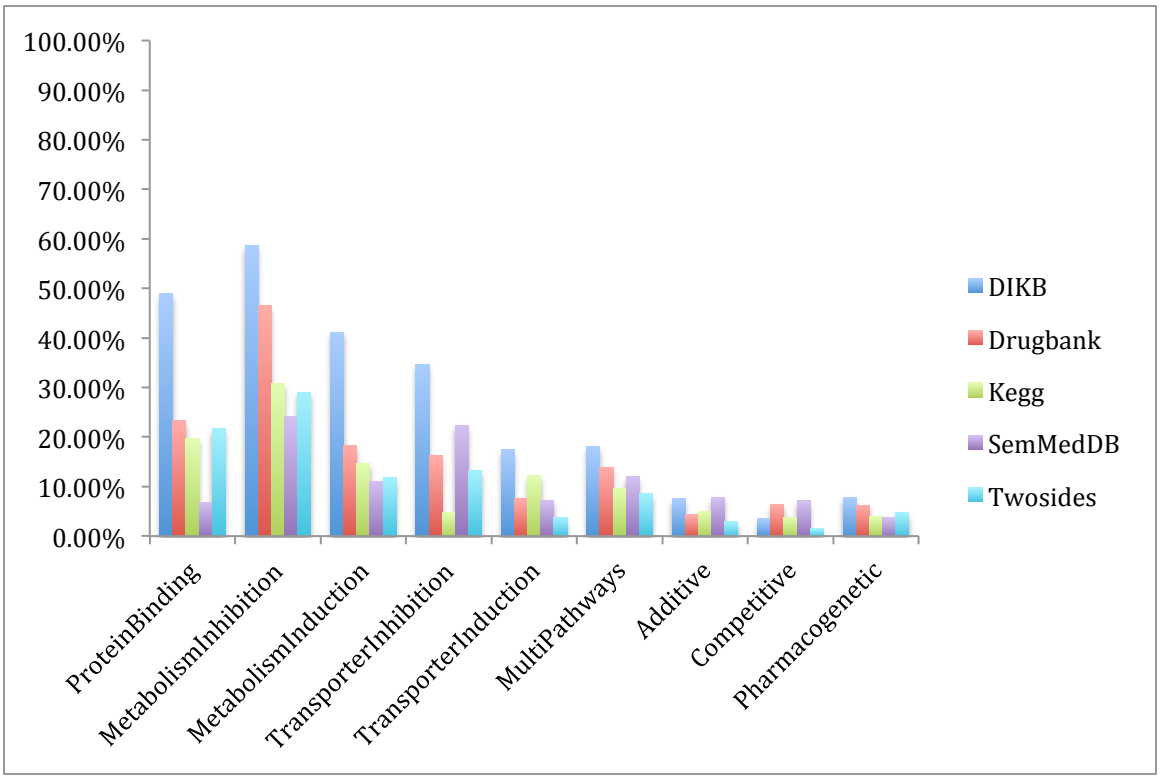


Figure 6.5: Mechanisms of interaction supporting levels by pharmacovigilance resources

Figures 6.3 – 6.5 show the frequency with which each specific mechanism of interaction was implicated as a mechanism for DDI among the different sources. As is clearly demonstrated in the charts, the 15 DDI resources vary greatly in the mechanisms they ascribe to their identified DDIs. In particular, the clinical resources **(Figure 6.3)** seem to mention fewer mechanisms of interaction than the other two categories of resources. Moreover, the mechanism with the largest support is inhibition of metabolism, being described as a mechanism in over seventy-five percent of DDIs, whereas pharmacological competition and additive pharmacodynamic effect mechanisms have the least support among all categories. The results of this analysis drive three important conclusions. First, our analysis shows the inconsistencies in terms of the types of interaction mechanisms supported among existing DDI resources; thus, there is a real need to accurately synthesize mechanistic information to elucidate the true mechanisms of interaction. Second, the clinical category (**Figure 6.3**), which includes six DDI resources, shows limited capacity to support all mechanisms overall and most notably the rare pathways such as additive and competition. Finally, though the overall support level for mechanistic information was not promising, the little support indicates that there is room for improvement in adding mechanistic information to what is already available for decision-making. In the next section, we attempt to overcome these limitations by annotating current DDI resources with mechanistic information.

VI.3 Utilizing Data Resources to Enhance Descriptions of DDI Mechanisms

The difficulty in discovering and studying DDIs occurs because DDIs, by their very nature, are complex processes that depend on many clinical, environmental, genetic, and physiological factors. Understanding the specific mechanisms that mediate DDIs, however, is an essential consideration for informing interventions to mitigate drug interactions. Throughout this thesis, we have highlighted the significant limitations in the contemporary research investigating DDI mechanisms. In **Chapter V**, we proposed a new system that is capable of studying DDIs across nine different mechanisms of interaction. In this section, we aim to utilize the D3 system to enhance the knowledge of existing and proven DDIs by more effectively integrating mechanistic information to characterize drug interactions. Since it is inefficient and impractical to confirm every DDI by means of a clinical trial, an expensive and slow process, providing information about the mechanism of the proposed interaction is key information that clinicians need to mitigate the occurrence of drug interactions. The process of adding the mechanistic information to existing DDIs involves three steps: (1) retrieving all proven DDIs from existing resources, (2) retrieving all possible mechanisms of DDI, and (3) annotating DDI reports with the most likely mechanisms of interaction.

Step 1: Retrieving All Proven DDIs from Existing Resources

Before annotating a given DDI with mechanistic information, we first need to retrieve all DDIs of interest. Therefore, from **Chapter V**, Step 1 of our learning

process (**Filtering Existing DDI Resources),** we retrieve a list of DDIs that contains 21,897 clinically important and non-duplicated DDIs. Those 21,897 DDIs are aggregated from ten different well-known clinical and text DDI resources.

Step 2: Retrieving All Possible Mechanisms of DDI

The next important step in enhancing DDIs with mechanistic information is to retrieve all possible mechanisms of DDI. This can be directly obtained via the learning process, **Step 3: Extracting All Possible Mechanisms of Interaction,** in **Chapter V**. This results in 9 different mechanism of interaction: (1) protein binding; (2) metabolism inhibition; (3) metabolism induction; (4) transporter inhibition; (5) transporter induction; (6) additive pharmacodynamic effect; (7) competitive pharmacological effect; (8) pharmacogenetic interactions; and finally, (9) multi-pathway interactions.

Step 3: Annotating DDIs with the Most Likely Mechanisms of Interaction

Once we retrieve the list of proven DDIs and all possible mechanisms of interaction, the annotation of the DDIs can be made by executing the novel **Ordering the Mechanisms of DDI** method from **Chapter V**. This method orders the mechanisms based on biomedical similarity between the DDI and the commonality of a mechanism within drug interaction profiles. A possible limitation of this method is that it assumes the correct mechanism to always be at the top of the list. Therefore, if it is not, the system does not consider it. In this step, we

overcome this limitation by introducing a margin of error for the DDI mechanisms. The goal of the margin of error is to improve our accuracy when assessing the mechanisms of interaction. Since we cannot be certain that the correct mechanism will be listed at the top of our created mechanism list, due to the previously demonstrated limitations in mechanistic information within the knowledge bases, we must determine the potential for error within our system. Moreover, applying the margin of error can assist in eliminating any extra mechanisms (i.e., those which are above the margin of error value) and reordering the mechanisms (i.e., by measuring how far each mechanism is from the margin of error value). This will eventually lead to retention of only the most likely mechanisms of DDI.

In order to identify the margin of error boundary, we collect a sample of 9 DDIs where we know in advance the mechanism of interaction, which is protein binding. Then we enter each DDI using **Method 3: Ordering the Mechanisms of DDI,** and compute the margin of error for each pair by subtracting the highest value from the correct value. For example, if the protein binding mechanism were ranked to be the highest for a DDI, then the result would be 0 (highest-correct). From the DDIs sampled, a margin of error sample was computed by finding the mean margins of error of all pairs of the nine DDIs. After computing the standard error of the mean for the collection sample, the upper confidence boundary equaled 0.025035745. Thus, there is a 97.5% chance that the population's true mean is below 0.025035745. **Table 6.2** shows the nine collection sample calculations.

Table 6.2: Nine protein binding samples are used to determine the averaged error rate

DDI Pair	**Error Result**
dexamethasone, phenytoin	0.02421753
rifampicin, phenytoin	0.024348518
aspirin, phenytoin	0.005775154
gemfibrozil, warfarin	0.01786845
doxycycline, warfarin	0.018178755
phenytoin, warfarin	0.011138968
clofibrate, warfarin	0.0
Sulphamethoxazole, warfarin	0.02000928
Sulphinpyrazone, warfarin	0.03797837
Mean	0.017723892
Standard deviation	0.011191613
Standard error of the mean	0.003730538
Upper 95% limit	0.025035745

Once we estimate the margin of error for the population mean, we use it to check the results from **Method 3: Ordering the Mechanisms of DDI.** We approach this by subtracting the value of the most likely mechanism from our list (0.13980994 in the example below) from the values given to the other mechanisms. Then, if the calculated difference (the value after subtraction) for each mechanism remains within our margin of error (i.e., less than 0.025035745), we consider it a possible mechanism of DDI; otherwise, we eliminate it. For example, the interaction between rivaroxaban and rifampin has been associated with an increased risk of

stroke (likely due to induction of CYP3A4 metabolism, the enzyme that metabolizes rivaroxaban, by rifampin) and, upon analysis by D3 ("Safety Information > Xarelto (Rivaroxaban) Tablets," n.d.), was found to possibly interact through 5 different pathways before applying the error margin:

A. 0.13980994 Metabolism Induction
B. 0.12819114 Protein Binding
C. 0.10624011 Multi Pathways
D. 0.10554226 Transporter Inhibition
E. 0.084756725 Transporter Induction

However, applying the margin of error enables us to eliminate three less important mechanisms of interaction. Two mechanisms are kept from the original list:

A. Metabolism Induction
B. Protein Binding

To identify the benefit of the margin of error in considering the most likely mechanisms of interaction, we collect a random sample of 40 proven DDIs that interact because of a great variety of mechanisms of interaction. Then we compute how often D3 is able to identify the correct mechanisms under three constraints:

1. **Listed:** How often does the correct mechanism show up on the list?
2. **Within margin of error**: How often does the correct mechanism still appear on the list after checking the margin of error threshold?
3. **Top Mechanism**: How often does the correct mechanism show up at the top of the list?

Figure 6.6 shows how the correct mechanisms fall within the margin of error: 92%. While the correct mechanism was not on the top, it was within the margin of error, and therefore, because it is, the correct mechanism could conceivably be at the top if the knowledge base was even more comprehensive.

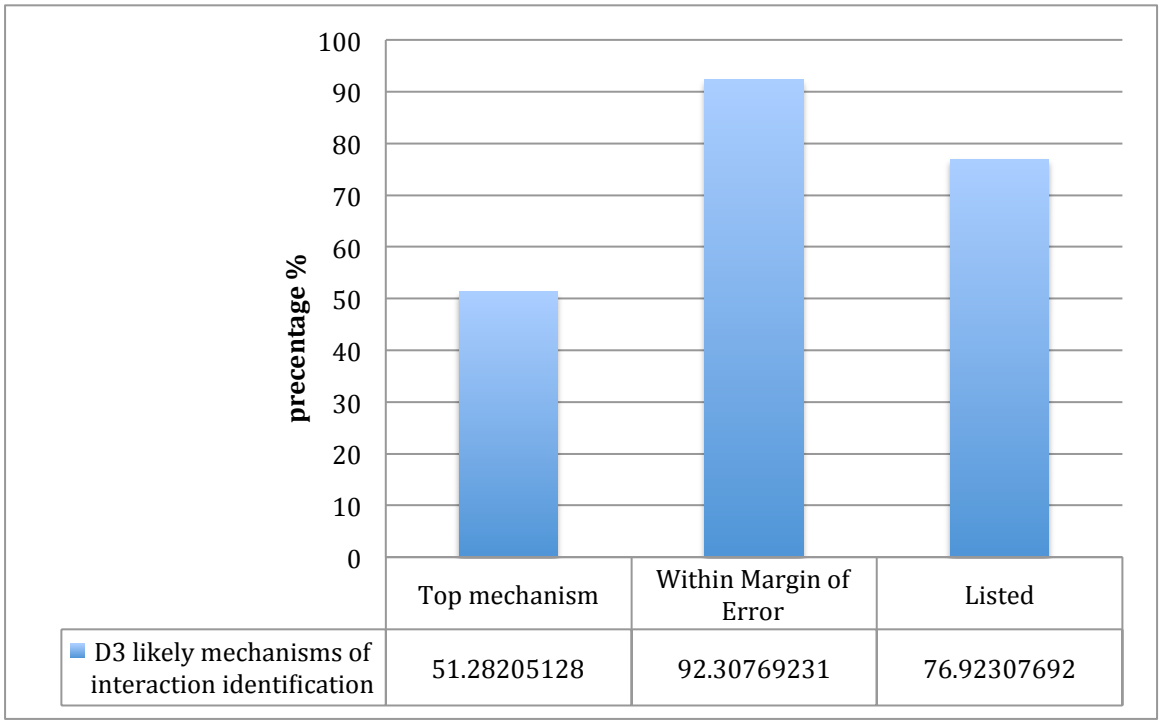


Figure 6.6: D3 most likely mechanisms of interaction identification

We apply the same methodologies to check for all possible mechanisms before annotating the proven DDIs. **Algorithm 12** shows how we add the mechanistic information to existing DDIs. Forty-two percent of the 21,897 DDIs were successfully annotated with mechanistic information, while 58% were reported to interact due to other mechanisms that are not covered by the D3 system.

Algorithm 12: Annotating DDIs with the Most Likely Mechanisms of Interaction

```
Input  : DDIs
Output: Possible mechanisms for DDI
comment: run Method 3: Ordering the Mechanisms of DDI
MoA = (DDI)
if MoA is empty then
    This DDI is porven but the mechansims is not covered by D3;
else
    comment: get the first (highest) mechansim form the list = MoA1
    comment: subtract the highest from others;
    newMoA(n) = MOA1 - MOA(n);
    comment: check if the mechansims value within margin error
    if newMoA(n) > marginErrorValue then
        remove MOA(n);
    else
        possible mechasnism = MoA(n);
    end
end
```

Figure 6.7 displays a pie chart of the annotation results, and **Figure 6.8** displays an example of an annotated DDI between aspirin and ibuprofen.

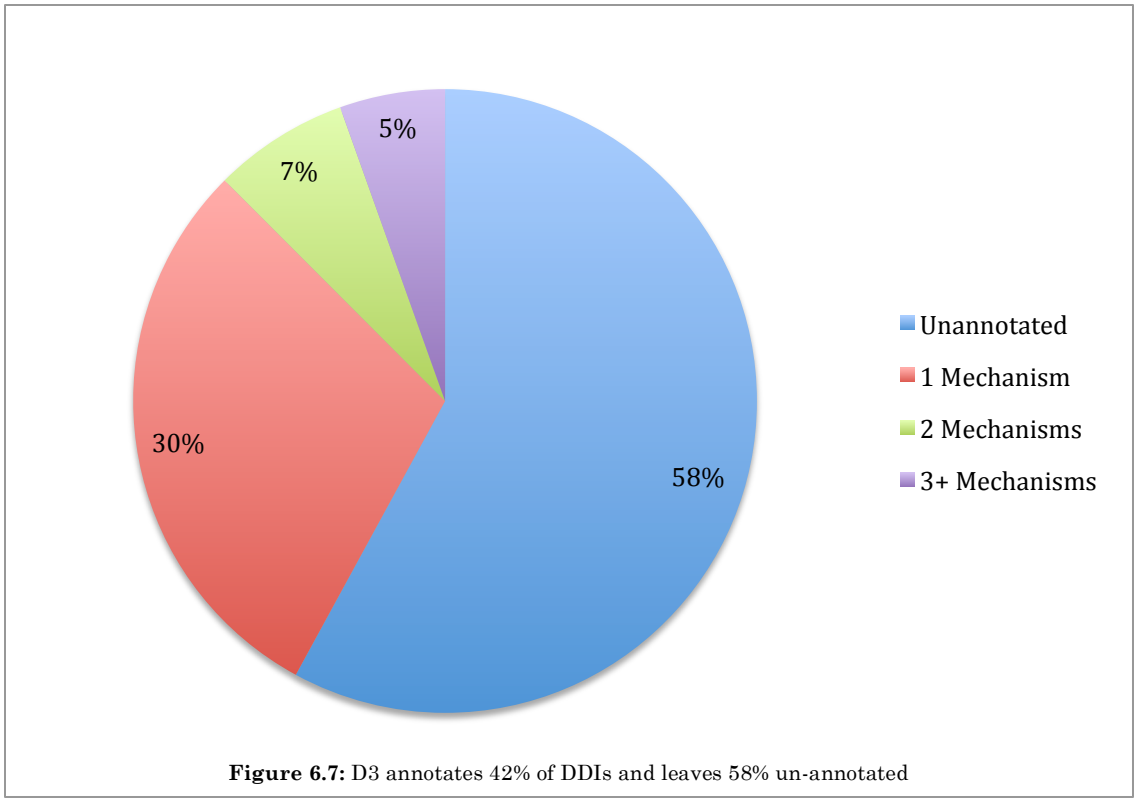


Figure 6.7: D3 annotates 42% of DDIs and leaves 58% un-annotated

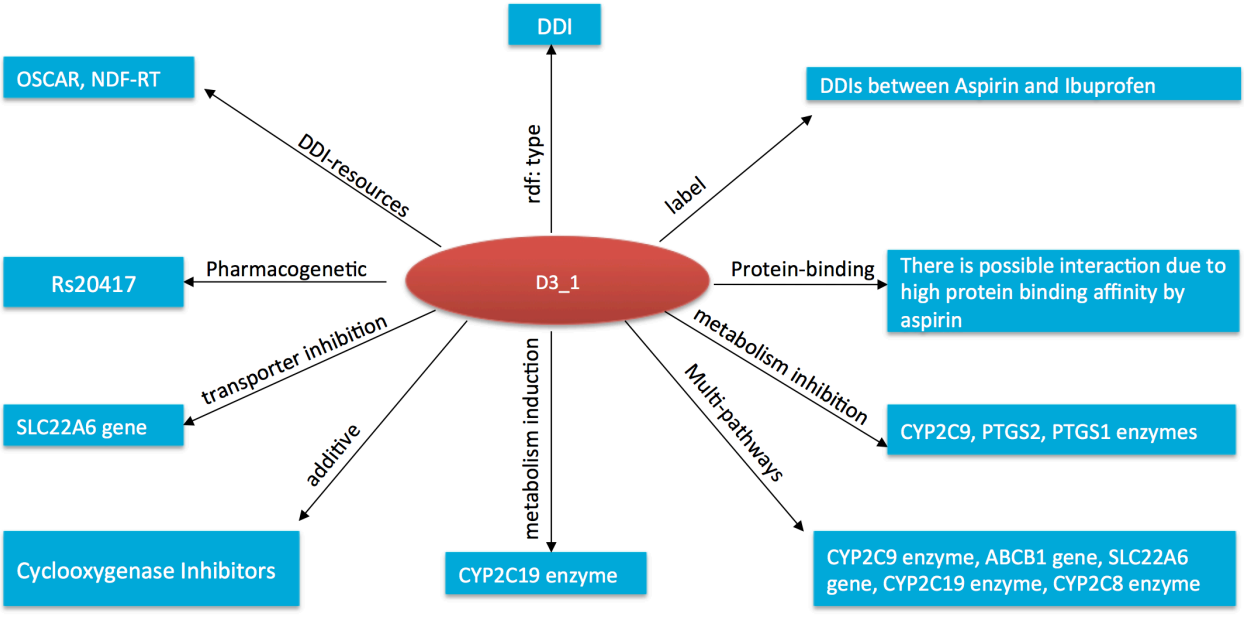


Figure 6.8: DDI between aspirin and ibuprofen with D3 annotation

The results of the annotating process can help to provide putative explanations for reported interactions, the majority of which are reported without an underlying mechanism. This technology will potentially aid researchers and clinicians in identifying the possible mechanisms for drug interactions of both pre-market and post-market-released drugs. Thus, it may improve the safety of medication use in the clinical setting and also provide a means of streamlining clinical trial design, which in turn will help to minimize the time and expense of designing clinical trials.

