

TRANSLATIONAL HIGH-DIMENSIONAL DRUG INTERACTION DISCOVERY AND VALIDATION
USING HEALTH RECORD DATABASES AND PHARMACOKINETICS MODELS

Chien-Wei Chiang

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the School of Informatics and Computing,

Indiana University

January 2018

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Huanmei Wu, Ph.D., Co-Chair

Lang Li, Ph.D., Co-Chair

Yunlong Liu, Ph.D.

October 31, 2017

Xiaowen Liu, Ph.D.

Acknowledgements

I would not have completed my Ph.D. dissertation without the tremendous supports of many people who have helped me in this complex research project.

First and the most, I would like to express my appreciation to my family, especially to my wife. I remembered how much support you gave me in the past six years during my graduate study.

I especially thank to my mentor, Dr. Lang Li, who provided me an opportunity to join his laboratory as programmer, and who supports me to be a Ph.D. candidate. Without his precious support it would not be possible to conduct this research.

I would also like to thank the members on my thesis committee, Dr. Huanmei Wu, Dr. Yunlong Liu and Dr. Xiaowen Liu. They provide critical evaluations and crucial suggestions to develop my dissertation.

Besides my committee members, I would like to give thanks to Dr. Jamie Renbarger and her laboratory members. It is been great to work with her on various pediatric cancer researches. I would also like to thank for Guanglong Jiang, Dr. Pengyue Zhang, Dr. Heng-yi Wu, Dr. Zhiping Wang, Dr. Lei Wang, Dr. Xu Han, Dr. Sara Quinney, Shijun Zhang, and all the laboratory members who have been helpful in many ways.

And finally, also to everyone in the Center for Computational Biology and Bioinformatics; it was great to work with all of you during last six years.

Chien-Wei Chiang

TRANSLATIONAL HIGH-DIMENSIONAL DRUG INTERACTION DISCOVERY AND VALIDATION
USING HEALTH RECORD DATABASES AND PHARMACOKINETICS MODELS

Polypharmacy leads to increased risk of drug-drug interactions (DDI's). In this dissertation, we create a database for quantifying fraction of metabolism (fm) of CYP450 isozymes for FDA approved drugs. A reproducible data collection protocol was developed to extract key information from publicly available in vitro selective CYP enzyme inhibition studies. The fm was then estimated from the curated data. Then, proposed a random control selection approach for nested case-control design for electronic health records (HER) and electronic medical records (EMR) databases. By relaxing the matching by case's index time restriction, random control dramatically reduces the computational burden compared with traditional control selection approaches. Using the Observational Medical Outcomes Partnership gold standard and an EMR database, random control is demonstrated to have better performances as well. Finally, combining epidemiological studies and pharmacokinetic modeling with fm database, we detected and evaluated high-dimensional drug-drug interactions among thirty high frequency drugs. Multi-drug combinations that increased risk of myopathy were identified in the FAERS and EMR databases by a mixture drug-count response model (MDCM) model. Twenty-eight 3-way and 43 4-way DDI's increased ratio of area under plasma concentration-time curve (AUCR) >2-fold and had significant myopathy risk in both databases. The predicted AUCR of omeprazole in the presence of fluconazole and clonidine was 9.35; and increased risk of myopathy was 6.41 (LFDR =

0.002) in FAERS and 18.46 (LFDR = 0.005) in EMR. We demonstrate that combining health record informatics and pharmacokinetic modeling is a powerful translational approach to detect high-dimensional DDI's.

Huanmei Wu, Ph.D., Co-Chair

Lang Li, Ph.D., Co-Chair

Table of Contents

| | |
|----------------------------------------------------------------------------------------|----|
| List of Tables | x |
| List of Figures | xi |
| Chapter 1. Introduction | 1 |
| 1.1 Adverse Drug event and Drug-drug interaction..... | 1 |
| 1.2 Fraction of metabolism and CYP P450 | 1 |
| 1.3 Evaluate Adverse Drug Event signal | 3 |
| 1.3.1 Evaluation method | 3 |
| 1.3.2 Epidemiology Designs..... | 4 |
| 1.4 Evaluate Drug-drug interaction..... | 5 |
| 1.4.1 Clinical pharmacokinetic studies | 5 |
| 1.4.2 Epidemiology and drug-drug-ADE associations | 6 |
| 1.4.3 In-vitro In-vivo extrapolation (IVIVE) pharmacokinetics model | 7 |
| 1.5 Proposed Solutions..... | 8 |
| 1.5.1 Scope of Aim 1: Drug fm database | 8 |
| 1.5.2 Scope of Aim 2: Random control approach based nested case-control design..... | 8 |
| 1.5.3 Scope of Aim 3: High-dimensional drug interaction | 9 |
| 1.6 Impact of Project | 9 |

| | |
|-------------------------------------------------------------------------------------------------------|----|
| 1.6.1 Impact of Aim 1 | 9 |
| 1.6.2 Impact of Aim 2 | 10 |
| 1.6.3 Impact of Aim 3 | 10 |
| Chapter 2. The Cancer Drug Fraction of Metabolism Database..... | 12 |
| 2.1 Introduction..... | 12 |
| 2.2 Construction and Content | 18 |
| 2.3 Utility | 26 |
| 2.4 Discussion | 32 |
| Chapter 3. A Computationally Efficient Design for Electronical Health/Medical Records Mining | 34 |
| 3.1 Introduction..... | 34 |
| 3.2 Methods | 38 |
| 3.2.1 Data source..... | 38 |
| 3.2.2 The OMOP Gold Standard | 39 |
| 3.2.3 Study Designs..... | 39 |
| 3.2.3.1 Case selection | 40 |
| 3.2.3.2 No Control Selection (All patients) | 40 |
| 3.2.3.3 Super Control and Dynamic control Selection..... | 40 |
| 3.2.3.4 Random control Selection..... | 41 |

| | |
|-----------------------------------------------------------------------------------|----|
| 3.2.4 Comparison analysis | 41 |
| 3.3 Results | 45 |
| 3.3.1 Performance evaluation under common signal detection rules..... | 45 |
| 3.3.2 Area under the receiver operating characteristic curve (AUC) analysis | 45 |
| 3.4 Conclusion and Discussion | 50 |
| Chapter 4. Translational high-dimensional drug Interaction discovery and | |
| validation using health record databases and pharmacokinetics models | 52 |
| 4.1 Introduction..... | 52 |
| 4.2 Materials and Methods..... | 56 |
| 4.2.1 Indiana Network of Patient Care (INPC) Electronic Medical Record..... | 56 |
| 4.2.2 FDA Adverse Event Reporting System (FAERS) | 57 |
| 4.2.3 Drug Selection Criteria and Data Curation | 57 |
| 4.2.4 A Mixture Drug-Count-Response Model for the High-Dimensional Drug | |
| Effect on Myopathy | 58 |
| 4.2.5 Sensitivity Data Analysis | 61 |
| 4.2.6 High Dimensional Drug Interaction In-Vitro to In-Vivo Extrapolation | |
| (IVIVE) | 61 |
| 4.3 Results | 62 |
| 4.3.1 Drug Selections | 62 |

| | |
|--------------------------------------------------------------------------------------------------------------|----|
| 4.3.2 INPC-CDM and FAERS Show Strikingly Different Drug-Count Myopathy Response Mixture Models | 63 |
| 4.3.3 Overlapping Drug Combinations in INPC-CDM and FAERS | 67 |
| 4.3.4 The Overall Trend of High Dimensional Drug Interactions in Pharmacokinetics Predicted by IVIVE..... | 67 |
| 4.3.5 Significant Three-way Drug Interactions and Their Sensitivity Analyses | 68 |
| 4.4 Discussion | 73 |
| Chapter 5. Conclusion | 81 |
| References | 84 |
| Curriculum Vitae | |

List of Tables

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 2.1 The validation Results | 25 |
| Table 2.2 The validation results of DDI prediction based on PBPK model | 28 |
| Table 3.1 2-by-2 contingency table for a drug-ADE pair..... | 42 |
| Table 3.2 Summary statistics for each ADE (each patient can have multiple new case events) | 43 |
| Table 3.3 PRR, ROR and IC | 44 |
| Table 3.4 Performances of different control selection approaches..... | 47 |
| Table 4.1 Myopathy phenotype definition | 60 |
| Table 4.2 Top 28 3-way drugs interaction combination. G12 and G23 means the p-value of odds ratio compared taken 1 of 3 drugs vs 2 of 3 drugs, and taken 2 of 3 drugs and 3 of 3 drugs. a. Combine analysis: treated omeprazole and esomeprazole as same drug. | 72 |
| Table 4.3 Drugs fm and fe | 77 |

List of Figures

| | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 2.1 The flow chart of data collection procedure | 20 |
| Figure 3.1 AUC values for four control selection strategies | 48 |
| Figure 3.2 AUC curves for all patients analysis and random control..... | 49 |
| Figure 4.1 Drugs selection flow chart | 65 |
| Figure 4.2 Mixture dose-response model curve for both INPC and FAERS data. Red dash line is the baseline risk estimated; Blue dash line is maximum risk estimated. | 66 |
| Figure 4.3 Overlapped drug combination between INPC and FAERS dataset. * LFDR < 0.05..... | 69 |
| Figure 4.4 Simulated AUOCR of 2-way to 5-way drugs combination. Only AUOCR>=1.25 are reported here. | 70 |
| Figure 4.5 Group-wise odds ratio analysis; OMZ is omeprazole or esomeprazole; CLN is clonidine; FLU is fluconazole | 71 |
| Figure 4.6 hepatic metabolism drugs interaction between omeprazole, clonidine, and fluconazole..... | 76 |

Chapter 1. Introduction

1.1 Adverse Drug event and Drug-drug interaction

Adverse drug events (ADEs) are considered to be a significant challenge for current clinical practices. Drug-drug interactions (DDIs) are a common cause of adverse drug events (ADE) (Hajjar, Cafiero, & Hanlon, 2007; U.S. Department of Health and Human Services; L. Zhang, Zhang, Zhao, & Huang, 2009). In the United States alone, each year an estimated 195,000 hospitalizations and 74,000 emergency room visits are the result of DDIs. National Health and Nutrition Examination Survey data published in 2010 (Gu, Dillon, & Burt, 2010) showed that patients, who used two or more prescription drugs, increased from 25.4% to 31.2% in ten years since 1999. In particular, among the older population, more than 64% took three or more prescription drugs, and 37% took five or more prescription drugs (Gu et al., 2010). In FDA Adverse Event Reporting System (FAERS) data, there are 35% reports that include three or more prescription drugs in each report, and 20% reports have five or more prescription drugs. Though ADEs can be detected in either pre-marketing clinical trials or post-marketing surveillances, most ADE knowledges are revealed in the post-marketing stage. This is because post-marketing stage allows larger population and prolonged follow up (Harpaz et al., 2012). Additionally, unlike pre-marketing clinical trials, drugs are administered without stringent inclusion/exclusion criteria.

1.2 Fraction of metabolism and CYP P450

There are some widely-used database for drug metabolism: a) DrugBank (Wishart et al., 2006), which is a comprehensive database which combines detailed drug

(i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information; b) Transformer (the former Super CYP) (Preissner et al., 2010), which integrates Cytochrome P450 enzyme interactions and some pharmacological information; c) DIDB (Hachad, Ragueneau-Majlessi, & Levy, 2010), which can evaluate the impact of DDI in the clinic by in vitro and in vivo DDI data, few fm data can be found.

The contributions of CYPs for a drug's metabolism can be estimated as the change in AUC or clearance in the absence and presence of a co-administered selective inhibitor through an in vivo approach (Creighton et al., 2008; Le Coutre et al., 2008). fm can be also estimated via a pharmacogenetics study where it can be calculated from the fold-change in exposure of a victim drug in extensive metabolizers (EMs) compared to poor metabolizers (PMs) (Ito, Hallifax, Obach, & Houston, 2005). Third, human ¹⁴C or ³H-ADME studies, which measure the concentration of the radiolabeled unchanged drug and its metabolites in plasma, urine and feces, is also regarded as a valuable approach and clinical pharmacokinetics study to estimate the metabolic pathways of a drug (Bohnert et al., 2016; Rodrigues, Winchell, & Dobrinska, 2001). Several other in vitro methods have been developed to determine the contribution of individual cytochrome P450 isozymes in a drug's metabolism. Substrate depletion in human liver microsomes (HLM) is one method that the drug is incubated with or without specific CYP450 selective inhibitors (Huskey, Dean, Miller, Rasmusson, & Chiu, 1995). Comparing to the metabolism rate, V_{max}/K_m , of the substrate without any inhibitor, the percent inhibition of this CYP pathway by the CYP-selective chemical inhibitor reflects the

contribution of this CYP towards the substrate's metabolism. Ideally chemical inhibitors should be potent, selective and metabolically stable. Substrate depletion can also be incubated with individual recombinant enzymes isoforms (Z. M. Li, Guo, & Ren, 2016). This approach estimates the metabolism rate, V_{max}/K_m , of the substrate in recombinant human (rh) CYP isoforms, and scales the rhCYP V_{max}/K_m to HLM CL_{int} via a RAF/ISEF approach (Bohnert et al., 2016). Each isozymes contribution is estimated as the percent contribution of each CYP enzyme towards the total $HLMCL_{int}$.

1.3 Evaluate Adverse Drug Event signal

1.3.1 Evaluation method

For a drug-ADE pair of interest, either univariate analyses or multivariate analyses can be used for signal detection. Under univariate analyses, available samples are usually summarized by a 2-by-2 contingency table, in which contains the frequencies classified by the usage of the drug (yes/no) and the occurrence of the ADE (yes/no). The outcome is the frequency that this drug-ADE pair is co-occurred, and the expectation is the expected frequency of this drug-ADE pair under the assumption of no association. Univariate analyses are also known as disproportion analysis (DPA), as they quantify ADE signals by the outcome to expectation ratio (i.e. relative risk) or its variants. Frequentist DPAs include proportional reporting ratio (PRR) and reporting odds ratio (ROR) (Evans, Waller, & Davis, 2001; van Puijenbroek et al., 2002). While, the empirical Bayesian geometric mean (EBGM) is an empirical Bayesian DPA and the information component (IC) is a Bayesian DPA (Bate et al., 1998; DuMouchel, 1999). DPAs are demonstrated to have promising performances and are computationally efficient (Harpaz et al., 2013).

However, confounding bias, especially confounding by co-medication, may cause biased disproportionality measurements of the true association. Typical multivariate analyses utilize multiple logistic regression (MLR) or regularized logistic regression (RLR) models to adjust potential confounding variables (i.e. co-medications). Besides this advantage, practices of MLR/RLR are often challenged by considerable larger sample sizes of pharmacovigilance databases. For instance, regular computers cannot fit MLR over millions of samples. In a summary, DPAs are powerful tools for large scale signal detection, as they can be applied to multiple ADEs together (Harpaz et al., 2012). Multivariate analyses that models an ADE and all co-medications are preferred for specific hypotheses testing.

1.3.2 Epidemiology Designs

In SRS, temporal information is omitted and each sample contains the ADE and drug status (Yes/No). Such a structure allows both multivariate analyses and DPAs to be applied directly to SRS. However, in EMR/HER database, patients are usually followed longitudinally with detailed temporal information about medications and phenotypes. As a consequence, EMR/EHR analyses typically require sophisticated epidemiology designs such as cohort design, self-controlled design, and nested case-control design (Hennessy et al., 2016). These epidemiology designs have been widely utilized for EMR/EHR analyses. Cohort design identifies exposure and non-exposure cohorts. ADE risks are estimated within each cohort, and the drug-ADE associations are estimated by the RR. The non-exposure cohort can be selected based on estimated propensity scores to ensure their similarities with the exposure cohort. For instance, Tatonetti *et al.*

assumes the latent cofounders can be characterized by co-medications, and propensity scores are then calculated by modeling the co-medications based on such assumption (Tatonetti, Ye, Daneshjou, & Altman, 2012). For the cases and their matched controls, nested case-control design uses a predefined window to examine drug exposures. For a case, usually 4 – 10 controls are selected. The controls are usually matched with the case's index date and risk factors (Schneeweiss, 2010). Additionally, they are case-free up to the case's index time. Thus, such controls are also named as at risk or dynamic control. If the controls are restricted to be case free for the entire follow up, this type of controls is known as super control. Some example of application of nested case-control designs includes Brauer *et al.*, La gamba *et al.* and Lee *et al.* (Brauer et al., 2014; La Gamba et al., 2017; Lee et al., 2011). In real applications, EMR/EHR databases may contain up to millions of patients. Such a sample size become an obstacle for the application of epidemiology designs to conduct large scale (i.e. drug wide and ADE wide) ADE signal detection. For instance, as we mentioned above, fitting propensity score model over millions of samples cannot be accomplished by regular computers. Moreover, in EMR/EHR analyses, propensity scores will be time dependent, which further increased the computational burden. Though nested case-control design do not require complicated modelling, the matching process still computationally expensive.

1.4 Evaluate Drug-drug interaction

1.4.1 Clinical pharmacokinetic studies

The translational significance of drug-drug interaction studies relies on both clinical and molecular pharmacology evidence. One salient example is that of breast

cancer hormonal therapy, tamoxifen. The formation of its active metabolite, endoxifen, was inhibited by concomitant selective serotonin reuptake inhibitor paroxetine in a clinical pharmacokinetics study (Stearns et al., 2003). In-vitro metabolism studies revealed that this is due to paroxetine's strong inhibition of the tamoxifen biotransformation to endoxifen via the CYP2D6 pathway (Desta, Ward, Soukhova, & Flockhart, 2004). In a follow-up pharmacogenetic study, breast cancer patients with CYP2D6 loss function variants have a higher risk of disease relapse and a lower incidence of hot flush (Goetz et al., 2005). The clinical consequence of treating breast cancer and depression using tamoxifen and SSRIs was reviewed (Henry, Stearns, Flockhart, Hayes, & Riba, 2008), and called for further investigation. Another example is the sedation agent midazolam. Co-administration of midazolam and ergosterol synthesis inhibitor ketoconazole has been identified to reduced subjects' cognitive function (Lam, Alfaro, Ereshefsky, & Miller, 2003). In clinical PK and in-vitro experiments, midazolam metabolism was inhibited by ketoconazole through the CYP3A pathway (Gascon & Dayer, 1991; Gorski, Hall, Jones, VandenBranden, & Wrighton, 1994), leading to increased midazolam exposure (Oikkola, Backman, & Neuvonen, 1994).

1.4.2 Epidemiology and drug-drug-ADE associations

As described by Hennessy and Flockhart (Hennessy & Flockhart, 2012), an integrated informatics, epidemiology, and pharmacology approach has the potential to accelerate the translational drug interaction studies. Pioneered by Tatonetti et al (Tatonetti et al., 2012), FAERS and electronic medical records were utilized to generate and validate Drug-ADE and drug-drug-ADE associations. In a follow-up study,

Lorberbaum *et al* demonstrated that patients co-administrated ceftriaxone and lansoprazole were 1.4 times as likely to have a prolong QT prolongation than the administrated single drug in both EMR and FAERS data. Further validation showed that ceftriaxone/lansoprazole drug interaction was due to hERG channel blocker in a patch-clamp experiment system (Lorberbaum *et al.*, 2016). Duke *et al* proposed a text mining strategy for DDI molecular pharmacology evidence discovery from the public literature (Duke *et al.*, 2012), which discovered 13,197 potential DDIs. In the follow-up in-vitro study, Han *et al* validated the loratadine-simvastatin myotoxicity interaction, and its increased myopathy risk in both EMR and FAERS databases (Han *et al.*, 2015). Similarly, Schelleman *et al* examined the increased risk of hypoglycaemia with co-administration of fibrates and statins in sulfonylurea users in a pharmaco-epidemiology study (Schelleman *et al.*, 2014). This DDI was further evaluated in an in-vitro in-vivo extrapolation (IVIVE) pharmacokinetics model.

1.4.3 In-vitro In-vivo extrapolation (IVIVE) pharmacokinetics model

To use IVIVE pharmacokinetics model, drugs pharmacokinetics properties have been studied, especially metabolism. The majority of small molecule drugs are catalyzed by Cytochrome P450 (CYP) enzymes, which are located in the hepatic endoplasmic reticulum (Renwick, 1999; Shen, Kunze, & Thummel, 1997). Many factors can alter hepatic drug metabolism, including genetic polymorphisms, disease and concomitant medications, foods etc.(Eichelbaum, Ingelman-Sundberg, & Evans, 2006; Ereshefsky, 1996; Shah & Smith, 2015). Among these factors, concomitant medications are very important because of poly-pharmacy (Admassie, Melese, Mequanent, Hailu, & Srikanth,

2013; Fitzgerald, 2009; Hanlon et al., 1996; Rosholm, Bjerrum, Hallas, Worm, & Gram, 1998; Sasaki et al., 2013). Therefore, it is highly important to quantify the contribution of different metabolism pathways in order to predict drug exposure change after inhibitions (D. Zhang, Zhu, & Humphreys, 2007). The term *fm*, is defined as the fraction of drug metabolized by an enzyme. There are multiple ways that *fm* can be estimated through clinical pharmacokinetics studies or in vitro pharmacokinetics experiments.

1.5 Proposed Solutions

1.5.1 Scope of Aim 1: Drug fm database

The purpose of the aim is to prepare the required pharmacokinetic parameters that can be used to in an in-vitro in-vivo extrapolation (IVIVE) pharmacokinetics model. The *fm* database was curated from published articles indexed in PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>). In drug selection, 237 FDA approved cancer drugs were identified in DrugBank and National Cancer Institute (NCI). The next stage is the key word search, including cancer drug names, “cytochrome P450”, “human liver microsomes” and/or “metabolism”. The detail of this aim will be described in Chapter 2.

1.5.2 Scope of Aim 2: Random control approach based nested case-control design

We propose a control selection approach for nested case-control design which will be computationally efficient for large scale ADE signal detection. We name the proposed controls as random control, as both the patients and their index time are randomly selected. The key difference between random control and super/dynamic control is the relaxation of matching by the index time condition. The performances of nested case-control design by using super, dynamic and random control will be

evaluated by the Observational Medical Outcomes Partnership (OMOP) gold standard (Ryan et al., 2012). The detail of this aim will be described in Chapter 3.

