

MODELING AND SIMULATION APPLICATIONS WITH POTENTIAL  
IMPACT IN DRUG DEVELOPMENT AND PATIENT CARE

Claire Li

Submitted to the faculty of the University Graduate School  
in partial fulfillment of the requirements  
for the degree  
Doctor of Philosophy  
in the Department of Medical and Molecular Genetics,  
Indiana University

August 2014

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

---

Robert R. Bies, Pharm.D., Ph.D.

---

Lang Li, Ph.D.

Doctoral Committee

---

Tatiana M. Foroud, Ph.D.

June 4, 2014

---

Jamie L. Renbarger, MD

## DEDICATION

This work is dedicated to my thesis mentor, Dr. Robert R. Bies and my friends and family, especially my husband, David Yan, my upcoming baby, Austin, my parents, Michael and Maggie Li, my sister, Elsy, my grandparents, Ren-Jen and Shi-Hou Li, my parents-in-law, Zhaowan Yan and Jing Zhang, and my uncle and aunt, Li-Kang Wang and Ivy Li.

## ACKNOWLEDGEMENTS

I would not have completed my Ph.D. dissertation without the tremendous support and guidance of many people who have influenced my life and shaped my experience both personally and professionally.

First and the most, I would like to express my appreciation to my family. To my husband, parents and parents-in-law who have always supported my decisions when I chose my career interests, encouraged me each step of the way and shared my ups and downs. I also want to give thanks to my upcoming baby who I have been carrying through the last few months of my Ph.D. Even though it is a hard and challenging task, I really appreciate his cooperation, so I could fully dedicate myself to accomplish my work.

I especially would like to express my sincere gratitude to my mentor, Dr. Robert Bies. Dr. Bies granted me the opportunity to grow into a mature professional and introduced me into the area of pharmacometrics and clinical research. He provided tremendous support in project developments, facilitated my projects with outside collaborations, gave me freedom to develop as an independent researcher, simulated my scientific thinking and supervised my scientific writing and presentation skills. I would also like to give thanks to Dr. Bies for his generous support for both internal and external pharmacometric trainings including Uppsala summer school, Buffalo PKPD courses etc. as well as his encouragements of my visibility in many scientific meetings. Through his support and recommendation, I had a great internship experience in Genentech during the summer 2013 and got to explore many job opportunities even before I graduate. It is honor to have Dr. Bies as my Ph.D. mentor, who not only helped me to grow as an independent scientist, but also taught me to be a family person.

I would like to give thanks to the members of my thesis committee, Dr. Lang Li, Dr. Tatiana Foroud and Dr. Jamie Renbarger for their guidance, suggestions, and constructive critiques. Especially, I would like to express my great appreciation to Dr. Lang Li who not only served as my committee member, but also played an important role in supervising few of my projects. His intellectual support and guidance in many ways including genetics, bioinformatics and statistics helped me to grow as a modeler with unique and diverse skills.

Besides my committee, I would like to give thanks to my colleagues of Indiana CTSI disease and therapeutic modeling program, the Department of Medical and Molecular Genetic, the Division of clinical pharmacology and the Department of Pediatrics for all their guidance and help. Especially, I would like to express my appreciation to Dr. Nieves Velez de Mendizabal for her support and friendship. She not only provided great scientific input and trained my skills during my Ph.D., but also encouraged me when I struggled. I would also like to give thanks to all the former and fellows including Dr. Eric Sherer, Dr. Jasper Stevens, Dr. Abhishek Gulati and Dr. Ayman Akil.

Last but not least, I would like to acknowledge all my collaborators who were willing to provide data, comments and inputs and support for project success and scientific discovery. I would like to acknowledge Dr. Michael Maitland and Sanja Karovic from the University of Chicago and CALGB (Alliance) group for their scientific contributions in “Data quality constrains utility of computational modeling of tumor burden in non-small cell lung cancer clinical trials” project. I would like to acknowledge Dr. Robert Stratford from Xavier University of Louisiana for his scientific contributions

in “Preclinical model-prediction of brain clozapine and norclozapine concentrations in humans using a scaled pharmacokinetic model for rat brain and plasma pharmacokinetic” project. I would like to acknowledge Dr. Constantine Lyketosos from John Hopkins University, Dr. Bruce Pollock from Centre for Addiction and Mental Health University of Toronto as well as DIADS-2 group for their scientific contributions in “Population pharmacokinetic modeling of sertraline treatment in Alzheimer’s disease patients: the DIADS-2 study” project. I would like to acknowledge Dr. Eric Sherer from Louisiana Tech University and Dr. Lionel Lewis from Dartmouth & Dartmouth-Hitchcock Medical center for their scientific contributions in “Clinical trial simulation to evaluate study design: capturing abiraterone and nilotinib exposures” project. I would also like to acknowledge Sriharsha Guntupalli and Pengyue Zhang for their scientific contributions in “Genetic determinants of vincristine neuropathy in pediatric acute lymphoblastic leukemia patients” project and “Genetic signature to predict vincristine neuropathy and relapse in children with acute lymphoblastic leukemia” project. My dissertation work would not be possible without all of their contributions.

Claire Li

Modeling and simulation applications with potential impact in drug development and patient care

Model-based drug development has become an essential element to potentially make drug development more productive by assessing the data using mathematical and statistical approaches to construct and utilize models to increase the understanding of the drug and disease. The modeling and simulation approach not only quantifies the exposure-response relationship, and the level of variability, but also identifies the potential contributors to the variability. I hypothesized that the modeling and simulation approach can: 1) leverage our understanding of pharmacokinetic-pharmacodynamic (PK-PD) relationship from pre-clinical system to human; 2) quantitatively capture the drug impact on patients; 3) evaluate clinical trial designs; and 4) identify potential contributors to drug toxicity and efficacy. The major findings for these studies included: 1) a translational PK modeling approach that predicted clozapine and norclozapine central nervous system exposures in humans relating these exposures to receptor binding kinetics at multiple receptors; 2) a population pharmacokinetic analysis of a study of sertraline in depressed elderly patients with Alzheimer's disease that identified site specific differences in drug exposure contributing to the overall variability in sertraline exposure; 3) the utility of a longitudinal tumor dynamic model developed by the Food and Drug Administration for predicting survival in non-small cell lung cancer patients, including an exploration of the limitations of this approach; 4) a Monte Carlo clinical trial simulation approach that was used to evaluate a pre-defined oncology trial with a sparse drug concentration sampling schedule with the aim to quantify how well individual drug

exposures, random variability, and the food effects of abiraterone and nilotinib were determined under these conditions; 5) a time to event analysis that facilitated the identification of candidate genes including polymorphisms associated with vincristine-induced neuropathy from several association analyses in childhood acute lymphoblastic leukemia (ALL) patients; and 6) a LASSO penalized regression model that predicted vincristine-induced neuropathy and relapse in ALL patients and provided the basis for a risk assessment of the population. Overall, results from this dissertation provide an improved understanding of treatment effect in patients with an assessment of PK/PD combined and with a risk evaluation of drug toxicity and efficacy.

Robert R Bies, Pharm.D., Ph.D.



## TABLE OF CONTENTS

LIST OF TABLES .....	xii
LIST OF FIGURES .....	xiv
LIST OF ABBREVIATIONS.....	xvii
Chapter I. <i>Introduction</i> .....	1
1. The need of modeling and simulation in drug development.....	1
2. Model based drug development (pharmacometrics and statistics) .....	2
3. The opportunities (applications) of modeling and simulation in drug development .....	4
4. Type of models .....	6
4.1. Population modeling .....	6
4.2. Pharmacokinetic modeling.....	7
4.3. Pharmacodynamic modeling.....	8
4.4. Biomarker modeling .....	9
4.5. Clinical trial simulation.....	11
4.6. Time to event modeling (GWAS study) .....	12
5. Hypothesis and specific aims.....	13
Chapter II. <i>Preclinical model-prediction of brain clozapine and norclozapine concentrations in humans using a scaled pharmacokinetic model for rat brain and plasma pharmacokinetic</i> .....	17
1. Introduction.....	17
2. Materials and Methods.....	20
3. Results.....	24
4. Discussion.....	38

Chapter III. <i>Population pharmacokinetic modeling of sertraline treatment in Alzheimer’s disease patients: the DIADS-2 study</i> .....	40
1. Introduction.....	40
2. Materials and Methods.....	41
3. Results.....	46
4. Discussion.....	52
Chapter IV. <i>Data quality constrains utility of computational modeling of tumor burden in non-small cell lung cancer clinical trials</i> .....	54
1. Introduction.....	54
2. Materials and Methods.....	56
3. Results.....	76
4. Discussion.....	86
Chapter V. <i>Clinical trial simulation to evaluate study design: capturing abiraterone and nilotinib exposures</i> .....	90
1. Introduction.....	90
2. Materials and Methods.....	91
3. Results.....	100
4. Discussion.....	107
Chapter VI. <i>Genetic determinants of vincristine neuropathy in pediatric acute lymphoblastic leukemia patients</i> .....	111
1. Introduction.....	111
2. Materials and Methods.....	114
3. Results.....	121

4. Discussion.....	151
Chapter VII. <i>Genetic signature to predict vincristine neuropathy and relapse in children with acute lymphoblastic leukemia</i> .....	155
1. Introduction.....	155
2. Materials and Methods.....	158
3. Results.....	163
4. Discussion.....	179
Chapter VIII. <i>Conclusions and future directions</i> .....	182
1. Conclusions and future directions.....	182
REFERENCES .....	192
CURRICULUM VITAE	

## LIST OF TABLES

Table 2.1: Parameter estimates of final population pharmacokinetic model.....	26
Table 3.1: Alzheimer’s disease patient demographics.....	48
Table 3.2: Population pharmacokinetic parameter estimates of sertraline in the final model .....	50
Table 4.1: Three U.S. National Cancer Institute (NCI)-sponsored studies conducted by the Cancer and Leukemia Group B CALGB.....	57
Table 4.2 Demographic and clinical characteristics of patients in CALGB 30203 and 30303 trial .....	67
Table 4.3: Tumor model parameter estimates and precision standard error (SE) of baseline (M_BASE), shrinkage rate (M_SR) and progression rate (M_PR) for FDA registration trials and CALGB 30203 and 30303 trial.....	77
Table 4.4: Comparison of tumor parameter estimates between CALGB30203 trial and a subset of E4599 trial which matched to CALGB30203 tumor information.....	79
Table 4.5: Comparisons of PFS and calculated TTG from the target lesion measurements by blinded radiologist evaluation (BE) and case report form (CRF).....	85
Table 5.1: Population pharmacokinetic parameters for abiraterone patients.....	93
Table 5.2: Population pharmacokinetic parameters for nilotinib patients .....	94
Table 5.3: Bias and precision of abiraterone oral clearance at the population level and individual level, between-individual variability, and between-occasion variability .....	104
Table 5.4: Bias and precision of nilotinib oral clearance at the population level and individual level, between-individual variability, and	

between-occasion variability .....	106
Table 6.1: Candidate genes selected from GWAS study .....	134
Table 6.2. Candidate genes selected from gene-enrichment analysis.....	137

## LIST OF FIGURES

Figure 1.1: Opportunities (applications) for modeling and simulation across the drug development process.....	6
Figure 1.2: Biomarker endpoint types .....	11
Figure 2.1: Compartmental representation of clozapine and norclozapine pharmacokinetics .....	25
Figure 2.2: Model diagnostic plots for clozapine .....	31
Figure 2.3: Model diagnostic plots for norclozapine .....	32
Figure 2.4: Visual predictive checks of clozapine in plasma and brain and norclozapine in plasma .....	33
Figure 2.5: Visual predictive checks of norclozapine in plasma and brain .....	34
Figure 2.6: Simulated clozapine human unbound concentrations vs. published human concentrations at steady state in brain with 200,300 and 400 mg OID from 12 to 24 hours.....	36
Figure 2.7: The predicted median percentage of receptor occupancy of D2, 5-HT2A, M1, $\alpha$ 1, $\alpha$ 2 and H1 for clozapine and D2 for norclozapine .....	37
Figure 3.1: Goodness-of-fit plots for the final sertraline model .....	51
Figure 4.1: Selection of patients contributing images from CALGB 30203 and 30303 ...	60
Figure 4.2: An example of a case report form from CALGB 9730 trial .....	61
Figure 4.3: A graphical representation of tumor matrices .....	65
Figure 4.4: Comparisons of identified target lesion measurements from original case report forms and by blinded radiologist reviewer .....	84
Figure 5.1: Study schema of abiraterone and nilotinib clinical trials .....	97