VI.4 Building the D3 Pharmacovigilance Resource

Even though the challenges posed by DDIs for effective medication therapy have been recognized for many years, there is still no well-trusted resource that can be universally used to check and validate DDIs. Accordingly, a recent study advises using more than one DDI resource for assessing the potential for interaction when using medications concomitantly *(*Conde-Estévez, Echeverría-Esnal, Tusquets, & Albanell, 2015). The lack of a single, comprehensive DDI resource exists because of two important reasons. First, the inconsistency among resources in reporting DDIs leads to extensive gaps in information. For instance, Zhang et al. suggest that using compendia for validation of DDIs could lead to many false positives due to the incompleteness of available curated data sources, and thus, precision can suffer (R. Zhang et al., 2014b). Tari et al. also report the limitations of using DrugBank as a gold standard for validation; they found only 11.0% of their results reported while 77% were found by searching the medical literature; therefore, weak true positive

rates are another problem when using these databases (Tari et al., 2010). Second, the astronomical number of reported potential DDIs distracts from efficient recognition of those listed DDIs that have true clinical relevance (Conde-Estévez et al., 2015). Therefore, the lack of agreement in reporting the clinical relevance of DDIs could lead to serious medical decision errors. For instance, a study shows vast conflicts between Micromedex and Drug Interaction Facts (DIF) in reporting the severity of DDIs.

In this section, we aim to solve the current issues of reporting DDIs by building a system that is superior to others and that can serve as a comprehensive go-to source for DDI information. We argue that most of the DDIs found in non-clinical resources (pharmacovigilance systems) are not clinically relevant and need to be reevaluated, and that applying the D3 system to evaluate existing DDI resources allows us to discover, infer and gain only the most clinically important DDIs. In **Chapter V**, we proposed a system to quantify the likelihood of a DDI. In that system, we determined the likelihood by computing the similarity of a potential DDI to other known DDIs with regards to biomedical features and interaction profiles. This process allows our pharmacovigilance system to discover additional information about a potential DDI, which leads to a better understanding of the mechanisms mediating the interaction and its clinical importance. In this section, however, we will apply the D3 system specifically to pharmacovigilance knowledge resources. To do this, we build our D3 pharmacovigilance resource by extracting and integrating DDIs from 5 existing pharmacovigilance DDI resources.

Specifically, we aim to evaluate these 5 non-clinical resources to capture DDIs with a high potential for clinical relevance according to analysis from our D3 system. The five pharmacovigilance DDI resources we consider for building our D3 resources are Drugbank, The Drug Interaction Knowledge Base (DIKB), KEGG, SemMed, and Twosides. In order to evaluate their DDI lists, we first need to remove the overlap from the 21,897 DDIs we identified in **Step 1: Retrieving All Proven DDIs from Existing Resources.** Once the overlap is removed, we are left with a unique list of DDIs (false positives) that have not been reported in any clinical resource. **Table 6.3** shows all resources with their DDI numbers before and after overlap removal.

Table 6.3: The unique positive DDIs retrieved from 5 pharmacovigilance resources

DDI source	Original DDI Number	Definition	After overlap removal
DrugBank	11840	List of DDIs along with their explanations.	8754
The Drug Interaction Knowledge Base	560	List of pharmacokinetic DDIs along with their evidence.	379
KEGG DDI	26328	DDI resource derived from Japanese product labels.	23,241
SemMedDB from PubMed	3536	DDI resource derived from PubMed abstracts.	3216
Twosides	63333	DDI resource derived from the analysis of spontaneously reported adverse events	59,922

These five DDI resources are used for this analysis because they do not contain strong clinical evidence to support their DDIs. Some of them also are a result of an inference process in which they identify many theoretical DDIs that have not been clinically proven. Once we get sets of false positives for each resource,

the next step is to use the D3 system to calculate the likelihood of each DDI in each resource set. The last step is to calculate the recall for the D3 system to obtain the appropriate threshold for generating the D3 resources. Therefore, recall of the D3 system is calculated with respect to the 10 different DDI resources used in this thesis. That is, we want to use recall rate as means of identifying how well the D3 system applies all the knowledge given. We calculate the recall of the 21,897 DDIs from **Step 1: Retrieving All Proven DDIs from Existing Resources** at different thresholds; **Figure 6.9** shows the performance of D3 system within in its knowledge sources.

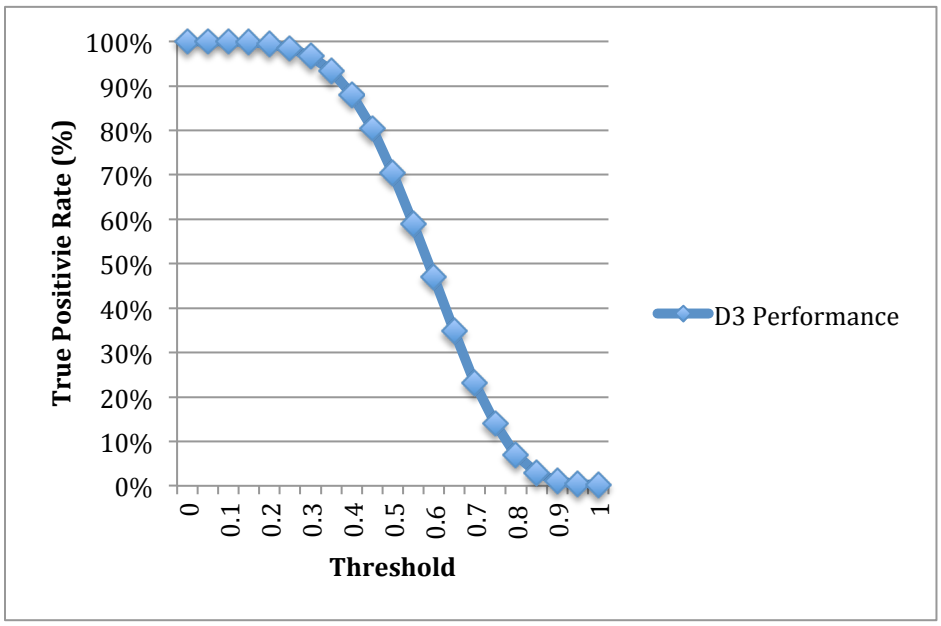


Figure 6.9: D3 recall using 10 DDI resources

Figure 6.9 confirms that the D3 system performs very well with its own knowledge sources at different thresholds in capturing most clinical DDIs. For example, at the 0.05 thresholds, D3 catches 99.9% of all clinical DDIs. Since we

want to have a low false positive rate and to only retrieve clinically relevant DDIs from the five resources, we set our threshold at 0.3, indicating that the D3 system would capture 97% of correct proven DDIs. **Table 6.4** shows the number of retrieved DDIs based on a 0.3 threshold value that would create the D3 pharmacovigilance resource.

Table 6.4 DDI pairs passed by D3 and used to create the D3 pharmacovigilance resource

DDI source	Original positive	Passed by D3 based on 0.3 threshold	False positive percentage
DrugBank	8,754	50%	47%
The Drug Interaction Knowledge Base	379	80%	17%
KEGG DDI	23,241	39%	58%
SemMedDB from PubMed	3,216	20%	77%
Twosides	59,922	25%	72%

The final step is to combine the extracted DDIs from the five resources and remove the overlap among them to create a clinically relevant DDI resource from existing DDIs. The result is 16,276 non-duplicated and potentially clinically relevant DDIs that are stored in the D3 pharmacovigilance resource. The list can be limited if we raise the threshold, or increased if the threshold is lowered. However, we want to keep our false positive rate as high as possible in order not to miss any clinically relevant proven DDIs.

VI.5 Conclusion

In this chapter, we highlight the immense need for a gold-standard mechanistic DDI resource as well as evaluate available DDI information. Then, we employ the D3 system to overcome these limitations. This is done by adding mechanistic information to existing DDI reports to generate a resource that can identify clinically relevant potential DDIs. For instance, D3 is able to annotate mechanistic information for 60% of DDIs encountered within 10 DDI resources. Such information will be able to not only assist clinicians in practice, but also direct research to predict potential DDIs before a drug is introduced to the market. Second, the D3 prediction performance value provides confidence in evaluating existing DDI resources. This is demonstrated by extracting clinically relevant potential DDIs from five pharmacovigilance resources. Our analysis indicates that these 5 resources report many theoretical false positive DDIs. Ultimately, the D3 pharmacovigilance resource is built by integrating important DDI information from a wide list of reputable knowledge sources, creating a comprehensive list of potential DDIs that assesses the probability that a pair of drugs will interact, the mechanism by which they interact, and the clinical relevance of potential interactions.

CHAPTER VII

BENCHMARKING AND EVALUATION OF D3

In the previous chapters of this thesis, we presented distinct evaluations that focused on examining specific aspects of our D3 system, such as evaluating a multi-pathway DDI and examining the capacity of the D3 query-based and probabilistic-based models. We evaluated such aspects in isolation via test cases to examine the potential of our system to effectively integrate and synthesize information to provide accurate inferences for characterizing DDIs. Different evaluation methodologies were employed to examine different capacities of our system. From these analyses, we have obtained positive results that demonstrate the potential of D3 to serve as a useful pharmacovigilance system. However, because these evaluations examined each aspect in isolation, further testing is needed to assess D3's ability to comprehensively characterize DDIs. Additionally, since most of our previous examples used in our test cases were literature-based rather than coming from clinical decision support tools, we aim to incorporate all the methods, techniques and algorithms previously proposed to assess D3's capacity to identify DDIs, define their mechanisms, and determine their clinical significance in

comparison with a trusted DDI decision support tool that is widely used in clinical practice (Micromedex). In particular, we aim to examine the effectiveness, efficiency and subjective satisfaction of using our system to answer the following essential questions:

- Does D3's integration of drug information provide a better pharmacovigilance service than other existing systems?
- Is D3's drug information knowledge base generalizable enough to be implemented for other applications (e.g., drug repurposing)?
- Can our system accurately assess the clinical relevance of DDIs?
- Can our system provide putative explanations for reported interactions, the majority of which are currently reported without an underlying mechanism?

We first discuss different evaluation methodologies from the literature, and then we introduce the methodology we will employ to evaluate our work.

VII.1 Evaluation of Pharmacovigilance Systems

The evaluation process for pharmacovigilance systems, especially those seeking to identify DDIs, is known to be difficult, time consuming and expensive (Abarca et al., 2006). This truth is attributed to a number of factors, including the enormous number of drug interaction pairs over which each pharmacovigilance system must be evaluated to gain a true assessment of their DDI identification

ability. Moreover, some systems are specialized to report only certain type of DDIs (e.g. metabolic DDIs), while other pharmacovigilance systems focus on more than one DDI interaction type. Therefore, each system may require different evaluation strategies and metrics. Another significant challenge in evaluating pharmacovigilance services is that definitive experiments are lacking for many theoretical DDIs, requiring time consuming and expensive clinical pharmacokinetic or pharmacodynamics studies to truly characterize the potential for drug interaction, its severity, and its potential clinical ramifications. These types of analyses typically include assessing how potential interactions alter the interacting drugs' concentrations and then determining the therapeutic index of the interacting medications, that is, how changes in drug concentrations will impact the drugs' therapeutic efficacy as well as their potential for adverse effects. Finally, drug interaction mechanistic information is rarely found in pharmacovigilance systems and the available information is often limited to one mechanism of interaction. Other mechanistic information is generally scattered across disparate literature sources and requires tedious manual work for effective extraction. As a result, there is no standard way to evaluate a pharmacovigilance system because each system proposes different strategies to identify DDIs, strategies that are tailored to a specific system's coverage and perspective. However, many researchers have recommended strategies to evaluate the pharmacovigilance system. For example, one study suggests that evaluation methods utilize drug compendia, while another study recommends utilizing more than one pharmacovigilance system to obtain

consensus information about specific interactions by which to evaluate new systems (Conde-Estévez et al., 2015; Scheife et al., 2015).

VII.2 Evaluation of the D3 System

The main aim of this chapter is to benchmark and evaluate our D3 system. Fundamentally, we want to know if using the existing resources about DDIs could provide improved pharmacovigilance services. To accomplish this task, we profile our system to figure out how it actually performs at each discrimination threshold. All interactions with likelihoods below this threshold are eliminated as false positives. Given these measurements of how well D3 performs at each threshold in the knowledge base, we can estimate what level of error should be expected when we compare D3 to other public or commercial pharmacovigilance systems. The evaluation process for the D3 system consists of three main steps: (1) testing DDIs contained within the D3 knowledge base, (2) testing DDIs that are new to the system, and (3) testing induction and inhibition of metabolic enzyme mechanisms of interaction as an example of how D3 is capable of finding the mechanism of interaction.

We first evaluate the D3 query-based model against 15 DDI resources and compute its recall rate. Then we examine the overall performance of the system by graphing the Detection Error Tradeoff (DET). After that, we compare the performance of D3 to five different free, publicly available pharmacovigilance systems. Then, the system is tested against a trusted, clinically utilized commercial

pharmacovigilance system (Micromedex). Finally, we test D3's predictions for two mechanisms of interaction related to the metabolic process.

VII.2.a. Evaluation of the D3 Query Model

The query-based model is designed to search for DDIs within 15 knowledge resources and to describe potential mechanisms for interaction across 9 different mechanistic types. To evaluate the query-based model, we calculate its recall with respect to the 15 different DDI resources used in this thesis. Recall is a helpful metric here **(Equation 7.1)** because it allows us to understand how well existing sources are covered by our system.

$$recall = \frac{relevantDDIs \cap retrievedDDIs}{relevantDDIs}$$

For each resource, we determine how many of its proven DDIs were inferred by our 9 inferences. We then determine the recall. We also determine the overall recall of the union of all existing sources to be 68%, which allows evaluation of the power of the system. For the recall of each source individually, see **Table 7.1.**

Table 7.1 Coverage of D3 for the extraction of explicit DDIs from 15 DDI resources

DDI resource	Source type	Number of DDIs	Recall (%)
Crediblemeds.org	Clinical	82	100
PK-Corpus	Text	165	93.9
DrugBank	Bioinformatics	11840	74.1

The Drug Interaction Knowledge Base	Bioinformatics	560	99.8
NDF-RT	Clinical	9883	49.9
ClinicalTrials.gov	Clinical	3645	36.7
KEGG DDI	Bioinformatics	26328	64.5
DDI-Corpus-2011	Text	569	88.2
DDI-Corpus-2013	Text	1282	83.2
National Library of Medicine cardiovascular Corpus	Text	246	91.4
ONC-High Priority	Clinical	1150	84.9
ONC-Non-interruptive	Clinical	2079	74.8
OSCAR EMR	Clinical	7753	55.1
SemMedDB from PubMed	Bioinformatics	3536	62.2
Twosides	Bioinformatics	63333	79.0
Overall	**Clinical, Bioinformatics & Text**	**114393**	**68.0**

To put these results into perspective, we must emphasize that these DDI resources are highly variable and many focus on reporting different types of interactions for different reasons. Thus, we examine the similarity between each of these 15 sources by computing their Jaccard Index (**Appendix C**). The average Jaccard Index was 1.7%, indicating a low level of overlap and a high level of diversity in reported interactions. The low average Jaccard Index and high overall recall indicate that our D3 query-based model is able to successfully integrate diverse knowledge bases to reliably predict a wide variety of reported DDIs.

VII.2.b Evaluation of the D3 Probabilistic Model

In this section we introduce a set of test cases and evaluate our work in a naturalistic environment, where D3 has no information about samples of DDIs. We identify the D3 true positive rate, compare D3's performance to that of free and commercial pharmacovigilance systems, and finally, discuss our results.

VII.2.b.1 Overall Performance of D3

Before evaluating our system, it is important to show how the system performs (how high its true positive rate is when compared to its true negative rate). Therefore, we use a Detection Error Tradeoff (DET) curve to identify the overall performance of the D3 system. DET is a useful metric here, because our system can be seen as a binary classification system (either identifying or not identifying a potential interaction). **Figure 7.1** shows the DET curves for the D3 system.

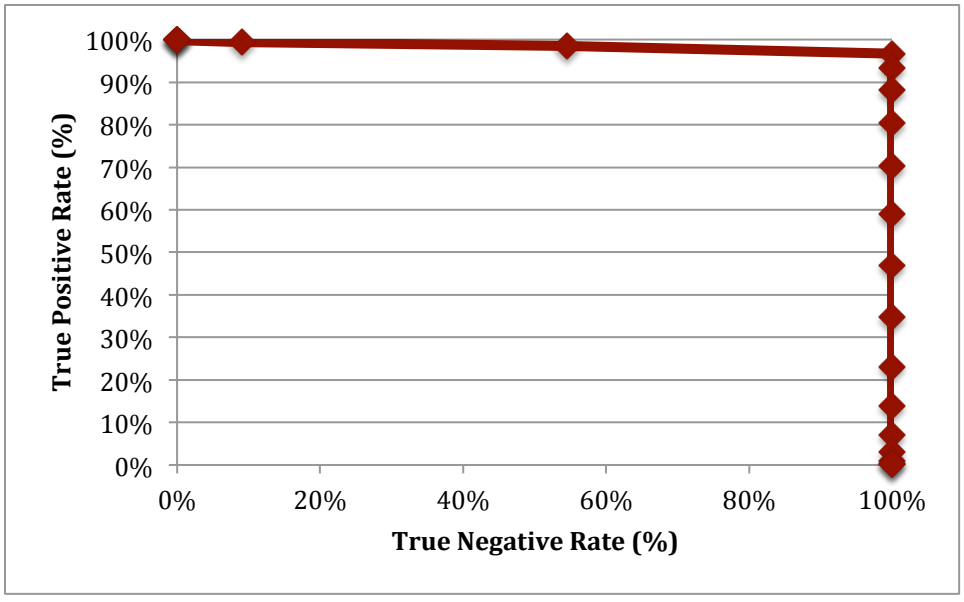


Figure 7.1: DET for D3 system

The **Figure 7.1** shows promising performance within its knowledge sources, as at 96% TPR, D3 is able to catch all true negative DDIs. It is important to point out that the sample's number of true negative interactions is only 11 DDIs, while the sample of TPR contains 21,897 DDIs. This small number of true negative samples is due to the non-existence of any gold-standard DDI resources that can be examined to ensure that two drugs do not interact.

VII.2.b.2 Test Case 1: Public Pharmacovigilance Systems Evaluation

Public pharmacovigilance systems refer to systems that report mostly non-clinical DDIs (i.e. usually theoretical) utilizing readily available data. To measure the D3's performance at different thresholds, we compare D3 to five different public pharmacovigilance systems: Drugbank, Keeg, The Drug Interaction Knowledge Base (DIKB), SemMedDB from PubMed and Twosides. This step is a performance comparison between our system and other systems. Similar to **Chapter VI (VI.4 Building the D3 Pharmacovigilance Resource)**, for each pharmacovigilance system we retrieve all of its DDIs and then remove the overlap with the D3's list of clinically relevant DDIs. This results in a unique DDI list for each resource. Then, for each list, we recall **Chapter V- (V.2.2 Phase 2: D3 Inferential Probabilistic Model)** to compute the likelihood for each DDI in the list. We repeat this same processes for all resources. **Figure 7.2** shows the benchmark graph (baseline) with curves for the D3 system and other public pharmacovigilance systems, where the x-

axis represents different threshold values and the y-axis show the true positive rate as a percentage.