1.5.3 Scope of Aim 3: High-dimensional drug interaction

We will use this newly developed mixture drug-count response model (MDCM) to detect high-dimensional drug-drug interactions (HDDIs) that lead to increased risk of myopathy in two independent databases: Indiana Network of Patient Care - CDM (INPC-CDM) electronic medical record and FAERS (Chiang et al., 2017). Using *in vitro* cytochrome P450 (CYP) inhibition data and mechanistic static *in vitro in vivo* DDI predictions, we evaluate the potential pharmacological mechanisms of these HDDIs. The detail of this aim will be described in Chapter 4.

1.6 Impact of Project

1.6.1 Impact of Aim 1

In this research, we present our initial effort in developing a drug fm database. We focus on the fm data collected and estimated from *in vitro* inhibition studies of human liver microsomes, including both metabolites formation and substrate depletion studies. There are additional data, such as pharmacogenetics clinical PK studies and drug interaction clinical PK studies, in which fm can also be estimated. These fm estimates could be more accurate than the fm estimates from *in vitro* studies.

This database has been demonstrated successfully to predict the drug-drug interactions regarding several CYP isozymes. We believe that the public availability of this database will facilitate pharmacokinetics research.

1.6.2 Impact of Aim 2

In this study, we propose a random control approach based nested case-control design for EMR/EHR analyses. In random controls, patients will be randomly selected to from a control pool, and index times are randomly selected as well. Hence, the risk factor evaluation is only limited to the control pool at the index times. Additionally, the control pools can be used for multiple ADEs. As a consequence, random control selection is efficient for large scale (i.e. all ADEs) signal detection. On the statistical prospect, dynamic/super control approaches estimate the frequency of a medication within a group of case free patients who have similar risk factors with the cases. Under random control, the frequency of drug is estimated within the population who has similar risk factors with the cases.

1.6.3 Impact of Aim 3

This study demonstrates the power to elucidate clinically significant high-dimensional drug interactions from clinical records. Using two unique data sets, ADE case reports from the FAERS and structured electronic medical record data from the INPC-CDM, we observed increasing trends in myopathy risk with higher medication burden. As a large number of DDIs are the result of PK interactions at the level of CYP enzymes, we also estimated the increased exposure of 9 substrate drugs in the presence of 2, 3, or more inhibitors. Although we demonstrated that decreased clearance of drugs due to CYP inhibition is one source of the increased myopathy risk among polypharmacy patients, this mechanism is unable to fully explain the increased risk of myopathy observed in subjects taking 4 or more medications. As our computational

efficiency expands to allow for the evaluation of greater number of drugs using our MDCM model, additional pharmacokinetic and pharmacodynamic mechanisms of interaction will need to be considered to further account for the increased risk of ADEs observed in polypharmacy patients.

Chapter 2. The Cancer Drug Fraction of Metabolism Database

Summary: The aim of this study is to create a database for quantifying fraction of metabolism (fm) of CYP450 isozymes for FDA approved cancer drugs. A reproducible data collection protocol was developed to extract key information including both substrate depletion and metabolites formation data from publicly available in vitro selective CYP enzyme inhibition studies. The fm was then estimated from the curated data. To demonstrate the utility of this database, we conduct an in vitro/in vivo drug interaction prediction among these 56 cancer drugs.

2.1 Introduction

Drugs are eliminated by excretion or metabolism after entering the human body (Rowland & Tozer, 2005). The majority of small molecule drugs are catalyzed by Cytochrome P450 (CYP) enzymes, which are located in the hepatic endoplasmic reticulum (Renwick, 1999; Shen et al., 1997). Many factors can alter hepatic drug metabolism, including genetic polymorphisms, disease and concomitant medications, foods etc. (Eichelbaum et al., 2006; Ereshefsky, 1996; Shah & Smith, 2015). Among these factors, concomitant medications are very important because of poly-pharmacy (Admassie et al., 2013; Fitzgerald, 2009; Hanlon et al., 1996; Rosholm et al., 1998; Sasaki et al., 2013).

Many enzymatic routes of elimination, including almost all of those occurring via the CYP450 enzymes, can be inhibited or induced by concomitant medications. Particularly, when the primary metabolic pathways of a drug are inhibited or induced by strong inhibitors or inducers, drug and metabolite concentrations in the blood and

tissue can be severely affected (Vose & Ings, 2014; Yu, Ritchie, Mulgaonkar, & Ragueneau-Majlessi, 2014). The dramatically changed drug exposure may result in unwanted adverse reactions or reduced efficacy (Williams et al., 2004). A randomized, open-label, parallel-group study indicated that after co-administration of ketoconazole for 12 days, the $AUC_{0-\infty}$ of midazolam was about 6.56 times higher (1280 ng·h/mL vs. 195 ng·h/mL) (Shoaf, Bricmont, & Mallikaarjun, 2012). Also, the SDMT (Symbol Digit Modalities Test) scores, the decrease in which represents an impairment of cognitive function, were reduced (-13.6 in SDMT score) by co-administration of ketoconazole and midazolam (Lam et al., 2003). Therefore, all these showed that midazolam and ketoconazole may have strong DDIs both on PD and PK. In another randomized study, co-administration of irbesartan with hydrochlorothiazide significantly decreased the hydrochlorothiazide AUC by 26.3% (1373 ng·h/ml vs. 1087 ng·h/ml). At the same time, the effect of irbesartan on systolic blood pressure when administered with hydrochlorothiazide were significantly different from those when irbesartan was administered alone. It was a 25% increase in E_{max} , and a 40% decrease in EC_{50} , which suggested a synergistic blood pressure lowering effect for the combination (Hedaya & Helmy, 2015).

Therefore, it is highly important to quantify the contribution of different metabolism pathways in order to predict drug exposure change after inhibitions (D. Zhang et al., 2007). The term *fm*, is defined as the fraction of drug metabolized by an enzyme. There are multiple ways that *fm* can be estimated through clinical pharmacokinetics studies or in vitro pharmacokinetics experiments. First, the

contributions of CYPs for a drug's metabolism can be estimated as the change in AUC or clearance in the absence and presence of a co-administered selective inhibitor through an in vivo approach (Creighton et al., 2008; Le Coutre et al., 2008). For example, Yeung et al. utilized clinical drug interaction studies, in which ketoconazole was used as the CYP3A4 probe inhibitor, and calculated a drug's f_m in the CYP3A4 pathway using the following equation (Yeung et al., 2015):

$$f_{m_{3A4}} = 1 - \frac{AUC_{control}}{AUC_{inhibite}}$$

where AUC is area under the concentration-time curve of victim drug. For example, cinacalcet is metabolized primarily by CYP3A4, and ketoconazole is a CYP3A4 probe inhibitor. The $AUC_{inhibited}/AUC_{control}$ is 2.03, and the $f_{m_{3A4}}$ is 0.51 using the above equation (Harris, Salfi, Sullivan, & Padhi, 2007).

Second, f_m can be estimated via a pharmacogenetics study where it can be calculated from the fold-change in exposure of a victim drug in extensive metabolizers (EMs) compared to poor metabolizers (PMs) (Ito et al., 2005). In Silas's study, a large population of patients were studied with respect to the metabolism of metoprolol (Silas, McGourty, Lennard, Tucker, & Woods, 1985). After a single 200 mg oral dose of metoprolol, the average AUC of metoprolol in blood over the 24 hours of six PMs was 7250 ng*ml⁻¹*h. While in the EM population, the average AUC was much lower which was 1246 ng*ml⁻¹*h. Thus, metoprolol's $f_{m_{CYP2D6}}$ can be calculated from the formula:

$$f_{m_{2D6}} = 1 - CL_{PM}/CL_{EM} = 1 - AUC_{EM}/AUC_{PM} = 0.828.$$

Third, human ^{14}C or ^3H -ADME studies, which measure the concentration of the radiolabeled unchanged drug and its metabolites in plasma, urine and feces, is also regarded as a valuable approach and clinical pharmacokinetics study to estimate the metabolic pathways of a drug (Bohnert et al., 2016; Rodrigues et al., 2001). For example, [^{14}C]faldaprevir was used for ADME study in which the formation rates of its metabolites by various rCYP isoforms were determined. The contribution of each cytochrome was determined by rates of metabolite formation after normalization by relative liver content of each cytochrome. The results showed that the normalized contributions by CYP3A4 were 94% and 97% for two kind of metabolites (Y. Li et al., 2014).

Several other in vitro methods have been developed to determine the contribution of individual cytochrome P450 isozymes in a drug's metabolism. Substrate depletion in human liver microsomes (HLM) is one method that the drug is incubated with or without specific CYP450 selective inhibitors (Huskey et al., 1995). Comparing to the metabolism rate, V_{max}/K_m , of the substrate without any inhibitor, the percent inhibition of this CYP pathway by the CYP-selective chemical inhibitor reflects the contribution of this CYP towards the substrate's metabolism. Ideally chemical inhibitors should be potent, selective and metabolically stable. Substrate depletion can also be incubated with individual recombinant enzymes isoforms (Z. M. Li et al., 2016). This approach estimates the metabolism rate, V_{max}/K_m , of the substrate in recombinant human (rh) CYP isoforms, and scales the rhCYP V_{max}/K_m to HLM CL_{int} via a RAF/ISEF approach (Bohnert et al., 2016). Each isozymes contribution is estimated as the percent contribution of each CYP enzyme towards the total $\text{HLMCL}_{\text{int}}$.

A metabolite formation study is another option. For example, after incubation, the mixtures of human liver microsomes and carbamazepine was analyzed by HPLC/MS. Rates of carbamazepine metabolites (2- and 3-hydroxycarbamazepine) formation were determined in microsomes and then compared with typical P450 enzyme activities. Data were analyzed by nonlinear regression and linear transformation to estimate apparent PK values and enzyme models, respectively. Then formation of 2- and 3-hydroxylated carbamazepine metabolites was evaluated in the presence or absence of known P450 inhibitors (Pearce, Vakkalagadda, & Leeder, 2002). Human liver microsomes from high/low CYP activity donors were used to estimate the inhibition percentage of carbamazepine metabolites formation. Also cDNA-expressed isoforms were examined for the affinity of different metabolites formation.

Recently, due to the success of the cryopreservation of human hepatocytes (Stephenson, Najimi, & Sokal, 2010), hepatocyte suspension model (Mao, Mohutsky, Harrelson, Wrighton, & Hall, 2011) becomes a new method to estimate f_m . Physiologically, cryopreserved human hepatocyte is closer to the human hepatic metabolism than the other in vitro system does. Desbans (Desbans et al., 2014) used cryopreserved human hepatocytes from 12 donors to estimate f_{mCYP3A} for five prototypical CYP3A substrates with varying degree of CYP3A-dependent in vivo clearance using intrinsic metabolic stability measurements in the presence and absence of a CYP3A probe inhibitor, ketoconazole. After hepatocytes are incubated with test compounds and/or the inhibitor, the intrinsic clearance was estimated from the parent

compound depletion profile. Then $f_{m_{CYP3A}}$ was calculated from the ratio between CL_{int} in absence and in presence of ketoconazole as:

$$f_{m_{3A}} = 1 - \frac{CL_{int} \text{ with ketoconazole}}{CL_{int} \text{ without ketoconazole}}$$

Although there are some widely-used database for drug metabolism: a) DrugBank (Wishart et al., 2006), which is a comprehensive database which combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information; b) Transformer (the former Super CYP) (Preissner et al., 2010), which integrates Cytochrome P450 enzyme interactions and some pharmacological information; c) DIDB (Hachad et al., 2010), which can evaluate the impact of DDI in the clinic by in vitro and in vivo DDI data, few fm data can be found.

However, there is no database that contains fm data systemically. In this chapter, we present our initial effort in developing a cancer drug fm database. We focus on the fm data collected and estimated from in vitro inhibition studies of human liver microsomes, including both metabolites formation and substrate depletion studies. Because in vitro assessment of the metabolic rate of drugs by each of the major CYPs is routinely carried out in drug discovery and development, we anticipate that the fm data of many cancer drugs are available in the literature. The other data types (e.g. PGx test and clinical drug interaction PK experiment) for fm estimation will be available in the future.

2.2 Construction and Content

The fm database was curated from published articles indexed in PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>). An overview of the data collection is available in Figure 2.1. In Cancer Drug Selection, 237 FDA approved cancer drugs were identified in DrugBank and National Cancer Institute (NCI). The next stage is the key word search, including cancer drug names, “cytochrome P450”, “human liver microsomes” and/or “metabolism”. Cancer drug generic names, their synonymous and brand names published in the DrugBank are included in the search. Cytochrome P450 enzyme names include their synonymous names in HUGO Gene Nomenclature Committee. CYP450 enzymes and metabolism are included in the search, because we focus only on the drug metabolism.

Identified PubMed abstracts were filtered. We purposely looked for whether this is a drug metabolism study; whether the contribution of CYP enzymes to the metabolisms is investigated; or whether the CYP enzyme inhibitors are discussed. If all these information are not reported in the abstract, this paper is then removed. If an abstract passes the filtering step, its full text receives further examination. In the Method section, substrate depletion in human liver microsomal study information and/or the metabolite formation study information are checked. We have searched and identified the following experimental information. In the metabolite formation studies, all incubations were performed at 37 °C; HLMs were incubated with sodium phosphate buffer and NADPH (usually NADP + glucose-6-phosphate dehydrogenase) before adding the substrates; and control samples containing no NADPH or substrates were also

included. Well established selective inhibitors of CYP enzymes were incubated with substrate. Incubations are carried out for a defined time. In final stage, drug metabolites were evaluated using HPLC in the presence or absence known CYP enzyme selective inhibitors. For substrate depletion studies, pre-incubation and incubation procedures are the same as metabolite formation studies. However, no NADPH or substrates samples was needed. After being vortexed, the concentration of the remaining parent drugs in the supernatants was measured using the HPLC assay in the presence or absence selective inhibitors. If this information were not reported in the method section, this paper is then removed.

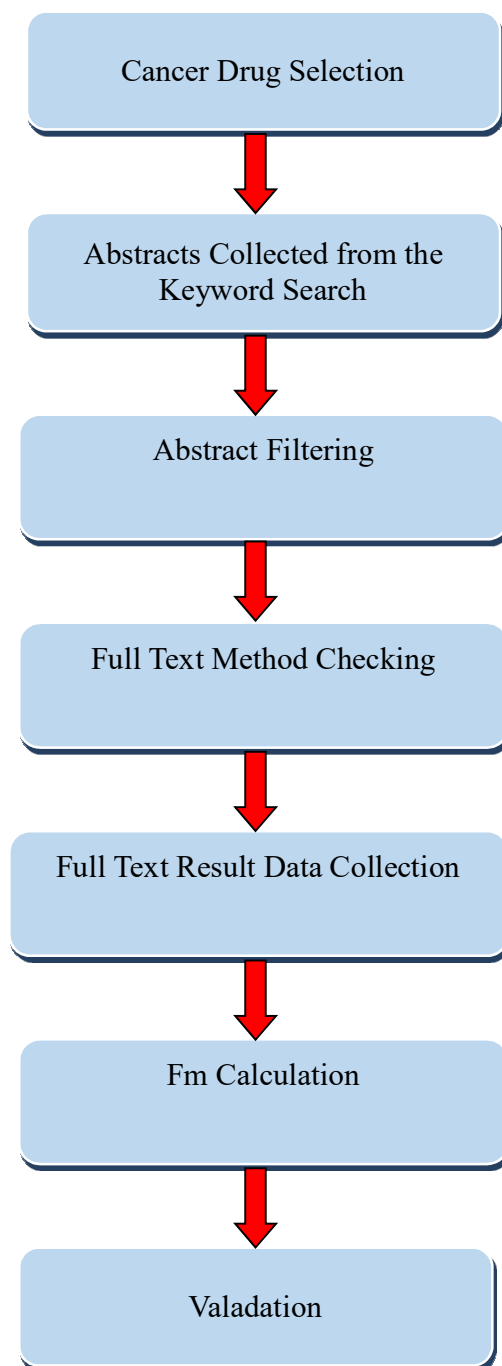


Figure 2.1 The flow chart of data collection procedure

The result section is then read through carefully, including the result narrative, figures and tables, including their legends, or supplemental materials. We collect the following data from the paper for the fm calculations: The metabolites for each drug and/or their relative contribution of the metabolisms; the percentage of inhibitions for each CYP enzyme and their related CYP inhibitor (sometimes, this percentage was directly reported, and other times, the metabolism rates under inhibition and control were reported); and the substrate concentrations, and inhibitor concentrations. After these data are collected, fm is then calculated by metabolite formation data according to following steps.

1. The metabolites from the drug by human liver microsomes are evaluated in the presence or absence (i.e., control sample) of known P450 isoform-selective inhibitors. The relative proportion of the metabolites formation in reaction mixtures with no inhibitor is set as 100%. If the inhibition percentage of a drug metabolite is not reported directly in the paper, it is calculated from the changes of control sample.
2. If multiple metabolite pathways exist, the contribution of each metabolite pathway to the total drug metabolism will be calculated.
3. If a CYP enzyme is inhibited by several inhibitors, the percent of inhibition for each inhibitor is calculated. Then, their mean value is calculated and taken as the inhibition percentage for this pathway.
4. If the substrate concentration of the inhibition experiment varies in the paper, the percentage of inhibition at each concentration is calculated. Then, their

mean value is calculated and taken.

5. The total inhibition percentage of one metabolite was normalized. The fraction of metabolite for the i th CYP enzyme in the j th metabolite is:

$$fm_{ij} = \frac{inhibition_i}{\sum_i inhibition_i} \times \text{percentage of pathway } j$$

where $inhibition_i$ refers to the percentage of inhibition for the i th enzyme. The sum of fm_{ij} over all metabolites is regarded as the fraction of metabolized for enzyme i .

On the other hand, f_m can be calculated through the substrate depletion data:

1. The inhibition percentage or the remaining proportions of substrates are evaluated in the presence or absence (i.e., control sample) of known P450 isoform-selective inhibitors. And it is known that the sum of inhibition percentage of certain substrate and its remaining proportion equals 100%. If the inhibition percentage of a drug metabolite is not reported directly in the paper, it is calculated from the changes of control sample or the remaining percentage.
2. If a CYP enzyme is inhibited by several inhibitors, the mean percent of inhibition for each inhibitor is taken as metabolite formation data.
3. If the substrate concentration of the inhibition experiment varies, the mean percentage of inhibition at each concentration is taken.
4. The metabolized fraction for the i th CYP enzyme is:

$$f_m = \frac{inhibition_i}{\sum_i inhibition_i}$$

where $inhibition_i$ refers to the percentage of inhibition for the i th enzyme.

To ensure the data integrity, two curators with biology background conducted data curating. Dr. Sara K. Quinney who has extensive pharmacology training background further checked any differentially annotated abstracts. Then, the data extraction from the full text was carried out by those two annotators again. Among the disagreed data collection between these two annotators, a group review was conducted by Drs Wang, Quinney and Li to reach the final agreement.

Data validation step was conducted as follows. Two sets of drugs were validated. In the positive set, 10 out of 57 random cancer drugs where fm data was identified are re-evaluated. In the negative set, 10 out of 179 random cancer drugs without identified fm data were re-evaluated. During this validation process, two independent annotators went through the total fm data curating process for these twenty drugs. These two validation annotators have master or PhD degrees in the computational biology and/or biology background. Drs. Quinney and Li who have the pharmacology background further evaluate the concordance among these three sets of annotations. The consistency of fm data is reported.

Among 237 cancer drugs, we have found fm data for 57 of them. During the validation, among the 10 negative cancer drugs (i.e. no fm data is identified from the first annotator), the other two annotators did not find their fm data either. Therefore, the validation overlap rate is 100% for the negative drug set. Among those 10 positive cancer drugs, each validation annotator manages to find fm data for 9 of them. The overlap rate is 90% for the positive drug set. Finally, among the overlapped positive drug set, the consistency of fm value are calculated. Within a 20% relative difference range,

60.0% and 30.0% of fm values are consistent between original annotator and validation annotator 1 and 2, respectively.

Among the discordant fm values for the positive cancer drug data set, the cited PubMed full text papers and fm values are further evaluated by two additional pharmacologists. They find the original annotation has the highest accuracy in collecting and calculating the Fm. She is right on 85.7% of the discordant values among three annotators. The validation results are shown in Table 2.1.

| | Drug | 2B6 | 2C8 | 2C19 | 2D6 | 3A | 2A6 | 2E1 | 2C9 | 1A1 | 1A2 |
|-------------------------------|-------------|------|------|------|------|------|------|------|-----|------|------|
| Validation annotator 1 | Anastrozole | 0 | 16.6 | 0 | 0 | 44.6 | 15.3 | 1.3 | 2.5 | 0 | 1.5 |
| | Dasatinib | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Aprepitant | 0 | 0 | 0 | 0.9 | 92.5 | 0 | 0 | 2.8 | 0 | 3.8 |
| | Azacitidine | 0 | 0 | 0 | 0 | 0 | 0 | 58.3 | 0 | 0 | 41.7 |
| | Bortezomib | 0 | 0 | 21.5 | 14.9 | 44.7 | 0 | 0 | 8.7 | 0 | 10.1 |
| | Colchicine | 0 | 0 | 0 | 4.2 | 70.2 | 0 | 17.2 | 8.3 | 0 | 0 |
| | Docetaxel | 0 | 0 | 0 | 0 | 45.8 | 0 | 13.2 | 0 | 0 | 11.5 |
| | Vinblastine | 0 | 1 | 0 | 6.4 | 81.6 | 0 | 1 | 0 | 0 | 0 |
| | Vinorelbine | 0 | 4.6 | 8.6 | 0.7 | 78.6 | 4.6 | 0 | 0 | 0 | 2.9 |
| | Irbesartan | 0 | 24.8 | 0 | 0 | 3.5 | 46 | 25.6 | 0 | 0 | 0 |
| Validation annotator 2 | Anastrozole | 18.6 | 21.8 | 0 | 0 | 35.4 | 24.2 | 0 | 0 | 0 | 0 |
| | Dasatinib | 0 | 0 | 0 | 0 | 28.2 | 0 | 0 | 0 | 22.5 | 0 |
| | Aprepitant | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| | Azacitidine | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | Bortezomib | 0 | 0 | 34.5 | 8.1 | 44 | 0 | 0 | 1.4 | 0 | 12 |
| | Colchicine | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| | Docetaxel | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| | Vinblastine | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| | Vinorelbine | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| | Irbesartan | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Original annotator | Anastrozole | 11.4 | 13.3 | 8.2 | 2.9 | 35.7 | 24.5 | 2 | 0 | 0 | 0 |
| | Dasatinib | 0 | 0 | 0 | 0 | 28.2 | 0 | 0 | 0 | 22.5 | 0 |
| | Aprepitant | 0 | 0 | 0 | 0.9 | 93.4 | 0 | 0 | 2.8 | 0 | 2.8 |
| | Azacitidine | 0 | 0 | 0 | 0 | 0 | 0 | 58.3 | 0 | 0 | 41.7 |
| | Bortezomib | 0 | 0 | 21.5 | 14.9 | 44.7 | 0 | 0 | 8.7 | 0 | 10.1 |
| | Colchicine | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 |
| | Docetaxel | 0 | 0 | 0 | 0 | 67.7 | 0 | 19.6 | 0 | 0 | 12.7 |
| | Vinblastine | 0 | 11 | 0 | 6.4 | 81.6 | 0 | 1 | 0 | 0 | 0 |
| | Vinorelbine | 0 | 4.6 | 8.6 | 0.7 | 78.6 | 4.6 | 0 | 0 | 0 | 2.9 |
| | Irbesartan | 0 | 24.8 | 0 | 0 | 3.5 | 46 | 25.6 | 0 | 0 | 0 |

Table 2.1 The validation Results

2.3 Utility

Fm database will help to predict drug interactions. These data characterize all the hepatic CYP450 metabolic pathways and their contributions in predicting drug interactions. We therefore further explore and predict the drug interactions among these 57 cancer drugs in the following case study.