Figure 5.2: Goodness-of-fit plots for the final model of abiraterone.....	102
Figure 5.3: Goodness-of-fit plots for the final model of nilotinib .....	103
Figure 6.1: Vincristine dosage regimens for COG P9904, 9905 and 9906 trials .....	115
Figure 6.2: Flowchart of ALL patient selection in the final GWAS analysis.....	122
Figure 6.3: Population stratification using principal component analysis .....	123
Figure 6.4: Manhattan plot and QQ plot of GWAS association analysis with primary neuropathy .....	125
Figure 6.5: Manhattan plot and QQ plot of GWAS association analysis with secondary neuropathy .....	126
Figure 6.6: Zoom-in plots of each candidate gene from GWAS study and their corresponding SNPs.....	128
Figure 6.7: Zoom-in plots of each candidate gene from Gene enrichment analysis study and their corresponding P-value and odds ratio .....	138
Figure 6.8: The network of cell-to cell signaling and interaction, infectious disease and inflammatory .....	141
Figure 6.9: The network of tissue development, tumor morphology and cancer .....	142
Figure 6.10: Fibrosarcoma cell line response to various cytostatic drugs in the GEO database.....	144
Figure 6.11: QQ-plots of genomic loci associated with ALL neuropathy risk in fibrosarcoma cell line.....	145
Figure 6.12: Genes and corresponding SNPs of genomic loci associated with ALL neuropathy risk in fibrosarcoma cell line.....	146

Figure 6.13: QQ-plots of genomic loci associated with ALL neuropathy risk in eQTL data of four brain tissues .....	148
Figure 6.14: Indirect association of ALL neuropathy risk with eQTL expression.....	150
Figure 7.1: Model strategies using the LASSO penalized regression approach.....	157
Figure 7.2: Cross-validation and LASSO coefficient path plots for primary neuropathy, secondary neuropathy and relapse .....	165
Figure 7.3: ROC curves of primary and secondary neuropathy in two genetic model strategies and the numbers of tested and selected biomarkers in the models .....	171
Figure 7.4: Time dependent ROC of relapse in two genetic models .....	173
Figure 7.5: Survival plot of high risk vs. low risk group of primary neuropathy with sensitivity=0.903081 (A) and specificity=0.901608 (B) .....	174
Figure 7.6: Survival plot of high risk vs. low risk group of secondary neuropathy with sensitivity=0.901325 (A) and specificity=0.900154 (B) .....	175
Figure 7.7: Survival plot of high risk vs. low risk group of relapse with sensitivity=0.900651 (A) and specificity=0.900068 (B) .....	176
Figure 7.8: A graphical representation of the clinical application using the signature model to identify ALL patient in four subgroups (HR RE/HR NE; HR RE/ LR NE; LR RE/HR NE; LR RE/ LR NE) based on relapse and primary neuropathy and their clinical recommendation.....	177
Figure 7.9: A graphical representation of the clinical application using the signature model to identify ALL patient in four subgroups (HR RE/HR NEP; HR RE/ LR NEP; LR RE/HR NEP; LR RE/ LR NEP based on relapse and secondary neuropathy endpoints and clinical recommendation .....	178



## LIST OF ABBREVIATIONS

-2LL	-2 x log likelihood
AD	Alzheimer's disease
ADD	Additive
ALL	Acute lymphoblastic leukemia
ASW	African ancestry in Southwest USA
AUC	Area under the curve
BAV	Between-animal variability
BBB	Blood brain barrier
BQL	Below the quantitation limit
CA-125	Cancer antigen-125
CALGB	Cancer and Leukemia Group B
CHB	Han Chinese in Beijing
CL	Clearance
CNS	Central nervous system\
COG	Children's Oncology Group
CR	Complete remission
CRFS	Case report forms
CSF	Cerebrospinal fluid
CT	Computed tomography
CTCAE	Common terminology criteria of adverse event
CV	Coefficient of variation
CYP	Cytochrome P450
DOM	Dominant
ECF	Extracellular fluid
EQTL	Expression quantitative trait locus
FDA	Food and drug administration
FOCE	First order condition estimation
GEO	Gene expression omnibus
GTEX	Genotype-Tissue Expression
GWAS	Genome wide association study
IIV	Inter-individual variability
KA	Absorption constant
KTR	Transit constant
LASSO	Least absolute shrinkage and selection operator
LLOQ	Lower limit of quantitation
MAF	Minor allele frequency
MBDD	Model based drug development
NA	No available
NC	No change

NCI	National cancer institute
NE	Neuropathy
NEP	Neuropathy+neuropathic pain
NNE	Nearest neighbor estimation
OFV	Objective function value
PD	Pharmacodynamics
PET	Positron emission tomography
PGP	P-glycoprotein
PK	Pharmacokinetics
PM	Pharmacometrics
POP-PK	Population pharmacokinetic
PSA	Prostate specific antigen
Q	Intercompartmental clearance
REC	Recessive
RECIST	Response evaluation criteria in solid tumor
ROC	Receiver operating characteristics
SAEM	Stochastic approximation expectation maximization
SC	Subcutaneously
SLD	Sum of longest diameter
SNP	Single nucleotide polymorphism
TRS	Treatment resistant schizophrenia
V	Volume of distribution
VIPN	Vincristine-induced peripheral neuropathy

## CHAPTER I: *Introduction*

### 1. The need for modeling and simulation in drug development

Despite significant expenditures, the drug development process requires a substantial amount of time and it is expensive and prone to failures. The average length of clinical development is quoted as 7-12 year at a cost of \$0.8-1.7 billion per approved agent (Kaitin et al., 2010; Dimasi et al., 2001; Connolly et al., 2001; Lesko et al., 2004). The cost of drug development takes into account those compounds that undergo clinical testing but do not make it to the marketing approval or fail NDA approval. Therefore, the rate at which pharmaceutical firms successfully develop investigational compounds for marketing approval by regulatory agencies is a critical indicator of the effectiveness of the drug development process (DiMasi et al., 2010). A retrospective analysis of both public and private drug pipeline database as well as from surveys comprising nearly 4000 drugs and biologics from the 50 largest pharmaceutical firms between the mid-1990s and the early 2000s by DiMasi's group pointed to clinical approval success rates that remained lower than 20% with an upward trend in costs (DiMasi et al., 2010). Estimated clinical approval success rates differed significantly by therapeutic class; for example, in oncology, the success rate was only ~5% (Kola et al., 2004). Furthermore, even with the massive amount of data that are generated and obtained from the clinical trials every year, lots of useful knowledge and information remains undiscovered or underutilized (Ette et al., 2007). Processes and technological innovations that can improve the predictability of outcomes for new compounds can therefore significantly increase the productivity of new drug innovation (DiMasi et al., 2010). Indeed, in the *white paper Challenge and Opportunity on the Critical Path to New Products* published in March of 2004, the Food

and Drug Administration (FDA) has addressed this concern and states “Not enough applied scientific work has been done to create new tools to get fundamentally better answers about how the safety and effectiveness of new product that can be demonstrated in faster time frames, with more certainty, and at lower cost. A new product development toolkit—containing powerful new scientific and technical methods such as animal or computer-based predictive models, biomarkers for safety and effectiveness, and new clinical evaluation techniques—is urgently needed to improve predictability and efficiency along the critical path from laboratory concept to commercial product. We need superior product development science to address these challenges.” To highlight all these issues regarding the high attrition rates and increasing costs in drug development, model-based drug development has been proposed as one methodology to potentially make drug development more productive by assessing the ongoing clinical trial studies as well as historical drug databases using mathematical and statistical approaches to construct and utilize models to increase the understanding of the drug and disease.

## 2. Model based drug development (pharmacometrics and statistics)

Model-based drug development (MBDD) is the concept of utilizing pharmacostatistical approaches to evaluate drug efficacy and safety from preclinical through the clinical data to improve drug development knowledge management and decision-making (Milligan et al., 2013; Lalonde et al., 2007). As Lalonde et al. described, the key components of MBDD can be divided into six categories: PK-PD and disease/placebo model; meta-analysis of candidate drug and competitor data, design considerations and trial execution models; data-analytic models; quantitative decision criteria; and trial performance metrics (Milligan et al., 2013; Lalonde et al., 2007). Those principles can

potentially overcome the common limitation of the current drug development strategy, specifically, that prior information is partly or completely ignored when analyzing and interpreting the results of the most recent clinical trials.

One of the essential parts of MBDD is the discipline of pharmacometrics. Pharmacometrics has been described as “the science of developing and applying mathematical and statistical methods to a) characterize, understand, and predict a drug’s pharmacokinetic and pharmacodynamic behavior, b) quantify uncertainty of information about that behavior and c) rationalize data-driven decision making in the drug development process and pharmacotherapy” (Ette et al., 2007). The modeling and simulation approach not only quantifies the exposure-response relationship, disease progression, and the level of variability, but also identifies the potential contributors to that variability. In addition, these models can be linked and applied to competing study designs and to customize patient drug therapy through therapeutic drug monitoring and improved population dosing strategies; therefore, facilitating the implementation of the personalized medicine (Bonate , 2011). The idea of personalized medicine can be further utilized to emphasize the significance of biomarker findings. As a result, biomarker identification and validation become one of the major foci of pharmacometricians across disciplines. Applying quantitative assessment approaches to understand biomarker dynamics can potentially: identify patients at risk for a disease; predict a patient response; and predict the risk of toxicity.

Other statistical tools have also become popular to describe the relationship of disease and clinical outcome which may or may not involve drugs or treatments using mathematical equations such as regression analysis and survival analysis. In particular,

survival analysis has been commonly used in clinical studies to analyze time to event data which can provide insight of patients' efficacy and safety profiles and facilitate decision making such as dose selection or patient care.

Since the quantity of the data expands as the improvement of data collection method, an efficient computer tool becomes critical and demanded to facilitate the access to large amount of information (ex: pharmacogenomics profiles). This can be overcome by the development and implementation of bioinformatics tool.

### 3. The opportunities (application) of modeling and simulation in drug development

As the role of modeling and simulation in drug development becomes more significant, there are many opportunities available for modeling across different phase of studies. In the journal of the pharmaceutical sciences 2002, Meibohm and Derendorf initially indicated the potential modeling and simulation applications from the discovery and preclinical phase to clinical setting and even the post-marketing phase in Figure 1.1. As shown in Figure 1.1, in the early stages of drug development, modeling and simulation can explore the pharmacokinetic and pharmacodynamic of drug and drug-drug interaction in *in vivo* system and leverage the understanding of biomarker to evaluate drug efficacy and toxicity profile. It can also be useful in selecting dosage regimen or form based on the previously developed PK/PD relationship and integrate the early decision making in candidate selection. Furthermore, with the modeling assessment of preclinical studies, a first to human dose can be suggested by extrapolating from the *in vivo* model, and the translational approach can be also beneficial for systematically assessing some pharmacological relationships which have not been assessable in the human system. A parallel approach has been suggested by combining preclinical and

early clinical development as a way to expand the learning process to all the phases of drug development. In the learning and confirming process of clinical trial, modeling and simulation serves as a tool to continuously evaluate exposure-response relationship across different therapeutic areas (Huang et al., 2013; Meibohm et al., 2001; Sheiner, 1997). The model can further evaluate contributors such as gender and food effects on exposure, predict PK/PD under a new dosing regimen, predict PK/PD in a special population such as children or elderly, and characterize drug-drug interaction and drug-disease interaction (Meibohm et al., 2002). In clinical drug development, predictive tools such as Monte Carlo simulation can explore various dosage regimens and optimize the trial design which might allow reducing the numbers and costs of the studies and improving drug development efficiency (Meibohm et al., 2002).