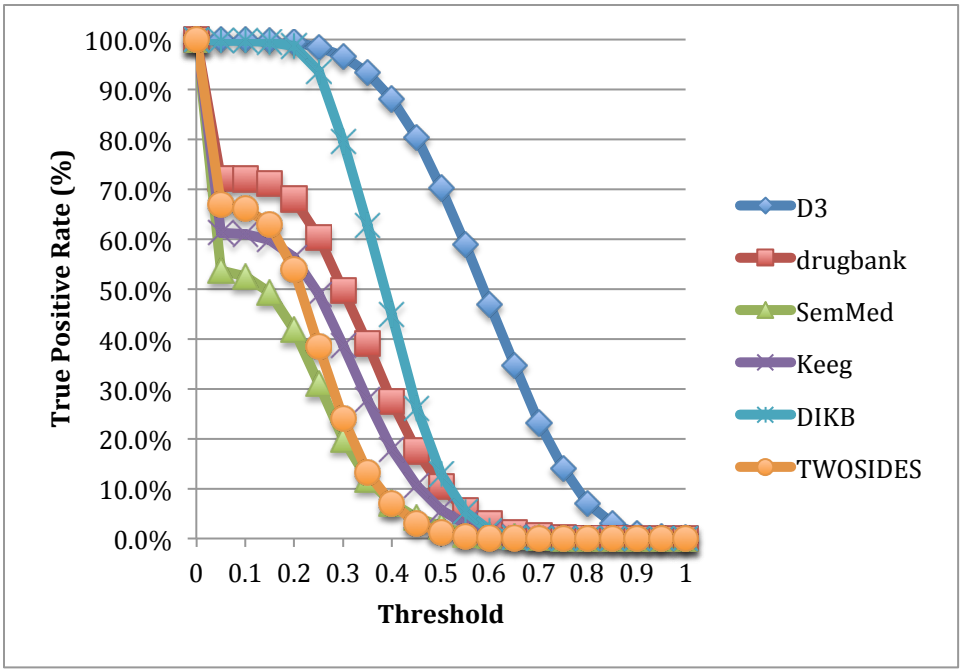


Figure 7.2: Comparison of D3 to five pharmacovigilance systems

The benchmark graph, **Figure 7.2,** shows the kind of performance we would expect from our system, which is based on a proven set of DDIs from clinical trials and medical literature search, per threshold. We see that a threshold of 0.3 (i.e. true positive rate = 95%) is the ideal value for a reliable DDI resource; if we select 0.3 as a threshold, we see from the graph that none of the public systems meet the threshold for reliability. The best competitor we find at the 0.3 threshold is the DIKB system, which still scores less than 80%. Additionally, Drugbank demonstrates a true positive rate of less than 50% at the 0.3 threshold while the other public systems perform even more poorly. Therefore, apart from DIKB, other

public pharmacovigilance resources realistically perform at only half the rate of a reliable system at the 0.3 threshold. As the discrimination threshold is increased, the performance of these public systems drops even more dramatically; **Table 7.2** shows the level of reliability of each system with a 0.5 discrimination threshold (70% true positive rate in the D3 system).

Table 7.2: Number of false positives found by D3 in current DDI resources

Pharmacovigilance system	Validated by D3 based on a 0.5 threshold	False positive percentage
Drugbank	10%	90%
KEGG	5%	95%
SemMedDB from PubMed	3%	97%
Twosides	2%	98%

While the reliability of the D3 system falls to approximately 70% at the 0.5 discrimination threshold (30% false positive rate), the public pharmacovigilance systems fare much worse, with Drugbank, the best performing of these systems, only correctly reporting 10% of clinically relevant DDIs. Therefore, it is clear that most the examined public pharmacovigilance systems do not accurately report clinically significant DDIs.

VII.2.b.3 Commercial Pharmacovigilance Systems Evaluation

Commercial pharmacovigilance systems attempt to report only clinically relevant DDIs based on clinical studies and evidence within the medical literature, and the data they use are not publicly available. In this evaluation step, we profile D3 to compare its performance at each threshold compared to commercial pharmacovigilance systems. Mainly, we consider Micromedex, a highly reliable DDI system that is widely utilized in clinical practice, as a gold standard for comparison. The question is whether D3 can perform well on additional samples of known DDIs when it has no preexisting information on a specific DDI. In order to answer this complex question, we test four different samples against Micromedex: (1) an interaction profile for a well-known drug, (2) a set of severe DDIs, (3) a combined list of all DDIs for 5 well-known drugs, and (4) a comparison of D3 to both public and commercial systems. Our hypothesis is that the D3 system will perform similarly to Micromedex in reporting clinically relevant DDIs.

Test Case 2: Comparison of D3 to an Interaction Profile: Atorvastatin as a Case Study

Atorvastatin is used therapeutically to lower blood cholesterol by inhibiting endogenous cholesterol synthesis. The medication is a substrate of both CYP3A4 and P-glycoprotein, meaning that inhibition or induction of these metabolism or transport systems may lead to an increase in adverse drug events, such as myopathy and rhabdomyolysis, or a decrease in therapeutic efficacy (“Drug Safety Labeling Changes > Lipitor (atorvastatin calcium) Tablets,” n.d.). Therefore,

atorvastatin has the potential for clinically important DDIs based on its pharmacokinetic and pharmacodynamic characteristics. Micromedex reports 89 interactions with atorvastatin. We retrieve all interactions and remove the overlap with the D3 clinical list. Then we recall **Chapter V's V.2.2 Phase 2: D3 Inferential Probabilistic Model** to compute the likelihood for each DDI in the list. **Figure 7.3** shows the benchmark graph for atorvastatin DDIs with two curves: one for the D3 system (baseline) and the other for Micromedex.

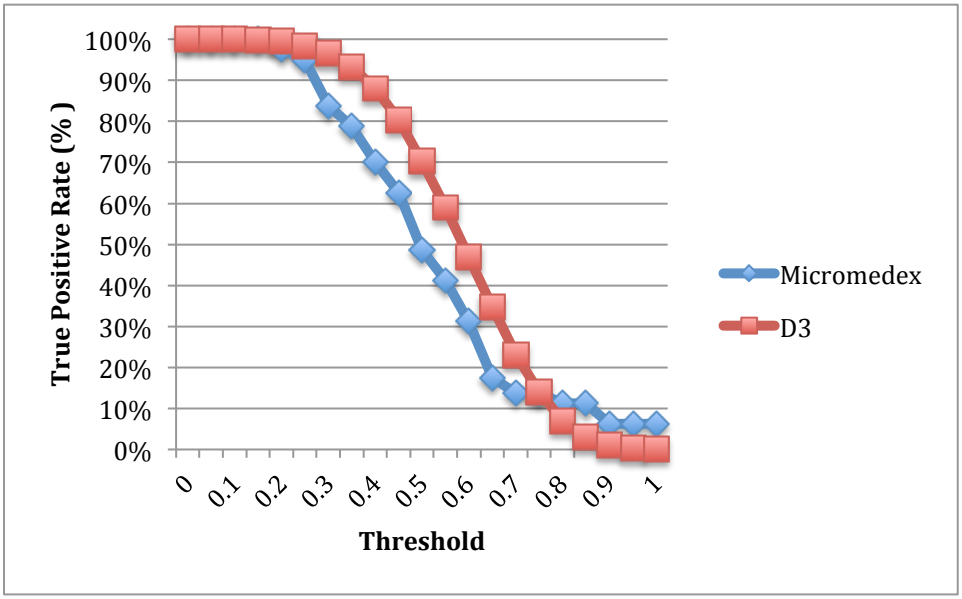


Figure 7.3: Comparison of D3 to atorvastatin interaction list from Micromedex

As demonstrated in **Figure 7.3,** the two curves are very close to each other, indicating that D3 has a very high recall rate with Micromedex in capturing the majority of atorvastatin DDIs, even at high discrimination thresholds. For example, with a low threshold of 0.15 (true positive rate = 99%), D3 captures all listed atorvastatin interactions. Moreover, even with a high threshold of 0.5 (true positive

rate = 70%), D3 performs well, evidenced by a 50% true positive rate. In this instance, D3's recall is impressive, compared to the well-regarded commercial resource Micromedex.

Test Case 3: Comparison of D3 to Micromedex in Evaluating Severe Interactions for Five Drugs

The main aim of this test case is to show how D3 can assist in determining the clinical relevance of DDIs. Five drugs (warfarin, phenytoin, digoxin, simvastatin and ketoconazole) were recommended for testing by a pharmacist at Purdue University based on their high potential of having clinically relevant DDIs. For each drug, we retrieve only the severe interactions from Micromedex (those deemed to be contraindications or major interactions) and remove the overlap with the D3 list. This resulting sample is 125 severe DDIs involving the five previously mentioned drugs. We then compute the likelihood for each DDI in the severity list. **Figure 7.4** shows the benchmark graph for the severe interactions.

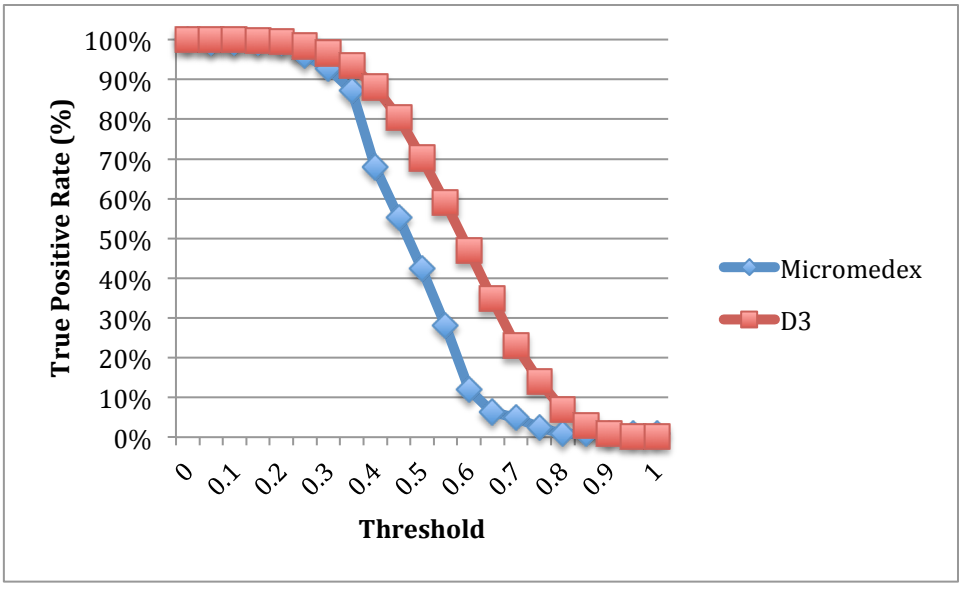


Figure 7.4: Comparison of D3 to severe DDIs for 5 drugs

As shown in **Figure 7.4**, D3 has high recall rate of DDIs for its knowledge base, as the system only misses one DDI (out of a possible 125) at the 0.15 discrimination threshold. At a more conservative 0.5 threshold, D3 still recalls 43% of the interactions for its knowledge base. While this may seem like there are many interactions being ignored, it is important to consider that the false-positive rate, or misidentified interactions from other sources, will be extremely-low at this conservative threshold. Because D3 is oriented towards culling clinically relevant interactions from pharmacovigilance systems, it would be expected that approximately 43% of the clinically relevant interactions would remain for those systems as well. This is particularly effective for choosing interactions in which to invest research and clinical testing. At a more liberal threshold of 0.3 as a warning system, D3 still recalls 93% of the proven interactions taken from Micromedex. A similar expectation could be made for other pharmacovigilance systems. It should be kept in mind that the interactions tested from Micromedex were not part of the D3 knowledge base and therefore acted as an independent test in the truest sense.

Test Case 4: Comparison of D3 to Combined List for Five Drugs

Unlike **Test Case 3,** which focused only on the severe DDIs for the five drugs, in this test case we investigate a larger and more complete list of all DDIs for the five previously mentioned drugs. From Micromedex, we retrieve 5 lists of interactions for warfarin, phenytoin, digoxin, simvastatin and ketoconazole. We then combine these lists into one unique list and remove the overlap with the D3

DDI list, obtaining 600 DDIs in all. The likelihood of each DDI is computed using the **Chapter V- V.2.2 Phase 2: D3 Inferential Probabilistic Model. Figure 7.5** shows the benchmark graph for D3 with Micromedex.

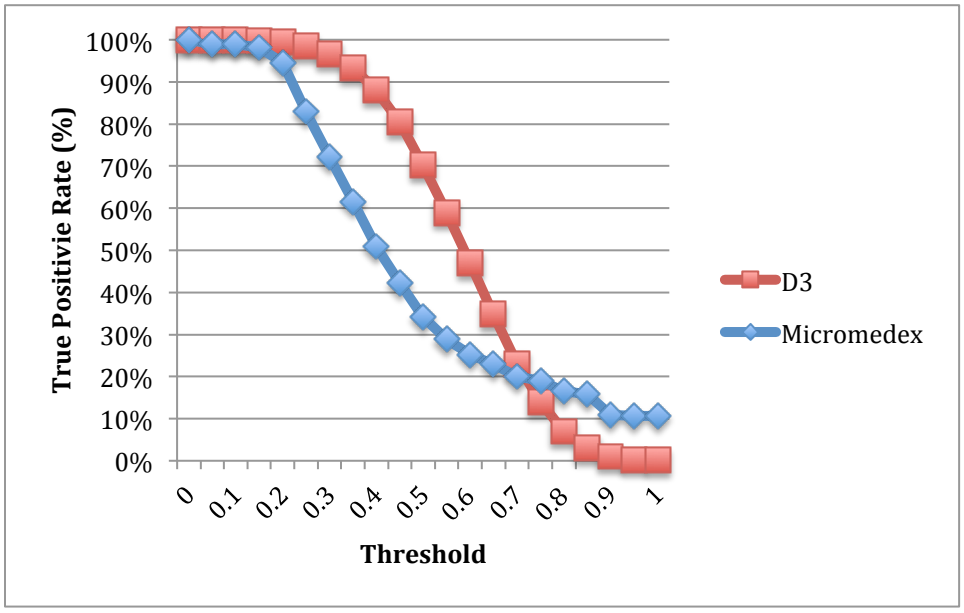


Figure 7.5: Comparison of D3 to a large sample from Micromedex

Similar to the previous benchmark graphs, **Figure 7.5** illustrates the excellent performance of D3 across the examined discrimination thresholds. At the previously mentioned 0.3 threshold, D3 identifies over 72% of the DDIs identified by Micromedex. As the threshold increases, D3's true positive rate declines in a manner similar to that seen with Micromedex.

Test Case 5: Comparison of D3 to Public and Commercial Pharmacovigilance Systems

The main aim of this test case is to combine **Test Case 1: Public** and **Test Case 4: Commercial** to assess the comparable performance of D3 for public and

commercial systems at each threshold. Because D3 was constructed from data drawn from public systems, it is necessary to independently validate its performance. Such validation would mean that D3 should perform comparably to commercial systems.

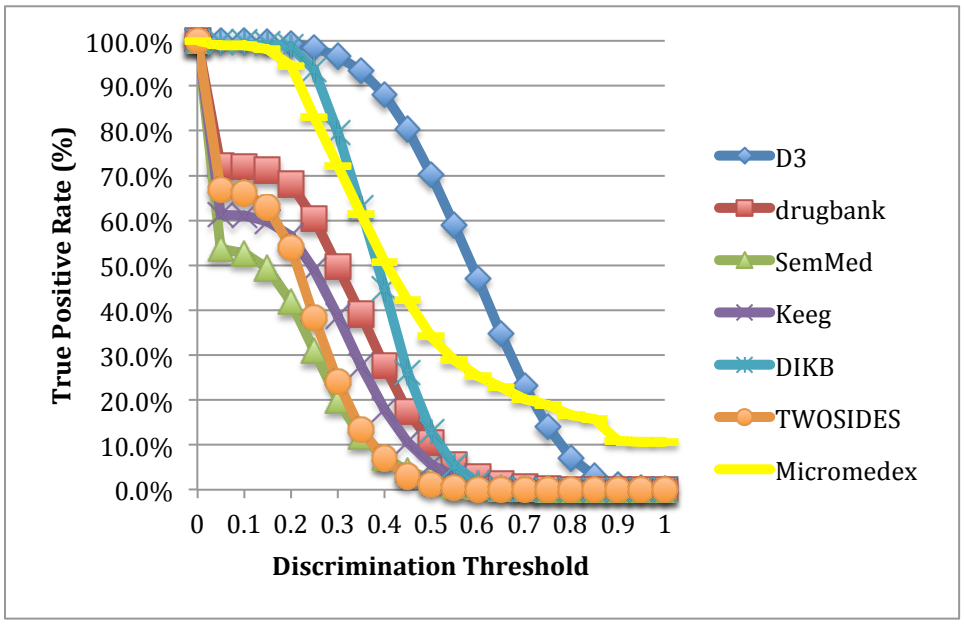


Figure 7.6: Comparison of D3 to public and commercial pharmacovigilance systems

The benchmark graph, **Figure 7.6**, shows that D3 demonstrates a similar characteristic for its public knowledge sources and Micromedex, a respected, commercial knowledge source. In contrast, the application of D3 to several existing pharmacovigilance systems showed the apparent, drastic over-reporting of the majority of these systems even for a conservative (low) threshold. Only DIKB performed with a comparable characteristic to D3 and Micromedex. In fact, DIKB reports only clinically relevant, evidence-based interactions. At a low threshold of

0.15 (true positive rate (recall rate) = 99% for D3 public sources), D3 corroborates 98% of the drug interactions identified by Micromedex. Apart from DIKB, at a 0.25 threshold, the other systems demonstrate at least a 30% over-reporting rate, indicating that many of the DDIs they report are clinically unlikely. On the other hand, with a high threshold 0.5 (true positive rate (recall rate) = 70% for D3 public sources), D3 corroborates 35% of Micromedex-identified DDIs whereas D3 corroborates 13% or less of the DDIs reported by all of the studied public pharmacovigilance systems. The marked difference in the characteristic of D3 for its public sources and Micromedex compared to the more liberal public sources shows the discrimination-power of D3 and the over-reporting of these public sources.

VII.2.c Evaluation of the Mechanistic Inference Method

Because there is no gold-standard source for drug interaction mechanistic information, in this section we design the evaluation methodology to consider two different metabolic interaction mechanisms: inhibition and induction. To do this, we manually construct two different test cases, one for metabolism inhibition and another for the metabolism induction. For both cases we choose classical examples of a strong CYP enzyme inhibitor ketoconazole and a strong CYP enzyme inducer rifampin (Böhmer, Drollmann, Gleiter, & Nave, 2008; Rae, Johnson, Lippman, & Flockhart, 2001). After that, we check Micromedex for both drugs' list of interactions and retrieve only drugs that interact because of metabolism reaction. The result is 20 induction interactions and 21 inhibition interactions. Then, we take each list and run the **VI.3 Utilizing Data Resources to Enhance Descriptions**

of DDI Mechanisms method from Chapter VI**. Figure 7.7** shows the induction inference while **Figure 7.8** shows the inhibition.