AUCratio (AUCR) is the key parameter to measure the drug interaction. The predicted AUCR is estimated as following:

$$\frac{AUC_i}{AUC} = \frac{1}{\sum \frac{fm}{1 + \frac{I_u}{K_{i,u}}} + (1 - \sum fm)}$$

where $\frac{AUC_i}{AUC}$ is the ratio of the area under the plasma concentration-time profile of the substrate drug in the presence (AUC_i) and absence (AUC) of the inhibitor drug; fm is the fraction of the total hepatic metabolism mediated through a CYP enzyme (from our fm database), I_u is the unbound inhibitor concentration, $K_{i,u}$ is the unbound inhibition constant. In this chapter, I_u equals to $C_{\max} \times f_u$, where C_{\max} is the maximum concentration, and f_u is the fraction of unbound drug in plasma. In this chapter, we have further identified f_u , C_{\max} and $K_{i,u}$ for 32 out of 57 cancer drugs through the PubMed literature review. Then, each drugs pair selected from 32 drugs are further evaluated twice in order to predict their interactions. Each time, one drug serves as substrate and the other one serves as inhibitor, and vice-versa. Following the FDA DDI guideline (Food & Administration, 2012) and expert experience, an AUCR >1.5 is

regarded as the moderate or strong DDI evidence. Based on our PBPK model based DDI predictions, we find 97 drug pairs with AUCR more than 1.5. After been validated in DrugBank, Drugs.com and PubMed, 33 pairs have at least one clear DDIs evidence mentioned in DrugBank or Drugs.com or in PubMed (Table 2.2).

| Drug Pair | AUCR | DrugBank | Drugs.com | Pubmed (in vitro) | Pubmed (in vivo) | Pubmed (clinical) |
|-------------------------|-------|-------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|----------------------|---------------------|----------------------|
| (vinblastine,clonidine) | 38.33 | The serum concentration of Clonidine can be decreased when it is combined with Vinblastine. | | | | |
| (vinoreblin,clonidine) | 38.33 | The serum concentration of Clonidine can be increased when it is combined with Vinorelbine. | | | | |
| (etoposide,clonidine) | 11.58 | The serum concentration of Clonidine can be increased when it is combined with Etoposide. | | | | |
| (aprepitant,clonidine) | 11.09 | The serum concentration of Clonidine can be increased when it is combined with Aprepitant. | | | | |
| (sorafenib,clonidine) | 7.67 | The serum concentration of Clonidine can be increased when it is combined with Sorafenib. | | | | |
| (colchicine,docetaxel) | 4.89 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Colchicine. | The risk of peripheral neuropathy may be increased during concurrent use of two or more agents that are associated with this adverse effect. | | | |
| (vinblastine,docetaxel) | 4.89 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Vinblastine. | The risk of peripheral neuropathy may be increased during concurrent use of two or more agents that are associated with this adverse effect. | | | |
| (vinoreblin,docetaxel) | 4.89 | | The risk of peripheral neuropathy may be increased during concurrent use of two or more agents that are associated with this adverse effect. | | | |
| (vincristine,clonidine) | 4.26 | The serum concentration of Clonidine can be decreased when it is combined with Vincristine. | | | | |

Table 2.2 The validation results of DDI prediction based on PBPK model

| Drug Pair | AUCR | DrugBank | Drugs.com | Pubmed (in vitro) | Pubmed (in vivo) | Pubmed (clinical) |
|-------------------------|------|-------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|------------------------------------------------|--------------------------------------------------|
| (everolimus,docetaxel) | 4.16 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Everolimus. | | | | |
| (etoposide,docetaxel) | 3.94 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Etoposide. | The risk of peripheral neuropathy may be increased during concurrent use of two or more agents that are associated with this adverse effect. | | | |
| (fenretinide,docetaxel) | 3.92 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Fenretinide. | | | | |
| (aprepitant,docetaxel) | 3.89 | The serum concentration of Docetaxel can be increased when it is combined with Aprepitant. | Coadministration with inhibitors of CYP450 3A4 may increase the plasma concentrations of docetaxel, which is a substrate of the isoenzyme. | | (Kaneta et al., 2014) (Nygren et al., 2005) | (Kaneta et al., 2014) (Nygren et al., 2005) |
| (sorafenib,docetaxel) | 3.4 | The serum concentration of Docetaxel can be increased when it is combined with Sorafenib. | Coadministration with sorafenib may increase the plasma concentrations of docetaxel. | | (Awada et al., 2012) | (Awada et al., 2012) (Mardjuadi et al., 2012) |
| (granisetron,docetaxel) | 2.73 | | | (Watanabe, Nakajima, Nozaki, Hoshiai, & Noda, 2003) | (Watanabe et al., 2003) | (Miyata et al., 2006) |
| (vincristine,docetaxel) | 2.67 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Vincristine. | The risk of peripheral neuropathy may be increased during concurrent use of two or more agents that are associated with this adverse effect. | | | |

Table 2.2 Continued

| Drug Pair | AUCR | DrugBank | Drugs.com | Pubmed (in vitro) | Pubmed (in vivo) | Pubmed (clinical) |
|-------------------------|------|---------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------|---------------------|---------------------------------------------------------------|
| (irinotecan,docetaxel) | 2.49 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Irinotecan. | | (Maekawa et al., 2010) | | (Argiris, Kut, Luong, & Avram, 2006) (Engels et al., 2007) |
| (ifosfamide,docetaxel) | 2.46 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Palifosfamide. | The concomitant or sequential administration of multiple antineoplastic agents may potentiate the risk and severity of additive toxicities, such as immunosuppression and myelotoxicity. | | | |
| (pazopanib,docetaxel) | 2.17 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Pazopanib. | Coadministration with inhibitors of CYP450 3A4 may increase the plasma concentrations of docetaxel, which is a substrate of the isoenzyme. | | | |
| (idarubicin,cytarabine) | 2.04 | | | (Colburn, Giles, Oladovich, & Smith, 2004) | | |
| (sunitinib,clonidine) | 2.02 | The serum concentration of Clonidine can be increased when it is combined with Sunitinib. | | | | |
| (lapatinib,clonidine) | 1.93 | The serum concentration of Clonidine can be increased when it is combined with Lapatinib. | | | | |
| (bortezomib,clonidine) | 1.77 | The metabolism of Clonidine can be decreased when combined with Bortezomib. | | | | |
| (bosutinib,clonidine) | 1.72 | The serum concentration of Clonidine can be increased when it is combined with Bosutinib. | | | | |
| (eribulin,docetaxel) | 1.71 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Eribulin. | The risk of peripheral neuropathy may be increased during concurrent use of two or more agents that are associated with this adverse effect. | | | |

Table 2.2 Continued

| Drug Pair | AUCR | DrugBank | Drugs.com | Pubmed (in vitro) | Pubmed (in vivo) | Pubmed (clinical) |
|-------------------------|------|-------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|----------------------------|-------------------------------------------------------------------------------|
| (sunitinib,docetaxel) | 1.70 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Sunitinib. | | | | (Bergh et al., 2012) |
| (paclitaxel,clonidine) | 1.68 | The serum concentration of Clonidine can be increased when it is combined with Paclitaxel. | | | | |
| (ixabepilone,docetaxel) | 1.68 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Ixabepilone. | The risk of peripheral neuropathy may be increased during concurrent use of two or more agents that are associated with this adverse effect | | | |
| (lapatinib,docetaxel) | 1.65 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Lapatinib. | Coadministration with inhibitors of CYP450 3A4 may increase the plasma concentrations of docetaxel, which is a substrate of the isoenzyme. | | | (LoRusso et al., 2008) |
| (bortezomib,docetaxel) | 1.55 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Bortezomib. | Bortezomib can cause peripheral neuropathy, and concurrent use of other agents that are also associated with this adverse effect can potentiate the risk and/or severity of nerve damage. | | (Messersmith et al., 2006) | (Messersmith et al., 2006) |
| (bosutinib,docetaxel) | 1.52 | The risk or severity of adverse effects can be increased when Docetaxel is combined with Bosutinib. | | | | |
| (paclitaxel,docetaxel) | 1.50 | The risk or severity of adverse effects can be increased when Paclitaxel is combined with Docetaxel. | The risk of peripheral neuropathy may be increased during concurrent use of two or more agents that are associated with this adverse effect | (Maekawa et al., 2010) (Watanabe et al., 2003) (Royer, Monsarrat, Sonnier, Wright, & Cresteil, 1996) | (Royer et al., 1996) | (Esposito et al., 1999) (Bahleda et al., 2014) (Izquierdo et al., 2006) |
| (paroxetine,clonidine) | 1.50 | The metabolism of Clonidine can be decreased when combined with Paroxetine. | | | | |

Table 2.2 Continued

2.4 Discussion

In this cancer drug fm database, the relative contributions of CYP isozyme are curated and calculated from drug metabolism studied using the human liver microsomes. These in vitro experiments include both metabolites formation and substrate depletion studies from the published articles. After a number of data filtering and data processing steps, this database curated 57 cancer drugs with fm data for primary CYP450 enzymes (i.e. CYP 3A4, CYP 2D6, CYP 2C9 etc.). A predefined data curating protocol was established to assure data quality and data reproducibility. Multiple annotators were employed in the data filtering stage and the data validation stage. Two independent validation annotators re-evaluated the random selection from positive and negative set. The overlap rate is 90% and 100% for the positive and negative set, respectively. Among the overlapped positive drug set, the consistency of fm are 60.0% and 30.0% for validation annotator 1 and 2. After evaluated by two additional pharmacologists, the original annotation has the highest accuracy in collecting and calculating the fm evaluated by two additional pharmacologists. They find the original annotation has the highest accuracy in collecting and calculating the fm among three annotators.

In our initial effort to curate fm data from the literature, we have focused on primarily the fm data calculate from the in vitro experiments. There are additional data, such as pharmacogenetics clinical PK studies and drug interaction clinical PK studies, in which fm can also be estimated. These fm estimates could be more accurate than the fm estimates from in vitro studies.

This database has been demonstrated successfully to predict the drug-drug interactions regarding several CYP isozymes. We believe that the public availability of this database will facilitate pharmacokinetics research.

Chapter 3. A Computationally Efficient Design for Electronical Health/Medical Records

Mining

Summary: Nested case-control design is a promising approach for detecting adverse drug events (ADEs) from electronical health records (HER) and electronical medical records (EMR) databases. Currently, computational burden is a significant challenge for the application of nested case-control design on EMR/HER databases, especially for investigating multiple ADEs. In this chapter, we propose a random control selection approach for nested case-control design. By relaxing the matching by case's index time restriction, random control dramatically reduces the computational burden compared with traditional control selection approaches. By using the Observational Medical Outcomes Partnership gold standard and an EMR database, random control is demonstrated to have better performances as well.

3.1 Introduction

Adverse drug events (ADEs) are considered to be a significant challenge for current clinical practices. In the United States alone, ADEs cause approximately 125,000 hospital admissions each year; up to 53% hospital stays prolonged; and as many as 4.6% of deaths (de Vries, Ramrattan, Smorenburg, Gouma, & Boermeester, 2008; Hall, DeFrances, Williams, Golosinskiy, & Schwartzman, 2010; Lazarou, Pomeranz, & Corey, 1998). Though ADEs can be detected in either pre-marketing clinical trials or post-marketing surveillances, most ADE knowledges are revealed in the post-marketing stage. This is because post-marketing stage allows larger population and prolonged follow up (Harpaz et al., 2012). Additionally, unlike pre-marketing clinical trials, drugs are

administered without stringent inclusion/exclusion criteria. During the past two decades, different types of databases have been maintained for drug safety surveillance such as Spontaneous Reporting System (SRS), Electronic Medical Records (EMR) and Electronic Health Records (EHR). SRS collects ADE reports from healthcare professionals, consumers, and manufacturers. EMR and EHR are administrative health databases. They usually contain standard medical and clinical data gathered in the provider's offices.

SRS and EMR/EHR analyses have gathered valuable ADE knowledges. For a drug-ADE pair of interest, either univariate analyses or multivariate analyses can be used for signal detection. Under univariate analyses, available samples are usually summarized by a 2-by-2 contingency table, in which contains the frequencies classified by the usage of the drug (yes/no) and the occurrence of the ADE (yes/no). The outcome is the frequency that this drug-ADE pair is co-occurred, and the expectation is the expected frequency of this drug-ADE pair under the assumption of no association. Univariate analyses are also known as disproportion analysis (DPA), as they quantify ADE signals by the outcome to expectation ratio (i.e. relative risk) or its variants. Frequentist DPAs include proportional reporting ratio (PRR) and reporting odds ratio (ROR) (Evans et al., 2001; van Puijenbroek et al., 2002). While, the empirical Bayesian geometric mean (EBGM) is an empirical Bayesian DPA and the information component (IC) is a Bayesian DPA (Bate et al., 1998; DuMouchel, 1999). DPAs are demonstrated to have promising performances and are computationally efficient (Harpaz et al., 2013). However, confounding bias, especially confounding by co-medication, may cause biased disproportionality measurements of the true association. Typical multivariate analyses

utilize multiple logistic regression (MLR) or regulated logistic regression (RLR) models to adjust potential confounding variables (i.e. co-medications). Besides this advantage, practices of MLR/RLR are often challenged by considerable larger sample sizes of pharmacovigilance databases. For instance, regular computers cannot fit MLR over millions of samples. In a summary, DPAs are powerful tools for large scale signal detection, as they can be applied to multiple ADEs together (Harpaz et al., 2012). Multivariate analyses that models an ADE and all co-medications are preferred for specific hypotheses testing.

In SRS, temporal information are omitted and each sample contains the ADE and drug status (Yes/No). Such a structure allows both multivariate analyses and DPAs to be applied directly to SRS. However, in EMR/HER database, patients are usually followed longitudinally with detailed temporal information about medications and phenotypes. As a consequence, EMR/EHR analyses typically require sophisticated epidemiology designs such as cohort design, self-controlled design, and nested case-control design (Hennessy et al., 2016). These epidemiology designs have been widely utilized for EMR/EHR analyses. Cohort design identifies exposure and non-exposure cohorts. ADE risks are estimated within each cohort, and the drug-ADE associations are estimated by the RR. The non-exposure cohort can be selected based on estimated propensity scores to ensure their similarities with the exposure cohort. For instance, Tatonetti *et al.* assumes the latent cofounders can be characterized by co-medications, and propensity scores are then calculated by modeling the co-medications based on such assumption (Tatonetti et al., 2012). For the cases and their matched controls, nested case-control

design uses a predefined window to examine drug exposures. For a case, usually 4 – 10 controls are selected. The controls are usually matched with the case's index date and risk factors (Schneeweiss, 2010). Additionally, they are case-free up to the case's index time. Thus, such controls are also named as at risk or dynamic control. If the controls are restricted to be case free for the entire follow up, this type of controls is known as super control. Some example of application of nested case-control designs includes Brauer *et al.*, La gamba *et al.* and Lee *et al.* (Brauer et al., 2014; La Gamba et al., 2017; Lee et al., 2011). In real applications, EMR/EHR databases may contain up to millions of patients. Such a sample size become an obstacle for the application of epidemiology designs to conduct large scale (i.e. drug wide and ADE wide) ADE signal detection. For instance, as we mentioned above, fitting propensity score model over millions of samples cannot be accomplished by regular computers. Moreover, in EMR/EHR analyses, propensity scores will be time dependent, which further increased the computational burden. Though nested case-control design do not require complicated modelling, the matching process still computationally expensive.

We propose a control selection approach for nested case-control design which will be computationally efficient for large scale ADE signal detection. Unlike the dynamic and super control, the proposed methods will select a pool of controls before cases are identified. Later, for a case, the matched controls will be patients with similar risk factors from the control pool. We name the proposed controls as random control, as both the patients and their index time are randomly selected. The key difference between random control and super/dynamic control is the relaxation of matching by the

index time condition. For super/dynamic control, at a case's index time, all available samples have to be evaluated for control selection. For instance, under dynamic control, the samples at risk are changing over time. While, random control only need to evaluate the control pool for once, and avoid the time consuming evaluation processes. In another word, the controls under super/dynamic control are case specific, but are not for random control. Additionally, random control can be applied to multiple ADEs. The performances of nested case-control design by using super, dynamic and random control will be evaluated by the Observational Medical Outcomes Partnership (OMOP) gold standard (Ryan et al., 2012).

3.2 Methods

3.2.1 Data source

The Indiana Network for Patient Care (INPC) is a health information infrastructure containing medical records for over 15 million patients from five major hospital systems (fifteen separate hospitals) (McDonald et al., 2005). INPC-Common Data Model (INPC-CDM) data was derived from INPC patients between 2004 and 2015, following CDM Version 5.0 guideline (<http://omop.org/CDM>). The INPC-CDM consists of structured longitudinal information including medical conditions and prescription medications for patients. It also includes lab tests and demographics for patients. In our analysis, to process these complex data, medical conditions and prescription medications were mapped into ICD-9 codes and drugbank ID (www.drugbank.ca) with starting and end date. Thus, for each patient, our final data included the patient's age,

gender, the starting and end date of each medical condition, and the starting and end date of each prescription medication.

3.2.2 The OMOP Gold Standard

The Observational Medical Outcomes Partnership (OMOP) gold standard was designed to establish a reference set for pharmacovigilance study (Ryan et al., 2012). It contains 399 drug-ADE pairs that were made up of 181 drugs and 4 ADEs (acute myocardial infarction, acute renal failure, acute liver injury, and gastrointestinal bleeding). These 399 drug-ADE pairs were classed as 165 true positive test cases and 234 true negative test cases.

In OMOP golden standard, 4 ADEs were defined by using Medical Dictionary for Regulatory Activities (MedDRA) (<https://www.meddra.org/>) Lower Level Terms (LLT). For our analysis, these MedDRA LLTs were mapped into ICD-9 codes. There were total 44, 6, 16, and 12 ICD-9 codes for acute myocardial infarction, acute renal failure, acute liver injury, and gastrointestinal bleeding respectively.

3.2.3 Study Designs

We aim for evaluating the performances of super, dynamic and random controls under a nested case-control design. The design is similar as existing EMR data analysis (Duke et al., 2012; Han et al., 2015; P. Zhang et al., 2015). Under the nested case-control design, we select 4 controls for each case. Each case and control will have an index time. At each index time, we examine the medications up to 30 days prior to the index time (one month drug window). Naturally, for a drug and an ADE, the data can be summarized as Table 3.1.

3.2.3.1 Case selection

For each ADE, two types of new case events were defined. The first type was the first case event where the first ADE occurred. The second case event type included any follow-up ADE event whose corresponding drug exposure was more than 6 months after the previous ADE event. In other words, the second type of new ADE case event required a “washout” period (i.e., no drug exposure) of more than 6 months. Summary statistics for each ADE are shown in Table 3.2. We identified total 204,780, 281,461, 261,874, and 221,330 ADE case events for acute myocardial infarction, acute renal failure, acute liver injury, and gastrointestinal bleeding with occurred percentage of 3.77%, 5.18%, 4.82%, and 4.07% respectively.

After the cases were identified, a and b in Table 3.1 can be determined for a selected drug. While, under different control selection approaches c and d are different.

3.2.3.2 No Control Selection (All patients)

We first conduct an analysis by using all patients without a control selection approach. Without control selection, c was the frequency of patients who experienced the ADE but unexposed to the drug in the entire follow up; d was the frequency of patients who neither experienced the ADE nor exposed to the drug in the entire follow up. These analyses will be served as reference for the following control selection approaches.

3.2.3.3 Super Control and Dynamic control Selection

The controls were matched with case’s index time, age and gender. Additionally, we restricted the controls and the case to have equal number of medications within the

one month drug window. At each case's index time, control patients were selected by these rules. Further, the dynamic controls were sampled from patients who were case free at the index time. On the other hand, the super controls were selected from patients who were case-free for their entire follow up.

3.2.3.4 Random control Selection

For random control selection, a control pool was randomly selected from all patients first. Similarly, we selected a random index time for each control patient in the pool. For each case, matched control were patients from the control pool with same age, gender and number of medications. In our analysis, we selected 1 million patients to from the control pool.

3.2.4 Comparison analysis

By using OMOP gold standard, DPAs and multivariate analyses were shown to have comparable performances (Harpaz et al., 2013). Hence, in this study, we selected 2 frequentist and 1 Bayesian DPAs for evaluation. They were PRR, ROR and IC (Bate et al., 1998; Evans et al., 2001; van Puijenbroek et al., 2002). Start from Table 3.1, the PRR, ROR and IC were given in Table 3.3.