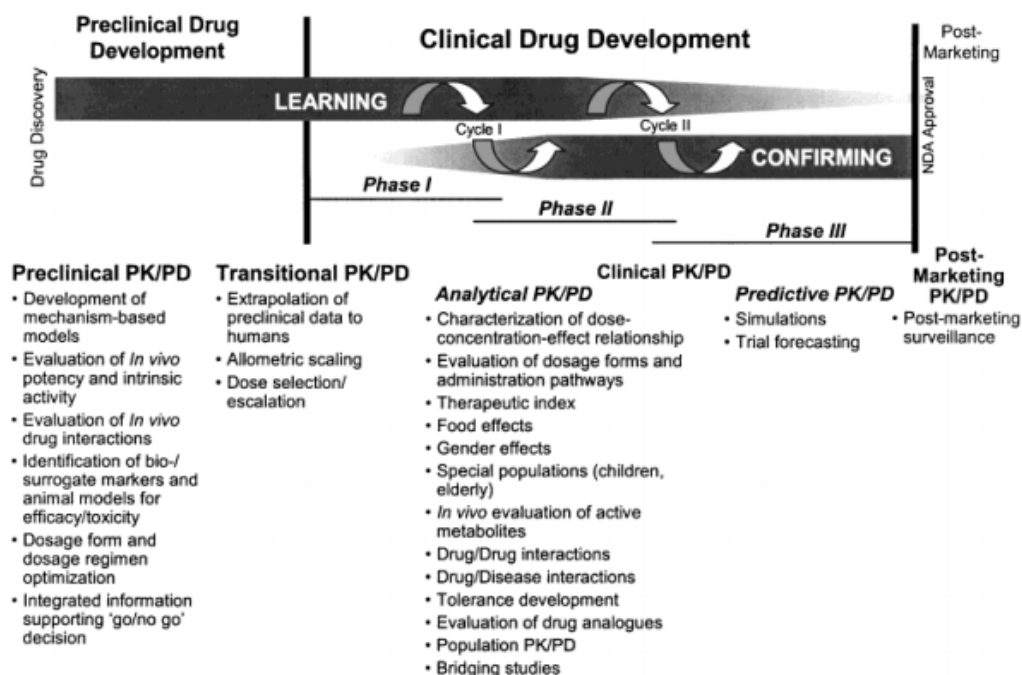


Figure 1.1. **Opportunities (applications) for modeling and simulation across the drug development process.** (This figure is reproduced with permission of John Wiley & Sons, Inc. Meibohm B and Derendorf H, Pharmacokinetic-pharmacodynamic studies in drug product

development, and copyright © 2002 Wiley-Liss, Inc. and the American Pharmaceutical Association).

#### 4. Type of models

As the demands and opportunities of modeling and simulation increases in each phase of drug development, various types of models have been advocated to address specific questions. Some of the model strategies are introduced in the following sections, and the applications are demonstrated in each section.

##### 4.1 Pharmacokinetic modeling

Pharmacokinetics began as a way to characterize the absorption, distribution, metabolism and elimination of a drug in the body and to reduce a concentration-time profile into a set of parameters that could be used for comparison, evaluation, and prediction (Teorell, 1937; Teorell 1937; Widmark, 1933; Wanger, 1973; Widmark, 1933; Derendorf et al., 2000). As the assays or the measurement tools becoming available not only in the plasma but also in some tissue levels, disposition of drug exposure pattern can be better understood. Furthermore, many forms of drug exposure were also characterized in the pharmacokinetic analysis. Depending on whether drug concentration measurements are available; sometimes, dose can be used as a representation of nominal exposure, and population average exposures after dose have been proposed. These are known as KPD models explore the dose-driven shape of the concentration vs time profile specifically the temporality without inter-individual variability in the context of exposure and pharmacodynamics relationship. Typically, drug concentrations measured in plasma represent the systemic exposure of the drug. Pharmacokinetic parameters such as clearance (CL), volume of distribution (Vd) or descriptors such as area under the curve (AUC), maximum concentration (C<sub>max</sub>) and elimination half-life ( $t_{1/2}$ ) can be derived



or calculated. Beyond the drug concentrations measured in the biologic fluid, with the innovation and improvement of the measurement techniques, drug concentrations now can be measured at the target site using microdialysis or positron-emission tomography scan so that the responses can be more precisely correlated to drug concentrations at the effect site (Derendorf et al., 2000; Chefer et al., 2009; Jacobson et al., 2013). In addition, understanding the covariate effect on systematic exposure and variability associated with exposure is also an important component in a pharmacokinetic model. Covariates are characteristics describing the patients, the conditions of the drug treatment or other factors potentially influencing the outcome. In late 90's, Vozeh's and Grasela's groups demonstrated how covariates can be used to predict patients' pharmacokinetics through the covariate model development in a population analysis (Vozeh et al., 1982; Grasela et al., 1984; Grasela et al., 1985; Grasela et al., 1987). The covariates can be constant within an individual or time-varying. They not only contribute to the estimates of structure model parameters but also impact the random effects distribution (Ribbing, 2007). In general, covariates can be classified as intrinsic factors (e.g. age, race, weight, sex, and genotype) or extrinsic factors (e.g. compliance, and smoking status) or can be categorized as categorical or continuous variables based on the measurement scale (Holford, 2013). In the process of model development, an assessment of covariate effect can improve predictive model performance for subject in the current data set, for trial simulation of future studies or for future patient population (Gastonguay MR, 2011).

#### 4.2. Pharmacodynamic modeling

Pharmacodynamics is often defined as what the drug does to the body (Ette et al., 2007). Based on the definition from Derendorf et al, "pharmacodynamics is a broad term

that is intended to include all of the pharmacological actions, pathophysiological effects and therapeutic responses, either beneficial or adverse of active drug ingredient, therapeutic moiety, and/or its metabolite(s) on the various systems of the body from subcellular effects to clinical outcomes” (Derendorf et al., 2000). In pharmacodynamic modeling, the pharmacological action is often modeled as a time dependent effect that is related to drug exposure, and the endpoints of pharmacodynamics can be physiological changes such as blood pressure and cholesterol level or clinical outcomes such as survival or toxicity. In addition, with the flexibility of mathematical functions, different types of pharmacodynamic endpoints including continuous, categorical, ordered categorical or count data can be described. Another type of pharmacodynamic response is the time to a specific event. A common approach for modeling, this type of event is survival analysis. Typically, a proportional hazard model or parametric survival model is likely to be used to describe the baseline hazard and the risk that depends on covariates in a predefined time period.

#### 4.3. Population PK/PD modeling

Sheiner and his colleagues published several articles illustrating a new mathematical approach to analyze pharmacokinetic data, which was later called population pharmacokinetics in the late 1970s and early 1980s (Sheiner et al., 1977; Sheiner et al., 1980; Sheiner et al., 1981; Sheiner et al., 1982; Bonate, 2005). In one of the Sheiner’s publications, he demonstrated that less-biased estimates of the population means and variances were found using the population approach than either the naïve-pooled or 2-stage approach (Sheiner et al., 1981; Bonate, 2005). Population methods were further applied to assess pharmacodynamics endpoints. Population PK/PD modeling

approaches represent a methodology of leveraging sparse PK and PD data. Commonly, this approach is applied to relatively sparse sampling but in a sufficient number of individuals so that inter-individual variability can be assessed and allows integrating data from single or multiple studies as well as studies from different phases (Ette et al., 2007; Gross et al., 2000). With the flexibility of analyzing combined data from different studies, PK or PD profile of the population can be better represented. However, analyzing combined data in a typical model-based meta-analysis may potentially result in bias estimates of parameters. Therefore, this strategy should be carefully evaluated and applied (Gastonguay et al., 2005; Gastonguay et al., 1999). This strategy also allows for the identification and quantification of the sources of inter-individual variability in response such as genetic polymorphisms in metabolic enzyme or transporter, clinical and demographic factors. In particular, the population modeling approach has become one of the standard assessments of clinical drug development that regulatory agencies recommend (Department of Health and Human Services, 1999).

#### 4.4. Biomarker modeling

The official National Institutes of Health (NIH) definition of a biological marker (Biomarker) is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention.” This definition is broad and though not explicitly stated, includes laboratory tests, radiologic studies as well as physical exam findings. Although biomarker is a relatively new term that dates back to the late 1960s, biologic assessments and measurements in the evaluation of human disease were practiced in antiquity (Ferguson et al., 2010). Clinically useful biomarkers have evolved over time and play an

increasingly important role in many aspects of pharmaceutical discovery and development including personalized medicine and the assessment of safety profile, reflecting the scientific and technologic progress made over the centuries.

As a result, an increasing number of clinically relevant tests and procedures are available to estimate organ injury and guide treatment. Recent discoveries in genetics and molecular biology have resulted in impressive advances in our understanding of the pathophysiologic processes of individual disease and yielded an abundance of prospective therapies directed against novel targets (Ferguson et al., 2010). This has brought about an increased focus on biomarker identification, validation, and quantification, as well as the development of analytical technologies for biomarker measurement. Efforts at biomarker discovery and validation have intensified since the twenty-first century (Ferguson et al., 2010). Advanced genomic, proteomic, and metabolomics techniques now permit comparative analysis of specimens from healthy and diseased individuals, facilitating biomarker identification. Generally, there are four types of biomarker endpoints (Figure 1.2) (Jenkins et al., 2011). The first is called a prognostic endpoint which is used to predict the likely disease prognosis independent of the mode of treatment. For instance, the number of circulating tumor cells (CTCs) in peripheral blood at baseline can give an indication of survival prognosis in prostate cancer (De Bono et al., 2008). The second type is a predictive biomarker that predicts the likelihood of response to a particular treatment or a class of treatments. As an illustration, CTCs can also be a predictive biomarker when compared wild type patients vs. patients with mutation in the same treatment. This acts more like a covariate to explain the variation in the outcome. The third type of biomarker is a pharmacodynamics biomarker.

The definition of a pharmacodynamic biomarker is a response over time to a treatment intervention. Simple biomarker examples include cancer antigen 125 for ovarian cancer (Schmidt et al., 2011) and PSA for prostate cancer (Romero et al., 2013), and tracking treatment over time is an important component. Pharmacodynamic biomarkers could reflect the safety and/or efficacy of a treatment and are typically measured at multiple time points. Markers that correlate well with a widely accepted clinical outcome at both an individual and group level could potentially act as a surrogate endpoint and substitute for a recognized clinical endpoint (Katz et al., 2004; Rothmann et al., 2012) and are the fourth type of biomarker. For example, LDL cholesterol can act as a surrogate for major cardiovascular events in the licensing of statins (Tardif et al., 2006). The model based assessment of the dynamics of a biomarker can provide insight into the disease progression and capture a therapeutic drug effect which might be useful for early decision making before the clinical endpoints are available.

### Biomarker Endpoint Types

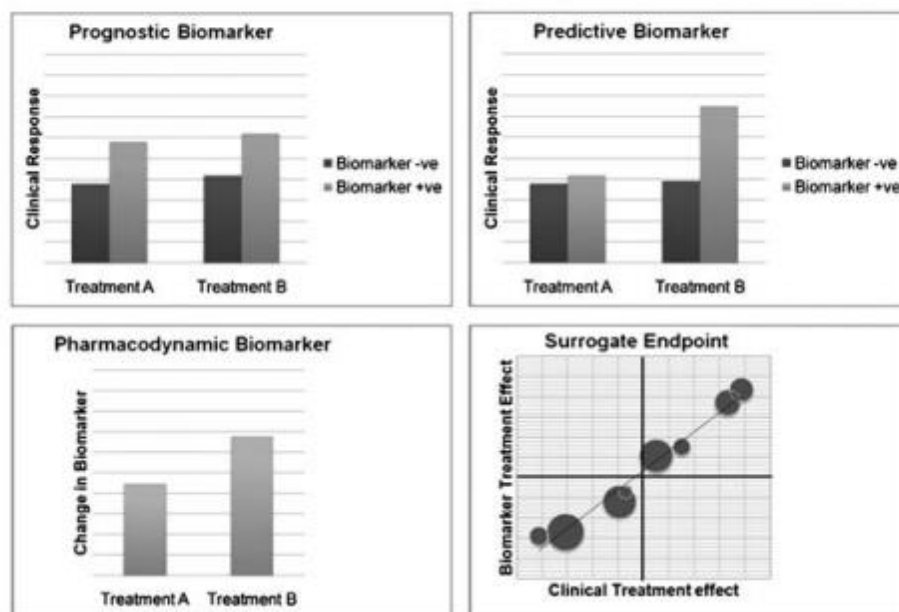


Figure 1.2. **Biomarker endpoint types** (This figure is reproduced with permission of John Wiley & Sons, Inc. Jenkins et al., A statistician's perspective on biomarkers in drug development, and copyright © 2011 John Wiley & Sons, Ltd.).

#### 4.5. Clinical Trial simulation

Another significant component of pharmacometrics is stochastic simulation. Simulation has been widely used in many disciplines such as engineering before being used in drug development. It can serve not only as a predictive tool to evaluate the proposed trial design and the outcome under certain assumptions and to exposure pattern resulting from various dosage regimens, sampling schemas and population characters but also as an analytical tool to characterize exposure-response relationship in different modeling circumstances in a virtual population (Kimko et al., 2002; Kimko et al., 2011; Meibohm et al., 2002). Clinical trial simulation also allows us to compare different study designs and optimize study design by updating the design continuously. With the simulation assessment, the drug development process can be more efficient and powerful and more informative in bridging studies from one phase to another.