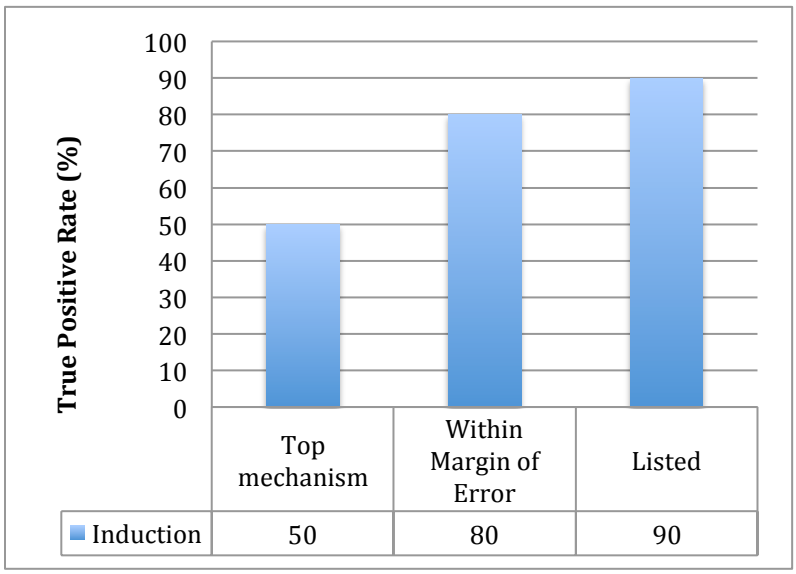


Figure 7.7: Performance of D3 in annotating DDIs with induction

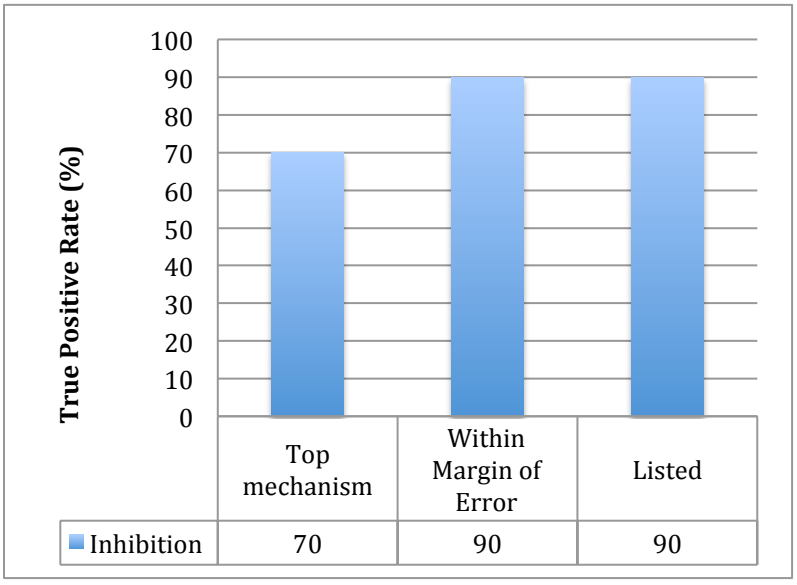


Figure 7.8: Performance of D3 in annotating DDIs with inhibition mechanism

Figures 7.8 and 7.9 show that D3 is able to predict the right mechanism of interaction with an 85% success rate. A corollary to this is that we can also use the same system to provide putative explanations for reported interactions, the majority of which are reported without an underlying mechanism.

VII.3 Conclusion

In this chapter, we integrated all the methods, algorithms and systems that were proposed in this thesis. This integration allowed us to evaluate the effectiveness of the D3 system in identifying drug interactions by way of comparison with other public and commercial pharmacovigilance services. The reported results from five test cases showed that our D3 system was able to provide novel and unique pharmacovigilance services by demonstrating: (1) increased performance (a higher true positive rate) when compared to publicly available pharmacovigilance systems when analyzing a complete list of unique drug interactions contained within each system; (2) performance similar to Micromedex when investigating interactions associated with the widely-used drug, atorvastatin; (3) a true positive recall rate of approximately 93% when identifying severe, clinically significant interactions involving five commonly used medications as identified by Micromedex; (4) a true positive recall rate of 72% relative to Micromedex when identifying all interactions with the five medications from the previous test case; and, finally, (5) superior performance relative to all public pharmacovigilance systems in recalling all interactions identified with the same five medications by Micromedex.

The strong performance of D3 in these test cases is attributed to specific advantages in the capacity of our system relative to other pharmacovigilance systems. In particular, the D3 system helped to assist in determining the clinical importance of DDIs and to provide the most likely mechanisms of interaction. These abilities are due to the capacity of D3 to consider not only the biomedical similarities between medications, but also to identify interaction profiles between DDIs that are used to improve the precision of recommendations. The evaluation of D3 against different pharmacovigilance systems using different interaction lists shows that our proposed system is generic enough to be applied in different applications. It also demonstrates the diverse drug dimensions that are utilized in our system to provide pharmacovigilance services. In addition, the proposed system can be used to evaluate current pharmacovigilance systems with regards to assessing the clinical relevance of their reported DDIs. Finally, D3 is also designed to add mechanistic information to enhance both clinical and non-clinical pharmacovigilance systems.

CHAPTER VIII

OVERALL CONCLUSIONS

Aiming to address the limitations in current pharmacovigilance systems, this thesis focuses on the process of aggregating information from drug-drug interaction (DDI) knowledge sources to perform inference-based predictions for statistical identification, validation, and classification of DDIs. The novelty of this work lies in the systematic approach in which DDI information is integrated into a comprehensive knowledge base and employed to build integrated data model from existing biomedical knowledge sources, automate validation of clinical relevance, automate identification of interaction mechanism(s), and evaluate known pharmacovigilance systems. In this chapter, we first summarize the demonstrated capacity of D3 as evidenced in multiple test cases throughout this thesis. Next, we discuss the achieved contributions to the field relative to other approaches present in the literature. Finally, we discuss the limitations of the D3 system and consider future work.

VIII.1 Summary and Achieved Contributions

VIII.1.a Constructing the Drug-Drug Interaction Discovery and Demystification Knowledge Base

One of the main challenges in current pharmacovigilance systems is the number of disparate, disconnected sources of information currently available for DDI discovery. Integrating these sources into a comprehensive resource that can aid in accurate and early discovery of DDIs continues to challenge researchers. Many pharmacovigilance studies do not address this challenge as they focus only on a single pharmacological aspect of potential DDIs (such as a single mechanism of interaction), but do not account for the multitude of factors that contribute to DDIs. A complete analysis, therefore, requires integration of drug information from multiple sources. A number of studies have attempted to develop sophisticated systems that are able to model and integrate DDI information (R. D. Boyce et al., 2007; Brochhausen et al., 2014). However, such studies face many challenges in accurately discovering DDIs, as they do not address an essential aspect of DDI, that interactions may occur as a result of two or more interactive mechanisms. Another common limitation in these studies is that they do not account for the many different types of interactive mechanisms. Therefore, these systems suffer from incompleteness and low performance when providing pharmacovigilance services because they do not adequately recognize the many different mechanisms of interaction that contribute to DDIs.

To address these limitations, this thesis proposes a novel knowledge base to model drug information that encompasses multiple interaction mechanisms. This knowledge base characterizes DDIs across nine mechanisms of interaction that we propose for improving the effectiveness of DDI study. Unlike other approaches in the literature, this knowledge base is able to model DDIs at multiple interaction levels. In particular, the contributions of our work can be summarized as follows:

- We develop a novel knowledge base that models DDI information across nine interaction levels: pharmacokinetic (interactions involving metabolism, transporters, and protein binding), pharmacodynamic (additive and competitive interactions), pharmacogenetic (single nucleotide polymorphisms that influence drug exposure), and multi-pathway (interactions involving more than one of the aforementioned pathways) interaction levels. This knowledge base takes advantage of the available biomedical knowledge resources to increase the accuracy of DDI identification.
- DDI information is represented as semantic networks, which are richer and more precise than traditional integration approaches. Such representation allows us to better understand DDIs as well as to exploit our improved semantic knowledge to provide better pharmacovigilance services.
- The proposed knowledge base is generic and flexible so it can be used in multiple domains.

VIII.1.b Building the D3 System

A primary shortcoming of most current pharmacovigilance systems is the inability to integrate drug information from multiple sources to identify and comprehensively characterize DDIs. Ultimately, this limitation impacts clinical practice negatively through suboptimal recommendations provided by pharmacovigilance-supported decision assistance tools. This shortcoming occurs for two important reasons: (1) the single-focus approach of current pharmacovigilance systems and (2) the overall lack of explanation regarding DDI mechanisms. This thesis seeks to overcome these limitations by proposing a new Semantic Web-based system to discover and explain DDIs based on information from 29 sources. In the D3 pharmacovigilance system, the hidden semantic knowledge and relationships between DDIs are discovered using two novel inferential models to infer DDIs and their mechanisms of interaction. The discovered knowledge is then utilized to provide a pharmacovigilance service that is capable of reasoning and can adapt itself to the different interaction types without the need of human intervention. The contributions of this system can be summarized as follows:

- Unlike other works that are limited to specific types of interaction, we develop a model to identify a DDI across nine interaction mechanisms. Such a model offers the benefits of enhancing the processes of discovering, inferring and gaining a better understanding of uncommon mechanisms of interaction, which results in a better pharmacovigilance system.

- We develop a novel query-based model to support DDI identification using information about different drugs from 15 DDI resources. In comparison to current systems described in the literature, our developed model is able to take advantage of richer DDI information to provide more accurate validation.
- We propose a mechanism to discover and exploit the hidden semantic knowledge from drug interaction profiles. Unlike many works in the literature which are incapable of inferring and exploiting hidden semantic information, our mechanism infers the semantic information between DDIs in biomedical resources and employs it to drive a better understanding of drug interactions, which can potentially improve the pharmacovigilance services provided to clinicians.
- We develop a novel probabilistic inference model to assist in determining the clinical relevance of DDIs. To our knowledge, none of the current systems in the literature have used probabilistic modeling in this way.
- We introduce a novel model for ordering the mechanisms of DDIs to provide their most likely causes. This approach is distinct from many other systems in the literature, which provide information for only one type of interaction mechanism.

VIII.1.c Exploiting the D3 System

As DDIs may have different and complex pathways of interaction, it is becoming increasingly important to consider all possible data when studying them. Most current pharmacovigilance systems collect only metabolic-based mechanistic information without considering any other mechanistic data. Such systems neglect the fact that DDIs may occur due to different mechanisms of interaction. The continued growth of drug data, such as information detailing therapeutic, chemical, genomic and phenotypic considerations in drug therapy, also increases the ability to study DDIs at an intricate level. Therefore, considering these types of data is becoming increasingly important when performing comprehensive pharmacovigilance services.

Attempts to incorporate this data have been made by a number of studies (Cheng & Zhao, 2014, p. –; Duke et al., 2012; Vilar et al., 2015). However, the main problem in these works is a lack of flexibility in modeling, integrating and exploiting DDI information for use in pharmacovigilance systems in a generic fashion. For example, some studies lack extensibility as they are limited to specific techniques (Percha & Altman, 2013). Another important limitation in predicting DDIs is assessing the clinical relevance so as to provide pharmacovigilance services that benefit clinical practice in an effective way. For instance, one study voiced the need for a standard characterization of clinical relevance due to the vast number of theoretical, clinically unimportant DDIs reported by many systems (Conde-Estévez et al., 2015). Overall, there is general lack of support for elucidating the mechanistic

reasons behind DDIs.

In this thesis, we present a novel pharmacovigilance system to address these limitations. We define query and probabilistic models to discover, learn, adapt and exploit DDIs. Unlike other works in the literature, this system is generic enough to consider diverse DDI dimensions, and to be applied in diverse domains of application. Another distinctive feature of this system is its ability to model and incorporate 15 DDI- and 14 biomedical sources to provide effective pharmacovigilance services. We incorporate these sources because not all DDI information is available in one single source, but, rather, information is scattered in different places and formats. Therefore, we incorporate those sources to provide not just clinically relevant DDIs but also to provide the most likely mechanisms of interaction for each DDI. To the best of our knowledge, this is the first work that attempts to model, exploit and incorporate all of these sources in a generic fashion to support the identification of highly clinically relevant DDIs associated with their most likely mechanisms of interaction. The contributions of this work can be summarized as follows:

- Many pharmacovigilance systems consider just specific types of interaction. In contrast, we proposed a method to study interaction across nine mechanistic levels. In this proposal, DDI profiles, which model all drug biomedical and interaction information, have a clear mechanistic distinction.
- Many pharmacovigilance systems do not take into account all of the details present within drug information from existing sources for proven DDIs. In

this thesis, the D3 model is built to incorporate and integrate information from resources about proven DDIs.

- Many pharmacovigilance systems are limited to specific pathways of interaction (Preissner et al., 2010). In contrast, the proposed D3 system considers diverse mechanistic information, and is also able to adopt new specific mechanistic dimensions based on the information provided.
- As DDI resources may identify different interaction lists, we propose a novel mechanism to combine and consider the most clinically relevant DDI in each resource. Unlike systems that do not assess the clinical importance of identified DDIs, D3 uses reliable DDIs that are extracted from clinical and literature-based resources to identify clinically significant DDIs (Tatonetti, Ye, et al., 2012b; R. Zhang et al., 2014b). In this case, we can evaluate existing DDI systems to identify their non-clinical DDIs (false positive rate).
- Many pharmacovigilance systems do not take into account rich mechanistic information with regards to DDIs. In contrast, in this thesis, we possess the capability to discover and exploit hidden mechanistic information from multiple diverse text formats, data schemas, and controlled vocabularies, and we have the ability to interlink this information with diverse datasets extracted from the biomedical domain to improve the quality of pharmacovigilance recommendations.
- Two major and distinctive contributions in this thesis are: (1) developing novel techniques to evaluate current non-clinical DDI systems to determine

the clinical relevance of proposed DDIs, and (2) utilizing data resources to enhance descriptions of DDI mechanisms.

VIII.1.d Development of a Comprehensive, Mechanism-based DDI Resource

Despite the fact that DDIs constitute a major cause of adverse drug events (ADEs), there remains no complete and comprehensive DDI resource. Moreover, there is also no single resource that is widely utilized to describe DDI mechanisms. These two challenges have created a vast need for a new resource in which only clinically relevant DDIs are considered along with their mechanisms of interaction.

In this thesis, we present a novel pharmacovigilance DDI resource that contains clinically significant DDIs associated with their mechanisms of interaction. We gather and incorporate DDI information from 15 publically available resources. We then normalize the information, filter it based on its clinical impact, and represent predictions in a formal and semantic representation. The contributions of this work in this area can be summarized as follows:

- We enhance current DDI resources by providing the most likely mechanisms of interaction.
- We build the D3 resource by extracting the most highly clinically impactful DDIs from five non-clinical DDI resources.
- We normalize data with UMLS and semantic representation.
- We record the original source of DDI information.

- We consider the external utility when developing our DDI resource to provide effective information for use in a wide range of studies.

VIII.1.e Evaluation of the D3 System

The process of evaluating pharmacovigilance systems is known to be difficult and expensive (Abarca et al., 2006). This truth is attributed to a number of factors, including the enormous number of drug interaction pairs through which each pharmacovigilance system must be evaluated to gain a true assessment of its DDI identification ability. Moreover, some systems are specialized to report only certain type of DDIs (e.g., metabolic DDIs), while other pharmacovigilance systems focus on more than one DDI interaction type. Therefore, each system requires different evaluation strategies and metrics. In addition, mechanistic information on drug interactions is rarely found in pharmacovigilance systems and the available information is often limited to one mechanism of interaction. Other mechanistic information is generally scattered across disparate literature sources and requires tedious manual work for effective extraction.

In this thesis, we implement two strategies to evaluate our work. First, we evaluate each model in our proposed system in isolation in order to probe different test cases and conditions that can affect its inferences. Then, we evaluate the complete system with both models to gain an understanding of the overall performance of our system in comparison to other works in the literature.

For the isolated evaluations, we use different methodologies, including

literature search, Fisher's exact test, and precision-recall evaluation strategies. For the literature, diverse scientific journals and articles are examined to validate different results and experiments to find the correct answers. For the Fisher's exact test and the precision-recall, we conduct statistical analyses to evaluate existing DDI resources, which allows us to gain a better understanding of how to design an effective pharmacovigilance system. Such experiments allow us to collect more reliable results as well as draw direct comparisons between different systems in the literature. The results of all the evaluations are encouraging, as they provide positive results, which support the validity of the proposed system. However, we found that there is a need to evaluate all of these models when they are integrated in one system in a real clinical environment in order to draw clear conclusions on the performance of our proposed system in comparison to other works in the literature. For this purpose, we developed three case studies that integrate all the proposed methods and models: (1) testing DDIs contained within the D3 knowledge base, (2) testing DDIs that are new to the system, and (3) testing induction and inhibition of metabolic enzyme mechanisms of interaction. This evaluation methodology allowed us to analyze the performance of different systems. The results obtained from this evaluation confirmed our hypotheses in this thesis, that our D3 system of identification, validation, and classification can provide significant benefits in improving pharmacovigilance services. The D3 system demonstrates the ability to identify only those DDIs that are highly clinically relevant. We also tested an external set of DDIs, which showed that clinicians who use our D3 system can

validate DDIs with a high level of accuracy. The contributions of these evaluations can be summarized as follows:

- In comparison to other works in the literature (DIKB, Twosides, Drugbank, SemMed, and Keeg), we provide more insights and a more accurate list of DDIs.
- The isolated evaluations of all the aspects of our system enable us to test different settings and inferences to find the effectiveness of each model, as well as to provide a direct comparison with other similar systems proposed in the literature.
- In comparison of our three test cases, D3 performed similarly to Micromedex, a respected commercial knowledge source, averaging an 92% detection rate for Micromedex-identified interactions.
- In comparing D3 to five public pharmacovigilance systems based on Micromedex-identified interactions, our system demonstrates a performance that is superior to all public pharmacovigilance systems.
- The diverse evaluations that are presented in this thesis show that our system is generic enough to be applied in a diverse domain of applications.

VIII.2 Discussion and Future Work

In this thesis, we have proposed different models, algorithms and techniques to model DDIs and their mechanisms in order to provide improved pharmacovigilance services. These models, algorithms and techniques cover a wide

range of areas including learning and adapting DDI profiles, developing a multi-pathway discovery model, creating a process of semantic inference and exploitation, and, finally, characterization of DDIs in order to generate recommendations. Although a wide range of problems have been addressed in this thesis, further improvements of the proposed system as well as incorporation of new information sources will be pursued. Now, we discuss some of the limitations of our work and provide possible solutions that may further improve our system.

Our first limitation is that the effectiveness of the techniques that we have proposed to infer and exploit semantic information from DDI profiles in this thesis depends heavily on the richness and the quality of the developed D3 knowledge base. Some DDI information in this thesis is transformed by different projects, such as Bio2rdf (Callahan, Cruz-Toledo, & Dumontier, 2013). That is, we have relied on the transformation quality for some original sources from the Bio2RDF website, which has been shown to have a high quality of transformation. However, in the future we plan to perform our own conversion of original sources to avoid any concerns about the quality and completeness the of Bio2rdf transformation process.

Another limitation of this thesis is that the system infers potential clinically relevant interactions, but it is not able to predict the seriousness of those interactions. In fact, determining the seriousness of any interaction based on mechanistic information to determine how the identified interactions may change the concentrations of the interacting drugs is a challenging task due to limited experimental data. Therefore, all D3 findings should be considered speculative until

clinically verified.

Another source of weakness in this thesis is associated with the computations of biomedical similarities between DDIs. When building the biomedical features similarity matrix for drug comparison, currently D3 is designed to treat all nine biomedical features similarly by assigning an equal weight. This was done so as to avoid arbitrary bias across the field of drugs being compared. Equal weighting implies that all biomedical features have an equal level of importance when determining similarity. In reality, this is not necessarily true. The problem is the statistical failure of any particular feature can be compensated for by a less-relevant similarity in the interaction being considered. Therefore, this may create a bias in many cases towards features whose similarity is far less relevant than others. This can be addressed by correlating individual biomedical feature similarity to actual rates of DDI occurrence for various classes of DDIs. Here we propose one way to cope with this limitation using logistic regression to pre-assign weight values to biomedical features. The main goal of the D3 biomedical similarity matrix is to statistically identify how two drugs are similar by comparing 9 different biomedical features. Moreover, the assumption is that if two pairs of drugs are similar to a degree (using predefined thresholds) and one of those pairs has been proven to interact, there is clinical relevance for a reaction predication within the other pair. By using logistic regression we can identify what essential biomedical features to consider for the comparison between two drugs. Logistic regression can be used here because we have a dichotomous (binary) variable (if a biomedical

feature is similar enough). In case of a dichotomous variable, our aim is to understand the relationship between the explanatory variables (presence or absence of a biomedical feature) and the dependent variables (proven DDI occurrence). However, not all interactions may occur simply due to the presence of a particular similarity, making an overall similarity-dissimilarity-irrelevance profile valuable for understanding classes of interactions. The explanatory variables that D3 uses to determine the similarity score of a clinically relevant DDI are: (1) indications, (2) side effects, (3) SNPs (4) mechanisms of action, (5) carriers, (6) enzymes, (7) transporters, (8) targets, and (9) physiologic features. Each of these variables has application in combination with others to particular interactions. Determining trust weights for them could be made more reliable by applying a priority order or filter. This would, in effect, lead to what may be considered understanding "patterns" of similarity. That would allow adjustment of the weights based on closeness to a pattern.