Normal approximation and delta method can be used to compute the variances of PRR, ROR and IC. The lower bound of 95% confidence intervals, PRR_025, ROR_025 and IC_025, will be used for evaluation as well. For this DPAs, common threshold, 1 for PRR, PRR_025, ROR and ROR_025; and 0 for IC and IC_025 were adopted for evaluation.

| | Case | Control | Total |
|---------|---------|---------|-----------------|
| Drug | a | b | $a + b$ |
| No Drug | c | d | $c + d$ |
| Total | $a + c$ | $b + d$ | $a + b + c + d$ |

Table 3.1 2-by-2 contingency table for a drug-ADE pair

| Adverse Event | Case Event | Percentage | Case Patients |
|-----------------------|------------|------------|---------------|
| Myocardial Infarction | 204,780 | 3.77% | 137,439 |
| GI Bleed | 281,461 | 5.18% | 235,056 |
| Liver Injury | 261,874 | 4.82% | 200,956 |
| Acute Renal Failure | 221,330 | 4.07% | 165,469 |

Table 3.2 Summary statistics for each ADE (each patient can have multiple new case events)

| DPA | Formula | Description |
|------------------------------------|----------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PRR (Evans et al., 2001) | $\frac{a}{(a+b)} / \frac{c}{(c+d)}$ | PRR measures the ratio of P(ADE Drug) over P(ADE No Drug). |
| ROR (van Puijenbroek et al., 2002) | $\frac{a/c}{b/d}$ | ROR measures the ratio of $\frac{P(\text{ADE} \text{Drug})}{P(\text{No ADE} \text{Drug})}$ over $\frac{P(\text{ADE} \text{No Drug})}{P(\text{No ADE} \text{No Drug})}$ |
| IC (Bate et al., 1998) | $\log_2 \left[\frac{a+1}{\frac{(a+c) \times (a+b)}{a+b+c+d} + 1} \right]$ | IC is the log 2 transformed outcome to expectation ratio $\left(\frac{P(\text{ADE \& Drug})}{P(\text{ADE}) \times P(\text{Drug})} \right)$. By adding 1 on both the numerator and the denominator, infrequent drug-ADE pairs will have penalized IC values. |

Table 3.3 PRR, ROR and IC

3.3 Results

3.3.1 Performance evaluation under common signal detection rules

The performances including specificity, sensitivity, precision, recall and F-score under common signal detection rules are shown in Table 3.4. Among all four control selection methods, random control approach has significant better F-scores (0.42-0.46) than the other three methods (0.21-0.30). Moreover, random control approach leads to better recall (0.36-0.41) compared to the other three approaches (0.12-0.23). All patients approach generates the best precision (0.79), followed by random control method (0.36-0.41), and dynamic control and super control methods have equally worst precision (0.42-0.45).

3.3.2 Area under the receiver operating characteristic curve (AUC) analysis

To further explore the properties of each control selection approach, we conduct AUC analysis. The AUCs are shown in Figure 3.1. After classification, there is no significant difference between dynamic control (AUC: 0.38-0.57) and super control (AUC: 0.36-0.58). All patients approach has the best overall AUC scores (0.59-0.62), followed by random control approach (0.57-0.60), and dynamic control and super control approaches have equally worst precision (0.48-0.54) in all three evaluation method (PRR, ROR, and IC). In sub-category phenotypes, all patients approach achieves the best AUC score (0.71-0.73 and 0.59-0.61) in acute liver injury and acute renal failure, but not performs well in acute myocardial infarction and gastrointestinal bleeding phenotypes with AUC (0.47-0.51 and 0.43-0.50). Random control approach is consistently generated good scores across all four phenotypes (0.46-0.62). In acute myocardial infarction, acute

liver injury and acute renal failure phenotypes, AUC (0.49-0.62) is comparable to all patients approach (0.47-0.73), but significant better than dynamic control and super control approach (0.35-0.51). In gastrointestinal bleeding phenotype, random control approach is equally good as dynamic control and super control methods with AUC (0.46-0.56) and (0.48-0.58) respectively.

Further, for all four ADEs, the AUC curves for random control approach and all patients approach are shown in Figure 3.2. Generally, two methods have comparable curves. While, random control have better true positive rates under lower false positive rates for all three matrices.

| Method | Design | Specificity | Sensitivity | Precision | Recall | F-score |
|------------|-----------------|-------------|-------------|-----------|--------|---------|
| PRR025 > 1 | Dynamic Control | 0.8058 | 0.1938 | 0.4366 | 0.1938 | 0.2684 |
| | Super Control | 0.7961 | 0.1938 | 0.4247 | 0.1938 | 0.2661 |
| | Random Control | 0.7330 | 0.3625 | 0.5133 | 0.3625 | 0.4249 |
| | All Patients | 0.9757 | 0.1188 | 0.7917 | 0.1188 | 0.2065 |
| ROR025 > 1 | Dynamic Control | 0.7864 | 0.2250 | 0.4500 | 0.2250 | 0.3000 |
| | Super Control | 0.7913 | 0.2125 | 0.4416 | 0.2125 | 0.2869 |
| | Random Control | 0.7136 | 0.4062 | 0.5242 | 0.4062 | 0.4577 |
| | All Patients | 0.9757 | 0.1188 | 0.7917 | 0.1188 | 0.2065 |
| IC025 > 0 | Dynamic Control | 0.8058 | 0.1938 | 0.4366 | 0.1938 | 0.2684 |
| | Super Control | 0.8010 | 0.1938 | 0.4306 | 0.1938 | 0.2672 |
| | Random Control | 0.7330 | 0.3625 | 0.5133 | 0.3625 | 0.4249 |
| | All Patients | 0.9757 | 0.1188 | 0.7917 | 0.1188 | 0.2065 |

Table 3.4 Performances of different control selection approaches

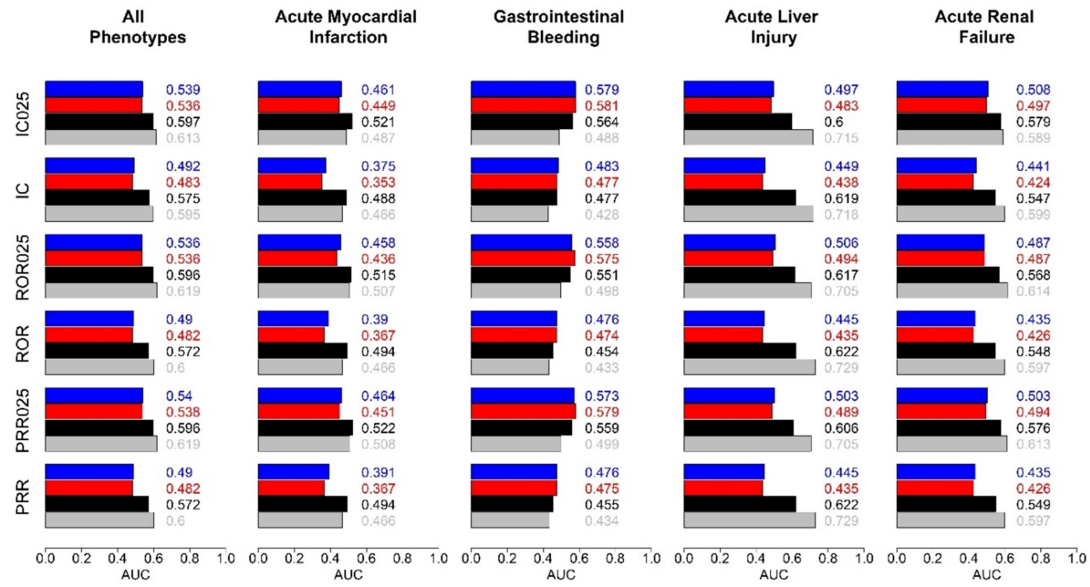


Figure 3.1 AUC values for four control selection strategies

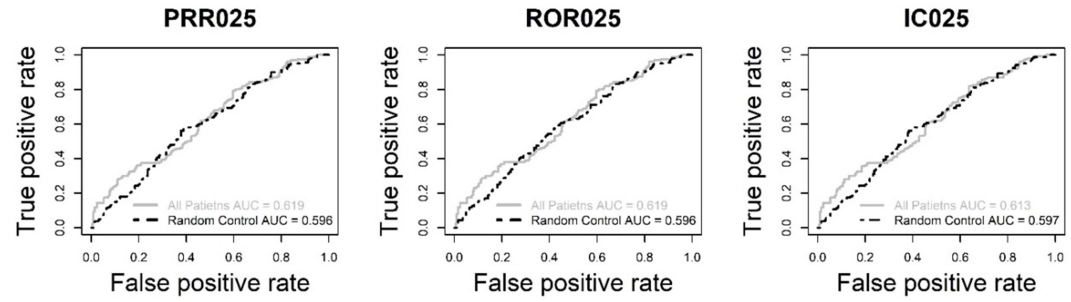


Figure 3.2 AUC curves for all patients analysis and random control

3.4 Conclusion and Discussion

In this study, we propose a random control approach based nested case-control design for EMR/EHR analyses. For a case, random controls are matched patients who have same risk factors as the case. While, compared with the dynamic/super control selection approach, random control selection relaxes the matching by the cases' index time restriction, which is the most computational expensive step. Under the matching by the cases' index time restriction, the risk factors for all patients will be examined at each distinct index time. Given the sample size and the amount of distinct case index times of EMR/HER data, such matching process generates a considerable computational burden. Moreover, for investigating multiple ADEs, the dynamic/super controls have to be different, which further increase the computational burden. For random controls, patients will be randomly selected to from a control pool, and index times are randomly selected as well. Hence, the risk factor evaluation is only limited to the control pool at the index times. Additionally, the control pools can be used for multiple ADEs. As a consequence, random control selection is efficient for large scale (i.e. all ADEs) signal detection. On the statistical prospect, dynamic/super control approaches estimate the frequency of a medication within a group of case free patients who have similar risk factors with the cases. Under random control, the frequency of drug is estimated within the population who has similar risk factors with the cases.

Additionally, we evaluate the performances of random control and other control selection approaches by using the OMOP gold standard and the INPC-CDM data. For each of the 399 drug-ADE pairs in the OMOP gold standard, samples in our data are

summarized into a 2-by-2 contingency table (Table 3.1). Six disproportionality measurements of drug-ADE association that generate from three ADE signal detection methods (2 frequentists and 1 Bayesian) are calculated. Finally, we compute the area under ROC curve (AUC) values by using these disproportionality measurements. We observe the random control (AUC: 0.572 – 0.597) and all patients analysis (AUC: 0.595 – 0.619) have modest well AUCs. While, the dynamic controls (AUC: 0.492 -0.539) and super controls (AUC: 0.482 – 0.538) have less powerful detection capability.

The goal of this study is to find a computational efficient study design for large scale EMR/EHR analysis. Through AUC analyses, the random control selection has a modest well detection capability. Its average AUC value is only less than the best average AUC value. The modest detection capability that we observed in this study coincides with a few other studies, which aimed to evaluate the performance of different ADE signal detection methods. For instance, Ryan *et al.* initially evaluated the signal detection methods on multiple EMR/EHR databases (Ryan et al., 2012). Their analysis is based on the OMOP gold standard. Results shown that case crossover design had AUC = 0.66 and case-control design had AUC = 0.62. Liu *et al.* perform an evaluation by using an EMR data from Vanderbilt University Medical Center and another gold standard (Liu et al., 2013). In their analysis, the evaluation is based on F-score. Results shown that PRR and ROR have F-score = 0.62, and IC has F-score = 0.60. In order to achieve better performance, multivariate methods may be used, as they are demonstrated to have better performances than DPAs. As we solely focused on study design, only DPAs are used for evaluation.

Chapter 4. Translational high-dimensional drug Interaction discovery and validation using health record databases and pharmacokinetics models

Summary: Polypharmacy increases the risk of drug-drug interactions (DDI's). Combining epidemiological studies with pharmacokinetic modeling, we detected and evaluated high-dimensional DDI's among thirty frequent drugs. Multi-drug combinations that increased risk of myopathy were identified in the FDA Adverse Event Reporting System (FAERS) and electronic medical record (EMR) databases by a mixture drug-count response model. CYP450 inhibition was estimated among the 30 drugs in the presence of 1 to 4 inhibitors using in vitro in vivo extrapolation. Twenty-eight 3-way and 43 4-way DDI's had significant myopathy risk in both databases and predicted increases in the area under the concentration time curve ratio (AUCR) >2-fold. The HD-DDI of omeprazole, fluconazole and clonidine was associated with a 6.41-fold (FAERS) and 18.46-fold (EMR) increase risk of myopathy (LFDR<0.005); the AUCR of omeprazole in this combination was 9.35. The combination of health record informatics and pharmacokinetic modeling is a powerful translational approach to detect high-dimensional DDI's.

4.1 Introduction

Drug-drug interactions (DDIs) are a common cause of adverse drug events (ADE) (Hajjar et al., 2007; U.S. Department of Health and Human Services; L. Zhang et al., 2009). In the United States alone, each year an estimated 195,000 hospitalizations and 74,000 emergency room visits are the result of DDIs. National Health and Nutrition Examination Survey data published in 2010 (Gu et al., 2010) showed that the number of

patients using two or more prescription drugs increased from 25.4% to 31.2% in ten years. In particular, more than 64% of elderly individuals took three or more prescription drugs, and 37% took five or more prescription drugs (Gu et al., 2010). In the FDA Adverse Event Reporting System (FAERS) data, 35% of reports include three or more prescription drugs, and 20% reports have five or more prescription drugs.

In order to evaluate clinical effects and molecular mechanisms of DDIs, clinical pharmacokinetic (PK) studies, pharmaco-epidemiologic studies, and *in vitro* PK experiments have been routinely utilized. One salient example is that of breast cancer hormonal therapy, tamoxifen. The formation of its active metabolite, endoxifen, was inhibited by concomitant selective serotonin reuptake inhibitor paroxetine in a clinical pharmacokinetics study (Stearns et al., 2003). *In vitro* metabolism studies revealed that this is due to paroxetine's strong inhibition of the tamoxifen bio-transformation to endoxifen via the CYP2D6 pathway (Desta et al., 2004). In a follow-up pharmacogenetic study, breast cancer patients with CYP2D6 loss function variants have a higher risk of disease relapse and a lower incidence of hot flush (Goetz et al., 2005). The clinical consequence of treating breast cancer and depression using tamoxifen and SSRIs was reviewed (Henry et al., 2008), and a call made for further investigation. Another example is the sedation agent midazolam. Co-administration of midazolam and ergosterol synthesis inhibitor ketoconazole has been identified to reduced subjects' cognitive function (Lam et al., 2003). In clinical PK and *in vitro* experiments, midazolam metabolism was inhibited by ketoconazole through the CYP3A pathway (Gascon &

Dayer, 1991; Gorski et al., 1994), leading to increased midazolam exposure (Olkola et al., 1994).

These examples clearly demonstrate that the translational significance of drug interaction studies relies on both clinical and molecular pharmacology evidence. As described by Hennessy and Flockhart (Hennessy & Flockhart, 2012), an integrated informatics, epidemiology, and pharmacology approach has the potential to accelerate the translational drug interaction studies. Pioneered by Tatonetti *et al.* (Tatonetti et al., 2012), FAERS and electronic medical records were utilized to generate and validate drug-ADE and drug-drug-ADE associations. In a follow-up study, Lorberbaum *et al.* demonstrated that patients co-administrated ceftriaxone and lansoprazole were 1.4 times as likely to have a prolong QT prolongation than the administrated single drug in both EMR and FAERS data. Further validation showed that ceftriaxone/lansoprazole drug interaction was due to hERG channel blocker in a patch-clamp experiment system (Lorberbaum et al., 2016). Duke *et al.* proposed a text mining strategy for DDI molecular pharmacology evidence discovery from the public literature (Duke et al., 2012), which discovered 13,197 potential DDIs. In the follow-up *in vitro* study, Han *et al.* validated the loratadine-simvastatin myotoxicity interaction, and its increased myopathy risk in both EMR and FAERS databases (Han et al., 2015). Similarly, Schelleman *et al.* examined the increased risk of hypoglycaemia with co-administration of fibrates and statins in sulfonylurea users in a pharmaco-epidemiology study (Schelleman et al., 2014). This DDI was further evaluated in an *in vitro in vivo* extrapolation (IVIVE) pharmacokinetic model.

High dimensional drug interactions (HDDIs), i.e. DDIs with three or more drugs, have not yet been broadly investigated. There are a few examples of clinical PK studies, *in vitro* PK experiments, and IVIVE PK models that evaluate interactions among 3 or more drugs. Co-administration of gemfibrozil and itraconazole were shown to increase repaglinide plasma exposure to a greater extent than either one alone (Niemi, Backman, Neuvonen, & Neuvonen, 2003). Zhang *et al.* found an additive PK model, including mechanism-based and competitive components, best described the *in vitro* inhibition of midazolam metabolism by erythromycin, diltiazem and their metabolites (X. Zhang, Jones, & Hall, 2009). Through IVIVE, this model was confirmed *in vivo* in mice (X. Zhang, Quinney, Gorski, Jones, & Hall, 2009). On the other hand, to our knowledge, there are no studies of three way drug interactions using pharmaco-epidemiology studies. Proportional reporting ratio (PRR) (Evans et al., 2001), the reporting odds ratio (ROR) (van Puijenbroek et al., 2002), the information component (IC) (Bate et al., 1998), and the empirical Bayes geometric mean (EBGM) (DuMouchel, 1999) have been proposed for detection of drug-ADE signals. However, these methods are focused on ADE detection for single drugs, not drug combination. To overcome this limitation, we have recently developed a new method, a mixture drug-count response model (MDCM). This model focuses on detecting high dimensional drug interactions, and characterizes the drug-count response relationship between the number of co-administered drugs and an ADE. We have successfully demonstrated its statistical and computational performance in a recent publication (P. Zhang et al., 2015).

In this chapter, we will use this newly developed MDCM to detect HDDIs that lead to increased risk of myopathy in two independent databases: Indiana Network of Patient Care - CDM (INPC-CDM) electronic medical record and FAERS. Using *in vitro* cytochrome P450 (CYP) inhibition data and mechanistic static *in vitro in vivo* DDI predictions, we evaluate the potential pharmacological mechanisms of these HDDIs.

4.2 Materials and Methods

4.2.1 Indiana Network of Patient Care (INPC) Electronic Medical Record

INPC-CDM data was derived from INPC patients between 2004 and 2015, following CDM Version 5.0 guideline (<http://omop.org/CDM>). The INPC-CDM consists of structured data detailing medical conditions, medications, and lab tests of patients. Using this data set, we have identified myopathy patients, as defined in our previous paper (Duke *et al.*, Table S6 (Duke *et al.*, 2012)). The myopathy case definition contains both severe symptoms, such as rhabdomyolysis, and mild symptoms, such as muscle weakness. The myopathy cases include the first myopathy event recorded for a patient and cases in which no other myopathy event occurred within the past 6 months. For each case, ten controls were selected from records within the same time-frame (anchor time-matched) and matching demographic criteria. These controls did not have any myopathy events recorded in the INPC-CDM data set. Under each anchor time in both cases and matched controls, a one-month drug exposure window was generated. Drugs were coded as present if their prescription time periods overlapped with the drug exposure window. Drug names were normalized to their DrugBank IDs. The INPC-CDM data set has 450,673 cases and 4,506,730 controls. This case/control design for the

myopathy using the INPC-CDM database was similar to that used in our previous publications (Fahmi et al., 2008; Han et al., 2015; P. Zhang et al., 2015).

4.2.2 FDA Adverse Event Reporting System (FAERS)

In FAERS, myopathy cases were similarly identified using the same terms we used in the INPC-CDM (Table 4.1). There were 136,791 myopathy cases, and 3,969,842 controls identified in the FAERS. The drug names were mapped to their DrugBank IDs.

4.2.3 Drug Selection Criteria and Data Curation

- Only FDA-approved drugs included
- Selected by frequency >0.5% in both INPC and FAERS databases
- Myopathy risk identified in SIDER
- Limited to 30 drugs due to computational expense of MDCM

Pharmacokinetic parameters required for *in vitro in vivo* extrapolation of AUCR (C_{max} , f_u , f_m , K_m , V_{max} , K_i) were curated from Goodman and Gilman (Goodman, Gilman, Brunton, Lazo, & Parker, 2006) or published literature identified in PubMed. C_{max} obtained from literature review (Goodman et al., 2006) was used as the inhibitor concentration $[I]$. This conservative approach to estimate maximal inhibition has been used by others (Lu, Miwa, Prakash, Gan, & Balani, 2007). If a pharmacokinetics parameter was reported in multiple sources, the sample mean was used.

The f_m data was curated from several different types of published studies. Most of the f_m data were estimated from substrate depletion studies in human liver microsomes, in which the substrate is incubated with or without CYP-selective inhibitors (Huskey et al., 1995). The percentage inhibition caused by the CYP-selective inhibitor

reflects the f_m of drug for this CYP. The f_m can also be estimated through *in vivo* pharmacokinetic studies comparing the AUC or clearance of a substrate in the presence and absence of a CYP-selective inhibitor (Creighton et al., 2008) or in a pharmacogenetic PK study where it can be calculated from the fold-change in exposure of a victim drug in extensive metabolizers compared to poor metabolizers (Ito et al., 2005; P. Zhang et al., 2015).