#### 4.6. Time to event model (GWAS study)

Clinical trials are often performed to evaluate the efficacy of new treatment regimens and the efficacy of existing treatments in a special population (National cancer institute, 2011). As mentioned briefly in section 4.2, the efficacy endpoints of the clinical trials are primary described as an event of interest such as death, relapse, and adverse drug reaction (Crom et al., 1994). The follow-up time of the patients may vary depending on the disease and treatment, the time component is an important assessment to understand the event occurrence. These statistical models consider the time course until an event occurs and compare the cumulative probability of events over time for two or

more cohorts (Lee et al., 1982; Singh et al., 2011). Therefore, a statistical procedure which is likely to be applied for analyzing these endpoints is called time to event analysis. Because survival is one of the most common endpoints in time to event analysis, people also describe this type of analysis as survival analysis. In time to event analysis, one of the essential components is how the dichotomous event of interest is pre-defined. For example, if an adverse drug reaction is classified using a grading scale from 0 to 5, the event of interest can be specified as binary or categorical outcome based on the question we are interested in (National Cancer Institute, 2009). Another key component of most time to event models is censoring. Censoring can arise under three circumstances: when an individual does not experience the event during the study period; when an individual discontinues the follow-up during the study period; and when an individual drops out from the study due to other events (Singh et al., 2011).

One of the most common time to event models is the Cox proportional hazards model (Cox et al., 1972). This model was first introduced by Cox, in 1972, for analysis of survival data with and without censoring, and for identifying the contributors to the difference of survival in clinical trials. These contributors include treatment effect, prognostic and clinical covariates. Different from parametric survival or the Kaplan-Meier method (Itman et al., 1992), the Cox regression model is considered a semi-parametric method which does not require a pre-specified hazard function for the baseline hazard but considers the hazard to be proportional at any point in time for the characteristics based on the model assumption. The hazard is calculated using the equation shown below:

$$H(t) = H_0(t) \times \exp(b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_jX_j)$$

where  $X_1 \dots X_j$  represent a group of predictor variables, and  $H_0(t)$  is the baseline hazard at time  $t$ , representing the hazard for a person when all the predictor variables are 0 (Cox et al., 1972).

The Cox model has been widely used in covariate selection for association studies such as Genome wide association study (GWAS), and it is very useful and robust tool in clinical research and provides valuable information about an intervention.

#### 5. Hypothesis and specific aims

Modeling and simulation has been shown to be a critical tool in drug development and patient care. In drug development, modeling and simulation provides an early assessment of a new drug entity in the initial discovery phase to facilitate: the compound selection; the prediction of pharmacokinetic-pharmacodynamic(PK-PD) relationships from in-vivo animal results to human; the characterization of exposure-response profiles at both the individual and population levels in the clinical studies; the evaluation of the trial design; and insights to inform decision making in drug development. Furthermore, as the studies go beyond drug approval, modeling and simulation can also be applied to the post-marketing evaluations for special populations and to predict drug toxicity and efficacy in a long-term patient care setting with the potential to inform clinical recommendations.

The main objectives of this proposal use a modeling and simulation approach: 1) to leverage our understanding of PK/PD relationship from pre-clinical system to human; 2) to quantitatively understand the drug impact on patients; 3) to evaluate clinical trial designs and; 4) to identify potential contributors (predictors) to drug toxicity and efficacy. This work explores the hypothesis that modeling and simulation can be useful to provide



insight into trial designs and facilitate decision making (Steimer et al., 2000). To test this hypothesis, we proposed the following specific aims:

1. Build a population pharmacokinetic model that accounts for both plasma and brain concentrations measured in rats and utilize this model to predict human concentrations of clozapine and its N-desmethyl metabolite, norclozapine. (Preclinical PK model)
2. Identify covariates that contribute to variability in sertraline concentration by performing a population pharmacokinetic analysis of sertraline in elderly patients with Alzheimer disease and generating population pharmacokinetic parameters for this population. (Population PK model in a special population)
3. Assess and refine the published FDA longitudinal tumor size model for predicting survival in NSCLC patients using archived tumor measurement data. (Pharmacodynamic /Biomarker model)
4. Evaluate whether a pre-defined oncology trial with a sparse drug concentration sampling schedule can adequately capture individual level drug exposures, random variability and the food effects of abiraterone and nilotinib. (Clinical trial simulation and trial design evaluation)
5. Establish a candidate gene database which included those polymorphisms associated with vincristine-induced neurotoxicity in childhood ALL patients from the GWAS study and gene enrichment analysis (Time to event model)
6. Create a signature comprising a combination of genetic and clinical markers that predict vincristine-induced neurotoxicity and relapse in childhood ALL patients using LASSO penalized regression model. (Signature model)

Successful completion of these aims would result in accelerated and information rich decision making by leveraging data from previous trials to simulate clinical outcomes as well as aid in the design of future clinical trials. Modeling and simulation allows one to explore the pharmacokinetic profile in a special population and exposure and response relationship in preclinical and clinical setting to facilitate drug development as well as optimal drug utilization and therefore provide better care for the patients. Furthermore, potential biomarkers for survival prediction and drug- induced adverse reactions identified by the modeling approach can help in guiding treatment decisions and facilitating the future protection for patients. Ideally, patients most likely to derive a benefit from treatment are more likely to be targeted; in contrast, those patients identified as non-responders are more likely to be excluded so that unnecessary harm to patients can be minimized. This may allow for the identification of ineffective treatments earlier in the clinical trial process thus also avoiding unnecessary patient exposures. In this study, modeling and simulation demonstrates the potential applications in drug development and patient care and the utility of clinical prediction.

**CHAPTER II:** *Preclinical model-prediction of brain clozapine and norclozapine concentrations in humans using a scaled pharmacokinetic model for rat brain and plasma pharmacokinetics (Published in Journal of translational medicine with open access license <http://www.biomedcentral.com/about/license>) (This material is reproduced with permission of BioMed Central)*

### *1. Introduction*

Schizophrenia is a debilitating disorder that affects approximately 1% of the global population without regard to race, sex or socioeconomic status (Mathers et al., 2006). It typically strikes in the late-teen years or early twenties and is characterized by a high rate of morbidity and mortality. Given these high personal and societal costs, investment in research aimed at understanding the biology of the disease, its genetic components and their interplay with environmental factors, continues on many levels. Over the past 50 years, pharmacotherapeutic support has been instrumental in managing primarily the positive symptoms of the disease and hinges on suppression of a central circuitry dysfunction that can be normalized by antagonism of dopamine D2 receptors in the striatum (Murray et al., 2008). Introduction of clozapine, the first so-called atypical antipsychotic approximately 25 years ago, represented a significant advance in our understanding of schizophrenia from a systems biology perspective in that this drug did not have the typical side effects of the 1<sup>st</sup> generation neuroleptics. This reduction in side effects was attributed to higher 5HT-2A than D2 binding (Meltzer et al., 1989).

However, clozapine pharmacology is not limited to D2 and 5HT2A antagonism. Albeit unintentionally, the drug binds to several other dopamine and serotonin receptor subtypes, as well as to muscarinic M1/M4 and alpha-1 adrenergic receptors with

pharmacologically relevant affinity (Horacek et al., 2006). From a clinical perspective, this broad receptor coverage may account for clozapine's unique superiority in treatment resistant schizophrenia (TRS), even amongst other atypical antipsychotics. From a research perspective, the broad receptor coverage of clozapine conceivably makes the drug a useful tool to advance our understanding of complex pharmacotherapy that incorporates multiple interacting receptor systems.

The use of positron emission tomography (PET) imaging to measure receptor occupancy of clozapine and other atypical antipsychotics in humans has been invaluable in demonstrating the importance of D2 and 5HT2A receptor antagonism contributing to the efficacy of these drugs (Takeuchi et al., 2013; Moriguchi et al., 2013; Tsuboi et al., 2013). However, broader application of this non-invasive technique has been limited by the lack of ligands specific for other receptors to which clozapine has affinity. In this regard, availability of other approaches that are complementary to PET imaging would be useful. One possibility is to link non-clinical measurements of clozapine disposition in the brain with clinical studies of clozapine systemic exposure using a translational PK modeling approach. Prediction of clozapine CNS exposure could then be related to its receptor binding kinetics at multiple receptors to impart a virtual predicted pharmacodynamic component to a model. This approach has been used recently to predict CNS concentrations of atomoxetine and duloxetine that were in the range of receptor affinities associated with therapeutic doses (Kielbasa et al., 2012). In a related manner, a population pharmacokinetic-pharmacodynamic (PK-PD) modeling approach was used to predict D2 receptor occupancy of olanzapine in humans (Johnson et al., 2011), and the D2 and 5HT2A receptor occupancy of risperidone and its active

metabolite paliperidone (9-OH risperidone) (Takeuchi et al., 2013; Moriguchi et al., 2013; Tsuboi et al., 2013; Kozleiska et al., 2012). These studies, as well as earlier PK-PD models applied to other CNS drugs (de Lange et al., 2012; de Lange et al., 2005), provide confidence in the ability of this approach to deepen our understanding of drug action in human brain.

A recent study measured clozapine and its N-desmethyl metabolite, norclozapine, in extracellular fluid (ECF) of rat medial prefrontal cortex using quantitative microdialysis, and these results provided evidence of net efflux from brain across the blood-brain barrier (BBB) (Cremers et al., 2012). This suggests that plasma concentrations may not be a good predictor of brain concentration for clozapine or norclozapine. Therefore, prediction of clozapine exposure in the ECF of human brain using a translational PK modeling approach could be cross-validated against PET results at D2 and 5HT2A receptor occupancy in humans, and subsequently used to estimate clozapine receptor occupancy at the drug's other receptor targets for which PET tracers do not exist. Such comprehensive PK-PD model could potentially support individualized dosing of clozapine to improve its efficacy and CNS tolerability. It would also support research aimed at discovering new approaches for the treatment of the schizophrenia in its different forms.

The purpose of this study was; (1) to build a PK model that accounted for both plasma and brain concentrations measured in rats; (2) to utilize this model to predict concentrations of clozapine and norclozapine in human brain. This would allow for the prediction of expected receptor occupancy in humans.

## *2. Materials and Methods*

### **2.1. Study Design**

A single dose of clozapine (10mg/kg) was administered subcutaneously to four male Wistar rats with an average weight of 0.35 kg purchased from Harlan (Zeist, The Netherlands). Three days prior to administration a microdialysis guide cannula was surgically implanted in the medial prefrontal cortex; at the same time, a catheter for blood sample collection was placed in the right jugular vein and was exteriorized through an incision at the top of the head. This vascular cannulation enabled an equivalent volume of saline replacement for each blood sample. A MetaQuant probe (6 mm, cellulose membrane, BrainLink, The Netherlands) was inserted into the guide cannula 24 hours prior to drug administration to enable sampling of brain extracellular fluid (ECF). Concentrations of clozapine and norclozapine were measured in plasma and brain (ECF) by HPLC with tandem mass spectrometry in the positive ion mode as previously described (Cremers et al., 2012). For each rat, the unbound concentrations in each compartment were measured at 9 time points (0, 15, 30, 60, 90, 120, 240, 360 and 480 minutes) in plasma and 18 time points (-30, 0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 360, 390, 420, 450 and 480 minutes) in brain. A single dose of norclozapine (10 mg/kg) was also administered subcutaneously to another five male Wistar rats with an average weight of 0.34 kg bought from the same Harlan laboratories. Concentrations were measured in plasma and brain ECF, and the same time points were used as those specified for clozapine.

## 2.2. Model Development

Different model structures were initially evaluated using the system dynamics software VENSIM (Ventana Systems, Inc., MA, US). Thereafter a population approach was used to describe the pharmacokinetics of clozapine and norclozapine. Population PK parameters were estimated using a nonlinear mixed effect modeling approach, as implemented in NONMEM version 7.2 (Icon Development Solutions, Hanover, Maryland) using Wings for NONMEM version 7 (Holford, 2013). The first-order conditional estimation method (FOCE) with interaction was used to estimate the structural PK parameters and the random effects parameters.

Model development was started with an assessment of clozapine PK in plasma. One and two compartment models with first order absorption for clozapine in plasma were tested. A peripheral compartment structure was subsequently implemented to represent the brain extracellular fluid concentrations. The transfer characteristics of clozapine between the plasma and the brain compartment were evaluated using an intercompartmental clearance,  $CL_{in}/CL_{out}$ , as well as incorporating delay functions (Savic et al., 2007). These delay functions included a lag time and transit compartment approaches. Once the structural model for clozapine was established, the plasma compartment of norclozapine was integrated and then connected to the brain compartment. The same strategy was utilized in building the structural model for norclozapine concentrations that were measured following norclozapine administration. Clozapine and norclozapine concentration measurements were then combined from the 9 rats and modeled simultaneously in the final structural model. The volume of distribution of clozapine and norclozapine in brain were tested with and without fixing this parameter

to a literature reported value (Fridén et al., 2007). A parallel metabolic pathway from the extravascular space was also explored.