Accomplishing this would require isolating drug interactions that share strong characteristic similarities of features to classify them. Building these samples would allow the prediction of classes and potentially their points of overlap. Regression could be performed on these class samples to determine appropriate levels of similarity based on likelihood of interaction occurrence, which could then be used to provide proper weighting for the feature similarity.

Another weakness of this study is that the system utilizes a patch update to revise the knowledge base. This process is slow and time consuming. In the future, we are planning to perform an incremental update model wherein we only change one component of the knowledge base at a time, rather than updating the whole system at once. Ideally, dynamically accessed information with real-time processing would be preferable to take into account the most recent developments per adequate vetting.

Two main difficulties have been continually documented throughout this thesis: (1) the non-existence of a resource for negative DDIs, which are DDIs that have no documented evidence for interaction on compendia and medical literature contained in MEDLINE(J. Huang et al., 2013) and (2) the incomplete and inconsistent documentation of mechanistic information within the field. As a result, when implementing and evaluating our work, we found it difficult to generate a suitable set of DDIs as a reference for validation. Similarly, there is no comprehensive mechanistic information resource. We overcame these limitations by manually searching the literature to generate DDI lists for reference and to identify sufficient mechanistic information.

Finally, when our system identified a new DDI with a high likelihood of interaction and a high potential for clinically relevant implications, the predicted interaction was still considered a false positive if there was no information in either the literature or compendia to validate the interaction. However, we are planning to integrate different patient health record resources as they often contain evidence for

potential DDIs that are not well characterized in other sources (Pathak et al., 2013). In this way, we can expand our search for information that confirms novel DDI pairs predicted by our system.

BIBLIOGRAPHY

Abarca, J., Colon, L. R., Wang, V. S., Malone, D. C., Murphy, J. E., & Armstrong, E. P. (2006). Evaluation of the performance of drug-drug interaction screening software in community and hospital pharmacies. Journal of Managed Care Pharmacy: JMCP, 12(5), 383–389.

Antoniou, G., & Harmelen, F. van. (2004). Web Ontology Language: OWL. In P. D. S. Staab & P. D. R. Studer (Eds.), Handbook on Ontologies (pp. 67–92). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-24750-0_4

Apache Jena - Reasoners and rule engines: Jena inference support. (n.d.). Retrieved October 13, 2015, from https://jena.apache.org/documentation/inference/

Asberg, A. (2003). Interactions between cyclosporin and lipid-lowering drugs: implications for organ transplant recipients. Drugs, 63(4), 367–378.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics, 25(1), 25–29. http://doi.org/10.1038/75556

At, M. (1989). The UMLS Semantic Network. Proceedings / the ... Annual Symposium on Computer Application [sic] in Medical Care. Symposium on Computer Applications in Medical Care, 503–507.

Awa, K., Satoh, H., Hori, S., & Sawada, Y. (2012). Prediction of time-dependent interaction of aspirin with ibuprofen using a pharmacokinetic/pharmacodynamic model. Journal of Clinical Pharmacy and Therapeutics, 37(4), 469–474. http://doi.org/10.1111/j.1365-2710.2011.01313.x

Ayvaz, S., Horn, J., Hassanzadeh, O., Zhu, Q., Stan, J., Tatonetti, N. P., … Boyce, R. D. (2015). Toward a complete dataset of drug-drug interaction information from publicly available sources. Journal of Biomedical Informatics, 55, 206–217. http://doi.org/10.1016/j.jbi.2015.04.006

Baciewicz, A. M., Chrisman, C. R., Finch, C. K., & Self, T. H. (2013). Update on rifampin, rifabutin, and rifapentine drug interactions. Current Medical Research and Opinion, 29(1), 1–12. http://doi.org/10.1185/03007995.2012.747952

Bates, D. W., Cullen, D. J., Laird, N., Petersen, L. A., Small, S. D., Servi, D., … Hallisey, R. (1995). Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. JAMA, 274(1), 29–34.

Bayer (Aspirin) Patient Information: Side Effects and Drug Images at RxList. (n.d.). Retrieved October 13, 2015, from http://www.rxlist.com/aspirin-drug/patient-images-side-effects.htm

Becker, M. L., Kallewaard, M., Caspers, P. W. J., Visser, L. E., Leufkens, H. G. M., & Stricker, B. H. C. (2007). Hospitalisations and emergency department visits due to drug-drug interactions: a literature review. Pharmacoepidemiology and Drug Safety, 16(6), 641–651. http://doi.org/10.1002/pds.1351

Berners-Lee, T., Fielding, R., & Masinter, L. (2005). Uniform Resource Identifier (URI): Generic Syntax (No. RFC3986). RFC Editor. Retrieved from https://www.rfc-editor.org/info/rfc3986

Berners-Lee, T., & Hendler, J. (2001). Publishing on the semantic web. Nature, 410(6832), 1023–1024. http://doi.org/10.1038/35074206

BioCarta_Pathways. (n.d.). Retrieved October 13, 2015, from http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways

Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Research, 32(Database issue), D267–D270. http://doi.org/10.1093/nar/gkh061

Bodenreider, O., & McCray, A. T. (2003). Exploring semantic groups through visual approaches. Journal of Biomedical Informatics, 36(6), 414–432. http://doi.org/10.1016/j.jbi.2003.11.002

Boguski, M. S., Mandl, K. D., & Sukhatme, V. P. (2009). Drug discovery. Repurposing with a difference. Science (New York, N.Y.), 324(5933), 1394–1395. http://doi.org/10.1126/science.1169920

Böhmer, G. M., Drollmann, A., Gleiter, C. H., & Nave, R. (2008). Effect of coadministered ketoconazole, a strong cytochrome P450 3A4 enzyme inhibitor, on the pharmacokinetics of ciclesonide and its active metabolite desisobutyryl-ciclesonide. Clinical Pharmacokinetics, 47(5), 343–349. http://doi.org/10.2165/00003088-200847050-00005

Bonnabry, P., Sievering, J., Leemann, T., & Dayer, P. (1999). Quantitative drug interactions prediction system (Q-DIPS): a computer-based prediction and management support system for drug metabolism interactions. European Journal of Clinical Pharmacology, 55(5), 341–347.

Böttiger, Y., Laine, K., Andersson, M. L., Korhonen, T., Molin, B., Ovesjö, M.-L., … Eiermann, B. (2009). SFINX-a drug-drug interaction database designed for clinical decision support systems. European Journal of Clinical Pharmacology, 65(6), 627–633. http://doi.org/10.1007/s00228-008-0612-5

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., & Bairoch, A. (2007). UniProtKB/Swiss-Prot. Methods in Molecular Biology (Clifton, N.J.), 406, 89–112.

Boyce, R., Collins, C., Horn, J., & Kalet, I. (2009a). Computing with evidence Part I: A drug-mechanism evidence taxonomy oriented toward confidence assignment.

Journal of Biomedical Informatics, 42(6), 979–989. http://doi.org/10.1016/j.jbi.2009.05.001

Boyce, R., Collins, C., Horn, J., & Kalet, I. (2009b). Computing with evidence part II: an evidential approach to predicting metabolic drug-drug interactions. Journal of Biomedical Informatics, 42(6), 990–1003. http://doi.org/10.1016/j.jbi.2009.05.010

Boyce, R. D., Collins, C., Horn, J., & Kalet, I. (2007). Modeling drug mechanism knowledge using evidence and truth maintenance. IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society, 11(4), 386–397.

Boyce, R., Gardner, G., & Harkema, H. (2012). Using Natural Language Processing to Identify Pharmacokinetic Drug-drug Interactions Described in Drug Package Inserts. In Proceedings of the 2012 Workshop on Biomedical Natural Language Processing (pp. 206–213). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=2391123.2391151

Brochhausen, M., Schneider, J., Daniel, M., Empey, P. E., Hogan, W. R., & Boyce, R. D. (2014). Towards a foundational representation of potential drug-drug interaction knowledge. In First International Workshop on Drug Interaction Knowledge Representation (DIKR-2014) at the International Conference on Biomedical Ontologies (ICBO 2014). Houston, Texas, United States. Retrieved from https://hal.archives-ouvertes.fr/hal-01076294

Brown, S. H., Elkin, P. L., Rosenbloom, S. T., Husser, C., Bauer, B. A., Lincoln, M. J., … Tuttle, M. S. (2004). VA National Drug File Reference Terminology: a cross-institutional content coverage study. Studies in Health Technology and Informatics, 107(Pt 1), 477–481.

Carroll, J. J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., & Wilkinson, K. (2004). Jena: Implementing the Semantic Web Recommendations. In Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers & Posters (pp. 74–83). New York, NY, USA: ACM. http://doi.org/10.1145/1013367.1013381

Caspases: Their Role in Cell Death and Cell Survival | Sigma-Aldrich. (n.d.). Retrieved October 13, 2015, from http://www.sigmaaldrich.com/catalog/product/sigma/z701785?lang=en&region=US

Cheng, F., & Zhao, Z. (2014). Machine learning-based prediction of drug–drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. Journal of the American Medical Informatics Association, 21(e2), e278–e286. http://doi.org/10.1136/amiajnl-2013-002512

Chen, X.-P., Tan, Z.-R., Huang, S.-L., Huang, Z., Ou-Yang, D.-S., & Zhou, H.-H. (2003). Isozyme-specific induction of low-dose aspirin on cytochrome P450 in healthy subjects. Clinical Pharmacology and Therapeutics, 73(3), 264–271. http://doi.org/10.1067/mcp.2003.14

Cheung, K.-H., Yip, K. Y., Townsend, J. P., & Scotch, M. (2008). HCLS 2.0/3.0: Health Care and Life Sciences Data Mashup Using Web 2.0/3.0. Journal of Biomedical Informatics, 41(5), 694–705. http://doi.org/10.1016/j.jbi.2008.04.001

Choi, S., & Cha, S. (2010). A survey of Binary similarity and distance measures. Journal of Systemics, Cybernetics and Informatics, 43–48.

Classen, D. C., Pestotnik, S. L., Evans, R. S., Lloyd, J. F., & Burke, J. P. (1997). Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. JAMA, 277(4), 301–306.

Conde-Estévez, D., Echeverría-Esnal, D., Tusquets, I., & Albanell, J. (2015). Potential clinical relevant drug–drug interactions: comparison between different compendia, do we have a validated method? Annals of Oncology, 26(6), 1272–1272. http://doi.org/10.1093/annonc/mdv151

Coulet, A., Shah, N. H., Garten, Y., Musen, M., & Altman, R. B. (2010). Using text to build semantic networks for pharmacogenomics. Journal of Biomedical Informatics, 43(6), 1009–1019. http://doi.org/10.1016/j.jbi.2010.08.005

CredibleMeds. (n.d.). Retrieved from https://crediblemeds.org/

de Coronado, S., Haber, M. W., Sioutos, N., Tuttle, M. S., & Wright, L. W. (2004). NCI Thesaurus: using science-based terminology to integrate cancer research results. Studies in Health Technology and Informatics, 107(Pt 1), 33–37.

Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., … Ashburner, M. (2008). ChEBI: a database and ontology for chemical

entities of biological interest. Nucleic Acids Research, 36(Database issue), D344–D350. http://doi.org/10.1093/nar/gkm791

Douillard, J. Y., Cunningham, D., Roth, A. D., Navarro, M., James, R. D., Karasek, P., … Rougier, P. (2000). Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. Lancet (London, England), 355(9209), 1041–1047.

Doulaverakis, C., Nikolaidis, G., Kleontas, A., & Kompatsiaris, I. (2014). Panacea, a semantic-enabled drug recommendations discovery framework. Journal of Biomedical Semantics, 5, 13. http://doi.org/10.1186/2041-1480-5-13

Drug Safety Labeling Changes > Lipitor (atorvastatin calcium) Tablets. (n.d.). Retrieved October 13, 2015, from http://www.fda.gov/Safety/MedWatch/SafetyInformation/Safety-RelatedDrugLabelingChanges/ucm133531.htm

Drugs.com | Prescription Drug Information, Interactions & Side Effects. (n.d.). Retrieved October 13, 2015, from http://www.drugs.com/

Duke, J. D., Han, X., Wang, Z., Subhadarshini, A., Karnik, S. D., Li, X., … Li, L. (2012). Literature based drug interaction prediction with clinical assessment using electronic medical records: novel myopathy associated drug interactions. PLoS Computational Biology, 8(8), e1002614. http://doi.org/10.1371/journal.pcbi.1002614

Dumontier, M., Baker, C. J., Baran, J., Callahan, A., Chepelev, L., Cruz-Toledo, J., … Hoehndorf, R. (2014). The Semanticscience Integrated Ontology (SIO) for biomedical research and knowledge discovery. Journal of Biomedical Semantics, 5(1), 14. http://doi.org/10.1186/2041-1480-5-14

Extensible Markup Language (XML) 1.0. (n.d.). Retrieved September 15, 2015, from http://www.w3.org/TR/1998/REC-xml-19980210

Facts & Comparisons® eAnswers | Clinical Drug Information. (n.d.). Retrieved October 13, 2015, from http://www.wolterskluwercdi.com/facts-comparisons-online/

FDA Adverse Events Reporting System (FAERS) > FAERS Reporting by Patient Outcomes by Year. (n.d.). Retrieved September 8, 2015, from http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070461.htm

FDA, P. D. S. I. for P. (n.d.). Postmarket Drug Safety Information for Patients and Providers - Information for Healthcare Professionals: Concomitant Use of Ibuprofen and Aspirin [WebContent]. Retrieved October 15, 2015, from http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm125222.htm

Feramisco, J. D., Sadreyev, R. I., Murray, M. L., Grishin, N. V., & Tsao, H. (2009). Phenotypic and Genotypic Analyses of Genetic Skin Disease through the Online Mendelian Inheritance in Man (OMIM) Database. Journal of

Investigative Dermatology, 129(11), 2628–2636. http://doi.org/10.1038/jid.2009.108

Flockhart DA. Drug Interactions: Cytochrome P450 Drug Interaction Table. (n.d.). Retrieved October 12, 2015, from http://medicine.iupui.edu/clinpharm/ddis/

Freifeld, A. G., Bow, E. J., Sepkowitz, K. A., Boeckh, M. J., Ito, J. I., Mullen, C. A., … Infectious Diseases Society of America, null. (2011). Clinical practice guideline for the use of antimicrobial agents in neutropenic patients with cancer: 2010 update by the infectious diseases society of america. Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America, 52(4), e56–93. http://doi.org/10.1093/cid/cir073

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. Knowl. Acquis., 5(2), 199–220. http://doi.org/10.1006/knac.1993.1008

Grymonpre, R. E., Mitenko, P. A., Sitar, D. S., Aoki, F. Y., & Montgomery, P. R. (1988). Drug-associated hospital admissions in older medical patients. Journal of the American Geriatrics Society, 36(12), 1092–1098.

Guthrie, B., Makubate, B., Hernandez-Santiago, V., & Dreischulte, T. (2015). The rising tide of polypharmacy and drug-drug interactions: population database analysis 1995–2010. BMC Medicine, 13(1), 74. http://doi.org/10.1186/s12916-015-0322-7

Haaz, M. C., Rivory, L., Riché, C., Vernillet, L., & Robert, J. (1998). Metabolism of irinotecan (CPT-11) by human hepatic microsomes: participation of cytochrome P-450 3A and drug interactions. Cancer Research, 58(3), 468–472.

Hahn, K. K., Wolff, J. J., & Kolesar, J. M. (2006). Pharmacogenetics and irinotecan therapy. American Journal of Health-System Pharmacy: AJHP: Official Journal of the American Society of Health-System Pharmacists, 63(22), 2211–2217. http://doi.org/10.2146/ajhp060155

Hamilton, R. A., Briceland, L. L., & Andritz, M. H. (1998). Frequency of hospitalization after exposure to known drug-drug interactions in a Medicaid population. Pharmacotherapy, 18(5), 1112–1120.

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., & McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Research, 33(suppl 1), D514–D517. http://doi.org/10.1093/nar/gki033

Hazlet, T. K., Lee, T. A., Hansten, P. D., & Horn, J. R. (2001). Performance of community pharmacy drug interaction software. Journal of the American Pharmaceutical Association (Washington,D.C.: 1996), 41(2), 200–204.

Health Care and Life Sciences (Semantic Web) Current Status - W3C. (n.d.). Retrieved October 13, 2015, from http://www.w3.org/standards/techs/lifesciences#w3c_all

Hendler, J. (2003). Science and the Semantic Web. Science, 299(5606), 520–521. http://doi.org/10.1126/science.1078874

Hennessy, S., & Flockhart, D. A. (2012). The need for translational research on drug-drug interactions. Clinical Pharmacology and Therapeutics, 91(5), 771–773. http://doi.org/10.1038/clpt.2012.39

Hinton, L. K., Galetin, A., & Houston, J. B. (2008). Multiple inhibition mechanisms and prediction of drug-drug interactions: status of metabolism and transporter models as exemplified by gemfibrozil-drug interactions. Pharmaceutical Research, 25(5), 1063–1074. http://doi.org/10.1007/s11095-007-9446-6

Holtzman, C. W., Wiggins, B. S., & Spinler, S. A. (2006). Role of P-glycoprotein in statin drug interactions. Pharmacotherapy, 26(11), 1601–1607. http://doi.org/10.1592/phco.26.11.1601

Home - ClinicalTrials.gov. (n.d.). Retrieved October 13, 2015, from https://clinicaltrials.gov/

Home - Gene - NCBI. (n.d.). Retrieved October 13, 2015, from http://www.ncbi.nlm.nih.gov/gene

Horn, J. R., Hansten, P. D., & Chan, L.-N. (2007). Proposal for a new tool to evaluate drug interaction cases. The Annals of Pharmacotherapy, 41(4), 674–680. http://doi.org/10.1345/aph.1H423

Horridge, M., Knublauch, H., Rector, A., Stevens, R., & Wroe, C. (2004). A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0. University of Manchester. Retrieved from ftp://gi29.geoinfo.tuwien.ac.at/courses/Ontology/ProtegeOWLTutorial.pdf

Huang, J., Niu, C., Green, C. D., Yang, L., Mei, H., & Han, J.-D. J. (2013). Systematic prediction of pharmacodynamic drug-drug interactions through protein-protein-interaction network. PLoS Computational Biology, 9(3), e1002998. http://doi.org/10.1371/journal.pcbi.1002998

Huang, S.-M., Temple, R., Throckmorton, D. C., & Lesko, L. J. (2007). Drug Interaction Studies: Study Design, Data Analysis, and Implications for Dosing and Labeling. Clinical Pharmacology & Therapeutics, 81(2), 298–304. http://doi.org/10.1038/sj.clpt.6100054

Hutzler, J. M., Cook, J., & Fleishaker, J. C. (2011). Drug–Drug Interactions: Designing Development Programs and Appropriate Product Labeling. In P. L. Bonate & D. R. Howard (Eds.), Pharmacokinetics in Drug Development (pp. 21–56). Springer US. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-7937-7_2

Ibuprofen Pathway, Pharmacokinetics. (n.d.). Retrieved September 1, 2015, from https://www.pharmgkb.org/pathway/PA166041114

Innocenti, F., Undevia, S. D., Ramírez, J., Mani, S., Schilsky, R. L., Vogelzang, N. J., … Ratain, M. J. (2004). A phase I trial of pharmacologic modulation of irinotecan with cyclosporine and phenobarbital. Clinical Pharmacology and Therapeutics, 76(5), 490–502. http://doi.org/10.1016/j.clpt.2004.07.016

Iyer, S. V., Harpaz, R., LePendu, P., Bauer-Mehren, A., & Shah, N. H. (2014). Mining clinical text for signals of adverse drug-drug interactions. Journal of

the American Medical Informatics Association: JAMIA, 21(2), 353–362. http://doi.org/10.1136/amiajnl-2013-001612

Johnsson, G., Svedmyr, N., & Thiringer, G. (1975). Effects of intravenous propranolol and metoprolol and their interaction with isoprenaline on pulmonary function, heart rate and blood pressure in asthmatics. European Journal of Clinical Pharmacology, 8(3-4), 175–180.