4.2.4 A Mixture Drug-Count-Response Model for the High-Dimensional Drug Effect on Myopathy

In 2015, our group developed a mixture drug-count-response model (MDCM) for identifying high-dimensional drug interaction-induced ADEs (P. Zhang et al., 2015). In the MDCM, i is denoted as the number of drugs for i -way drug combinations; j is the j th i -way drug combinations; n_{ij} is the total number of patients taking j th i -way drug combination; and y_{ij} is the number of cases among those n_{ij} patients. The parameter π represents the proportion of drug combinations that follow the drug-count-response model. The probability distribution of y_{ij} is defined as the following mixture model

$$P(y_{ij}) = (1 - \pi) \text{Bin}(n_{ij}, y_{ij}, P_{const}) + \pi \text{Bin}(n_{ij}, y_{ij}, P_{count}) \quad (\text{Eq. 1})$$

where P_{const} and P_{count} represent a constant ADE risk probability and a drug-count-response ADE risk probability respectively:

$$P_{const} = \frac{c \times \exp(\beta_0)}{1 + \exp(\beta_0)}, P_{count} = \frac{c \times \exp(\beta_0 + \beta_1(i-1))}{1 + \exp(\beta_0 + \beta_1(i-1))}.$$

After applying MDCM to both FAERS and INPC-CDM data, a local false discovery rate (LFDR) is calculated for each drug combination. A LFDR demonstrates the significance of

a drug combination that follows the drug-count-response model. Then, all the drug combinations are ranked based on the LFDR accordingly.

$$LFDR = \frac{(1-\pi)Bin(n_{ij}, y_{ij}, P_{const})}{(1-\pi)Bin(n_{ij}, y_{ij}, P_{const}) + \pi Bin(n_{ij}, y_{ij}, P_{const})} \quad (\text{Eq. 2})$$

This MDCM allows different drug combinations to share the same risk probabilities, either a constant risk or a drug-count-response risk. This strategy overcomes the small sample size in each high dimensional drug combination. In this chapter, we evaluate the dimension of the drugs taken from single drug to 5 co-administered drugs.

| INPC – CDM | FAERS |
|--------------------------------------------------------|-------------------|
| Disorder of skeletal muscle (Myopathy) | Myopathy |
| Muscle pain (Myalgia and myositis) | Myositis |
| | Myalgia |
| Muscle, ligament and fascia disorders (Rhabdomyolysis) | Rhabdomyolysis |
| Muscle weakness | Muscular weakness |
| Polymyositis | Polymyositis |
| Myoglobinuria | Myoglobinuria |
| | Muscle spasms |
| | Muscle injury |
| | Muscle fatigue |

Table 4.1 Myopathy phenotype definition

4.2.5 Sensitivity Data Analysis

Overlapping and mutually validated significant (LFDR < 0.05) high dimensional drug combinations between FAERS and INPC-CDM data were further evaluated. In the follow-up sensitivity analyses, a sequential logistic regression was conducted to compare myopathy risk between any two adjacent dimensions of drug combinations, for example 1-way vs 2-way, 2-way vs 3-way, etc. The logistic regression model included the demographic, and the number of other co-medications as covariates to adjust for the confounding effects.

4.2.6 High Dimensional Drug Interaction In-Vitro to In-Vivo Extrapolation (IVIVE)

A substrate's clearance (Cl_{total}) is composed of hepatic clearance (Cl_H) and renal clearance (Cl_R), $Cl_{total} = Cl_H + Cl_R$. The ratio of area of under the concentration-time curve in the presence and absence of inhibitor (AUCR) is defined in (Eq 3), where Cl_H' is the hepatic clearance of the substrate drug after inhibition (Ito, Brown, & Houston, 2004).

$$AUCR = \frac{AUC'}{AUC} = \frac{Cl_{total}}{Cl_{total}'} = \frac{Cl_H + Cl_R}{Cl_H' + Cl_R} \quad (\text{Eq. 3})$$

Let the fraction of the renal clearance (f_e) = $\frac{Cl_R}{Cl_H + Cl_R}$, and $Cl_R = Cl_H \left(\frac{f_e}{1 - f_e} \right)$, and

the AUCR can be further defined as

$$AUCR = \frac{1}{(1 - f_e) \frac{Cl_H'}{Cl_H} + f_e}$$

For high clearance drugs in which Cl_{int} approaches hepatic blood flow, Cl_H is limited by hepatic blood flow rate, and significant hepatic inhibition is unlikely. For the low

clearance drugs, $\frac{Cl_H'}{Cl_H} = \frac{f_u Cl_{int}'}{f_u Cl_{int}} = \frac{Cl_{int}'}{Cl_{int}} = \sum f m_i \frac{Cl_{int,i}'}{Cl_{int,i}}$, where $f m_i$ is the fraction of metabolism by CYP enzyme (i), and $\sum_{i=1}^n f m_i = 1$ where n is the number of metabolism routes.

When there is only one inhibitor and one metabolizing enzyme, Ito *et al.* (Ito et al., 2004; Ito et al., 2005) showed that the change in clearance in the presence of a DDI can be predicted as $\frac{Cl_{int,i}'}{Cl_{int,i}} = \frac{1}{1 + \frac{[I]}{K_i}}$, where $[I]$ is fraction unbound concentration of the inhibitor and K_i is the competitive inhibition constant. If there are inhibitors on the same enzyme, the inhibitors' effects are assumed to fit an additive model (Rostami-Hodjegan & Tucker, 2007). The AUCR for inhibition of a single enzyme then becomes

$$AUCR = \frac{Cl_{int,i}'}{Cl_{int,i}} = \frac{1}{1 + \left(\sum_{j=1}^J \frac{[I_j]}{K_{i_j}} \right)}$$

When considering multiple inhibitors and multiple enzyme pathways, the AUCR is predicted by equation 4 (Lu et al., 2007):

$$AUCR = \frac{1}{(1-f_e) \sum_{i=1}^n f m_i \times \frac{1}{1 + \left(\sum_{j=1}^J \frac{[I_j]}{K_{i_j}} \right)} + f_e} \quad (\text{Eq. 4})$$

We used this method to predict the AUCR for each substrate drug in the presence of 1- to 4- inhibitors.

4.3 Results

4.3.1 Drug Selections

As our MDCM is computationally expensive, we limited this analysis to top 30 drugs. After normalizing drug names by their DrugBank IDs, 1,238 and 1,716 FDA approved drugs were identified in the INPC-CDM and FAERS datasets. Of these, after

selecting drugs 268 and 172 drugs in INPC-CDM and FAERS, respectively, had a relative frequency $\geq 0.5\%$, with 119 drugs overlapping both databases. Among these 119 drugs, 95 have reported myopathy risk according to SIDER (<http://sideeffects.embl.de/>). We curated the *in vitro* pharmacokinetic data for these 119 drugs to determine their potential to interact via CYP450 inhibition. Nine drugs had reported fraction metabolism (f_m) by human CYP450 enzymes. These drugs were evaluated as substrates for the prediction of area under the concentration-time curve ratio in the presence to absence of inhibitor (AUCR). Published competitive inhibition (k_i) values for at least one CYP450 enzyme were available for 64 of the drugs. The top 23 drugs with inhibition ranked by $\frac{f_u \times C_{max}}{k_i}$ were selected as inhibitors (Figure 4.1).

4.3.2 INPC-CDM and FAERS Show Strikingly Different Drug-Count Myopathy Response Mixture Models

Under the MDCM, each drug combination has a probability of being assigned to either a drug-count myopathy response model or a constant myopathy risk model. The drug-count response model and constant model share the same myopathy risk when subjects take only one drug. The drug-count myopathy response model captures the overall trend of the nonlinear relationship between the number of co-administered drugs and myopathy risk. It also characterizes a constant myopathy risk and the maximum myopathy risk as the number of co-administered drugs increases. Strikingly, the drug-count response model demonstrated very different trends between drug combinations and myopathy risk in INPC-CDM and FAERS databases. INPC-CDM had constant myopathy risk of 0.42, and maximum risk of 0.74; while FAERS data shows

constant risk of 0.07 and a maximum risk of 1.0 (Figure 4.2). In the INPC-CDM, the maximum myopathy risk occurs when the four drugs are co-administered. However, in the FAERS, the myopathy risk continues to increase beyond 5-drug combinations.

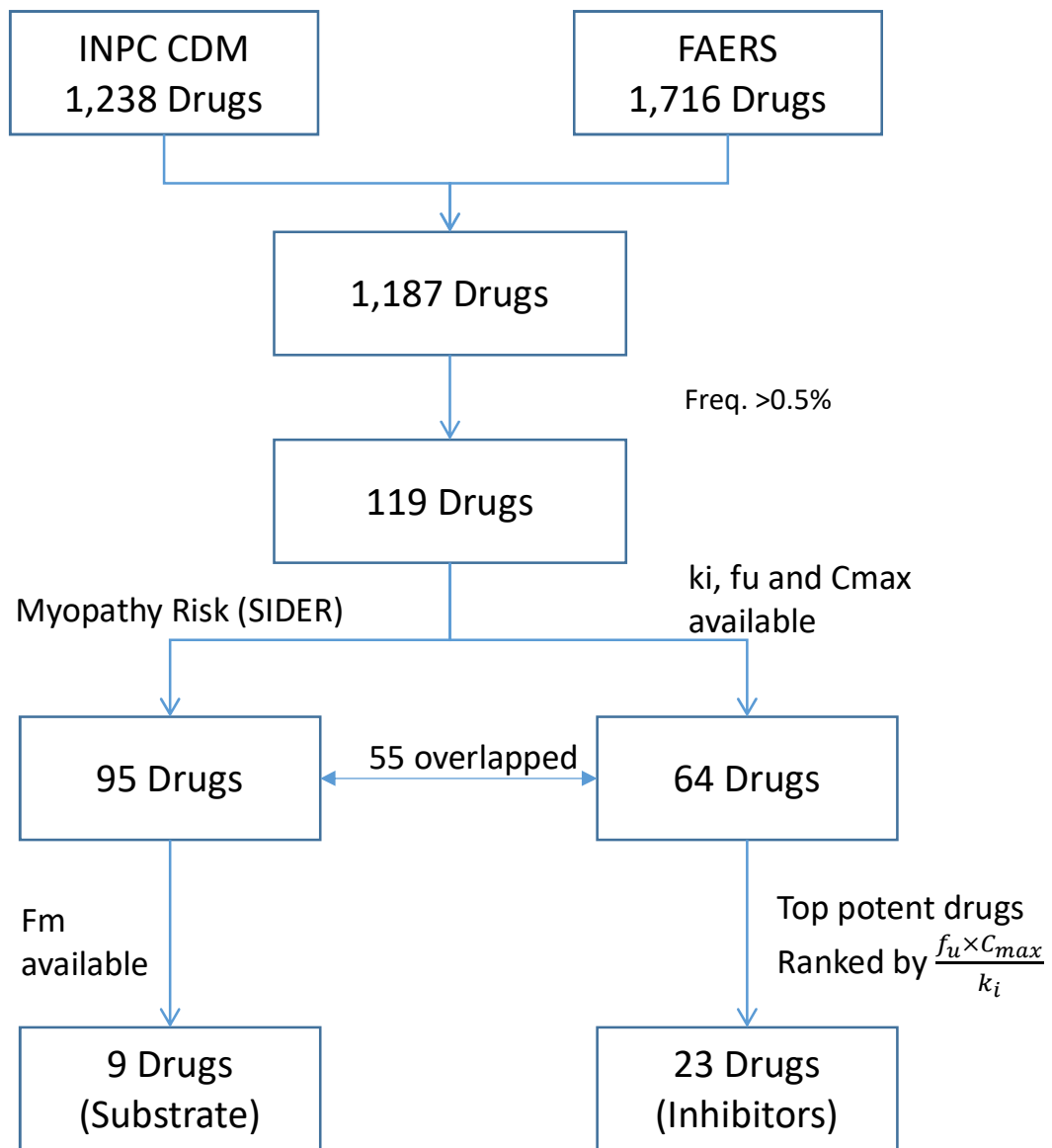


Figure 4.1 Drugs selection flow chart

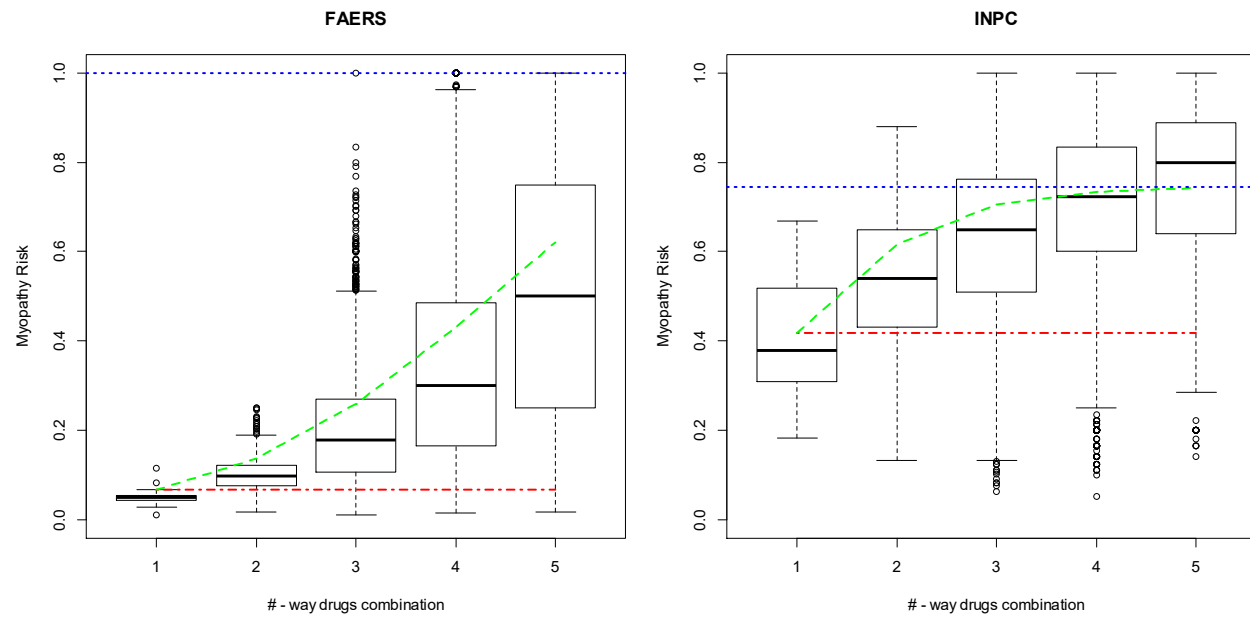


Figure 4.2 Mixture dose-response model curve for both INPC and FAERS data. Red dash line is the baseline risk estimated; Blue dash line is maximum risk estimated.

4.3.3 Overlapping Drug Combinations in INPC-CDM and FAERS

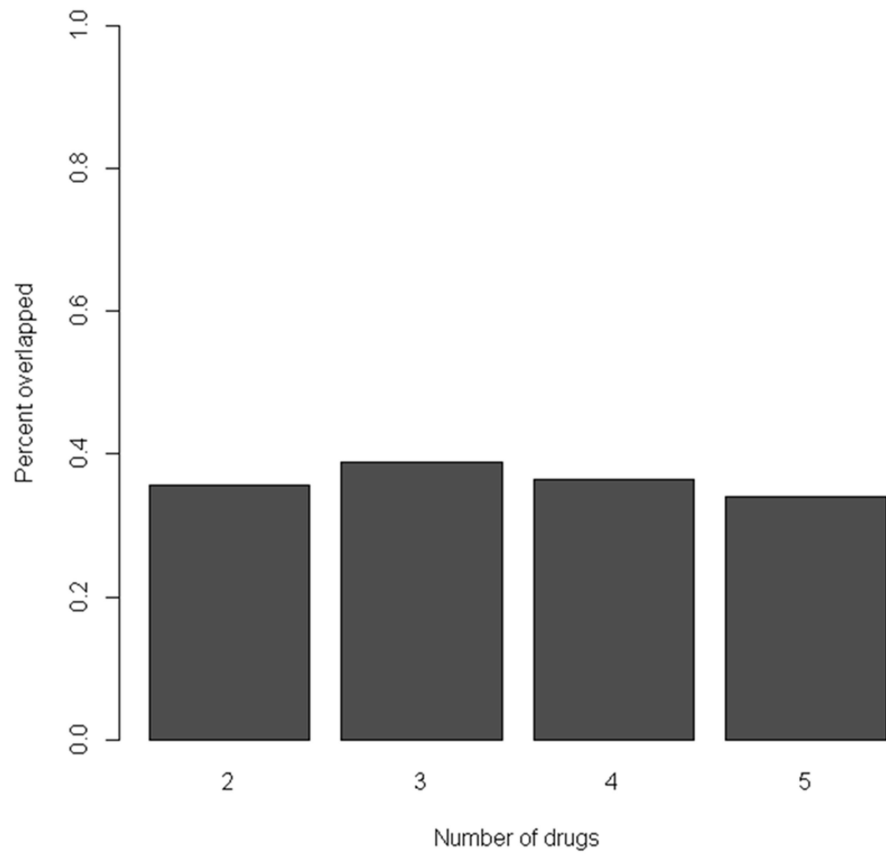
The drug-count-response mixture model generates a local false discovery rate (LFDR) statistic that allows us to differentiate drug combinations that are more likely to follow the drug-response myopathy risk model than the constant risk model. Figure 4.3 shows the percentage of overlapping drug combinations that are shared between two databases, with a LFDR of 0.05. We see strong and consistent evidence of increased myopathy risk with an overlap of 37-40% of 2-way to 5-way drug combinations shown to increase myopathy risk in both databases.

4.3.4 The Overall Trend of High Dimensional Drug Interactions in Pharmacokinetics Predicted by IVIVE

One common mechanism of drug-drug interaction occurs through inhibition of drug metabolism. In order to assess whether the potential DDIs identified by pharmaco-epidemiology evidence are due to pharmacokinetic drug interactions, we used IVIVE to evaluate the change in drug exposure between a drug administered alone and co-administered with 2-, 3-, or more drugs. In the IVIVE prediction, CYP substrates and CYP inhibitors are differentially defined. Among our 30 selected drugs, nine are CYP substrates with curated f_m , f_e , and f_u data from the literature. We also obtained the k_i , f_u , and C_{max} for 23 drugs identified as CYP inhibitors. These data were combined through IVIVE to predict the area under the concentration time curve ratio (AUCR) for 156 two-way, 1302 three-way, 6971 four-way, and 26901 five-way drug interaction combinations (Figure 4.4). When we look at all the AUCR data under different dimensions of drug interactions, their medians, 75th percentiles, and the maximum AUCRs all reach a maximum plateau with the co-administration of 3 drugs.

4.3.5 Significant Three-way Drug Interactions and Their Sensitivity Analyses

Among the 9 substrates and 23 inhibitors, 28 three-way drug interactions have a predicted AUCR >2 and a LFDR from the MDCM model < 0.05. We hypothesized that among these drugs, the risk of myopathy would increase as the number of co-occurring drugs increased. Therefore, for each 3-drug combination we evaluated myopathy risk between individuals taking 1 drug vs any 2-drug combination, and between 2-drug and 3-drug combinations. One drug-triplet showed a strong increasing trend in myopathy risk: omeprazole, clonidine, and fluconazole (Figure 4.5) after adjusting for the number of additional co-medications, age, and gender. In the FAERS data set, taking any two of these three drugs together increased the risk of myopathy by 1.88-fold compared to taking any of the drugs alone ($p = 0.012$). Taking the 3-drug combination of omeprazole, clonidine and fluconazole increased the risk of myopathy by 5.01-fold compared to the 2-drug combinations ($p = 0.000012$). In the INPC-CDM, taking any two of these drugs concurrently increased the risk of myopathy by 1.75-fold ($p=7.8E-13$) and taking all three together increased the risk by 4.18-fold ($p=0.0069$) compared to taking one drug and two drugs, respectively. Similar comparisons of 2- and 3-drug combinations to 1- and 2 drug combinations for the other 27 drug triplets examined did not reveal significantly increased myopathy risk in both data sets for other drug combinations (Table 4.2).



| | | | | |
|--------------------|------------|-------------|-------------|--------------|
| Overlapped* | 109 | 1044 | 1482 | 444 |
| Common | 435 | 3449 | 5220 | 10020 |
| INPC* | 235 | 1891 | 2975 | 1042 |
| FAERS* | 180 | 1997 | 8574 | 20615 |

Figure 4.3 Overlapped drug combination between INPC and FAERS dataset. * LFDR < 0.05

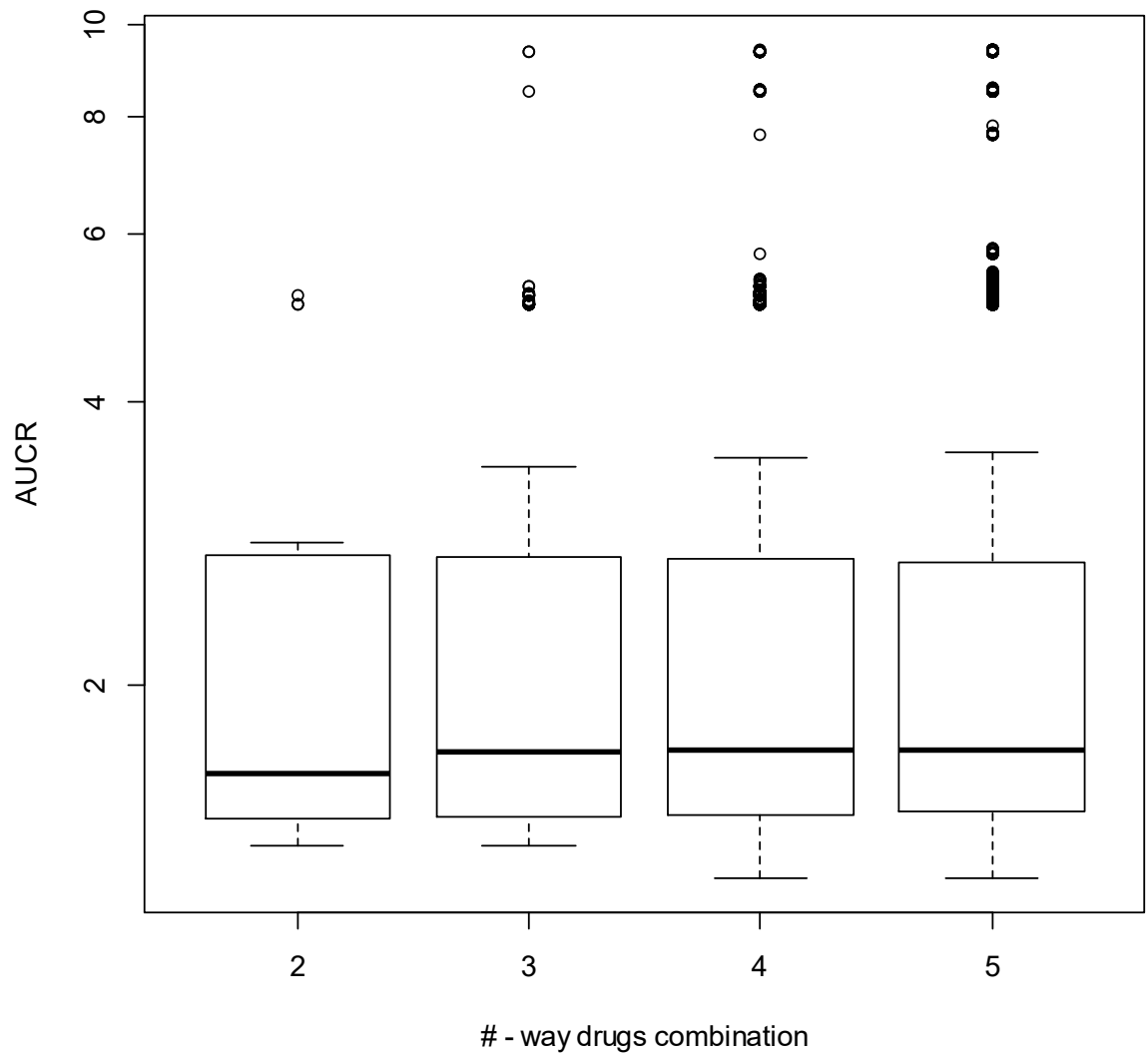


Figure 4.4 Simulated AUCR of 2-way to 5-way drugs combination. Only AUCR \geq 1.25 are reported here.