Between-animal variability (BAV) for PK parameters was assumed to be log-normally distributed and evaluated using an exponential model  $P_i = P_{TV} \times e^{\eta_p}$  where  $P_i$  is the parameter estimate for the  $i^{\text{th}}$  animal, and  $P_{TV}$  is the typical parameter value at the population level. The difference between  $i^{\text{th}}$  individual and population parameter values was described by  $\eta_p$ , which was identically distributed with mean equal to 0 and variance,  $\omega_{\eta}^2$  (Feng et al., 2006). A combined additive and proportional model was first used to describe the intra-animal variability. If one of the elements of the model was found to be negligible and not significant, it was then removed from the residual error model. Residual error parameters were assumed to be normally distributed with mean equal to 0 and variance,  $\sigma^2$ .

### **2.3 Model selection and evaluation**

Model evaluation was based on a likelihood ratio test using the objective function value (OFV) provided by NONMEM. The minimum OFV returned by NONMEM is approximately equal to  $-2 \times \log$  likelihood ( $-2LL$ ) and served as a guide during model design. A decrease in  $-2LL$  of 6.63 points for 1 degree of freedom was regarded as a significant model improvement, corresponding to a p value of 0.01 for nested models. The final model was further examined using goodness-of-fit plots generated using R version 2.13 based on the conditional weighted residuals distribution and the predicted versus observed concentrations at both the population and individual levels. Furthermore, the final pharmacokinetic model was also evaluated using a visual predictive check



(VPC), and the uncertainty on each parameter was determined using a non-parametric bootstrap sampling with replacement 1000 times from the original dataset.

## **2.4 Prediction of Human Clozapine, Norclozapine Brain Concentrations and Expected Receptor Occupancy**

After the pharmacokinetic model of clozapine in rat was finalized, the PK model framework was adapted by scaling PK parameters with allometric principles to predict human concentrations in brain. The following exponents were utilized scaling body weight to: clearance 0.75; volume of distribution 1; and first order rate constants 0.25 (Sharma et al., 2009). A 50% of conversion from clozapine to norclozapine in humans was assumed in the model based on prior reports (Centorrino et al., 1994; Raedler et al., 2008; Couchman et al., 2010), and this was implemented in the simulated model assuming  $CL_{clo}/F$  is equal to  $CL_{clo-p}/F$ . The model performance was evaluated comparing model simulated plasma concentrations to published human clozapine concentrations (Ismail et al., 2012) in plasma at steady state following 200, 300 and 400 mg daily doses. The published human clozapine data were reported as total concentrations, and these concentrations were converted to free concentration using 3% unbound fraction (Clozapine Product Insert, 2013) prior to the comparison. After model validation, the simulated human clozapine and norclozapine concentrations were used to calculate the expected human receptor occupancy for multiple receptors. Receptor occupancies were predicted for: dopamine 2(D2); serotonin 2A (5-HT2A); muscarinic-1 (M1); alpha-1 adrenergic ( $\alpha 1$ ); alpha-2 adrenergic ( $\alpha 2$ ); and histamine-1 (H1) using published equilibrium dissociation constants ( $K_d$ ) for clozapine (Seeman et al., 2002; Bumaster et al., 1996; Kroeze et al., 2003) and norclozapine (D2 only) (Lidow et al., 2000).

### 3. Results

#### 3.1 Rat Population Pharmacokinetics

A two-compartment model with first order absorption best described clozapine pharmacokinetics in rats using a central compartment for plasma concentrations and a peripheral compartment for brain concentrations. Between plasma and brain, an apparent delay in the distribution of clozapine was identified. Several structural models were tested to capture the observed delay. A transit compartment model with two compartments best described flow from plasma to brain, and inter-compartment clearance described the return from brain to plasma (Figure 2.1). Population pharmacokinetic estimates are given in Table 1. Norclozapine exposures in plasma and brain following clozapine administration were adequately described using a similar structure, but with one fewer transit compartment ( $K_{tr2}$ ), which was estimated to be approximately 40% of the clozapine value. The volume of distribution of clozapine ( $V_{clo-p}/F$ ) and norclozapine in brain ( $V_{met-p}/F$ ) were fixed to the previously estimated values in the final model. Although a significant reduction in the OFV was observed when both parameters were estimated, they were estimated with very poor precision. The elimination of clozapine converted to norclozapine was  $CL_{clo-met}$  at 0.055 L/min, which is approximately 10% of total clozapine systemic clearance. The NONMEM control stream with the selected model is also included in the supplementary material (Additional file 1).

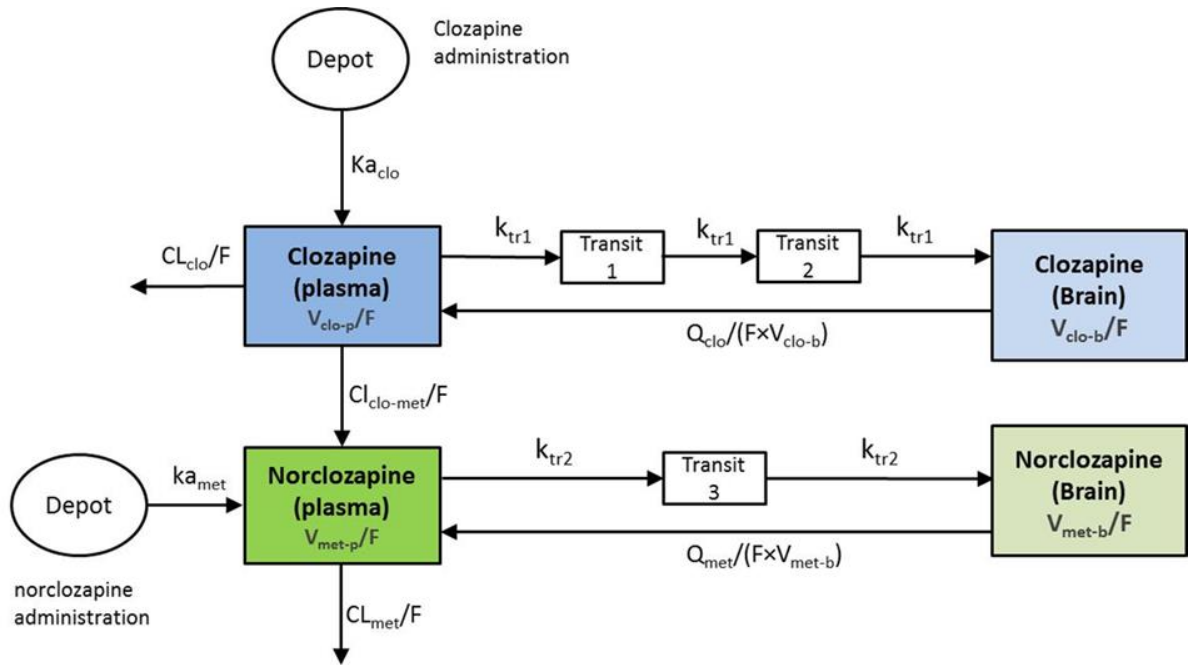


Figure 2.1. **Compartmental representation of clozapine and norclozapine pharmacokinetics.**

Two compartments in blue represented the plasma and brain compartment of clozapine.

Two compartments in green represented the plasma and brain compartment of norclozapine.

Clo (Clozapine):  $CL_{clo}/F$ =clearance of clozapine;  $V_{clo-p}/F$ =volume of distribution of clozapine in plasma;  $Ka_{clo}$ =absorption rate of clozapine;  $Q_{clo}/F$ =intercompartmental clearance of clozapine;

$V_{clo-b}/F$ = volume of distribution of clozapine in brain;  $Ktr1$ = transit rate constant of clozapine.

Met (Norclozapine):  $CL_{clo-met}/F$ =clearance of clozapine to norclozapine;  $CL_{met}/F$ =clearance of norclozapine;

$V_{met-p}/F$ =volume of distribution of norclozapine in plasma;  $Ka_{met}$ =absorption rate of norclozapine;

$Q_{met}/F$ =intercompartmental clearance of norclozapine;

$V_{met-b}/F$ = volume of distribution of norclozapine in brain;  $Ktr2$ = transit rate constant of norclozapine.

Parameters	Estimates (RSE%)	BAV (RSE%)	Bootstrap analysis Median [5-95th percentiles]	
			Estimates	BAV
Clozapine				
CL <sub>clo</sub> /F (L/min)	0.5(20.3)		0.463[0.358-0.661]	
V <sub>clo-p</sub> /F (L)	19.4(40.3)		18.9[5.43-31.59]	
Ka <sub>clo</sub> (1/min)	0.00801(14.7)		0.00815[0.0051-0.0095]	
Q <sub>clo</sub> /F (L/min)	2.01(40.8)	0.193(33.6)	2.08[0.58-3.43]	0.2[0.07-0.29]
V <sub>clo-b</sub> /F (L)	0.214 FIXED			
Ktr1 (1/min)	0.0125(9.4)	0.05 (48.9)	0.0129[0.011-0.015]	0.047[0.0033-0.069]
F <sub>clo</sub>	1 FIXED	0.259(37.1)		0.243[0.062-0.35]
Norclozapine				
CL <sub>clo-met</sub> /F (L/min)	0.055(24.8)		0.0584[0.04-0.0855]	
CL <sub>met</sub> /F (L/min)	0.419(18.8)	0.111(74.8)	0.43[0.33-0.61]	0.08[0.0038-0.18]
V <sub>met-p</sub> /F (L)	2.95(38.6)	0.168(51.1)	3.02[1.91-5.55]	0.149[0.000047-0.27]
Ka <sub>met</sub> (1/min)	0.00277(31)	0.371(44.2)	0.00296[0.0015-0.0047]	0.319[0.065-0.54]
Q <sub>met</sub> /F (L/min)	0.388(45)		0.386[0.229-0.768]	
V <sub>met-b</sub> /F (L)	0.25 FIXED			
Ktr2 (1/min)	0.00517(12.3)		0.00521[0.0045-0.0064]	
F <sub>met</sub>	1 FIXED			
Residual error (proportional)				
Parent-plasma	0.109(70)		0.084[0.029-0.19]	
Parent-brain	0.0367(27)		0.0371[0.023-0.055]	
Metabolite-plasma	0.0762(24.1)		0.0714[0.051-0.11]	
Metabolite-brain	0.014(18)		0.0133[0.0095-0.017]	

Table 2.1. **Parameter estimates of final population pharmacokinetic model.**

Clo (Clozapine): CL<sub>clo</sub> /F=clearance of clozapine; V<sub>clo-p</sub> /F=volume of distribution of clozapine in plasma; Ka<sub>clo</sub> =absorption rate of clozapine; Q<sub>clo</sub> /F=intercompartmental clearance of clozapine; V<sub>clo-b</sub> /F= volume of distribution of clozapine in brain; Ktr1= transit rate constant of clozapine.

Met (Norclozapine):  $CL_{clo-met}/F$ =clearance of clozapine to norclozapine;  $CL_{met}/F$ =clearance of norclozapine;  $V_{met-p}/F$ =volume of distribution of norclozapine in plasma;  $K_{a_{met}}$ =absorption rate of norclozapine;  $Q_{met}/F$ =intercompartmental clearance of norclozapine;  $V_{met-b}/F$ = volume of distribution of norclozapine in brain;  $K_{tr2}$ = transit rate constant of norclozapine.

BAV=between animal variability

RSE%= percent relative standard error

Additional file1: NONMEM control stream

...