Jupp, S., Malone, J., Bolleman, J., Brandizi, M., Davies, M., Garcia, L., … Jenkinson, A. M. (2014). The EBI RDF platform: linked open data for the life sciences. Bioinformatics (Oxford, England), 30(9), 1338–1339. http://doi.org/10.1093/bioinformatics/btt765

Juty, N., Novère, N. L., & Laibe, C. (2012). Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. Nucleic Acids Research, 40(D1), D580–D586. http://doi.org/10.1093/nar/gkr1097

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28(1), 27–30.

Kilicoglu, H., Shin, D., Fiszman, M., Rosemblat, G., & Rindflesch, T. C. (2012). SemMedDB: a PubMed-scale repository of biomedical semantic predications. Bioinformatics (Oxford, England), 28(23), 3158–3160. http://doi.org/10.1093/bioinformatics/bts591

Klein, T. E., Chang, J. T., Cho, M. K., Easton, K. L., Fergerson, R., Hewett, M., … Altman, R. B. (2001). Integrating genotype and phenotype information: an

overview of the PharmGKB project. The Pharmacogenomics Journal, 1(3), 167–170. http://doi.org/10.1038/sj.tpj.6500035

Knublauch, H., Fergerson, R. W., Noy, N. F., & Musen, M. A. (2004). The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications. In S. A. McIlraith, D. Plexousakis, & F. van Harmelen (Eds.), The Semantic Web – ISWC 2004 (pp. 229–243). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-30475-3_17

König, J., Müller, F., & Fromm, M. F. (2013). Transporters and drug-drug interactions: important determinants of drug disposition and effects. Pharmacological Reviews, 65(3), 944–966. http://doi.org/10.1124/pr.113.007518

Korhonen, L. E., Rahnasto, M., Mähönen, N. J., Wittekindt, C., Poso, A., Juvonen, R. O., & Raunio, H. (2005). Predictive three-dimensional quantitative structure-activity relationship of cytochrome P450 1A2 inhibitors. Journal of Medicinal Chemistry, 48(11), 3808–3815. http://doi.org/10.1021/jm0489713

Krishna, G., Ma, L., Prasad, P., Moton, A., Martinho, M., & O'Mara, E. (2012). Effect of posaconazole on the pharmacokinetics of simvastatin and midazolam in healthy volunteers. Expert Opinion on Drug Metabolism & Toxicology, 8(1), 1–10. http://doi.org/10.1517/17425255.2012.639360

Kuhn, M., Campillos, M., Letunic, I., Jensen, L. J., & Bork, P. (2010). A side effect resource to capture phenotypic effects of drugs. Molecular Systems Biology, 6, 343. http://doi.org/10.1038/msb.2009.98

Latest Medical News, Clinical Trials, Guidelines – Today on Medscape. (n.d.). Retrieved October 13, 2015, from http://www.medscape.com/

Lee, Y.-S., Kim, H., Wu, T.-X., Wang, X.-M., & Dionne, R. A. (2006). Genetically mediated interindividual variation in analgesic responses to cyclooxygenase inhibitory drugs. Clinical Pharmacology and Therapeutics, 79(5), 407–418. http://doi.org/10.1016/j.clpt.2006.01.013

LePendu, P., Iyer, S. V., Bauer-Mehren, A., Harpaz, R., Mortensen, J. M., Podchiyska, T., … Shah, N. H. (2013). Pharmacovigilance Using Clinical Notes. Clinical Pharmacology & Therapeutics, 93(6), 547–555. http://doi.org/10.1038/clpt.2013.47

Lewis, L. D. (2010). Drug–drug interactions: is there an optimal way to study them? British Journal of Clinical Pharmacology, 70(6), 781–783. http://doi.org/10.1111/j.1365-2125.2010.03829.x

Lexicomp® Online | Clinical Drug Information. (n.d.). Retrieved October 13, 2015, from http://www.wolterskluwercdi.com/lexicomp-online/

Life Science Resource Name (LSRN). (n.d.). Retrieved October 13, 2015, from http://lsrn.org/

Lipscomb, C. E. (2000). Medical Subject Headings (MeSH). Bulletin of the Medical Library Association, 88(3), 265–266.

Liu, Y., & Unni, E. (2014). Methodological issues of retrospective studies assessing health outcomes of potential clopidogrel-statin interaction. The International Journal of Pharmacy Practice, 22(5), 360–362. http://doi.org/10.1111/ijpp.12085

Lynch, T., & Price, A. (2007). The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. American Family Physician, 76(3), 391–396.

MacDonald, T. M., & Wei, L. (2003). Effect of ibuprofen on cardioprotective effect of aspirin. Lancet (London, England), 361(9357), 573–574.

Maglott, D., Ostell, J., Pruitt, K. D., & Tatusova, T. (2005). Entrez Gene: gene-centered information at NCBI. Nucleic Acids Research, 33(Database issue), D54–58. http://doi.org/10.1093/nar/gki031

Ma, M. K., & McLeod, H. L. (2003). Lessons learned from the irinotecan metabolic pathway. Current Medicinal Chemistry, 10(1), 41–49.

Mathijssen, R. H., van Alphen, R. J., Verweij, J., Loos, W. J., Nooter, K., Stoter, G., & Sparreboom, A. (2001). Clinical pharmacokinetics and metabolism of irinotecan (CPT-11). Clinical Cancer Research: An Official Journal of the American Association for Cancer Research, 7(8), 2182–2194.

Micromedex® Healthcare Series. (2015). Thomson Micromedex. Retrieved from http://micromedex.com/

MIRIAM Registry. (n.d.). Retrieved October 13, 2015, from http://www.ebi.ac.uk/miriam/main/

Mullokandov E, A. J. (2014). Protein Binding Drug-Drug Interaction between Warfarin and Tizoxanide in Human Plasma. Austin Journal of Pharmacology and Therapeutics, 2(7).

Murphy, S. N., Weber, G., Mendis, M., Gainer, V., Chueh, H. C., Churchill, S., & Kohane, I. (2010). Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). Journal of the American Medical Informatics Association: JAMIA, 17(2), 124–130. http://doi.org/10.1136/jamia.2009.000893

National Drug File – Reference Terminology (NDF-RT™) Documentation. (n.d.). Retrieved from http://evs.nci.nih.gov/ftp1/NDF-RT/NDF-RT%20Documentation.pdf

Nolin, R., Ansell, P., Belleau, F., Idehen, K., Tourigny, N., Roe, P., … Laval, U. (n.d.). Bio2RDF Network Of Linked Data.

Novère, N. L., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., … Wanner, B. L. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). Nature Biotechnology, 23(12), 1509–1515. http://doi.org/10.1038/nbt1156

Okkam - Thinking identifiers! (n.d.). Retrieved October 13, 2015, from http://www.okkam.org/

OSCAR Electronic Medical Record. (n.d.). Retrieved from http://oscar-emr.com/

Owens, A., Seaborne, A., Gibbins, N., & schraefel, mc. (2008, November). Clustered TDB: A Clustered Triple Store for Jena [Monograph]. Retrieved September 1, 2015, from http://eprints.soton.ac.uk/266974/

OWL Web Ontology Language Overview. (n.d.). Retrieved September 15, 2015, from http://www.w3.org/TR/owl-features/

Patel, P. S., Rana, D. A., Suthar, J. V., Malhotra, S. D., & Patel, V. J. (2014). A study of potential adverse drug-drug interactions among prescribed drugs in medicine outpatient department of a tertiary care teaching hospital. Journal of Basic and Clinical Pharmacy, 5(2), 44–48. http://doi.org/10.4103/0976-0105.134983

Peng, C. C., Glassman, P. A., Marks, I. R., Fowler, C., Castiglione, B., & Good, C. B. (2003). Retrospective drug utilization review: incidence of clinically relevant potential drug-drug interactions in a large ambulatory population. Journal of Managed Care Pharmacy: JMCP, 9(6), 513–522.

Percha, B., & Altman, R. B. (2013). Informatics confronts drug–drug interactions. Trends in Pharmacological Sciences, 34(3). http://doi.org/10.1016/j.tips.2013.01.006

Percha, B., Garten, Y., & Altman, R. B. (2012). Discovery and explanation of drug-drug interactions via text mining. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 410–421.

Peters, L. B., Bahr, N., & Bodenreider, O. (2015). Evaluating drug-drug interaction information in NDF-RT and DrugBank. Journal of Biomedical Semantics, 6. http://doi.org/10.1186/s13326-015-0018-0

Phansalkar, S., Desai, A. A., Bell, D., Yoshida, E., Doole, J., Czochanski, M., … Bates, D. W. (2012). High-priority drug-drug interactions for use in electronic health records. Journal of the American Medical Informatics Association: JAMIA, 19(5), 735–743. http://doi.org/10.1136/amiajnl-2011-000612

Phansalkar, S., van der Sijs, H., Tucker, A. D., Desai, A. A., Bell, D. S., Teich, J. M., … Bates, D. W. (2013). Drug-drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records. Journal of the American Medical Informatics Association: JAMIA, 20(3), 489–493. http://doi.org/10.1136/amiajnl-2012-001089

Pharmacology Condensed, 2nd Edition | Maureen Dale, Dennis Haylett | ISBN 9780443067730. (n.d.). Retrieved September 15, 2015, from http://store.elsevier.com/Pharmacology-Condensed/Maureen-Dale/isbn-9780443067730/

Pleuvry, B. J. (2005). Pharmacodynamic and pharmacokinetic drug interactions. Anaesthesia & Intensive Care Medicine, 6(4), 129–133. http://doi.org/10.1383/anes.6.4.129.63634

Pommier, Y. (2006). Topoisomerase I inhibitors: camptothecins and beyond. Nature Reviews. Cancer, 6(10), 789–802. http://doi.org/10.1038/nrc1977

Pommier, Y., Leo, E., Zhang, H., & Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. Chemistry & Biology, 17(5), 421–433. http://doi.org/10.1016/j.chembiol.2010.04.012

Preissner, S., Kroll, K., Dunkel, M., Senger, C., Goldsobel, G., Kuzman, D., … Preissner, R. (2010). SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. Nucleic Acids Research, 38(Database issue), D237–243. http://doi.org/10.1093/nar/gkp970

Prud'hommeaux, E., & Seaborne, A. (n.d.). SPARQL Query Language for RDF. Retrieved from http://www.w3.org/TR/rdf-sparql-query/

Qu, X. A., Gudivada, R. C., Jegga, A. G., Neumann, E. K., & Aronow, B. J. (2007). ailSemantic Web-based data representation and reasoning applied to disease mechanism and pharmacology. In IEEE International Conference on Bioinformatics and Biomedicine Workshops, 2007. BIBMW 2007 (pp. 131–143). http://doi.org/10.1109/BIBMW.2007.4425411

Rae, J. M., Johnson, M. D., Lippman, M. E., & Flockhart, D. A. (2001). Rifampin is a selective, pleiotropic inducer of drug metabolism genes in human hepatocytes: studies with cDNA and oligonucleotide expression arrays. The Journal of Pharmacology and Experimental Therapeutics, 299(3), 849–857.

RDF Schema 1.1. (n.d.). Retrieved October 13, 2015, from http://www.w3.org/TR/rdf-schema/

Reference Guide For Foreign Pharmacy Licensing Exam Pharmacy Management & Pharmacoeconomics Question And Answers. (n.d.). Retrieved September 15, 2015, from http://www.goodreads.com/book/show/2396603.Reference_Guide_For_Foreign_Pharmacy_Licensing_Exam_Pharmacy_Management_Pharmacoeconomics_Question_And_Answers

Reis, A. M. M., & Cassiani, S. H. D. B. (2010). Evaluation of three brands of drug interaction software for use in intensive care units. Pharmacy World & Science: PWS, 32(6), 822–828. http://doi.org/10.1007/s11096-010-9445-2

Resource Description Framework (RDF) Model and Syntax Specification. (n.d.). Retrieved September 15, 2015, from http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/

Rice, P. J., Perry, R. J., Afzal, Z., & Stockley, I. H. (2003). Antibacterial prescribing and warfarin: a review. British Dental Journal, 194(8), 411–415. http://doi.org/10.1038/sj.bdj.4810049

Rogatko, A., Babb, J. S., Tighiouart, M., Khuri, F. R., & Hudes, G. (2005). New paradigm in dose-finding trials: patient-specific dosing and beyond phase I. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research, 11(15), 5342–5346. http://doi.org/10.1158/1078-0432.CCR-05-0458

Rosholm, J. U., Bjerrum, L., Hallas, J., Worm, J., & Gram, L. F. (1998). Polypharmacy and the risk of drug-drug interactions among Danish elderly. A prescription database study. Danish Medical Bulletin, 45(2), 210–213.

Ruttenberg, A., Rees, J. A., Samwald, M., & Marshall, M. S. (2009). Life sciences on the Semantic Web: the Neurocommons and beyond. Briefings in Bioinformatics, 10(2), 193–204. http://doi.org/10.1093/bib/bbp004

RxNorm Overview. (n.d.). Retrieved October 13, 2015, from https://www.nlm.nih.gov/research/umls/rxnorm/overview.html

Safety Information > Xarelto (Rivaroxaban) Tablets. (n.d.). Retrieved October 13, 2015, from http://www.fda.gov/Safety/MedWatch/SafetyInformation/ucm367392.htm

Sahoo, S., Bodenreider, O., Zeng, K., & Sheth, A. (2007). An Experiment in Integrating Large Biomedical Knowledge Resources with RDF: Application to Associating Genotype and Phenotype Information. Kno.e.sis Publications. Retrieved from http://corescholar.libraries.wright.edu/knoesis/692

Samowitz, W. S., Wolff, R. K., Curtin, K., Sweeney, C., Ma, K.-N., Andersen, K., … Slattery, M. L. (2006). Interactions between CYP2C9 and UGT1A6 polymorphisms and nonsteroidal anti-inflammatory drugs in colorectal cancer prevention. Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association, 4(7), 894–901. http://doi.org/10.1016/j.cgh.2006.04.021

Santos, A., Zanetta, S., Cresteil, T., Deroussent, A., Pein, F., Raymond, E., … Vassal, G. (2000). Metabolism of irinotecan (CPT-11) by CYP3A4 and CYP3A5 in humans. Clinical Cancer Research: An Official Journal of the American Association for Cancer Research, 6(5), 2012–2020.

Saxena, A., Balaramnavar, V. M., Hohlfeld, T., & Saxena, A. K. (2013). Drug/drug interaction of common NSAIDs with antiplatelet effect of aspirin in human platelets. European Journal of Pharmacology, 721(1-3), 215–224. http://doi.org/10.1016/j.ejphar.2013.09.032

Scheife, R. T., Hines, L. E., Boyce, R. D., Chung, S. P., Momper, J. D., Sommer, C. D., … Malone, D. C. (2015). Consensus recommendations for systematic evaluation of drug-drug interaction evidence for clinical decision support. Drug Safety, 38(2), 197–206. http://doi.org/10.1007/s40264-014-0262-8

Seaborne, A., & Manjunath, G. (2007, April 24). {SPARQL/Update}: A language for updating {RDF} graphs. Retrieved September 1, 2015, from http://jena.hpl.hp.com/~afs/SPARQL-Update.html

Segura Bedmar, I., Martínez, P., & Sánchez Cisneros, D. (2011). The 1st DDIExtraction-2011 Challenge Task: Extraction of Drug-Drug Interactions from Biomedical Texts. Retrieved from http://e-archivo.uc3m.es/handle/10016/20389

Shad, M. U., Marsh, C., & Preskorn, S. H. (2001). The economic consequences of a drug-drug interaction. Journal of Clinical Psychopharmacology, 21(1), 119–120.

Sherr, C. J., & McCormick, F. (2002). The RB and p53 pathways in cancer. Cancer Cell, 2(2), 103–112.

Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical OWL-DL reasoner. Web Semantics: Science, Services and Agents on the World Wide Web, 5(2), 51–53. http://doi.org/10.1016/j.websem.2007.03.004

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., … Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology, 25(11), 1251–1255. http://doi.org/10.1038/nbt1346

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., … Rosse, C. (2005). Relations in biomedical ontologies. Genome Biology, 6(5), R46. http://doi.org/10.1186/gb-2005-6-5-r46

Spina, E., & de Leon, J. (2007). Metabolic drug interactions with newer antipsychotics: a comparative review. Basic & Clinical Pharmacology & Toxicology, 100(1), 4–22. http://doi.org/10.1111/j.1742-7843.2007.00017.x

Stan, J. (2014). A Machine-Learning Approach for Drug-Drug Interaction Extraction from FDA Structured Product Labels.

Statistics for Online Mendelian Inheritance in Man. (n.d.). Retrieved October 13, 2015, from http://download.bio2rdf.org/release/3/omim/omim.html

Takarabe, M., Shigemizu, D., Kotera, M., Goto, S., & Kanehisa, M. (2011). Network-based analysis and characterization of adverse drug-drug interactions.

Journal of Chemical Information and Modeling, 51(11), 2977–2985. http://doi.org/10.1021/ci200367w

Tari, L., Anwar, S., Liang, S., Cai, J., & Baral, C. (2010). Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism. Bioinformatics (Oxford, England), 26(18), i547–553. http://doi.org/10.1093/bioinformatics/btq382

Tatonetti, N. P., Fernald, G. H., & Altman, R. B. (2012). A novel signal detection algorithm for identifying hidden drug-drug interactions in adverse event reports. Journal of the American Medical Informatics Association: JAMIA, 19(1), 79–85. http://doi.org/10.1136/amiajnl-2011-000214

Tatonetti, N. P., Ye, P. P., Daneshjou, R., & Altman, R. B. (2012a). Data-Driven Prediction of Drug Effects and Interactions. Science Translational Medicine, 4(125), 125ra31–125ra31. http://doi.org/10.1126/scitranslmed.3003377

Tatonetti, N. P., Ye, P. P., Daneshjou, R., & Altman, R. B. (2012b). Data-driven prediction of drug effects and interactions. Science Translational Medicine, 4(125), 125ra31. http://doi.org/10.1126/scitranslmed.3003377

The Description Logic Handbook. (n.d.). Retrieved October 13, 2015, from http://www.cambridge.org/us/academic/subjects/computer-science/programming-languages-and-applied-logic/description-logic-handbook-theory-implementation-and-applications-2nd-edition

The Semantic Web. (n.d.). Retrieved September 15, 2015, from http://www.scientificamerican.com/article/the-semantic-web/

Thummel, K. E., & Wilkinson, G. R. (1998). In vitro and in vivo drug interactions involving human CYP3A. Annual Review of Pharmacology and Toxicology, 38, 389–430. http://doi.org/10.1146/annurev.pharmtox.38.1.389

UMLS - Release Notes. (n.d.). Retrieved October 13, 2015, from https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/notes.html

van der Heijden, P. G. M., van Puijenbroek, E. P., van Buuren, S., & van der Hofstede, J. W. (2002). On the assessment of adverse drug reactions from spontaneous reporting systems: the influence of under-reporting on odds ratios. Statistics in Medicine, 21(14), 2027–2044. http://doi.org/10.1002/sim.1157

Venkatakrishnan, K., von Moltke, L. L., Obach, R. S., & Greenblatt, D. J. (2003). Drug metabolism and drug interactions: application and clinical value of in vitro models. Current Drug Metabolism, 4(5), 423–459.

Vilar, S., Harpaz, R., Chase, H. S., Costanzi, S., Rabadan, R., & Friedman, C. (2011). Facilitating adverse drug event detection in pharmacovigilance databases using molecular structure similarity: application to rhabdomyolysis. Journal of the American Medical Informatics Association: JAMIA, 18 Suppl 1, i73–80. http://doi.org/10.1136/amiajnl-2011-000417

Vilar, S., Lorberbaum, T., Hripcsak, G., & Tatonetti, N. P. (2015). Improving Detection of Arrhythmia Drug-Drug Interactions in Pharmacovigilance Data

through the Implementation of Similarity-Based Modeling. PloS One, 10(6), e0129974. http://doi.org/10.1371/journal.pone.0129974

Vizenor, L., Bodenreider, O., Peters, L., & McCray, A. T. (2006). Enhancing biomedical ontologies through alignment of semantic relationships: exploratory approaches. AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium, 804–808.