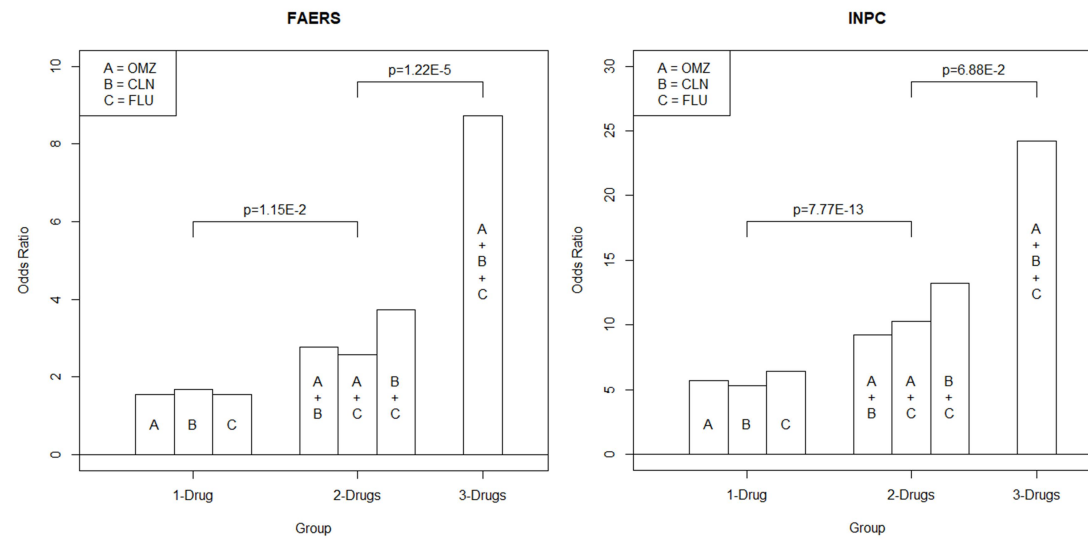


Figure 4.5 Group-wise odds ratio analysis; OMZ is omeprazole or esomeprazole; CLN is clonidine; FLU is fluconazole

| Drugs | | | LFDR | | AUCR | | | INPC-CDM | | | | | FAERS | | | | |
|--------------------------------|-------------|----------------|----------|----------|------|-------|-------|----------|-------|------|----------|----------|-------|-------|------|----------|----------|
| Substrate | Inhibitor 1 | Inhibitor 2 | INPC | FAERS | AUCR | AUCR1 | AUCR2 | OR | Total | Case | G12 | G23 | OR | Total | Case | G12 | G23 |
| omeprazole | Fluconazole | clonidine | 5.67E-03 | 2.68E-03 | 9.35 | 5.06 | 1.45 | 18.46 | 37 | 24 | 2.64E-12 | 4.18E-01 | 6.41 | 105 | 19 | 1.38E-01 | 2.00E-01 |
| esomeprazole | Fluconazole | clonidine | 1.40E-04 | 7.14E-12 | 8.51 | 5.17 | 1.35 | 36.67 | 28 | 22 | 1.34E-04 | 4.35E-02 | 13.04 | 100 | 31 | 1.09E-01 | 7.64E-06 |
| a- omeprazole/ esomeprazole | Clonidine | fluconazole | | | | | | 24.21 | | | 7.77E-13 | 6.88E-02 | 8.73 | | | 1.15E-02 | 1.22E-05 |
| esomeprazole | Fluconazole | acetaminophen | 4.22E-14 | 5.38E-04 | 5.20 | 5.17 | 1.01 | 18.77 | 279 | 182 | 5.13E-05 | 5.06E-02 | 5.20 | 922 | 140 | 3.75E-07 | 9.67E-02 |
| esomeprazole | Fluconazole | trazodone | 6.57E-04 | 2.78E-06 | 5.19 | 5.17 | 1.01 | 17.31 | 71 | 45 | 3.83E-24 | 8.04E-01 | 10.45 | 68 | 18 | 1.22E-01 | 1.04E-01 |
| esomeprazole | fluconazole | fentanyl | 7.06E-08 | 2.28E-15 | 5.18 | 5.17 | 1.01 | 39.09 | 54 | 43 | 2.88E-24 | 4.96E-01 | 8.09 | 303 | 66 | 6.21E-18 | 2.30E-01 |
| esomeprazole | fluconazole | metoclopramide | 3.40E-06 | 1.29E-09 | 5.17 | 5.17 | 1.00 | 36.67 | 42 | 33 | 1.20E-09 | 9.91E-02 | 6.46 | 368 | 67 | 1.66E-01 | 9.00E-01 |
| esomeprazole | fluconazole | tramadol | 1.76E-02 | 3.69E-10 | 5.17 | 5.17 | 1.00 | 15.88 | 44 | 27 | 4.70E-03 | 4.40E-01 | 8.69 | 165 | 38 | 1.42E-01 | 2.76E-02 |
| esomeprazole | fluconazole | levofloxacin | 9.02E-03 | 1.41E-23 | 5.17 | 5.17 | 1.00 | 16.88 | 43 | 27 | 1.29E-16 | 6.44E-01 | 9.15 | 363 | 87 | 7.19E-01 | 8.97E-01 |
| esomeprazole | fluconazole | ciprofloxacin | 1.06E-03 | 4.06E-14 | 5.17 | 5.17 | 1.00 | 18.42 | 54 | 35 | 1.68E-05 | 1.00E-01 | 7.02 | 411 | 80 | 4.42E-04 | 2.95E-01 |
| omeprazole | fluconazole | acetaminophen | 1.25E-21 | 4.09E-13 | 5.10 | 5.06 | 1.02 | 18.67 | 450 | 293 | 3.78E-48 | 4.57E-02 | 5.67 | 1109 | 181 | 1.28E-01 | 4.15E-01 |
| omeprazole | fluconazole | trazodone | 4.12E-04 | 6.72E-16 | 5.08 | 5.06 | 1.01 | 18.26 | 65 | 42 | 4.48E-52 | 7.03E-01 | 12.74 | 141 | 43 | 4.50E-01 | 6.23E-04 |
| omeprazole | fluconazole | fentanyl | 1.07E-05 | 3.97E-16 | 5.08 | 5.06 | 1.01 | 27.69 | 49 | 36 | 4.33E-49 | 5.47E-01 | 7.85 | 343 | 73 | 1.21E-03 | 7.31E-01 |
| omeprazole | fluconazole | paroxetine | 1.69E-02 | 1.88E-06 | 5.06 | 5.06 | 1.00 | 18.00 | 28 | 18 | 1.00E-04 | 1.70E-01 | 7.38 | 148 | 30 | 1.94E-03 | 9.55E-01 |
| omeprazole | fluconazole | diclofenac | 2.42E-02 | 2.04E-03 | 5.06 | 5.06 | 1.00 | 17.78 | 25 | 16 | 8.62E-08 | 2.30E-01 | 6.61 | 97 | 18 | 3.57E-01 | 7.10E-03 |
| omeprazole | fluconazole | levofloxacin | 2.09E-03 | 1.93E-37 | 5.06 | 5.06 | 1.00 | 17.50 | 55 | 35 | 6.35E-25 | 4.23E-01 | 9.94 | 502 | 128 | 6.16E-02 | 5.28E-07 |
| loratadine | fluconazole | tramadol | 2.11E-03 | 1.37E-03 | 3.13 | 2.65 | 1.06 | 22.00 | 32 | 22 | 4.50E-09 | 6.91E-01 | 7.52 | 68 | 14 | 8.59E-01 | 5.49E-01 |
| loratadine | clonidine | metoclopramide | 2.70E-02 | 2.40E-02 | 2.86 | 2.84 | 1.00 | 26.67 | 11 | 8 | 1.74E-05 | 6.80E-01 | 7.81 | 33 | 7 | 3.91E-04 | 8.40E-01 |
| loratadine | clonidine | levofloxacin | 1.27E-03 | 4.50E-03 | 2.84 | 2.84 | 1.00 | N/A | 10 | 10 | 7.61E-05 | 6.96E-01 | 8.06 | 46 | 10 | 4.58E-01 | 7.91E-01 |
| loratadine | fluconazole | fentanyl | 2.32E-02 | 2.64E-04 | 2.68 | 2.65 | 1.02 | 35.00 | 9 | 7 | 3.63E-16 | 6.77E-01 | 7.46 | 88 | 18 | 1.61E-05 | 6.22E-01 |
| loratadine | fluconazole | acetaminophen | 2.57E-19 | 1.31E-04 | 2.67 | 2.65 | 1.04 | 39.34 | 148 | 118 | 3.97E-86 | 3.21E-03 | 5.86 | 262 | 44 | 1.19E-03 | 3.68E-01 |
| loratadine | fluconazole | metoclopramide | 2.50E-03 | 1.17E-04 | 2.67 | 2.65 | 1.00 | 100.00 | 11 | 10 | 1.49E-07 | 2.06E-01 | 8.88 | 64 | 15 | 1.28E-01 | 4.15E-01 |
| loratadine | fluconazole | levofloxacin | 2.14E-04 | 1.02E-09 | 2.65 | 2.65 | 1.00 | 80.00 | 18 | 16 | 9.40E-07 | 6.35E-02 | 10.24 | 115 | 30 | 4.97E-02 | 1.16E-01 |
| venlafaxine | tramadol | clonidine | 5.74E-03 | 2.47E-07 | 2.30 | 2.03 | 1.06 | 36.67 | 14 | 11 | 5.97E-08 | 1.18E-01 | 11.38 | 71 | 20 | 6.26E-03 | 5.49E-03 |
| venlafaxine | tramadol | fentanyl | 1.83E-05 | 2.69E-29 | 2.06 | 2.03 | 1.03 | 73.34 | 25 | 22 | 1.30E-22 | 4.55E-02 | 15.96 | 203 | 72 | 9.58E-21 | 1.33E-09 |
| venlafaxine | tramadol | trazodone | 6.48E-06 | 7.86E-13 | 2.06 | 2.03 | 1.03 | 18.00 | 112 | 72 | 2.39E-25 | 2.05E-01 | 7.72 | 281 | 59 | 1.64E-05 | 1.97E-03 |
| venlafaxine | tramadol | metoclopramide | 7.86E-04 | 8.81E-16 | 2.05 | 2.03 | 1.02 | 28.57 | 27 | 20 | 3.39E-16 | 3.34E-01 | 12.16 | 149 | 44 | 1.56E-09 | 4.52E-05 |
| venlafaxine | tramadol | ciprofloxacin | 1.60E-02 | 3.45E-12 | 2.05 | 2.03 | 1.00 | 16.92 | 35 | 22 | 1.30E-20 | 5.76E-01 | 14.51 | 90 | 30 | 8.54E-05 | 5.01E-02 |
| venlafaxine | tramadol | acetaminophen | 7.15E-17 | 1.05E-14 | 2.05 | 2.03 | 1.00 | 17.87 | 390 | 250 | 1.52E-59 | 1.07E-01 | 6.10 | 766 | 133 | 2.09E-16 | 1.04E-04 |

Table 4.2 Top 28 3-way drugs interaction combination. G12 and G23 means the p-value of odds ratio compared taken 1 of 3 drugs vs

2 of 3 drugs, and taken 2 of 3 drugs and 3 of 3 drugs. a. Combine analysis: treated omeprazole and esomeprazole as same drug.

The AUCRs of omeprazole due to the inhibition by clonidine and fluconazole alone or together were estimated from literature data using mechanistic static interaction models (Ito et al., 2005). Omeprazole is a racemic mixture of R- and S-omeprazole. While both enantiomers are metabolized by CYP2C19 and CYP3A (Figure 4.6), the f_m 's for each pathway are slightly different (Table 4.3). Fluconazole is a well-known CYP2C19 and CYP3A inhibitor with a median k_i value of 5.075 μM and 13.25 μM , respectively (Moody, Griffin, Mather, McGinnity, & Riley, 1999; Niwa, Shiraga, & Takagi, 2005; Wienkers et al., 1996). The k_i of clonidine for CYP3A is 0.15 μM (Tanaka, Nakamura, Inomata, & Honda, 2006). Incorporating these values into the IVIVE model (Equation 4), predicts AUCRs of R- and S-omeprazole of 1.45 and 1.35, respectively, after the clonidine inhibition, and 5.06 and 5.17 after fluconazole inhibition. Following co-administration of the 3 drugs, the AUCRs of R- and S-omeprazole are predicted to be 9.35 and 8.51, respectively. Alternatively, the f_m of omeprazole can also be estimated from pharmacogenetics PK study. Venkatakrisnan *et al.* calculated the f_m of omeprazole from AUCR of CYP2C19 extensive metabolizer and poor metabolizer (Venkatakrisnan, Obach, & Rostami-Hodjegan, 2007; Yasui-Furukori et al., 2004) ($f_m = 0.87$ for CYP2C19). Using this f_m , the AUCR estimated from IVIVE is predicted to be 1.14 after the clonidine inhibition; 5.99 after fluconazole inhibition; and 7.69 after inhibition by both drugs.

4.4 Discussion

Using the MDCM, we mined the high dimensional drug interactions among 30 common drugs in two independent health record databases. We identified a number of

high-dimensional drug interactions that have increased risk of myopathy in both databases. This model further reveals interesting differences in the drug-count response relationship in the two databases. In the FAERS, the maximum myopathy risk goes to almost 1.0 as the dimension of drug combinations increases. Thus, a patient would be expected to experience a myopathy event if he/she takes a large number of drugs. On the other hand, the maximum myopathy risk in the INPC-CDM data is 0.72. However, because it was estimated from a case-control study, this number cannot be simply interpreted as a population maximum myopathy risk. Our 1:10 case control design has a higher myopathy case frequency (0.09) than the INPC-CDM population myopathy risk (0.067) reported previously (Duke et al., 2012). Hence, we anticipate that the population maximum HDDI myopathy risk in the INPC-CDM data is less than 0.72. While we observe that the maximum myopathy risk is higher in the FAERS database than in the INPC-CDM database, it is extremely striking that the frequency in FAERS approaches one. Although the maximum myopathy risk in the INPC-CDM data is lower than that of FAERS, it is still an extremely common ADE.

Among the statistically significant three-way drug interactions identified from MDCM and validated by two databases, more stringent sequential comparisons are further conducted between 1-drug vs 2-drug and 2-drug vs 3-drugs. We further demonstrated that omeprazole, clonidine, and fluconazole, exhibited statistically significantly increased myopathy risk from single drug to three-drug combination in both databases.

In parallel to our MDCM model, we evaluated the potential CYP PK interactions among the 30 drugs. In the IVIVE high dimensional drug interaction prediction, one drug acts as the substrate and the other drugs are assumed to be reversible inhibitors of CYP enzymes. Substrate AUCR changes were predicted based the curated *in vitro* PK data. The maximum predicted AUCR was reached with three-drug combinations. In contrast, the maximum HDDI myopathy risk observed in the two medical record databases did not occur until five drugs were co-administered. This suggests that additional mechanisms are responsible for the increased risk of myopathy observed with HDDIs. For instance, we have previously shown that the increased risk myopathy with co-administration of loratadine and simvastatin is due to a pharmacodynamic mechanism in the muscle cell (Han et al., 2015). We have also shown that the increased risk of myopathy observed when chloroquine is co-administered with simvastatin is the result of increased lysosomal OATP1B1 protein degradation (Alam et al., 2016).

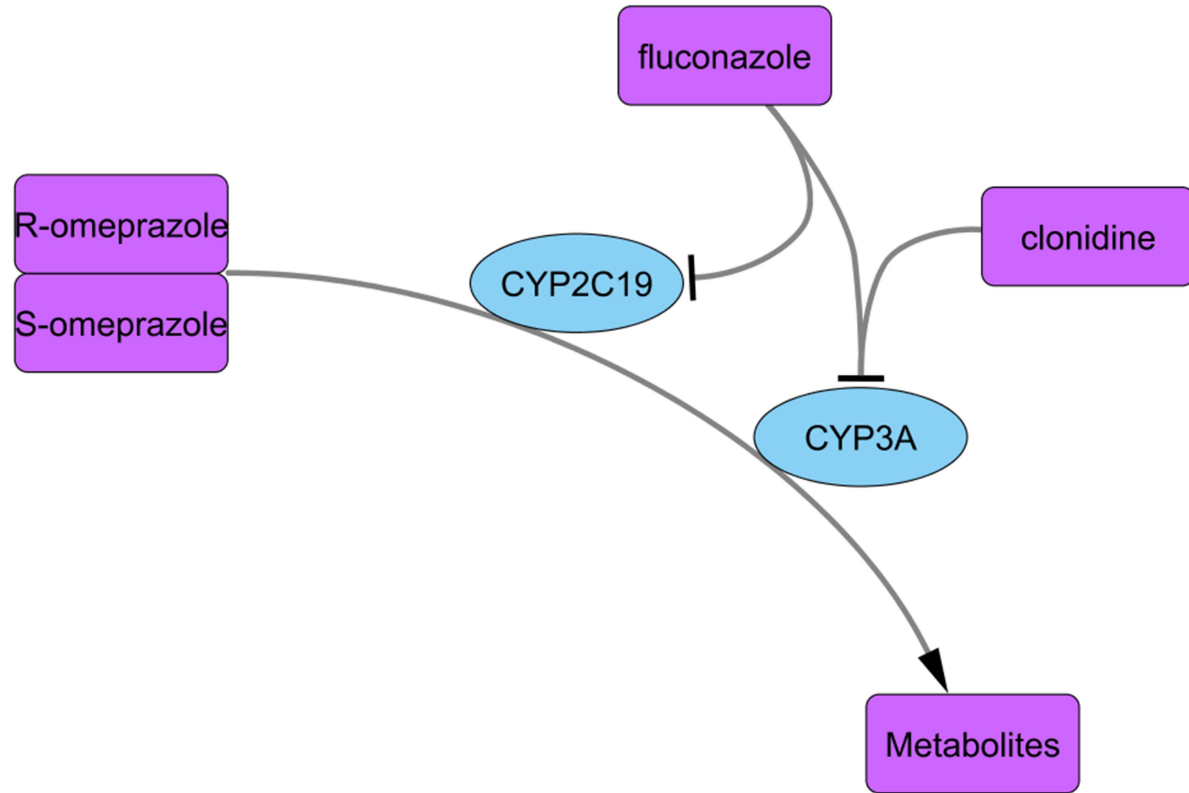


Figure 4.6 hepatic metabolism drugs interaction between omeprazole, clonidine, and fluconazole.

| Drug | DBID | 1A2 | 2A6 | 2B6 | 2C8 | 2C9 | 2C19 | 2D6 | 2E1 | 3A | unknown CYP pathway | Fe |
|--------------|---------|--------|-------|-----|-----|-------|--------|--------|--------|--------|---------------------|--------|
| Alprazolam | db00404 | - | 4.50% | - | - | 9.80% | 1.20% | - | - | 53.00% | 31.50% | 20.00% |
| Duloxetine | db00476 | 41.27% | - | - | - | - | - | 50.06% | - | - | 8.67% | - |
| esomeprazole | db00736 | - | - | - | - | - | 73.00% | - | - | 27.00% | - | 0.50% |
| lansoprazole | db00448 | - | - | - | - | - | 67.90% | - | - | 32.10% | - | - |
| Loratadine | db00455 | - | - | - | - | - | 20.00% | 10.00% | - | 70.00% | - | 5.00% |
| omeprazole | db00338 | - | - | - | - | - | 68.00% | - | - | 32.00% | - | - |
| simvastatin | db00641 | - | - | - | - | - | - | - | - | 52.50% | 47.50% | 10.00% |
| Tramadol | db00193 | 14.08% | - | - | - | 8.18% | 8.89% | 23.89% | 10.32% | 34.63% | - | 20.00% |
| Venlafaxine | db00285 | 0.47% | - | - | - | 6.01% | - | 87.39% | - | 6.13% | - | 5.50% |

Table 4.3 Drugs fm and fe

As the risk of myopathy consistently increased among 2-way and 3-way combinations of omeprazole, clonidine, and fluconazole, we closely examined the predicted AUCR using IVIVE for these combinations. Compared to the 1.5-5 fold increase in AUCR of R- and S-omeprazole when inhibited by only one drug, either clonidine or fluconazole, the co-administration of both drugs with omeprazole led to an AUCR of 8.5-9.4. As f_m is a critical component of predicting the extent of interaction, we evaluated the interaction using both *in vitro* data and data from clinical pharmacogenetic study of CYP2C19 to determine omeprazole's f_m . We estimated $f_{mCYP2C19}$ (0.68) from *in vitro* data mining (McGinnity, Parker, Soars, & Riley, 2000); Roman *et al.* estimated omeprazole clearance is decreased by 50% in CYP2C19 heterozygous patients ($f_{mCYP2C19} = 0.5$) (Roman *et al.*, 2014); Venkatakrisnan *et al.* estimated $f_{mCYP2C19}$ (0.87) from AUCR of CYP2C19 PMs (Venkatakrisnan *et al.*, 2007; Yasui-Furukori *et al.*, 2004). The different $f_{mCYP2C19}$ estimation results AUCR of omeprazole inhibit by clonidine and fluconazole range from 7.55-11.74. Although the mechanism of omeprazole induced myopathy has not yet been understood, various hypotheses have been proposed such as the induction of auto-immune antibodies (Clark & Strandell, 2006; Sivakumar & Dalakas, 1994), and the irreversible inhibition of potassium-hydrogen ATPase in the skeletal muscles (Schonhofer, Werner, & Troger, 1997).