\$SUBS ADVAN13 TOL=6

\$MODEL

COMP(DEPOT) ; ORAL DOSE OF CLOZAPINE  
COMP(CENTRAL) ; PLASMA COMPARTMENT FOR CLOZAPINE  
COMP(PER) ; BRAIN COMPARTMENT FOR CLOZAPINE  
COMP(TRANSIT1) ; TRANSIT COMPARTMENT 1 FOR CLOZAPINE  
COMP(TRANSIT2) ; TRANSIT COMPARTMENT 2 FOR CLOZAPINE  
COMP(MET) ; PLASMA COMPARTMENT FOR NORCLOZAPINE  
COMP(DEPOT2) ; ORAL DOSE OF NORCLOZAPINE  
COMP(PER2) ; BRAIN COMPARTMENT FOR NORCLOZAPINE  
COMP(TRANSIT3) ; TRANSIT COMPARTMENT 1 FOR NORCLOZAPINE

\$PK

CLclo=THETA(1)

CLclo-met=THETA(2)

Vclo-p=THETA(3)

KAclo=THETA(4)

S2=Vclo-p

Vclo-b=THETA(5)

Qclo=THETA(6)\*EXP(ETA(1))

Fclo=THETA(7)\*EXP(ETA(2))

KTR1=THETA(8)\*EXP(ETA(3))

CLmet-p=THETA(9)\*EXP(ETA(4))

KAmet=THETA(10)\*EXP(ETA(5))

Vmet-p=THETA(11)\*EXP(ETA(6))

Vmet-b=THETA(12)

Qmet=THETA(13)

KTR2=THETA(14)

Fmet=THETA(15)

A\_0(1)=0

A\_0(2)=0

A\_0(3)=0

$$A_0(4)=0$$

$$A_0(5)=0$$

$$A_0(6)=0$$

$$A_0(7)=0$$

$$A_0(8)=0$$

$$A_0(9)=0$$

-----

\$DES

-----PK ODES-----

$$DADT(1)=-KA_{clo} * A(1)$$

$$DADT(2)=KA_{clo} * A(1) - (CL_{clo-met}/V_{clo-p}) * A(2) - (CL_{clo}/V_{clo-p}) * A(2) + (Q_{clo}/V_{clo-b}) * A(3) - KTR1 * A(2)$$

$$DADT(3)=KTR1 * A(5) - (Q_{clo}/V_{clo-b}) * A(3)$$

$$DADT(4)=KTR1 * A(2) - KTR1 * A(4)$$

$$DADT(5)=KTR1 * A(4) - KTR1 * A(5)$$

$$DADT(6)=(CL_{clo-met}/V_{clo-p}) * A(2) - (CL_{met}/V_{met-p}) * A(6) + KA_{met} * A(7) + (Q_{met}/V_{met-b}) * A(8) - KTR2 * A(6)$$

$$DADT(7)=-KA_{met} * A(7)$$

$$DADT(8)=KTR2 * A(9) - (Q_{met}/V_{met-b}) * A(8)$$

$$DADT(9)=KTR2 * A(6) - KTR2 * A(9)$$

...

Figures 2.2 and 2.3 show the goodness of fit plots of the final model for the parent drug and metabolite in plasma and brain, respectively. Population and individual predictions as well as the conditional weighted residuals distribution are shown in these figures. In the case of norclozapine, imprecision in population predicted plasma concentrations was evident and is attributed to the inter-animal variability observed in the context of the limited number of animals available to support these predictions. The majority of the fixed effects were estimated with less than 40% relative standard error (Table 1). BAV was estimated for several of the structural parameters and ranged from 5% ( $K_{tr1}$ ) to 75% ( $Cl_{met}$ ). Residual variability for clozapine in plasma and brain were 10.9% and 3.7%, respectively, and residual variability of norclozapine in plasma and brain were 7.6% and 1.4%, respectively. VPC results are shown in Figures 2.4 and 2.5, the observed medians (dashed black lines) concentrations were adequately captured by the corresponding simulation based 90% predicted intervals of median concentrations for clozapine and norclozapine (shaded areas). Median, 5<sup>th</sup> and 95<sup>th</sup> percentiles of the parameters derived from the bootstrap analysis of 1000 replicates are shown in Table 2.1.



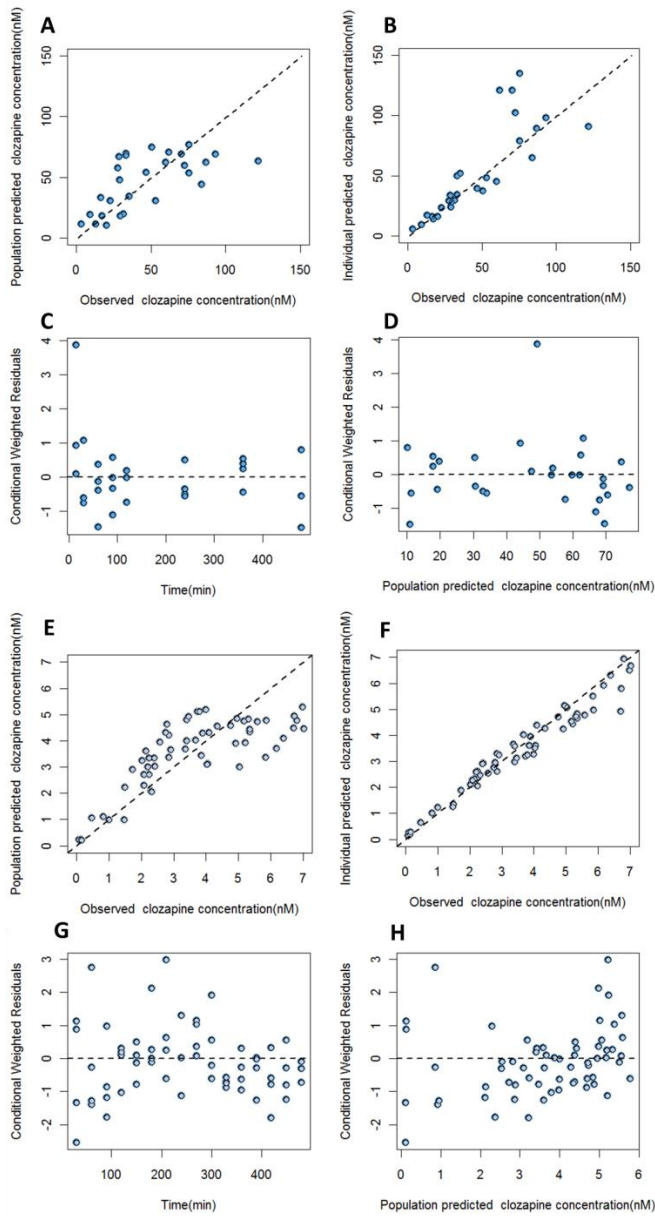


Figure 2.2. **Model diagnostic plots for clozapine.**

A and E: population prediction vs. observation plots for clozapine in plasma and brain ECF, respectively, solid line is line of identity. B and F: individual prediction vs. observation plots for clozapine in plasma and brain ECF, respectively. C and G: conditional weighted residuals vs. time for clozapine in plasma and brain ECF, respectively. D and H: conditional weighted residuals vs. population prediction for clozapine in plasma and brain ECF, respectively.

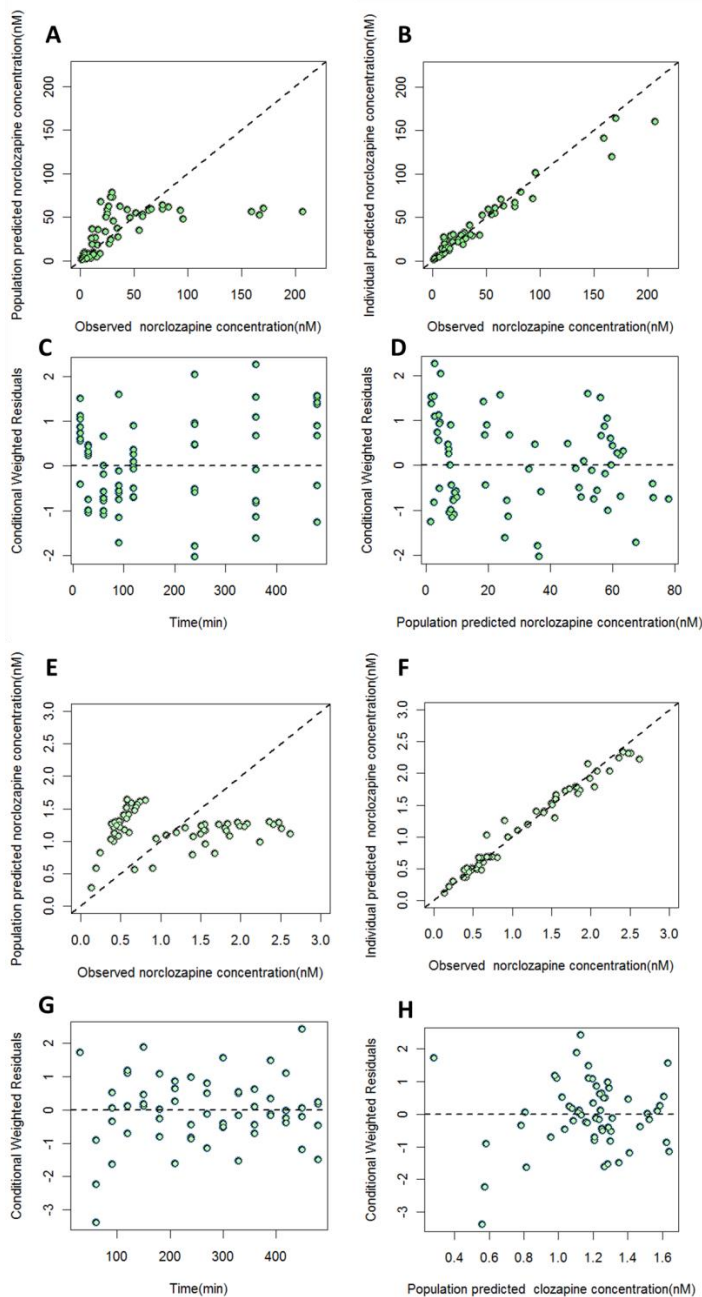


Figure 2.3. **Model diagnostic plots for norclozapine.**

A and E: population prediction vs. observation plots for norclozapine in plasma and brain ECF, respectively, dashed line is the line of identity. B and F: individual prediction vs. observation plots for norclozapine in plasma and brain ECF, respectively. C and G: conditional weighted residuals vs. time for norclozapine in plasma and brain ECF, respectively. D and H: conditional weighted residuals vs. population prediction for norclozapine in plasma and brain ECF, respectively.

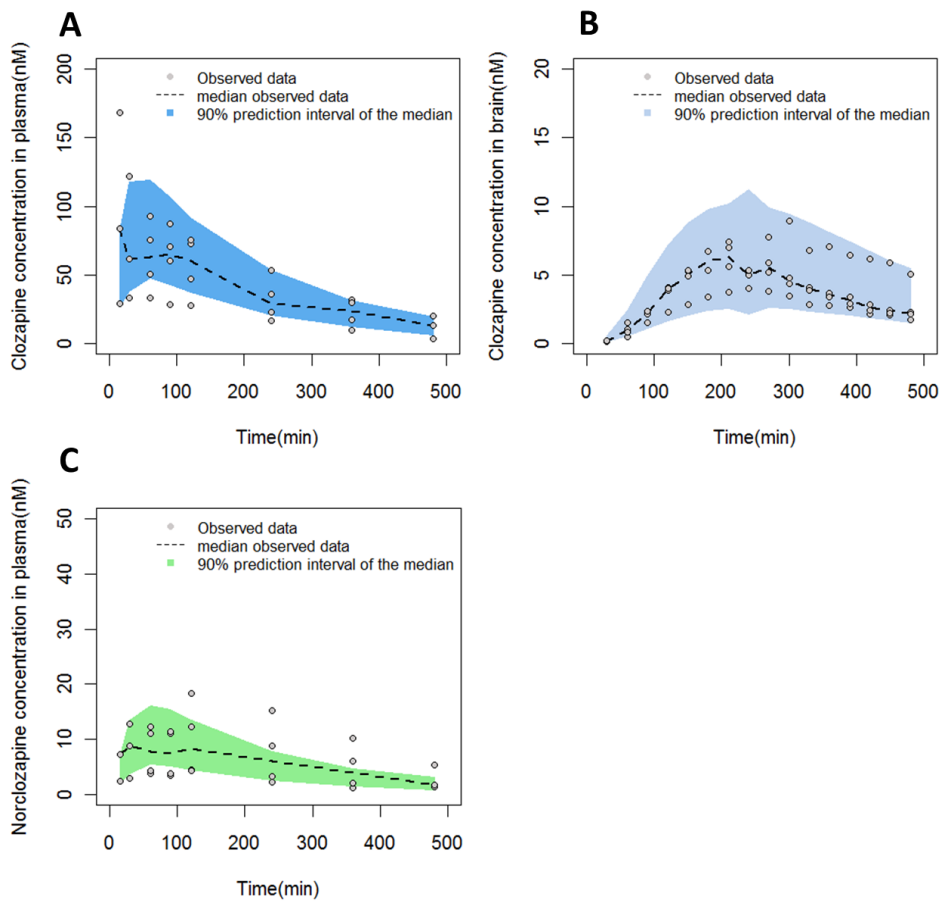


Figure 2.4. **Visual predictive checks of clozapine in plasma and brain and norclozapine in plasma.**

Visual predictive check of clozapine concentrations in plasma (A) and brain (B), and norclozapine concentrations in plasma (C) following a 10 mg/kg subcutaneous dose of clozapine. Norclozapine concentrations in brain following a 10 mg/kg subcutaneous dose of clozapine were not measurable.

Dashed line is the median of observed concentrations, the shape represents the 90% predicted interval of the median, and the dots represent the observed concentrations.