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., … Wang, J. (2008). The diploid genome sequence of an Asian individual. Nature, 456(7218), 60–65. http://doi.org/10.1038/nature07484

Wang, Q., Cai, Y., Van de Casteele, M., Pipeleers, D., & Ling, Z. (2011). Interaction of glibenclamide and metformin at the level of translation in pancreatic β cells. The Journal of Endocrinology, 208(2), 161–169. http://doi.org/10.1677/JOE-10-0372

WHOCC - ATC/DDD Index. (n.d.). Retrieved October 13, 2015, from http://www.whocc.no/atc_ddd_index/

Wienkers, L. C., & Heath, T. G. (2005). Predicting in vivo drug interactions from in vitro drug discovery data. Nature Reviews. Drug Discovery, 4(10), 825–833. http://doi.org/10.1038/nrd1851

Wiley: Semantic Web Technologies: Trends and Research in Ontology-based Systems - John Davies, Rudi Studer, Paul Warren. (n.d.). Retrieved September 15, 2015, from http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470025964.html

Williams, D. D., & Feely, J. (2012). Pharmacokinetic-Pharmacodynamic Drug Interactions with HMG-CoA Reductase Inhibitors. Clinical Pharmacokinetics, 41(5), 343–370. http://doi.org/10.2165/00003088-200241050-00003

Williams, J. A., Hyland, R., Jones, B. C., Smith, D. A., Hurst, S., Goosen, T. C., … Ball, S. E. (2004). Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. Drug Metabolism and Disposition: The Biological Fate of Chemicals, 32(11), 1201–1208. http://doi.org/10.1124/dmd.104.000794

Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., … Hassanali, M. (2008). DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Research, 36(Database issue), D901–906. http://doi.org/10.1093/nar/gkm958

Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ (Clinical Research Ed.), 318(7182), 527–530.

World Wide Web Consortium (W3C). (n.d.). Retrieved September 15, 2015, from http://www.w3.org/

Xu, Y., & Villalona-Calero, M. A. (2002). Irinotecan: mechanisms of tumor resistance and novel strategies for modulating its activity. Annals of Oncology: Official Journal of the European Society for Medical Oncology / ESMO, 13(12), 1841–1851.

Yap, C. W., & Chen, Y. Z. (2005). Prediction of cytochrome P450 3A4, 2D6, and 2C9 inhibitors and substrates by using support vector machines. Journal of Chemical Information and Modeling, 45(4), 982–992. http://doi.org/10.1021/ci0500536

Yasui-Furukori, N., Kubo, K., Ishioka, M., Tsuchimine, S., & Inoue, Y. (2013). Interaction between paliperidone and carbamazepine. Therapeutic Drug Monitoring, 35(5), 649–652. http://doi.org/10.1097/FTD.0b013e3182966c2f

Zhang, L. (2010, March 17). Transporter-Mediated Drug-Drug Interactions (DDIs). FDA.

Zhang, R., Cairelli, M. J., Fiszman, M., Rosemblat, G., Kilicoglu, H., Rindflesch, T. C., … Melton, G. B. (2014a). Using semantic predications to uncover drug-drug interactions in clinical data. Journal of Biomedical Informatics, 49, 134–147. http://doi.org/10.1016/j.jbi.2014.01.004

Zhang, R., Cairelli, M. J., Fiszman, M., Rosemblat, G., Kilicoglu, H., Rindflesch, T. C., … Melton, G. B. (2014b). Using semantic predications to uncover drug-drug interactions in clinical data. Journal of Biomedical Informatics, 49, 134–147. http://doi.org/10.1016/j.jbi.2014.01.004

APPENDIX A

MULTI-PATHWAY INTERACTION VALIDATION

In this appendix, we show the validation process for irinotecan's multi-pathway findings from Chapter IV. Three different tables are illustrated including: (1) interactions found in compendia, (2) interactions found in literature and on clinical websites, and (3) potential interactions.

Table A1: irinotecan interactions found in compendia

Drug	Compendia	Mechanisms of interaction
Amiodarone	Lexi-Comp – Medscape- PDDIs	P-gp Transporter
Carvedilol	Lexi-Comp	P-gp Transporter
Citalopram	PDDIs - Micromedex	P-gp Transporter
Cyclosporine	Medscape	P-gp Transporter
Dasatinib	Lexi-Comp - Drugs.com	P-gp Transporter
Gefitinib	PDDIs	P-gp Transporter
Itraconazole	Lexi-Comp – Medscape- PDDIs	P-gp Transporter
Loperamide	PDDIs	P-gp Transporter
Mefloquine	Lexi-Comp	P-gp Transporter

Morphine	PDDIs	P-gp Transporter
Propranolol	Lexi-Comp - Drugs.com	P-gp Transporter
Quinidine	Lexi-Comp - Medscape	P-gp Transporter
Tacrolimus	Lexi-Comp - Drugs.com - Medscape	P-gp Transporter
Tamoxifen	Lexi-Comp	P-gp Transporter
Trazodone	Medscape	P-gp Transporter
Vinblastine	Lexi-Comp	P-gp Transporter
Boceprevir	Lexi-Comp	CYP3A4 Metabolism
Carbamazepine	PDDIs - Lexi-Comp - Drugs.com – Micromedex	CYP3A4 Metabolism
Cimetidine	PDDIs - Lexi-Comp	CYP3A4 Metabolism
Clotrimazole	Lexi-Comp	CYP3A4 Metabolism
Clozapine	Lexi-Comp- PDDIs	CYP3A4 Metabolism
Desipramine	Lexi-Comp	CYP3A4 Metabolism
Dexamethasone	PDDIs - Lexi-Comp – Micromedex	CYP3A4 Metabolism
Diazepam	PDDIs	CYP3A4 Metabolism
Docetaxel	PDDIs	CYP3A4 Metabolism
Haloperidol	Lexi-Comp	CYP3A4 Metabolism
Imatinib	PDDIs - Lexi-Comp	CYP3A4 Metabolism
Ketoconazole	Lexi-Comp - Drugs.com – Micromedex – Medscape- PDDIs	CYP3A4 Metabolism
Methylprednisolone	PDDIs	CYP3A4 Metabolism
Mifepristone	Lexi-Comp - Drugs.com - Medscape	CYP3A4 Metabolism
Nelfinavir	Lexi-Comp- PDDIs	CYP3A4 Metabolism

Olanzapine	PDDIs	CYP3A4 Metabolism
Pazopanib	Drugs.com- PDDIs	CYP3A4 Metabolism
Phenobarbital	Lexi-Comp - Drugs.com – Micromedex- PDDIs	CYP3A4 Metabolism
Phenytoin	Lexi-Comp - NDF-RT – Micromedex- PDDIs	CYP3A4 Metabolism
Prednisolone	PDDIs	CYP3A4 Metabolism
Pravastatin	Lexi-Comp - Micromedex	CYP3A4 Metabolism
Prednisolone	PDDIs	CYP3A4 Metabolism
Saquinavir	Lexi-Comp- PDDIs	CYP3A4 Metabolism
Sertraline	PDDIs - Lexi-Comp	CYP3A4 Metabolism
Telaprevir	Lexi-Comp - Drugs.com	CYP3A4 Metabolism
Topotecan	PDDIs - Drugs.com	CYP3A4 Metabolism
Acetaminophen	PDDIs	P-gp & CYP3A4
Amlodipine	PDDIs	P-gp & CYP3A4
Amprenavir	PDDIs	P-gp & CYP3A4
Astemizole	PDDIs	P-gp & CYP3A4
Atorvastatin	Lexi-Comp - Medscape	P-gp & CYP3A4
Clarithromycin	Lexi-Comp – Medscape- PDDIs	P-gp & CYP3A4
Clopidogrel	PDDIs - Lexi-Comp	P-gp & CYP3A4
Cyclophosphamide	PDDIs	P-gp & CYP3A4
Diltiazem	PDDIs - Lexi-Comp	P-gp & CYP3A4
Doxorubicin	PDDIs - Lexi-Comp	P-gp & CYP3A4
Fentanyl	PDDIs	P-gp & CYP3A4
Hydrocortisone	PDDIs	P-gp & CYP3A4
Ivacaftor	Lexi-Comp - Drugs.com - Medscape	P-gp & CYP3A4
Indinavir	Lexi-Comp – Medscape- PDDIs	P-gp & CYP3A4
Lapatinib	Lexi-Comp - Drugs.com – Medscape- PDDIs	P-gp & CYP3A4
Lidocaine	PDDIs	P-gp & CYP3A4
Losartan	PDDIs - Drugs.com	P-gp & CYP3A4
Levofloxacin	Drugs.com	P-gp & CYP3A4
Midazolam	PDDIs	P-gp & CYP3A4
Nilotnib	Lexi-Comp - Medscape	P-gp & CYP3A4

Nefazodone	Lexi-Comp – Medscape- PDDIs	P-gp & CYP3A4
Nicardipine	Lexi-Comp - Medscape	P-gp & CYP3A4
Nifedipine	PDDIs	P-gp & CYP3A4
Omeprazole	PDDIs	P-gp & CYP3A4
Paclitaxel	PDDIs	P-gp & CYP3A4
Pantoprazole	PDDIs	P-gp & CYP3A4
Prednisone	PDDIs	P-gp & CYP3A4
Sorafenib	Drugs.com - Medscape	P-gp & CYP3A4
Ranitidine	PDDIs	P-gp & CYP3A4
Rifampin	Lexi-Comp – Micromedex - Medscape	P-gp & CYP3A4
Ritonavir	Lexi-Comp -Medscape- PDDIs	P-gp & CYP3A4
Simvastatin	PDDIs	P-gp & CYP3A4
Sirolimus	Drugs.com	P-gp & CYP3A4
Venlafaxine	PDDIs	P-gp & CYP3A4
Verapamil	PDDIs - Lexi-Comp	P-gp & CYP3A4
Vincristine	PDDIs	P-gp & CYP3A4
80 true positives: 16 could interact at the P-gp transporter level, 26 could interact at the CYP3A4 metabolism level, and 38 could interact at the transporter and metabolism levels		

Table A2: irinotecan interactions found in literature and clinical websites

Drug	Literature and clinical websites	Mechanisms of Interaction
Testosterone	http://www.ncbi.nlm.nih.gov/pubmed/11901092	P-gp Transporter
Trimethoprim	http://www.ncbi.nlm.nih.gov/pubmed/9458091	P-gp Transporter
Nitrendipine	http://www.ncbi.nlm.nih.gov/pubmed/12019202	P-gp & CYP3A4
Fluoxetine	http://www.ncbi.nlm.nih.gov/pubmed/21395523	P-gp & CYP3A4
Fluvoxamine	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2949913	P-gp & CYP3A4
Lovastatin	http://meeting.ascopubs.org/cgi/content/abstract/28/15_suppl/3020	P-gp & CYP3A4
Methadone	http://www.ncbi.nlm.nih.gov/pubmed/15260917	P-gp & CYP3A4
Probenecid	http://www.ncbi.nlm.nih.gov/pubmed/15618649	P-gp & CYP3A4
Alfentanil	http://www.fda.gov/drugs/developmentapprovalprocess/developmentresources/druginteractionslabeling/ucm093664.htm	P-gp & CYP3A4

Clobazam	https://online.epocrates.com/u/10a1494/irinotecan	Other
Erlotinib	http://www.ncbi.nlm.nih.gov/pubmed/15475439	Other
Axitinib	http://www.ncbi.nlm.nih.gov/pubmed/21390185	Other
12 interactions with irinotecan from literature and other data sources. 2 could interact at the P-gp transporter level, 7 could interact at the transporter and CYP3A4 metabolism levels, and 3 for other reasons		

Table A3: irinotecan potential interactions

Drug	Mechanisms of Interaction
Imipramine	P-gp Transporter
Amitriptyline	P-gp Transporter
Dextromethorphan	P-gp Transporter
Doxepin	P-gp Transporter
Buspirone	P-gp Transporter
Ticagrelor	P-gp Transporter
Clomipramine	P-gp Transporter
Propafenone	P-gp Transporter
Brentuximab vedotin	CYP3A4 Metabolism
Risperidone	CYP3A4 Metabolism
Nisoldipine	P-gp & CYP3A4
Caffeine	P-gp & CYP3A4
Pimozide	P-gp & CYP3A4
Bromocriptine	P-gp & CYP3A4
Colchicine	P-gp & CYP3A4
Daunorubicin	P-gp & CYP3A4
Felodipine	P-gp & CYP3A4
Lansoprazole	P-gp & CYP3A4
Mitoxantrone	P-gp & CYP3A4
Ethinyl estradiol	Other
Domperidone	Other
Zidovudine	Other
24 potential interactions with irinotecan were found. 9 could interact at the P-gp transporter level, 2 could interact at the CYP3A4 metabolism level, 9 could interact at the transporter and metabolism levels, and 4 for other reasons	

APPENDIX B

D3 INFERENTIAL QUERY MODEL INFERENCES

In this appendix, we aim to show how the D3 query-based model that was presented in Chapter V works. As an illustration for the proposed query-based model, we consider nine inferences for discovering and explaining a DDI. That is, the D3 inferential query-based model works by querying the D3 knowledge base looking for an interaction between two drugs. The system then returns a determination of whether there is a reported interaction or not, along with the common pathways it finds to account for the interactions. Here we provide an example of each inference where the D3 query-based model was capable of identifying the correct interactions based on predefined semantic inferences. We confirm our results by providing a clinical citation from the literature confirming the interaction. The target mechanism is colored in blue.

1. **Protein-binding interaction:**

Drug A: Warfarin
Drug B: Sulfamethoxazole
Result: Reduce the Warfarin concentrations
Clinical Evidence: http://www.ncbi.nlm.nih.gov/pubmed/7053283
D3 output:

```
Starting D3 Inferential Query Model
Check the interaction between [Warfarin] and [Sulfamethoxazole]

Check if Drugs are in UMLS database
First Drug CUI is: C0043031
Second Drug CUI is: C0038689

Check if Drugs are in D3 knowledge base
[Warfarin] and [Sulfamethoxazole] are in the D3 knowledge base

DDI sources report DDI:
[NDF-RT, Drugbank, CredibleMeds, OSCAR]

**** This DDI is proven ****
**** Mechanisms contribute for DDI ****

Pharmacogenetic Mechanism
Assertion facts are collected from pharmgkb the pharmacogenomics
knowledge base
[rs1057910]

Protein-Binding Mechanism
Assertion facts are collected from Drugbank database
[There is possible interaction due to high protein binding affinity by
Warfarin]

Additive Mechanism
Assertion facts are collected from The National Drug File - Reference
Terminology (NDF-RT) and NCI Thesaurus - National Institutes of Health
**** No Additive Mechanism is found  ****

Competition Mechanism
Assertion facts are collected from Gene Ontology (GO), NCI Thesaurus -
National Institutes of Health and National Center for Biotechnology
Information (NCBI)
```

```
**** No Competition Mechanism is found  ****

Metabolism inhibition Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
[CYP3A4 gene, CYP2C9 gene, CYP2C8 gene]
Both drugs can cause the interaction Warfarin AND Sulfamethoxazole

Metabolism induction Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
**** No Metabolism induction Mechanism is found ****

Transporter inhibition Mechanism
Assertion facts are collected from DrugBank
**** No Transporter inhibition Mechanism is found ****

Transporter induction Mechanism
Assertion facts are collected from DrugBank
**** No Transporter induction Mechanism is found ****

Multi-pathways Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
**** No Multi-pathways Mechanism is found ****
```

2. **Metabolism induction-based interaction:**

Drug A: Rivaroxaban
Drug B: Rifampin
Result: Increase the risk of stroke
Clinical Evidence: http://www.ncbi.nlm.nih.gov/pubmed/23136913
D3 output:

```
Starting D3 Inferential Query Model

Check the interaction between [Rivaroxaban] and [Rifampin]

Check if Drugs are in UMLS database
First Drug CUI is: C1739768
```

Second Drug CUI is: C0035608

Check if Drugs are in D3 knowledge base
[Rivaroxaban] and [Rifampin] are in the D3 knowledge base

DDI sources report DDI:
[NDF-RT, Kegg]

**** This DDI is proven ****
**** Mechanisms contribute for DDI ****

Pharmacogenetic Mechanism
Assertion facts are collected from pharmgkb the pharmacogenomics knowledge base
**** No Pharmacogenetic Mechanism is found ****

Protein-Binding Mechanism
Assertion facts are collected from Drugbank database
[There is possible interaction due to high protein binding affinity by Rivaroxaban]

Additive Mechanism
Assertion facts are collected from The National Drug File - Reference Terminology (NDF-RT) and NCI Thesaurus - National Institutes of Health
**** No Additive Mechanism is found ****

Competition Mechanism
Assertion facts are collected from Gene Ontology (GO), NCI Thesaurus - National Institutes of Health and National Center for Biotechnology Information (NCBI)
**** No Competition Mechanism is found ****

Metabolism inhibition Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
**** No Metabolism inhibition Mechanism is found ****

Metabolism induction Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
[CYP3A4 gene, CYP3A5 gene]
The induction happens by Rifampin

Transporter inhibition Mechanism

```
Assertion facts are collected from DrugBank
[ABCB1 gene]
The inhibition happens by Rifampin

Transporter induction Mechanism
Assertion facts are collected from DrugBank
[ABCB1 gene]
The induction happens by Rifampin

Multi-pathways Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
[CYP3A4 gene, ABCB1 gene, CYP3A5 gene]
```

3. **Metabolism inhibition-based interaction:**

Drug A: Warfarin
Drug B: Erythromycin
Result: Lead to a consequent risk because of warfarin anti-clotting
Clinical Evidence: http://www.ncbi.nlm.nih.gov/pubmed/12778089
D3 output:

```
Starting D3 Inferential Query Model

Check the interaction between [Warfarin] and [Erythromycin]

Check if Drugs are in UMLS database
First Drug CUI is: C0043031
Second Drug CUI is: C0014806

Check if Drugs are in D3 knowledge base
[Warfarin] and [Erythromycin] are in the D3 knowledge base

DDI sources report DDI:
[NDF-RT, Twosides, Drugbank, OSCAR]

**** This DDI is proven ****
**** Mechanisms contribute for DDI ****

Pharmacogenetic Mechanism
Assertion facts are collected from pharmgkb the pharmacogenomics
```

knowledge base
**** No Pharmacogenetic Mechanism is found ****

Protein-Binding Mechanism
Assertion facts are collected from Drugbank database
[There is possible interaction due to high protein binding affinity by Warfarin]

Additive Mechanism
Assertion facts are collected from The National Drug File - Reference Terminology (NDF-RT) and NCI Thesaurus - National Institutes of Health
**** No Additive Mechanism is found ****

Competition Mechanism
Assertion facts are collected from Gene Ontology (GO), NCI Thesaurus - National Institutes of Health and National Center for Biotechnology Information (NCBI)
**** No Competition Mechanism is found ****

Metabolism inhibition Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
[CYP1A2 gene, CYP3A4 gene]
The inhibition happens by Erythromycin

Metabolism induction Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
**** No Metabolism induction Mechanism is found ****

Transporter inhibition Mechanism
Assertion facts are collected from DrugBank
**** No Transporter inhibition Mechanism is found ****

Transporter induction Mechanism
Assertion facts are collected from DrugBank
**** No Transporter induction Mechanism is found ****

Multi-pathways Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
**** No Multi-pathways Mechanism is found ****