Recently, Sansone *et al.* surveyed the Italian National Network of Pharmacovigilance Database and found that omeprazole was more frequently involved in reports of myopathy than any other non-statin drug (Capogrosso Sansone *et al.*, 2017). From these evidences, omeprazole is the primary candidate that induced

myopathy. In addition, there were sixteen case reports in the literature (Jeon et al., 2016; Visruthan, Boo, Kader, Ping, & Ong, 2012)) and summarized in a case series that associate the use of omeprazole with myopathy (Clark & Strandell, 2006). Among these sixteen cases, two were identified as Asians, but the race were not reported among the other cases. Similarly, in our database, the race data were incomplete and sometimes inaccurate. Hence, the race effect of omeprazole induces myopathy remains unknown.

One limitation of the MDCM model is its large computational expense. Because of this, only a limited number of drugs (i.e. 30) could be screened for high-dimensional drug interactions. Thus, screening steps are required to reduce the number of drugs. More research is needed to increase the computational efficiency of MDCM. Future research will also apply the MDCM model to evaluate the drug-count response patterns for other ADEs.

Although the mechanistic static model is a well-accepted screening tool for pharmacokinetic drug interactions, it may over- or under-estimate the extent of interaction. Of note, our model only included reversible inhibition and did not consider gut wall metabolism effects of inhibitors. Thus, we may have under-estimated the AUCR for orally administered drugs. While mechanistic static models have been established to predict the effect of mechanism based inhibitors and CYP inducers (Fahmi et al., 2008; Mayhew, Jones, & Hall, 2000), these models have not been validated with respect to high-dimensional interactions among drugs that inhibit the same enzyme. Additionally, interactions in other pathways, such as drug transporters, could be included using more

mechanistic and time-dependent modeling techniques. In addition, we utilized C_{\max} data from the literature without respect to the doses of drugs observed in the clinical records.

This study demonstrates the power to elucidate clinically significant high-dimensional drug interactions from clinical records. Using two unique data sets, ADE case reports from the FAERS and structured electronic medical record data from the INPC-CDM, we observed increasing trends in myopathy risk with higher medication burden. As a large number of DDIs are the result of PK interactions at the level of CYP enzymes, we also estimated the increased exposure of 9 substrate drugs in the presence of 2, 3, or more inhibitors. Although we demonstrated that decreased clearance of drugs due to CYP inhibition is one source of the increased myopathy risk among polypharmacy patients, this mechanism is unable to fully explain the increased risk of myopathy observed in subjects taking 4 or more medications. As our computational efficiency expands to allow for the evaluation of greater number of drugs using our MDCM model, additional pharmacokinetic and pharmacodynamic mechanisms of interaction will need to be considered to further account for the increased risk of ADEs observed in polypharmacy patients.

Chapter 5. Conclusion

In this research, we create a fm database that will help to predict drug interactions. These data characterize all the hepatic CYP450 metabolic pathways and their contributions in predicting drug interactions. We therefore further explore and predict the drug interactions among these 57 cancer drugs in the following case study. In this dissertation, we have further identified f_u , C_{max} and $K_{i,u}$ for 32 out of 57 cancer drugs through the PubMed literature review. Then, each drugs pair selected from 32 drugs are further evaluated twice in order to predict their interactions. Each time, one drug serves as substrate and the other one serves as inhibitor, and vice-versa. Following the FDA DDI guideline(Food & Administration, 2012) and expert experience, an AUCR >1.5 is regarded as the moderate or strong DDI evidence. Based on our PBPK model based DDI predictions, we find 97 drug pairs with AUCR more than 1.5. After been validated in DrugBank, Drugs.com and PubMed, 33 pairs have at least one clear DDIs evidence mentioned in DrugBank or Drugs.com or in PubMed.

Using the MDCM with IVIVE, we mined the high dimensional drug interactions among 30 common drugs in two independent health record databases. We identified a number of high-dimensional drug interactions that have increased risk of myopathy. The predicted AUCR demonstrated the maximum inhibition in CYP enzymes when the dimension of drug combination reaches three. In contrast, the maximum HDDI myopathy risk observed in the two medical record databases did not occur until five drugs were co-administered. This suggests that additional mechanisms are responsible for the increased risk of myopathy observed with HDDIs. Using more stringent

sequential comparisons between 1-drug vs 2-drug and 2-drug vs 3-drugs, we further demonstrated that omeprazole, clonidine, and fluconazole, exhibited statistically significantly increased myopathy risk from single drug to three-drug combination in both databases. Compared to the 1.5-5 fold increase in AUCR of R- and S-omeprazole when inhibited by only one drug, either clonidine or fluconazole, the co-administration of both drugs with omeprazole led to an AUCR of 8.5-9.4. As f_m is a critical component of predicting the extent of interaction, we evaluated the interaction using both in-vitro data and data from clinical pharmacogenetic study of CYP2C19 to determine omeprazole's f_m (Venkatakrishnan et al., 2007; Yasui-Furukori et al., 2004). We further review the risk of myopathy with omeprazole. There are case reports since 1992 and case series published by the World Health Organization (WHO) that the use of omeprazole may be associated with myopathy (Clark & Strandell, 2006). Recently, Sansone *et al* use Italian National Network of Pharmacovigilance Database to identify that omeprazole could be involved in reports of myopathy more frequently than any non-statin drug (Capogrosso Sansone et al., 2017). For the first time, we demonstrate both clinical impact and pharmacologic mechanism for the omeprazole, clonidine, and fluconazole interaction induced myopathy where myopathy has been report associated with omeprazole and fluconazole alone in SIDER.

Finally, we propose a random control approach based nested case-control design for EMR/EHR analyses. For a case, random controls are matched patients who have same risk factors as the case. While, compared with the dynamic/super control selection approach, random control selection relaxes the matching by the cases' index time

restriction, which is the most computational expensive step. We evaluate the performances of random control and other control selection approaches by using the OMOP gold standard and the INPC-CDM data. The area AUC values by using these disproportionality measurements. We observe the random control (AUC: 0.572 – 0.597) and all patients analysis (AUC: 0.595 – 0.619) have modest well AUCs. While, the dynamic controls (AUC: 0.492 -0.539) and super controls (AUC: 0.482 – 0.538) have less powerful detection capability. Through AUC analyses, the random control selection has a modest well detection capability. Its average AUC value is only less than the best average AUC value.

In the future, with random control selection, all other ADEs can share same controls; we are looking forward to investigating other drug-count responsive patterns for other ADEs. Also, because of the limitation of the computation expense MDCM, it computationally can handle only a limited number of drugs (i.e. 30). Thus, screening steps are required to reduce the number of drugs. More research is needed to increase the computational efficiency of MDCM.

References

- Admassie, E., Melese, T., Mequanent, W., Hailu, W., & Srikanth, B. A. (2013). Extent of poly-pharmacy, occurrence and associated factors of drug-drug interaction and potential adverse drug reactions in Gondar Teaching Referral Hospital, North West Ethiopia. *J Adv Pharm Technol Res*, 4(4), 183-189. doi:10.4103/2231-4040.121412
- Alam, K., Pahwa, S., Wang, X., Zhang, P., Ding, K., Abuznait, A. H., . . . Yue, W. (2016). Downregulation of Organic Anion Transporting Polypeptide (OATP) 1B1 Transport Function by Lysosomotropic Drug Chloroquine: Implication in OATP-Mediated Drug-Drug Interactions. *Mol Pharm*, 13(3), 839-851. doi:10.1021/acs.molpharmaceut.5b00763
- Argiris, A., Kut, V., Luong, L., & Avram, M. J. (2006). Phase I and pharmacokinetic study of docetaxel, irinotecan, and celecoxib in patients with advanced non-small cell lung cancer. *Invest New Drugs*, 24(3), 203-212. doi:10.1007/s10637-005-3259-4
- Awada, A., Hendlisz, A., Christensen, O., Lathia, C. D., Bartholomeus, S., Lebrun, F., . . . Gil, T. (2012). Phase I trial to investigate the safety, pharmacokinetics and efficacy of sorafenib combined with docetaxel in patients with advanced refractory solid tumours. *Eur J Cancer*, 48(4), 465-474. doi:10.1016/j.ejca.2011.12.026
- Bahleda, R., Sessa, C., Del Conte, G., Gianni, L., Capri, G., Varga, A., . . . Soria, J. C. (2014). Phase I clinical and pharmacokinetic study of ombrabulin (AVE8062) combined with cisplatin/docetaxel or carboplatin/paclitaxel in patients with advanced solid tumors. *Invest New Drugs*, 32(6), 1188-1196. doi:10.1007/s10637-014-0119-0
- Bate, A., Lindquist, M., Edwards, I. R., Olsson, S., Orre, R., Lansner, A., & De Freitas, R. M. (1998). A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*, 54(4), 315-321. doi:DOI 10.1007/s002280050466
- Bergh, J., Mariani, G., Cardoso, F., Liljegren, A., Awada, A., Vigano, L., . . . Gianni, L. (2012). Clinical and pharmacokinetic study of sunitinib and docetaxel in women with advanced breast cancer. *Breast*, 21(4), 507-513. doi:10.1016/j.breast.2012.01.012
- Bohnert, T., Patel, A., Templeton, I., Chen, Y., Lu, C., Lai, G., . . . Quality in Pharmaceutical Development Victim Drug-Drug Interactions Working, G. (2016). Evaluation of a New Molecular Entity as a Victim of Metabolic Drug-Drug Interactions-an Industry Perspective. *Drug Metab Dispos*, 44(8), 1399-1423. doi:10.1124/dmd.115.069096
- Brauer, R., Douglas, I., Rodriguez, L. A. G., Bate, A., Smeeth, L., Reynolds, R., . . . Ruigomez, A. (2014). The Risk of Acute Liver Injury among Users of Antibiotic Medications in the PROTECT Project: The Results of a Nested Case-Control Study Using European Outpatient Healthcare Data. *Pharmacoepidemiology and Drug Safety*, 23, 372-373.
- Capogrosso Sansone, A., Convertino, I., Galiulo, M. T., Salvadori, S., Pieroni, S., Knezevic, T., . . . Tuccori, M. (2017). Muscular Adverse Drug Reactions Associated with

- Proton Pump Inhibitors: A Disproportionality Analysis Using the Italian National Network of Pharmacovigilance Database. *Drug Saf.* doi:10.1007/s40264-017-0564-8
- Chiang, C. W., Zhang, P., Wang, X., Wang, L., Zhang, S., Ning, X., . . . Li, L. (2017). Translational high-dimensional drug Interaction discovery and validation using health record databases and pharmacokinetics models. *Clin Pharmacol Ther.* doi:10.1002/cpt.914
- Clark, D. W., & Strandell, J. (2006). Myopathy including polymyositis: a likely class adverse effect of proton pump inhibitors? *Eur J Clin Pharmacol*, 62(6), 473-479. doi:10.1007/s00228-006-0131-1
- Colburn, D. E., Giles, F. J., Oladovich, D., & Smith, J. A. (2004). In vitro evaluation of cytochrome P450-mediated drug interactions between cytarabine, idarubicin, itraconazole and caspofungin. *Hematology*, 9(3), 217-221. doi:10.1080/10245330410001701585
- Creighton, C. J., Massarweh, S., Huang, S., Tsimelzon, A., Hilsenbeck, S. G., Osborne, C. K., . . . Schiff, R. (2008). Development of resistance to targeted therapies transforms the clinically associated molecular profile subtype of breast tumor xenografts. *Cancer Res*, 68(18), 7493-7501. doi:10.1158/0008-5472.CAN-08-1404
- de Vries, E. N., Ramrattan, M. A., Smorenburg, S. M., Gouma, D. J., & Boermeester, M. A. (2008). The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care*, 17(3), 216-223. doi:10.1136/qshc.2007.023622
- Desbans, C., Hilgendorf, C., Lutz, M., Bachellier, P., Zacharias, T., Weber, J. C., . . . Ungell, A. L. (2014). Prediction of fraction metabolized via CYP3A in humans utilizing cryopreserved human hepatocytes from a set of 12 single donors. *Xenobiotica*, 44(1), 17-27. doi:10.3109/00498254.2013.809617
- Desta, Z., Ward, B. A., Soukhova, N. V., & Flockhart, D. A. (2004). Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: prominent roles for CYP3A and CYP2D6. *J Pharmacol Exp Ther*, 310(3), 1062-1075. doi:10.1124/jpet.104.065607
- 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 Comput Biol*, 8(8), e1002614. doi:10.1371/journal.pcbi.1002614
- DuMouchel, W. (1999). Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. *The American Statistician*, 53(3), 177-190. doi:10.2307/2686093
- Eichelbaum, M., Ingelman-Sundberg, M., & Evans, W. E. (2006). Pharmacogenomics and individualized drug therapy. *Annu Rev Med*, 57, 119-137. doi:10.1146/annurev.med.56.082103.104724
- Engels, F. K., de Jong, F. A., Sparreboom, A., Mathot, R. A., Loos, W. J., Kitzen, J. J., . . . Mathijssen, R. H. (2007). Medicinal cannabis does not influence the clinical pharmacokinetics of irinotecan and docetaxel. *Oncologist*, 12(3), 291-300. doi:10.1634/theoncologist.12-3-291

- Ereshefsky, L. (1996). Drug-drug interactions involving antidepressants: focus on venlafaxine. *J Clin Psychopharmacol*, 16(3 Suppl 2), 37S-50S; discussion 50S-53S.
- Esposito, M., Venturini, M., Vannozi, M. O., Tolino, G., Lunardi, G., Garrone, O., . . . Rosso, R. (1999). Comparative effects of paclitaxel and docetaxel on the metabolism and pharmacokinetics of epirubicin in breast cancer patients. *J Clin Oncol*, 17(4), 1132. doi:10.1200/JCO.1999.17.4.1132
- Evans, S. J., Waller, P. C., & Davis, S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf*, 10(6), 483-486. doi:10.1002/pds.677
- Fahmi, O. A., Maurer, T. S., Kish, M., Cardenas, E., Boldt, S., & Nettleton, D. (2008). A combined model for predicting CYP3A4 clinical net drug-drug interaction based on CYP3A4 inhibition, inactivation, and induction determined in vitro. *Drug Metab Dispos*, 36(8), 1698-1708. doi:10.1124/dmd.107.018663
- Fitzgerald, R. J. (2009). Medication errors: the importance of an accurate drug history. *Br J Clin Pharmacol*, 67(6), 671-675. doi:10.1111/j.1365-2125.2009.03424.x
- Food, U., & Administration, D. (2012). Drug Interaction Studies—Study Design, Data Analysis, Implications for Dosing, and Labeling Recommendations (Draft Guidance) 2012. *Maryland, USA: US Food and Drug Administration*.
- Gascon, M. P., & Dayer, P. (1991). In vitro forecasting of drugs which may interfere with the biotransformation of midazolam. *Eur J Clin Pharmacol*, 41(6), 573-578. doi:10.1007/BF00314987
- Goetz, M. P., Rae, J. M., Suman, V. J., Safgren, S. L., Ames, M. M., Visscher, D. W., . . . Ingle, J. N. (2005). Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J Clin Oncol*, 23(36), 9312-9318. doi:10.1200/JCO.2005.03.3266
- Goodman, L. S., Gilman, A., Brunton, L. L., Lazo, J. S., & Parker, K. L. (2006). *Goodman & Gilman's the pharmacological basis of therapeutics* (11th ed.). New York: McGraw-Hill.
- Gorski, J. C., Hall, S. D., Jones, D. R., VandenBranden, M., & Wrighton, S. A. (1994). Regioselective biotransformation of midazolam by members of the human cytochrome P450 3A (CYP3A) subfamily. *Biochem Pharmacol*, 47(9), 1643-1653.
- Gu, Q., Dillon, C. F., & Burt, V. L. (2010). Prescription drug use continues to increase: U.S. prescription drug data for 2007-2008. *NCHS Data Brief*(42), 1-8.
- Hachad, H., Ragueneau-Majlessi, I., & Levy, R. H. (2010). A useful tool for drug interaction evaluation: the University of Washington Metabolism and Transport Drug Interaction Database. *Hum Genomics*, 5(1), 61-72.
- Hajjar, E. R., Cafiero, A. C., & Hanlon, J. T. (2007). Polypharmacy in elderly patients. *Am J Geriatr Pharmacother*, 5(4), 345-351. doi:10.1016/j.amjopharm.2007.12.002
- Hall, M. J., DeFrances, C. J., Williams, S. N., Golosinskiy, A., & Schwartzman, A. (2010). National Hospital Discharge Survey: 2007 summary. *Natl Health Stat Report*(29), 1-20, 24.
- Han, X., Quinney, S. K., Wang, Z., Zhang, P., Duke, J., Desta, Z., . . . Li, L. (2015). Identification and Mechanistic Investigation of Drug-Drug Interactions Associated

- With Myopathy: A Translational Approach. *Clin Pharmacol Ther*, 98(3), 321-327. doi:10.1002/cpt.150
- Hanlon, J. T., Weinberger, M., Samsa, G. P., Schmader, K. E., Uttech, K. M., Lewis, I. K., . . . Feussner, J. R. (1996). A randomized, controlled trial of a clinical pharmacist intervention to improve inappropriate prescribing in elderly outpatients with polypharmacy. *Am J Med*, 100(4), 428-437. doi:10.1016/S0002-9343(97)89519-8
- Harpaz, R., DuMouchel, W., LePendu, P., Bauer-Mehren, A., Ryan, P., & Shah, N. H. (2013). Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. *Clin Pharmacol Ther*, 93(6), 539-546. doi:10.1038/clpt.2013.24
- Harpaz, R., DuMouchel, W., Shah, N. H., Madigan, D., Ryan, P., & Friedman, C. (2012). Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther*, 91(6), 1010-1021. doi:10.1038/clpt.2012.50
- Harris, R. Z., Salfi, M., Sullivan, J. T., & Padhi, D. (2007). Pharmacokinetics of cinacalcet hydrochloride when administered with ketoconazole. *Clin Pharmacokinet*, 46(6), 495-501. doi:10.2165/00003088-200746060-00003
- Hedaya, M. A., & Helmy, S. A. (2015). Modeling of the pharmacokinetic/pharmacodynamic interaction between irbesartan and hydrochlorothiazide in normotensive subjects. *Biopharm Drug Dispos*, 36(4), 216-231. doi:10.1002/bdd.1935
- Hennessy, S., & Flockhart, D. A. (2012). The need for translational research on drug-drug interactions. *Clin Pharmacol Ther*, 91(5), 771-773. doi:10.1038/clpt.2012.39
- Hennessy, S., Leonard, C. E., Gagne, J. J., Flory, J. H., Han, X., Brensinger, C. M., & Bilker, W. B. (2016). Pharmacoepidemiologic Methods for Studying the Health Effects of Drug-Drug Interactions. *Clin Pharmacol Ther*, 99(1), 92-100. doi:10.1002/cpt.277
- Henry, N. L., Stearns, V., Flockhart, D. A., Hayes, D. F., & Riba, M. (2008). Drug interactions and pharmacogenomics in the treatment of breast cancer and depression. *Am J Psychiatry*, 165(10), 1251-1255. doi:10.1176/appi.ajp.2008.08040482
- Huskey, S. W., Dean, D. C., Miller, R. R., Rasmusson, G. H., & Chiu, S. H. (1995). Identification of human cytochrome P450 isozymes responsible for the in vitro oxidative metabolism of finasteride. *Drug Metab Dispos*, 23(10), 1126-1135.
- Ito, K., Brown, H. S., & Houston, J. B. (2004). Database analyses for the prediction of in vivo drug-drug interactions from in vitro data. *Br J Clin Pharmacol*, 57(4), 473-486. doi:10.1111/j.1365-2125.2003.02041.x
- Ito, K., Hallifax, D., Obach, R. S., & Houston, J. B. (2005). Impact of parallel pathways of drug elimination and multiple cytochrome P450 involvement on drug-drug interactions: CYP2D6 paradigm. *Drug Metab Dispos*, 33(6), 837-844. doi:10.1124/dmd.104.003715
- Izquierdo, M. A., Garcia, M., Ponton, J. L., Martinez, M., Valenti, V., Navarro, M., . . . Germa-Lluch, J. R. (2006). A phase I clinical and pharmacokinetic study of paclitaxel and docetaxel given in combination in patients with solid tumours. *Eur J Cancer*, 42(12), 1789-1796. doi:10.1016/j.ejca.2005.10.031