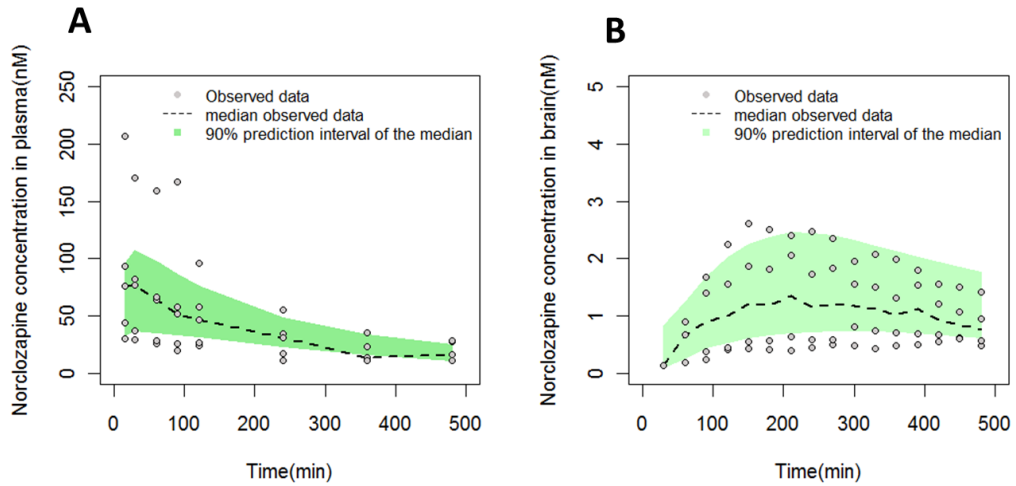


Figure 2.5. **Visual predictive checks of norclozapine in plasma and brain.**

Visual predictive check of norclozapine concentrations in plasma (A) and brain (B) following a 10 mg/kg subcutaneous dose of norclozapine.

Dashed line is the median of observed concentrations, the shape represents the 90% predicted interval of the median, and the dots represent the observed concentrations.

### 3.2 Human PK Simulation

Simulated unbound clozapine concentrations in plasma from 12 to 24 hours after administration were compared with published human data and are shown in Figures 2.6A, 2.6B and 2.6C for 3 doses (200, 300 and 400mg/day). Predicted occupancy of D2, 5-HT2A, M1,  $\alpha$ 1,  $\alpha$ 2 and H1 receptors were calculated using simulated unbound clozapine and norclozapine (D2 only) brain concentrations for the three dose levels. For clozapine, the predicted median percentage of receptor occupancy of D2 ranged from 6-42%, 9-52% and 11-59% for the 200, 300 and 400mg daily doses, respectively, across the inter-dose time interval. The median percentage of 5-HT2A receptor occupancy decreased from 93% to 52%, 95% to 62% and 96% to 69% from 6 to 24 hours after 200,300 and 400mg daily doses, respectively. For M1,  $\alpha$ 1 and H1 receptors, occupancies ranged from 74 to 99% across the dosage interval. In addition, the median percentage occupancy of  $\alpha$ 2 receptors was predicted to be in the range of 3-40% across the dosage interval. For norclozapine, the predicted median percentage of receptor occupancy of D2 ranged from 1.1-17.3% across the dosage interval. Receptor occupancy results across the dosage interval are summarized in Figure 2.7.

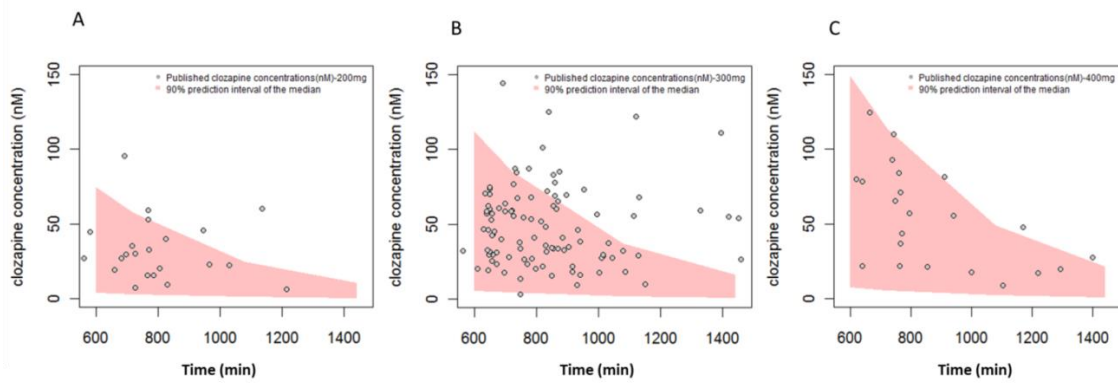
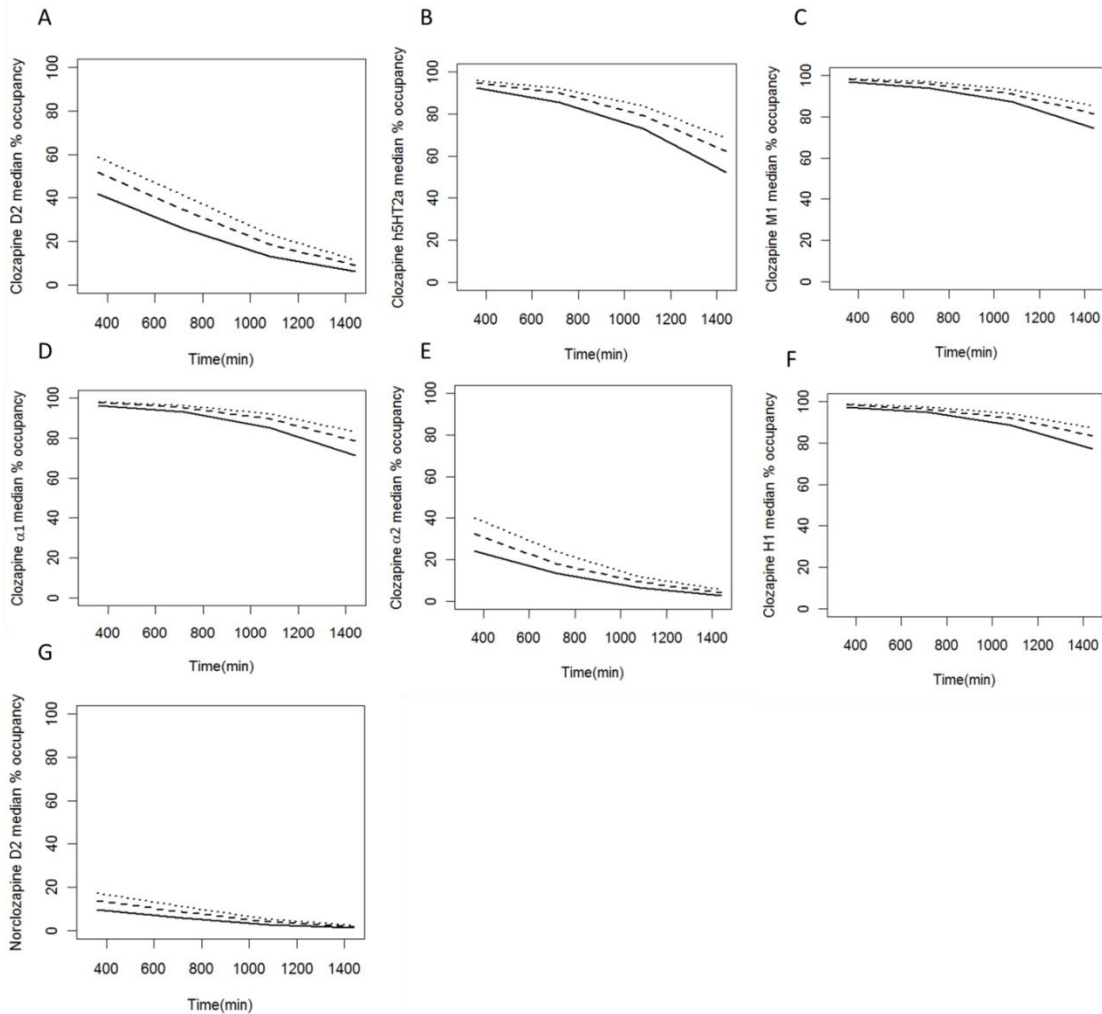


Figure 2.6. **Simulated clozapine human unbound concentrations vs. published human concentrations at steady state in brain with 200,300 and 400 mg OID from 12 to 24 hours.**

The shape represents the 90% predicted interval of the median and the dots represent the observed data.



**Figure 2.7. The predicted median percentage of receptor occupancy of D2, 5-HT2A, M1,  $\alpha_1$ ,  $\alpha_2$  and H1 for clozapine and D2 for norclozapine.**

The predicted median percentage of receptor occupancy of D2, 5-HT2A, M1,  $\alpha_1$ ,  $\alpha_2$  and H1 for clozapine were shown in A to F, respectively, and predicted median percentage of receptor occupancy of D2 for norclozapine was shown in G between 6 to 24 hour after dose. The solid line, dash line and dot line represent the predicted median percentage of receptor occupancy following 200, 300 and 400 mg daily doses, respectively.

#### 4. Discussion

The model presented represents a unique PK model developed from directly measured concentrations of clozapine and norclozapine in rat plasma and brain ECF. A multiple transit compartment model was used to account for a delay in the transport of clozapine and norclozapine from plasma to brain across the rat blood brain barrier. Some evidence suggests that Pgp may be involved in the process of clozapine transport (Cremers et al., 2012; Doran et al., 2005) across the blood brain barrier. The need to incorporate transit compartments in the present model is consistent with a Pgp role in clozapine transport across this barrier. Using an animal model, drug exposure can be measured by microdialysis at the target site. Based on a previously published non-compartmental analysis, the ratio of AUC between parent and metabolite in the rat indicated that only about 10% of parent drug was eliminated through metabolism (Olsen et al., 2008). This is consistent with the ratio of norclozapine to clozapine clearance ( $Cl_{clo-met}$  is 9.91% of  $Cl_{clo}$ .) As the results revealed, even with relatively rich sampling profiles, the uncertainty of some parameters, in particular of between-animal variability, was large likely because of the small number of animals in this study.

Simulated human plasma concentrations were based on previously published human plasma concentration data (Ismail et al., 2012). The unbound plasma concentrations at steady state after a range of doses overlapped with published data corrected for the unbound fraction of clozapine (3%) (Clozapine Product Insert, 2013). Subsequently, plasma exposures were linked to the plasma–brain structural PK parameters, using allometric scaling, that described clozapine and norclozapine transport between plasma and brain in the rat to ultimately predict human brain ECF exposure. This population



pharmacokinetic approach, based on a transit compartmental approach as opposed to explicit assumption of a Pgp role and its associated interspecies scaling, enabled translational representation of the system across species to predict human brain ECF concentrations. As an atypical antipsychotic drug, clozapine targets D2 receptors as well as acts as an agonist or antagonist at several other receptors found in the CNS. In order to get a more complete profile of PK-PD linkage, percentage receptor occupancy of each receptor was calculated from 6 to 24 hour after three dosage levels. Our results show that the median percentage D2 receptor occupancy was in a range of 42% to 59% 6 hours after administration of a daily dose of 200-400 mg. This range is congruent with the 33% to 67% range reported by Nordström et al (Nordström et al., 1995), and agrees with the widely recognized understanding of low D2 receptor occupancy of therapeutic doses of clozapine relative to those obtained with therapeutic doses of other antipsychotics (first and second generation). In addition to D2, percent 5-HT<sub>2a</sub> receptor occupancy also overlapped with the results of Nordström et al.

The proposed PK model thus demonstrated the ability to extrapolate human systemic exposure to predict clozapine brain concentrations and associated receptor occupancy profiles in humans at clinically relevant doses. In addition, the model simultaneously captured parent and metabolite in the system, which is relevant since norclozapine also has activity at multiple receptors (Bishara et al., 2008). However, the model can be improved in the precision of the PK parameter estimates by increasing the sample size. With this limitation taken into consideration, the model framework reported shows promise in predicting clozapine receptor occupancy at multiple receptors in human CNS, which can then be probed as a correlate to response and/or toxicity.