4. **Transporter induction-based interaction:**

Drug A: Digoxin
Drug B: Rifampin
Result: Avoid combination
Clinical Evidence: http://www.ncbi.nlm.nih.gov/pubmed/10411543
D3 output:

```
Starting D3 Inferential Query Model

Check the interaction between [Digoxin] and [Rifampin]

Check if Drugs are in UMLS database
First Drug CUI is: C0012265
Second Drug CUI is: C0035608

Check if Drugs are in D3 knowledge base
[Digoxin] and [Rifampin] are in the D3 knowledge base

DDI sources report DDI:
[NDF-RT, Twosides, Kegg, DDI-Corpus-2011, DDI-Corpus-2013, NLM-Corpus]

**** This DDI is proven ****
**** Mechanisms contribute for DDI ****

Pharmacogenetic Mechanism
Assertion facts are collected from pharmgkb the pharmacogenomics
knowledge base
[rs1045642]

Protein-Binding Mechanism
Assertion facts are collected from Drugbank database
**** No Protein-Binding Mechanism is found  ****

Additive Mechanism
Assertion facts are collected from The National Drug File - Reference
Terminology (NDF-RT) and NCI Thesaurus - National Institutes of Health
**** No Additive Mechanism is found  ****

Competition Mechanism
Assertion facts are collected from Gene Ontology (GO), NCI Thesaurus -
National Institutes of Health and National Center for Biotechnology
Information (NCBI)
**** No Competition Mechanism is found  ****
```

```
Metabolism inhibition Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
**** No Metabolism inhibition Mechanism is found ****

Metabolism induction Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
[CYP3A4 gene]
The induction happens by Rifampin

Transporter inhibition Mechanism
Assertion facts are collected from DrugBank
[ABCB1 gene, ABCB11 gene, SLCO1A2 gene, SLCO1B1 gene, SLCO2B1 gene,
SLCO1B3 gene]
The inhibition happens by Rifampin

Transporter induction Mechanism
Assertion facts are collected from DrugBank
[ABCB1 gene]
Both drugs can cause the interaction Digoxin AND Rifampin

Multi-pathways Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
[CYP3A4 gene, ABCB1 gene, ABCB11 gene, SLCO1A2 gene, SLCO1B1 gene,
SLCO2B1 gene, SLCO1B3 gene]
```

5. **Transporter inhibition-based interaction:**

Drug A: Digoxin
Drug B: Quinidine
Result: Quinidine showed to significantly increase the serum level of digoxin; thus serious and complicated DDI can be occurred.
Clinical Evidence: http://www.ncbi.nlm.nih.gov/pubmed/23686349
D3 output:

```
Starting D3 Inferential Query Model
```

```
Check the interaction between [Digoxin] and [Quinidine]

Check if Drugs are in UMLS database
First Drug CUI is: C0012265
Second Drug CUI is: C0034414

Check if Drugs are in D3 knowledge base
[Digoxin] and [Quinidine] are in the D3 knowledge base

DDI sources report DDI:
[NDF-RT, Drugbank, DDI-Corpus-2011, DDI-Corpus-2013, NLM-Corpus,
OSCAR, SemMedDB]

**** This DDI is proven ****
**** Mechanisms contribute for DDI ****

Pharmacogenetic Mechanism
Assertion facts are collected from pharmgkb the pharmacogenomics
knowledge base
**** No Pharmacogenetic Mechanism is found  ****

Protein-Binding Mechanism
Assertion facts are collected from Drugbank database
**** No Protein-Binding Mechanism is found  ****

Additive Mechanism
Assertion facts are collected from The National Drug File - Reference
Terminology (NDF-RT) and NCI Thesaurus - National Institutes of Health
**** No Additive Mechanism is found  ****

Competition Mechanism
Assertion facts are collected from Gene Ontology (GO), NCI Thesaurus -
National Institutes of Health and National Center for Biotechnology
Information (NCBI)
**** No Competition Mechanism is found  ****

Metabolism inhibition Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
[CYP3A4 gene]
The inhibition happens by Quinidine

Metabolism induction Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
```

```
National Institutes of Health
[CYP3A4 gene]
The induction happens by Quinidine

Transporter inhibition Mechanism
Assertion facts are collected from DrugBank
[ABCB1 gene, ABCB11 gene, SLCO1A2 gene, SLCO1B1 gene, SLC22A8 gene]
Both drugs can cause the interaction Digoxin AND Quinidine

Transporter induction Mechanism
Assertion facts are collected from DrugBank
[ABCB1 gene]
The induction happens by Digoxin

Multi-pathways Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
[CYP3A4 gene, ABCB1 gene, ABCB11 gene, SLCO1A2 gene, SLCO1B1 gene,
SLC22A8 gene]
```

6. **Multi-pathway-based interaction:**

Drug A: Cyclosporine
Drug B: Atorvastatin
Result: Cyclosporine inhabits both the metabolism and transporter of statin.
Clinical Evidence: http://www.ncbi.nlm.nih.gov/pubmed/12558459
D3 output:

```
Starting D3 Inferential Query Model

Check the interaction between [Cyclosporine] and [Atorvastatin]

Check if Drugs are in UMLS database
First Drug CUI is: C0010592
Second Drug CUI is: C0286651

Check if Drugs are in D3 knowledge base
[Cyclosporine] and [Atorvastatin] are in the D3 knowledge base

DDI sources report DDI:
```

[NDF-RT, Twosides, Drugbank, Kegg, NLM-Corpus, PK-Corpus]

**** This DDI is proven ****
**** Mechanisms contribute for DDI ****

Pharmacogenetic Mechanism
Assertion facts are collected from pharmgkb the pharmacogenomics knowledge base
[rs1128503, rs2032582, rs776746]

Protein-Binding Mechanism
Assertion facts are collected from Drugbank database
[There is possible interaction due to high protein binding affinity by Atorvastatin]

Additive Mechanism
Assertion facts are collected from The National Drug File - Reference Terminology (NDF-RT) and NCI Thesaurus - National Institutes of Health
**** No Additive Mechanism is found ****

Competition Mechanism
Assertion facts are collected from Gene Ontology (GO), NCI Thesaurus - National Institutes of Health and National Center for Biotechnology Information (NCBI)
**** No Competition Mechanism is found ****

Metabolism inhibition Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
[CYP3A4 gene, CYP2C9 gene, CYP2C19 gene, CYP2C8 gene, CYP2D6 gene]
Both drugs can cause the interaction Cyclosporine AND atorvastatin

Metabolism induction Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
[CYP3A4 gene, CYP2C9 gene, CYP3A7 gene, CYP3A5 gene]
The induction happens by Cyclosporine

Transporter inhibition Mechanism
Assertion facts are collected from DrugBank
[ABCB1 gene, SLCO1A2 gene, SLCO1B1 gene, ABCC2 gene, ABCC1 gene]
The inhibition happens by Cyclosporine

Transporter induction Mechanism

```
Assertion facts are collected from DrugBank
[ABCB1 gene]
The induction happens by Cyclosporine

Multi-pathways Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
[CYP3A4 gene, ABCB1 gene, ABCC2 gene, ABCC1 gene, SLCO1A2 gene,
SLCO1B1 gene, CYP2C9 gene, CYP2C19 gene, CYP3A7 gene, CYP3A5 gene,
CYP2C8 gene, CYP2D6 gene]
```

7. **Competitive-based interaction:**

Drug A: Propranolol
Drug B: Albuterol
Result: Propranolol and albuterol target the beta-2 receptors, but propranolol blocks the beta-2 receptors while albuterol catalyzes them; thus DDI occurs.
Clinical Evidence: http://www.ncbi.nlm.nih.gov/pubmed/1233216
D3 output:

```
Starting D3 Inferential Query Model

Check the interaction between [Propranolol] and [Albuterol]

Check if Drugs are in UMLS database
First Drug CUI is: C0033497
Second Drug CUI is: C0001927

Check if Drugs are in D3 knowledge base
[Propranolol] and [Albuterol] are in the D3 knowledge base

DDI sources report DDI:
[Twosides, Drugbank, OSCAR]

**** This DDI is proven ****
**** Mechanisms contribute for DDI ****

Pharmacogenetic Mechanism
Assertion facts are collected from pharmgkb the pharmacogenomics
knowledge base
**** No Pharmacogenetic Mechanism is found  ****
```

Protein-Binding Mechanism
Assertion facts are collected from Drugbank database
**** No Protein-Binding Mechanism is found ****

Additive Mechanism
Assertion facts are collected from The National Drug File - Reference Terminology (NDF-RT) and NCI Thesaurus - National Institutes of Health
**** No Additive Mechanism is found ****

Competition Mechanism
Assertion facts are collected from Gene Ontology (GO), NCI Thesaurus - National Institutes of Health and National Center for Biotechnology Information (NCBI)
[Receptors, Adrenergic, beta-1, beta-2 Adrenergic Receptors]

Metabolism inhibition Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
[CYP3A4 gene]
The inhibition happens by Albuterol

Metabolism induction Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
**** No Metabolism induction Mechanism is found ****

Transporter inhibition Mechanism
Assertion facts are collected from DrugBank
**** No Transporter inhibition Mechanism is found ****

Transporter induction Mechanism
Assertion facts are collected from DrugBank
**** No Transporter induction Mechanism is found ****

Multi-pathways Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
**** No Multi-pathways Mechanism is found ****

8. **Additive-based interaction:**

Drug A: Glibenclamide
Drug B: Metformin
Result: the combination of glibenclamide and metformin could result in hypoglycemic.
Clinical Evidence: http://www.ncbi.nlm.nih.gov/pubmed/21084384
D3 output:

```
Starting D3 Inferential Query Model

Check the interaction between [Glibenclamide] and [Metformin]

Check if Drugs are in UMLS database
First Drug CUI is: C0017628
Second Drug CUI is: C0025598

Check if Drugs are in D3 knowledge base
[Glyburide] and [Metformin] are in the D3 knowledge base

DDI sources report DDI:
[Twosides, Kegg, Clinicaltrials]

**** This DDI is proven ****
**** Mechanisms contribute for DDI ****

Pharmacogenetic Mechanism
Assertion facts are collected from pharmgkb the pharmacogenomics
knowledge base
**** No Pharmacogenetic Mechanism is found  ****

Protein-Binding Mechanism
Assertion facts are collected from Drugbank database
**** No Protein-Binding Mechanism is found  ****

Additive Mechanism
Assertion facts are collected from The National Drug File - Reference
Terminology (NDF-RT) and NCI Thesaurus - National Institutes of Health
[Insulin Receptor Agonists [MoA]]

Competition Mechanism
Assertion facts are collected from Gene Ontology (GO), NCI Thesaurus -
National Institutes of Health and National Center for Biotechnology
Information (NCBI)
```

```
**** No Competition Mechanism is found  ****

Metabolism inhibition Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
**** No Metabolism inhibition Mechanism is found ****
Metabolism induction Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
**** No Metabolism induction Mechanism is found ****

Transporter inhibition Mechanism
Assertion facts are collected from DrugBank
**** No Transporter inhibition Mechanism is found ****

Transporter induction Mechanism
Assertion facts are collected from DrugBank
**** No Transporter induction Mechanism is found ****

Multi-pathways Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus -
National Institutes of Health
**** No Multi-pathways Mechanism is found ****
```

9. **Pharmacogenetic-based interaction:**

Drug A: Ibuprofen
Drug B: Rofecoxib
Result: COX-1 gene (PTGS1) and the COX-2 gene (PTGS2) polymorphism could lead to inefficacy of the both drugs**.**
Clinical Evidence: http://www.ncbi.nlm.nih.gov/pubmed/16678543
D3 output:

```
Starting D3 Inferential Query Model

Check the interaction between [Ibuprofen] and [Rofecoxib]

Check if Drugs are in UMLS database
First Drug CUI is: C0020740
```

Second Drug CUI is: C0762662

Check if Drugs are in D3 knowledge base
[Ibuprofen] and [Rofecoxib] are in the D3 knowledge base

DDI sources report DDI:
[Twosides]

**** This DDI is proven ****
**** Mechanisms contribute for DDI ****

Pharmacogenetic Mechanism
Assertion facts are collected from pharmgkb the pharmacogenomics knowledge base
[rs20417]

Protein-Binding Mechanism
Assertion facts are collected from Drugbank database
[There is possible interaction due to high protein binding affinity by Ibuprofen]

Additive Mechanism
Assertion facts are collected from The National Drug File - Reference Terminology (NDF-RT) and NCI Thesaurus - National Institutes of Health
**** No Additive Mechanism is found ****

Competition Mechanism
Assertion facts are collected from Gene Ontology (GO), NCI Thesaurus - National Institutes of Health and National Center for Biotechnology Information (NCBI)
**** No Competition Mechanism is found ****

Metabolism inhibition Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
[PTGS2 gene]
Both drugs can cause the interaction Ibuprofen AND rofecoxib

Metabolism induction Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
**** No Metabolism induction Mechanism is found ****

Transporter inhibition Mechanism

Assertion facts are collected from DrugBank
[ABCC4 gene]
Both drugs can cause the interaction Ibuprofen AND rofecoxib

Transporter induction Mechanism
Assertion facts are collected from DrugBank
**** No Transporter induction Mechanism is found ****

Multi-pathways Mechanism
Assertion facts are collected from DrugBank and NCI Thesaurus - National Institutes of Health
[PTGS1 gene, ABCC4 gene, CYP2C9 gene, CYP2C8 gene]

APPENDIX C

IDENTIFYING LEVELS OF AGREEMENT AMONG DDI RESOURCES USING JACCARD INDEX

In this appendix, we prove the large disagreements among current DDI resources in reporting the interactions by computing the Jaccard Index. In particular, fifteen DDI resources used by this thesis are being examined as pairs, which result in 105 comparisons. These DDI resources are: (1) Clinicaltrials.gov, (2) NDF-RT, (3) DIKB, (4) ONC-HighPriority, (5) CredibleMeds, (6) KEGG, (7) Drugbank, (8) Twosides, (9) DDI-Corpus-2011, (10) SemMedDB, (11) PK-Corpus, (12) DDI-Corpus-2013, (13) NLM-Corpus, (14) OSCAR, and (15) ONC-NonInteruptive. In Chapter VII we computed the averaged Jaccard Index among them; in this appendix on the other hand, for each of these 15 resources, we computed the Jaccard Index between it and the other 14 resources. **Table C.1** shows the results of the comparisons.

The average Jaccard Index is 1.7%. **See Table C.1** for the Jaccard Index for each of these resources. Such a low average Jaccard Index confirms that across a

large number of DDI resources, there is no level of agreement between DDI resources in reporting the interactions.

Table C1: Jaccard Index for 15 DDI resources

DDI Resources	Jaccard Result
Clinicaltrials.gov, NDF-RT	0.0019998518
DIKB, ONCHighPriority	0.019677997
CredibleMeds, KEGG	0.0010233863
Clinicaltrials.gov, Drugbank	0.00401997
CredibleMeds, Twosides	9.312457E-4
DDI-Corpus-2011, SemMedDB	0.014833127
SemMedDB, Twosides	0.0054883915
DDI-Corpus-2011, Twosides	0.0028405078
ONC-HighPriority, SemMedDB	0.0025673942
CredibleMeds, ONC-HighPriority	0.006535948
OSCAR, SemMedDB	0.01137789
PK-Corpus, DDI-Corpus-2011	0.039660055
DIKB, DDI-Corpus-2011	0.022644928
Clinicaltrials, DDI-Corpus-2013	0.0038712306
NLM-Corpus, ONC-HighPriority	0.0064888247
Drugbank, DDI-Corpus-2013	0.035429653
PK-Corpus,NLM-Corpus	0.067532465
NLM-Corpus, OSCAR	0.0045209094
OSCAR, Twosides	0.017258156
Clinicaltrials.gov, PK-Corpus	0.0015772871
DDI-Corpus-2013, OSCAR	0.012665322
KEGG, Twosides	0.043419063
PK-Corpus, Drugbank	0.0062866723
PK-Corpus, KEGG	0.0029149
Drugbank, Twosides	0.02306813
NDF-RT, ONC-NonInteruptive	0.006902357
NDF-RT, DDI-Corpus-2011	0.016929364
DDI-Corpus-2013, Twosides	0.005336694
Clinicaltrials.gov, SemMedDB	0.0050384887
Drugbank, DDI-Corpus-2011	0.017548176
DDI-Corpus-2011, ONC-HighPriority	0.007620164
CredibleMeds, DIKB	0.033816423
PK-Corpus, NDF-RT	0.005202081
PK-Corpus, DIKB	0.022566997
KEGG, NLM-Corpus	0.004004836

NDF-RT, SemMedDB	0.005846638
KEGG, ONC-HighPriority	0.0058201253
ONC-NonInteruptive, Twosides	0.0073923487
Drugbank, NLM-Corpus	0.009437902
Clinicaltrials.gov, OSCAR	0.0010539258
Drugbank, KEGG	0.058781102
ONC-HighPriority, OSCAR	0.0049667004
Drugbank, ONC-NonInteruptive	0.013101391
PK-Corpus, DDI-Corpus-2013	0.03653295
Drugbank, ONC-HighPriority	0.02525651
DDI-Corpus-2013, ONC-NonInteruptive	0.0014898689
DIKB,KEGG	0.0056852186
Drugbank, OSCAR	0.025489375
Clinicaltrials.gov, CredibleMeds	0.0
KEGG, SemMedDB	0.01240762
DIKB, OSCAR	8.427643E-4
PK-Corpus, OSCAR	0.002786221
CredibleMeds, Drugbank	0.0048040454
DDI-Corpus-2013, SemMedDB	0.024234693
DIKB, Drugbank	0.015477848
NDF-RT, DDI-Corpus-2013	0.029506685
ONC-HighPriority, ONC-NonInteruptive	6.1977067E-4
NLM-Corpus, ONC-NonInteruptive	8.609557E-4
CredibleMeds, OSCAR	0.0029441884
CredibleMeds, SemMedDB	0.0
CredibleMeds, ONC-NonInteruptive	0.0018544274
NDF-RT, OSCAR	0.019068532
CredibleMeds, DDI-Corpus-2013	0.0036791759
KEGG, DDI-Corpus-2013	0.015372168
NLM-Corpus, Twosides	0.0016068813
Clinicaltrials.gov, ONC-NonInteruptive	0.0028030833
PK-Corpus, SemMedDB	0.0032529142
DDI-Corpus-2011, ONC-NonInteruptive	0.0015128592
KEGG, DDI-Corpus-2011	0.007906768
Drugbank, NDF-RT	0.116289824
KEGG, ONC-NonInteruptive	0.018135551
ONC-HighPriority, Twosides	0.0015532051
PK-Corpus, ONC-HighPriority	7.61035E-4
Clinicaltrials.gov, KEGG	0.0064808596
DDI-Corpus-2013, ONC-HighPriority	0.00954753
NDF-RT, NLM-Corpus	0.0071591926
DIKB, Twosides	0.004970351
NLM-Corpus, SemMedDB	0.009340806
NDF-RT, KEGG	0.04303367

DIKB, NLM-Corpus	0.010025063
DIKB, SemMedDB	4.885198E-4
CredibleMeds, NDF-RT	0.0050428645
DIKB, NDF-RT	0.012703646
DDI-Corpus-2011, DDI-Corpus-2013	0.40653494
PK-Corpus, Twosides	0.0012930491
NDF-RT, ONC-HighPriority	0.03343949
ONC-NonInteruptive, SemMedDB	0.0014267879
KEGG, OSCAR	0.025362536
Clinicaltrials.gov, DIKB	4.758506E-4
NDF-RT, Twosides	0.015422168
DDI-Corpus-2011, NLM-Corpus	0.06258149
CredibleMeds, PK-Corpus	0.0
ONC-NonInteruptive, OSCAR	0.09597592
Clinicaltrials.gov, Twosides	0.0070062545
Clinicaltrials, ONC-HighPriority	2.0859408E-4
DDI-Corpus-2011, OSCAR	0.00823843
Clinicaltrials, NLM-Corpus	0.0015444015
DIKB, DDI-Corpus-2013	0.019933555
PK-Corpus, ONC-NonInteruptive	4.4583148E-4
DIKB, ONC-NonInteruptive	0.0
DDI-Corpus-2013, NLM-Corpus	0.055248618
CredibleMeds, DDI-Corpus-2011	0.00618238
CredibleMeds, NLM-Corpus	0.009230769
Clinicaltrials, DDI-Corpus-2011	0.0016638935
Drugbank, SemMedDB	0.009984235