- Jeon, D. H., Kim, Y., Kim, M. J., Cho, H. S., Bae, E. J., Chang, S. H., & Park, D. J. (2016). Rhabdomyolysis associated with single-dose intravenous esomeprazole administration: A case report. *Medicine (Baltimore)*, *95*(29), e4313. doi:10.1097/MD.0000000000004313
- Kaneta, T., Fujita, K., Akiyama, Y., Kawara, K., Sunakawa, Y., Kawachi, A., . . . Sasaki, Y. (2014). No pharmacokinetic alteration of docetaxel following coadministration of aprepitant 3 h before docetaxel infusion. *Cancer Chemother Pharmacol*, *74*(3), 539-547. doi:10.1007/s00280-014-2528-3
- La Gamba, F., Corrao, G., Romio, S., Sturkenboom, M., Trifiro, G., Schink, T., & de Ridder, M. (2017). Combining evidence from multiple electronic health care databases: performances of one-stage and two-stage meta-analysis in matched case-control studies. *Pharmacoepidemiol Drug Saf*, *26*(10), 1213-1219. doi:10.1002/pds.4280
- Lam, Y. W., Alfaro, C. L., Ereshefsky, L., & Miller, M. (2003). Pharmacokinetic and pharmacodynamic interactions of oral midazolam with ketoconazole, fluoxetine, fluvoxamine, and nefazodone. *J Clin Pharmacol*, *43*(11), 1274-1282. doi:10.1177/0091270003259216
- Lazarou, J., Pomeranz, B. H., & Corey, P. N. (1998). Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA*, *279*(15), 1200-1205.
- Le Coutre, P., Ottmann, O. G., Giles, F., Kim, D.-W., Cortes, J., Gattermann, N., . . . O'Brien, S. G. (2008). Nilotinib (formerly AMN107), a highly selective BCR-ABL tyrosine kinase inhibitor, is active in patients with imatinib-resistant or-intolerant accelerated-phase chronic myelogenous leukemia. *Blood*, *111*(4), 1834-1839.
- Lee, A. S., Macedo-Vinas, M., Francois, P., Renzi, G., Schrenzel, J., Vernaz, N., . . . Harbarth, S. (2011). Impact of combined low-level mupirocin and genotypic chlorhexidine resistance on persistent methicillin-resistant *Staphylococcus aureus* carriage after decolonization therapy: a case-control study. *Clinical Infectious Diseases*, *52*(12), 1422-1430. doi:10.1093/cid/cir233
- Li, Y., Zhou, J., Ramsden, D., Taub, M. E., O'Brien, D., Xu, J., . . . Tweedie, D. J. (2014). Enzyme-transporter interplay in the formation and clearance of abundant metabolites of faldaprevir found in excreta but not in circulation. *Drug Metab Dispos*, *42*(3), 384-393. doi:10.1124/dmd.113.055863
- Li, Z. M., Guo, L. H., & Ren, X. M. (2016). Biotransformation of 8:2 fluorotelomer alcohol by recombinant human cytochrome P450s, human liver microsomes and human liver cytosol. *Environ Sci Process Impacts*, *18*(5), 538-546. doi:10.1039/c6em00071a
- Liu, M., McPeck Hinz, E. R., Matheny, M. E., Denny, J. C., Schildcrout, J. S., Miller, R. A., & Xu, H. (2013). Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records. *J Am Med Inform Assoc*, *20*(3), 420-426. doi:10.1136/amiajnl-2012-001119
- Lorberbaum, T., Sampson, K. J., Chang, J. B., Iyer, V., Woosley, R. L., Kass, R. S., & Tatonetti, N. P. (2016). Coupling Data Mining and Laboratory Experiments to Discover Drug Interactions Causing QT Prolongation. *J Am Coll Cardiol*, *68*(16), 1756-1764. doi:10.1016/j.jacc.2016.07.761

- LoRusso, P. M., Jones, S. F., Koch, K. M., Arya, N., Fleming, R. A., Loftiss, J., . . . Burris, H. A., 3rd. (2008). Phase I and pharmacokinetic study of lapatinib and docetaxel in patients with advanced cancer. *J Clin Oncol*, *26*(18), 3051-3056. doi:10.1200/JCO.2007.14.9633
- Lu, C., Miwa, G. T., Prakash, S. R., Gan, L. S., & Balani, S. K. (2007). A novel model for the prediction of drug-drug interactions in humans based on in vitro cytochrome p450 phenotypic data. *Drug Metab Dispos*, *35*(1), 79-85. doi:10.1124/dmd.106.011346
- Maekawa, K., Harakawa, N., Yoshimura, T., Kim, S. R., Fujimura, Y., Aohara, F., . . . Saito, Y. (2010). CYP3A4*16 and CYP3A4*18 alleles found in East Asians exhibit differential catalytic activities for seven CYP3A4 substrate drugs. *Drug Metab Dispos*, *38*(12), 2100-2104. doi:10.1124/dmd.110.034140
- Mao, J., Mohutsky, M. A., Harrelson, J. P., Wrighton, S. A., & Hall, S. D. (2011). Prediction of CYP3A-mediated drug-drug interactions using human hepatocytes suspended in human plasma. *Drug Metab Dispos*, *39*(4), 591-602. doi:10.1124/dmd.110.036400
- Mardjuadi, F., Medioni, J., Kerger, J., D'Hondt, L., Canon, J. L., Duck, L., . . . Machiels, J. P. (2012). Phase I study of sorafenib in combination with docetaxel and prednisone in chemo-naive patients with metastatic castration-resistant prostate cancer. *Cancer Chemother Pharmacol*, *70*(2), 293-303. doi:10.1007/s00280-012-1914-y
- Mayhew, B. S., Jones, D. R., & Hall, S. D. (2000). An in vitro model for predicting in vivo inhibition of cytochrome P450 3A4 by metabolic intermediate complex formation. *Drug Metab Dispos*, *28*(9), 1031-1037.
- McDonald, C. J., Overhage, J. M., Barnes, M., Schadow, G., Blevins, L., Dexter, P. R., . . . Committee, I. M. (2005). The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. *Health Aff (Millwood)*, *24*(5), 1214-1220. doi:10.1377/hlthaff.24.5.1214
- McGinnity, D. F., Parker, A. J., Soars, M., & Riley, R. J. (2000). Automated definition of the enzymology of drug oxidation by the major human drug metabolizing cytochrome P450s. *Drug Metab Dispos*, *28*(11), 1327-1334.
- Messersmith, W. A., Baker, S. D., Lassiter, L., Sullivan, R. A., Dinh, K., Almuete, V. I., . . . Armstrong, D. K. (2006). Phase I trial of bortezomib in combination with docetaxel in patients with advanced solid tumors. *Clin Cancer Res*, *12*(4), 1270-1275. doi:10.1158/1078-0432.CCR-05-1942
- Miyata, M., Yasuda, K., Burioka, N., Takane, H., Suyama, H., Shigeoka, Y., . . . Shimizu, E. (2006). The influence of granisetron on the pharmacokinetics and pharmacodynamics of docetaxel in Asian lung cancer patients. *Cancer J*, *12*(1), 69-72.
- Moody, G. C., Griffin, S. J., Mather, A. N., McGinnity, D. F., & Riley, R. J. (1999). Fully automated analysis of activities catalysed by the major human liver cytochrome P450 (CYP) enzymes: assessment of human CYP inhibition potential. *Xenobiotica*, *29*(1), 53-75. doi:10.1080/004982599238812

- Niemi, M., Backman, J. T., Neuvonen, M., & Neuvonen, P. J. (2003). Effects of gemfibrozil, itraconazole, and their combination on the pharmacokinetics and pharmacodynamics of repaglinide: potentially hazardous interaction between gemfibrozil and repaglinide. *Diabetologia*, *46*(3), 347-351. doi:10.1007/s00125-003-1034-7
- Niwa, T., Shiraga, T., & Takagi, A. (2005). Effect of antifungal drugs on cytochrome P450 (CYP) 2C9, CYP2C19, and CYP3A4 activities in human liver microsomes. *Biol Pharm Bull*, *28*(9), 1805-1808.
- Nygren, P., Hande, K., Petty, K. J., Fedgchin, M., van Dyck, K., Majumdar, A., . . . van Belle, S. (2005). Lack of effect of aprepitant on the pharmacokinetics of docetaxel in cancer patients. *Cancer Chemother Pharmacol*, *55*(6), 609-616. doi:10.1007/s00280-004-0946-3
- Olkola, K. T., Backman, J. T., & Neuvonen, P. J. (1994). Midazolam should be avoided in patients receiving the systemic antimycotics ketoconazole or itraconazole. *Clin Pharmacol Ther*, *55*(5), 481-485.
- Pearce, R. E., Vakkalagadda, G. R., & Leeder, J. S. (2002). Pathways of carbamazepine bioactivation in vitro I. Characterization of human cytochromes P450 responsible for the formation of 2- and 3-hydroxylated metabolites. *Drug Metab Dispos*, *30*(11), 1170-1179.
- 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 Res*, *38*(Database issue), D237-243. doi:10.1093/nar/gkp970
- Renwick, A. G. (1999). The metabolism of antihistamines and drug interactions: the role of cytochrome P450 enzymes. *Clin Exp Allergy*, *29 Suppl 3*, 116-124.
- Rodrigues, A. D., Winchell, G. A., & Dobrinska, M. R. (2001). Use of In Vitro Drug Metabolism Data to Evaluate Metabolic Drug-Drug Interactions in Man: The Need for Quantitative Databases. *The Journal of Clinical Pharmacology*, *41*(4), 368-373.
- Roman, M., Ochoa, D., Sanchez-Rojas, S. D., Talegon, M., Prieto-Perez, R., Rivas, A., . . . Cabaleiro, T. (2014). Evaluation of the relationship between polymorphisms in CYP2C19 and the pharmacokinetics of omeprazole, pantoprazole and rabeprazole. *Pharmacogenomics*, *15*(15), 1893-1901. doi:10.2217/pgs.14.141
- 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. *Dan Med Bull*, *45*(2), 210-213.
- Rostami-Hodjegan, A., & Tucker, G. T. (2007). Simulation and prediction of in vivo drug metabolism in human populations from in vitro data. *Nat Rev Drug Discov*, *6*(2), 140-148. doi:10.1038/nrd2173
- Rowland, M., & Tozer, T. N. (2005). *Clinical pharmacokinetics/pharmacodynamics*: Lippincott Williams and Wilkins Philadelphia.
- Royer, I., Monsarrat, B., Sonnier, M., Wright, M., & Cresteil, T. (1996). Metabolism of docetaxel by human cytochromes P450: interactions with paclitaxel and other antineoplastic drugs. *Cancer Res*, *56*(1), 58-65.

- Ryan, P. B., Madigan, D., Stang, P. E., Overhage, J. M., Racoosin, J. A., & Hartzema, A. G. (2012). Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. *Stat Med*, *31*(30), 4401-4415. doi:10.1002/sim.5620
- Sasaki, T., Fujita, K., Sunakawa, Y., Ishida, H., Yamashita, K., Miwa, K., . . . Sasaki, Y. (2013). Concomitant polypharmacy is associated with irinotecan-related adverse drug reactions in patients with cancer. *Int J Clin Oncol*, *18*(4), 735-742. doi:10.1007/s10147-012-0425-5
- Schelleman, H., Han, X., Brensinger, C. M., Quinney, S. K., Bilker, W. B., Flockhart, D. A., . . . Hennessy, S. (2014). Pharmacoepidemiologic and in vitro evaluation of potential drug-drug interactions of sulfonylureas with fibrates and statins. *Br J Clin Pharmacol*, *78*(3), 639-648. doi:10.1111/bcp.12353
- Schneeweiss, S. (2010). A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf*, *19*(8), 858-868. doi:10.1002/pds.1926
- Schonhofer, P. S., Werner, B., & Troger, U. (1997). Ocular damage associated with proton pump inhibitors. *BMJ*, *314*(7097), 1805.
- Shah, R. R., & Smith, R. L. (2015). Inflammation-induced phenoconversion of polymorphic drug metabolizing enzymes: hypothesis with implications for personalized medicine. *Drug Metab Dispos*, *43*(3), 400-410. doi:10.1124/dmd.114.061093
- Shen, D. D., Kunze, K. L., & Thummel, K. E. (1997). Enzyme-catalyzed processes of first-pass hepatic and intestinal drug extraction. *Adv Drug Deliv Rev*, *27*(2-3), 99-127.
- Shoaf, S. E., Bricmont, P., & Mallikaarjun, S. (2012). Effects of CYP3A4 inhibition and induction on the pharmacokinetics and pharmacodynamics of tolvaptan, a non-peptide AVP antagonist in healthy subjects. *Br J Clin Pharmacol*, *73*(4), 579-587. doi:10.1111/j.1365-2125.2011.04114.x
- Silas, J. H., McGourty, J. C., Lennard, M. S., Tucker, G. T., & Woods, H. F. (1985). Polymorphic metabolism of metoprolol: clinical studies. *Eur J Clin Pharmacol*, *28 Suppl*(1), 85-88. doi:10.1007/bf00543716
- Sivakumar, K., & Dalakas, M. C. (1994). Autoimmune syndrome induced by omeprazole. *Lancet*, *344*(8922), 619-620.
- Stearns, V., Johnson, M. D., Rae, J. M., Morocho, A., Novielli, A., Bhargava, P., . . . Flockhart, D. A. (2003). Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. *J Natl Cancer Inst*, *95*(23), 1758-1764.
- Stephene, X., Najimi, M., & Sokal, E. M. (2010). Hepatocyte cryopreservation: is it time to change the strategy? *World J Gastroenterol*, *16*(1), 1-14.
- Tanaka, E., Nakamura, T., Inomata, S., & Honda, K. (2006). Effects of premedication medicines on the formation of the CYP3A4-dependent metabolite of ropivacaine, 2', 6'-Pipicoloxylidide, on human liver microsomes in vitro. *Basic Clin Pharmacol Toxicol*, *98*(2), 181-183. doi:10.1111/j.1742-7843.2006.pto_265.x

- Tatonetti, N. P., Ye, P. P., Daneshjou, R., & Altman, R. B. (2012). Data-driven prediction of drug effects and interactions. *Sci Transl Med*, 4(125), 125ra131. doi:10.1126/scitranslmed.3003377
- U.S. Department of Health and Human Services, F. a. D. A., Center for Drug Evaluation and Research (CDER). . Guidance for industry, drug interaction studies—study design, data analysis, implications for dosing, and labeling recommendations.
- van Puijenbroek, E. P., Bate, A., Leufkens, H. G., Lindquist, M., Orre, R., & Egberts, A. C. (2002). A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. *Pharmacoepidemiol Drug Saf*, 11(1), 3-10. doi:10.1002/pds.668
- Venkatakrisnan, K., Obach, R. S., & Rostami-Hodjegan, A. (2007). Mechanism-based inactivation of human cytochrome P450 enzymes: strategies for diagnosis and drug-drug interaction risk assessment. *Xenobiotica*, 37(10-11), 1225-1256. doi:10.1080/00498250701670945
- Visruthan, N. K., Boo, P. K., Kader, A., Ping, C. H., & Ong, C. (2012). Omeprazole-induced myositis in a child receiving triple therapy for Helicobacter pylori infection. *J Pediatr Gastroenterol Nutr*, 55(3), 338-339. doi:10.1097/MPG.0b013e318255117b
- Vose, C., & Ings, R. (2014). *Drug Metabolism*: Royal Society of Chemistry: London.
- Watanabe, Y., Nakajima, H., Nozaki, K., Hoshiai, H., & Noda, K. (2003). The effect of granisetron on in vitro metabolism of paclitaxel and docetaxel. *Cancer J*, 9(1), 67-70.
- Wienkers, L. C., Wurden, C. J., Storch, E., Kunze, K. L., Rettie, A. E., & Trager, W. F. (1996). Formation of (R)-8-hydroxywarfarin in human liver microsomes. A new metabolic marker for the (S)-mephenytoin hydroxylase, P4502C19. *Drug Metab Dispos*, 24(5), 610-614.
- 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 (AUC_i/AUC) ratios. *Drug Metab Dispos*, 32(11), 1201-1208. doi:10.1124/dmd.104.000794
- Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., . . . Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*, 34(Database issue), D668-672. doi:10.1093/nar/gkj067
- Yasui-Furukori, N., Takahata, T., Nakagami, T., Yoshiya, G., Inoue, Y., Kaneko, S., & Tateishi, T. (2004). Different inhibitory effect of fluvoxamine on omeprazole metabolism between CYP2C19 genotypes. *Br J Clin Pharmacol*, 57(4), 487-494. doi:10.1111/j.1365-2125.2003.02047.x
- Yeung, C. K., Yoshida, K., Kusama, M., Zhang, H., Ragueneau-Majlessi, I., Argon, S., . . . Huang, S. M. (2015). Organ Impairment-Drug-Drug Interaction Database: A Tool for Evaluating the Impact of Renal or Hepatic Impairment and Pharmacologic Inhibition on the Systemic Exposure of Drugs. *CPT Pharmacometrics Syst Pharmacol*, 4(8), 489-494. doi:10.1002/psp4.55

- Yu, J., Ritchie, T. K., Mulgaonkar, A., & Ragueneau-Majlessi, I. (2014). Drug disposition and drug-drug interaction data in 2013 FDA new drug applications: a systematic review. *Drug Metab Dispos*, 42(12), 1991-2001. doi:10.1124/dmd.114.060392
- Zhang, D., Zhu, M., & Humphreys, W. G. (2007). *Drug metabolism in drug design and development*: John Wiley & Sons.
- Zhang, L., Zhang, Y. D., Zhao, P., & Huang, S. M. (2009). Predicting drug-drug interactions: an FDA perspective. *AAPS J*, 11(2), 300-306. doi:10.1208/s12248-009-9106-3
- Zhang, P., Du, L., Wang, L., Liu, M., Cheng, L., Chiang, C. W., . . . Li, L. (2015). A Mixture Dose-Response Model for Identifying High-Dimensional Drug Interaction Effects on Myopathy Using Electronic Medical Record Databases. *CPT Pharmacometrics Syst Pharmacol*, 4(8), 474-480. doi:10.1002/psp4.53
- Zhang, X., Jones, D. R., & Hall, S. D. (2009). Prediction of the effect of erythromycin, diltiazem, and their metabolites, alone and in combination, on CYP3A4 inhibition. *Drug Metab Dispos*, 37(1), 150-160. doi:10.1124/dmd.108.022178
- Zhang, X., Quinney, S. K., Gorski, J. C., Jones, D. R., & Hall, S. D. (2009). Semiphysiologically based pharmacokinetic models for the inhibition of midazolam clearance by diltiazem and its major metabolite. *Drug Metab Dispos*, 37(8), 1587-1597. doi:10.1124/dmd.109.026658

Curriculum Vitae

Chien-Wei Chiang

Research Interests:

- Clinical Pharmacokinetics and Pharmacodynamics for Drug-Drug Interaction
- Network Modeling for Genetics Regulatory Network
- Genome-wide analysis and Next generation sequence

Education:

- Ph.D. in Bioinformatics (Minor: Pharmacometrics)
Indiana School of Informatics and Computing, IUPUI, Indianapolis, IN
- M.S. in Mathematics
Purdue School of Science, IUPUI, Indianapolis, IN
- B.S. in Biomedical Engineering
Purdue School of Engineering & Technology, IUPUI, Indianapolis, IN

Research and Training Experience:

- Eli Lilly pharmacometrics Summer internship (2015)
- Literature based drug interaction prediction with clinical assessment using electronic medical records: novel myopathy associated drug interaction.
 - CYP450 recombinant enzyme inhibitor screening
 - Task start with building a drug dictionary covers all the brand/generic drug names. Based on the drug dictionary, drugs inhibition constant are mined from PubMed literatures.

- Conducts CYP450 2D6 and CYP3A5 enzyme screening assay, and test their inhibitors
- Workshop on monoclonal antibody – **Pharmacokinetics and pharmacodynamics** – by William J. Jusko, PhD, University at Buffalo (2012)
- Hands-on Experience with Automated Model-based drug development: Incorporating population variability into mechanistic prediction of PK and modelling PK-PD – by simCYP (2010)

Computer skills:

- SimCYP and NONMEM analysis
- Perform genome-wide association study
- Complete statistics analysis by R-programing
- Building databases with SQL-like language
- Protein docking with autoduck-tool
- Using advance programing language (C, Matlab) to complete tasks

Laboratory skills:

- Cell culture (human epithelial cell)
- DNA extraction, PCR, and western blot
- Antibody testing, and gel electrophoresis
- Gas chromatography, infrared spectroscopy, and NMR

- CYP450 enzyme recombinant system based substrate, inhibitor, and inducer screening

Publications:

Chiang, C. W., Zhang, P., Wang, X., Wang, L., Zhang, S., Ning, X., . . . Li, L. (2017).

Translational high-dimensional drug Interaction discovery and validation using health record databases and pharmacokinetics models. *Clin Pharmacol Ther.* doi:10.1002/cpt.914

Du, L., Chakraborty, A., Chiang, C. W., Cheng, L., Quinney, S. K., Wu, H., . . . Shen, L.

(2015). Graphic Mining of High-Order Drug Interactions and Their Directional Effects on Myopathy Using Electronic Medical Records. *CPT Pharmacometrics Syst Pharmacol*, 4(8), 481-488. doi:10.1002/psp4.59

Han, X., Chiang, C., Leonard, C. E., Bilker, W. B., Brensinger, C. M., Li, L., & Hennessy, S.

(2017). Biomedical Informatics Approaches to Identifying Drug-Drug Interactions: Application to Insulin Secretagogues. *Epidemiology*, 28(3), 459-468. doi:10.1097/EDE.0000000000000638

Kreutz, R. P., Nystrom, P., Kreutz, Y., Miao, J., Desta, Z., Breall, J. A., . . . Jin, Y. (2012).

Influence of paraoxonase-1 Q192R and cytochrome P450 2C19 polymorphisms on clopidogrel response. *Clin Pharmacol*, 4, 13-20. doi:10.2147/CPAA.S27822

Kreutz, R. P., Owens, J., Jin, Y., Nystrom, P., Desta, Z., Kreutz, Y., . . . Flockhart, D. A.

(2013). Cytochrome P450 3A4*22, PPAR-alpha, and ARNT polymorphisms and clopidogrel response. *Clin Pharmacol*, 5, 185-192. doi:10.2147/CPAA.S53151

- Lavoie Smith, E. M., Li, L., Chiang, C., Thomas, K., Hutchinson, R. J., Wells, E. M., . . . Renbarger, J. (2015). Patterns and severity of vincristine-induced peripheral neuropathy in children with acute lymphoblastic leukemia. *J Peripher Nerv Syst*, 20(1), 37-46. doi:10.1111/jns.12114
- Wang, L., Chiang, C., Liang, H., Wu, H., Feng, W., Quinney, S. K., . . . Li, L. (2015). How to Choose In Vitro Systems to Predict In Vivo Drug Clearance: A System Pharmacology Perspective. *Biomed Res Int*, 2015, 857327. doi:10.1155/2015/857327
- Wu, H. Y., Chiang, C. W., & Li, L. (2014). Text mining for drug-drug interaction. *Methods Mol Biol*, 1159, 47-75. doi:10.1007/978-1-4939-0709-0_4
- Wu, H. Y., Karnik, S., Subhadarshini, A., Wang, Z., Philips, S., Han, X., . . . Li, L. (2013). An integrated pharmacokinetics ontology and corpus for text mining. *BMC Bioinformatics*, 14, 35. doi:10.1186/1471-2105-14-35
- Zhang, P., Du, L., Wang, L., Liu, M., Cheng, L., Chiang, C. W., . . . Li, L. (2015). A Mixture Dose-Response Model for Identifying High-Dimensional Drug Interaction Effects on Myopathy Using Electronic Medical Record Databases. *CPT Pharmacometrics Syst Pharmacol*, 4(8), 474-480. doi:10.1002/psp4.53