**CHAPTER III: Population Pharmacokinetic Modeling of Sertraline Treatment in Alzheimer's disease Patients: The DIADS-2 Study (Published in The Journal of Clinical Pharmacology) (This material is reproduced with permission of John Wiley & Sons, Inc.)**

*1. Introduction*

Alzheimer's disease (AD) is a neurodegenerative disease associated with a number of neuropsychiatric symptoms (NPS). One commonly found NPS is depression, affecting as many as 60% of AD patients (Steinberg et al., 2008). The antidepressant sertraline, a selective serotonin reuptake inhibitor (SSRI), has been used for the treatment of depression in AD patients (Martin et al., 2006). It is the second most potent inhibitor of serotonin reuptake (Hiemake et al., 2000). In a study of 117 randomized controlled trials from 1991 to 2007, sertraline was proposed as the best first line treatment for moderate to severe depression in adults based on an overall evaluation of benefits, acceptability and other factors (Cipriani et al., 2009). Sertraline is orally administered with high plasma protein binding affinity (Owen et al., 1997). The average elimination half-life of sertraline is approximately 26 hours and the peak plasma concentration ( $C_{max}$ ) is reached at 6-8 hours (Warrington et al., 1991). Sertraline is mainly eliminated by hepatic metabolism to its major metabolite, N-desmethylsertraline, by multiple cytochrome p450 enzymes including CYP2B6, CYP2D6, CYP2C9, CYP2C19 and CYP3A4 (Kobayashi et al., 1999). This metabolite has 5 -10% of sertraline's serotonin reuptake inhibitor potency; thus its clinical effect on sertraline response is negligible (Sprouse et al., 1996).

The pharmacokinetic profile of sertraline has been broadly explored in previous clinical studies where patient ages spanned broad ranges (Muijsers et al., 2002; Schneider et al., 2003; Axelson et al., 2002). In a pharmacokinetic study of 16 elderly ( $\geq 65$  years of

age) patients treated with 100 mg sertraline once daily for 14 days, plasma sertraline clearance was approximately 40% lower compared to similarly studied younger ( 25-32 years ) patients (Warrington et al., 1991; Zoloft® product information New York: Roeg, 2001). A comparable result was found in a 21 day study (n=44), with the elimination rate constant (0.019 / hr) in elderly individuals 16 to 63% lower than that observed in young adults (Ronfeld et al., 1997).

In the elderly, sertraline's effectiveness is comparable to the SSRI fluoxetine as well as the tricyclic antidepressants (TCAs) nortriptyline, amitriptyline and imipramine. It has lower rates of adverse side effects than the TCAs (Muijsers et al., 2002). Although many studies have examined the sertraline pharmacokinetic profile in elderly subjects, non-compartmental methods were employed that have limitations in assessing sources of inter-individual variability in sertraline concentration. In fact, this is the first population pharmacokinetic (PPK) study focusing on AD patients with depression. While this analysis was based on data from a null clinical study, it provided an opportunity to capture the pharmacokinetic characteristics in elderly individuals of sertraline. In these analyses, we aim to gain insights relating to inter-individual variability in the pharmacokinetics of sertraline in AD patients. The objective is to identify covariates that contribute to variability in sertraline concentration by performing a PPK analysis of sertraline in elderly patients with AD and generating PPK parameters for this population.

## *2. Materials and Methods*

### **2.1 Participants and Study design**

The design of the multicenter Depression in Alzheimer's Study-2 (DIADS-2) has been described in detail elsewhere (Martin et al., 2006; Rosenberg et al., 2010; Weintraub

et al., 2010; Drye et al., 2011). Briefly, DIADS-2 enrolled 131 AD patients with mild-to-moderate AD. Patients were randomized in a, 12-week, double-blind, placebo-controlled (n=64) antidepressant trial of sertraline (n=67; range: 25-125 mg per day). An initial treatment regimen of sertraline 50 mg QD or identical-appearing placebo was prescribed. The dosage of sertraline in the active treatment arm was increased to 100 mg QD after one week. The daily dose was adjusted depending on the response and tolerability of the treatment in the first four weeks post randomization. Single concentration samples of sertraline were collected in individual patients at weeks 4 and 12. The time of last dose and exact time of collection were available for each of these samples.

The study was approved by Institutional Review Boards at all five study sites and the coordinating center: Johns Hopkins School of Medicine, Baltimore, MD; University of Pennsylvania School of Medicine, Philadelphia, PA; Medical University of South Carolina, Charleston, NC; University of Rochester School of Medicine, Rochester, NY; University of Southern California Keck School of Medicine, Los Angeles, CA; Johns Hopkins Bloomberg School of Public Health. In addition, the PPK analysis was approved by the institutional Review Board of Indiana University School of Medicine, Indianapolis, IN.

## **2.2 Analytical Procedure to Measure Sertraline Concentration**

Plasma sertraline concentration was determined using high-performance liquid chromatography (HPLC). The extraction of plasma was done in the mobile phase at a 60:40 ratio of 0.025 M potassium phosphate buffer, pH 2.5, and acetonitrile with a flow rate of 1 mL/min. The separation of plasma occurred in a Knauer nucleosil, C18, 100 angstrom, 150 mm x 4.6 mm column with a Supelco pelliguard LC-18, 2 cm X 4.6 mm

pre-column cartridge. Ninety microliters of n-octylamine was added as a modifier to the mobile phase and degassed. The sertraline and D-sertraline assays were linearized by an internal standard, 200 ng/mL of clomipramine, in the range of 10 to 600 ng/mL. The intra-assay and inter-assay variability in the coefficient of variation (CV) ranged from 6.6% to 9.4% and ranged from 2.8% to 6.3%, respectively. With 1 mL of plasma, recovery for sertraline and metabolite ranged from 87% to 93% with a lower limit of quantification (LLOQ) of 10ng/mL. The analysis was carried out using a Turbo Chrom data system.

### **2.3 Population Pharmacokinetic Model Development**

The population pharmacokinetics of sertraline were analyzed using nonlinear mixed effect modeling software, NONMEM, Version VII (GloboMax\_LLC, Ellicott City, MD, USA) using Wings for NONMEM, Version 7 (Holford,2012). The initial model development focused on a base model structure on PK parameter assessment. One and two compartment models with first order absorption and elimination were evaluated using subroutine ADVAN2 TRANS2 and ADVAN4 TRANS4, respectively. A likelihood based approach (Method 3) was used to handle measurements below the quantitation limit (BQL) at 10 ng/mL (Ahn et al., 2008).

PPK analyses used the first-order conditional estimation (FOCE) LAPLACIAN method. Inter-individual variability (IIV) for PK parameters was assumed to be log-normally distributed and evaluated using an exponential model  $P_i = P_{TV} \times e^{\eta_p}$  where  $P_i$  is the parameter estimate for the *i*th individual, and  $P_{TV}$  is the typical value for the parameter at the population level. The variability between *i*th individual and population parameter values was described by  $\eta_p$  which was identically distributed with a mean of 0 and a

variance of  $\omega_{\eta}^2$  (Feng et al., 2006). In addition to the IIV, intraindividual variability, system noise, experimental error and/or model misspecifications was described by a residual error model. The residual error models evaluated were: additive ( $y_{ij} = \hat{y}_{ij} + \varepsilon_{ij}$ ); proportional ( $y_{ij} = \hat{y}_{ij}(1 + \varepsilon_{ij})$ ); and combined ( $y_{ij} = \hat{y}_{ij}(1 + \varepsilon_{ij}) + \varepsilon_{ij}'$ ); where  $y_{ij}$  and  $\hat{y}_{ij}$  represents the jth observed sertraline concentration, and its corresponding model predicted concentration with the difference described by  $\varepsilon_{ij}$  or  $\varepsilon_{ij}'$ .  $\varepsilon_{ij}$  was assumed to be normally distributed with a mean of 0 and a variance of  $\sigma^2$ . The absorption rate constant (Ka) was fixed to 0.5 based on the literature values of  $t_{max}$  and an elimination constant (Ronfeld et al., 1997). This was done because the estimation of Ka in this dataset resulted in unstable model runs.

To evaluate the inter-individual variability estimated by the nonlinear mixed effects modeling approach, patients' demographic characteristic (weight, height, age, sex, race and study site) were evaluated to see if these explained this variability. These factors were assessed independently in a step-wise forward addition approach. Covariates such as weight, height, age, sex and race were included to examine potential physiologic differences that could contribute to difference in drug elimination rate or distribution volume across this population. Given the possibility of differences in adherence to the protocol by either subjects or the study site in sample collections, etc, clinical site was also tested as a covariate.

For continuous covariates, the effects of the covariates on PK parameter estimates were tested in the following model structures:

$$(1) P_{TV} = \theta_1 + \theta_2 * Cov$$

$$(2) P_{TV} = \theta_1 + \theta_2 * (Cov - Med_{cov})$$

$$(3) P_{TV} = \theta_1 * (\text{Cov} / \text{Med}_{\text{cov}})^{\theta_2}$$

$$(4) P_{TV} = \theta_1 * \exp[\theta_2 * (\text{Cov} / \text{Med}_{\text{cov}})]$$

where  $P_{TV}$  is the typical population estimate of a particular PK parameter, and  $\theta_1$  and  $\theta_2$  are fixed effect estimates for a corresponding covariate, Cov normalized by the median value of the covariate,  $\text{Med}_{\text{cov}}$ . Missing data were found in both weight and height covariates. A naïve substitution approach (Yang et al) was carried out to simply replace the missing value with the median value based on sex.

$$(5) \text{ IF (WT=0 and SEX=1) THEN WT= Med}_{\text{WT,Sex=Male}}$$

$$\text{ IF (WT=0 and SEX=2) THEN WT= Med}_{\text{WT,Sex=Female}}$$

A common allometric function of scaling the PK parameters to the 0.75 power of body weight was also tested after the missing weight values were replaced.

$$(6) P_{TV} = \theta_1 * (\text{WT} / \text{Med}_{\text{WT}})^{0.75}$$

Categorical variables such as sex, race and site were tested in the following model structure. Each category was evaluated in a separate fashion.

$$(7) \text{ IF (Cov.EQ.1) THEN } P_{TV} = \theta_1$$

$$\text{ Else } P_{TV} = \theta_2$$

For example, each ethnic group was divided into a category, African American=1, White=4 and Hispanic/Latinos=5. Each ethnic group was examined separately from other races. In addition, groupings of racial categories were affected when the individual race effects were not uniquely identifiable. In this case,  $\theta_1$  and  $\theta_2$  are the population PK parameter estimates for groupings of race that were uniquely identifiable, in this case African American and all others.

Model evaluation was based on a likelihood ratio test using the objective function value (OFV) from NONMEM. The change in the OFV returned by NONMEM is approximately equal to  $-2 \times \log$  likelihood. The difference in  $-2 \times \log$  likelihood between two models that are nested follows a  $\chi^2$  distribution. Covariates were added to the model in a step-wise addition fashion and remained in the final model if the OFV decreased by greater than 3.84 ( $p\text{-value} \leq 0.05$ ,  $df=1$ ). The final model was further examined using goodness-of-fit plots generating using R (Version 2.13) based on the conditional weighted residuals distribution and the predicted versus observed sertraline concentrations at both the population and individual levels.

### 3. Results

#### 3.1 Patient Characteristics

A total of 131 participants entered the trial with 67 randomized to sertraline and 64 to placebo. Only the concentration measurements taken from patients in the active treatment arm ( $n=67$ ) were utilized for this analysis. An average of 1.7 sertraline concentration measurements per individual was available, and 5 of the measurements were found below the LLOQ. Seventeen individuals were removed from the analysis. Specifically, 16 individuals lacked sertraline concentration information at both weeks 4 and 12, and one individual had missing dosage information for both occasions. In the remaining 50 individuals, other single observations were removed as follows: 14 individuals only had a single sertraline concentration measurement from one of the two visits (non-measured visit removed). Of these 14 observations, 11 were missing a concentration measurement, 1 was missing a dosage time associated with a concentration sample, and 2 were missing dosage amount information associated with that



concentration sample. The PPK analysis was conducted using the remaining 85 PK observations from 50 individuals.

As shown in Table 3.1, the median age of the patients was 75 (range: 53-89). There were 20 males and 30 females in the analysis broken down as follows: 18% African American; 72% White; and 10% Hispanic. This analysis included 4 patients with missing weight information and 9 with missing height information. The median values of weight and height without considering the missing values were 147 lb (range: 107-245) and 64 in (range: 57-72), respectively.

**Table 3.1. Alzheimer’s disease patient demographics**

	N (%)	Mean (SD)	Median(range)
Sample size	50		
Number of observations	85		
Gender			
Male	20 (40)		
Female	30(60)		
Race			
Black/African American	9 (18)		
White	36 (72)		
Hispanic /Latino	5 (10)		
Clinical sites			
A	15 (30)		
B	11 (22)		
C	6 (12)		
D	8(16)		
E	10(20)		
Baseline Age (years)		75 (7.76)	76 (53-89)
Baseline Weight (lb) Without 4 missing values		159 (37.08)	147 (107-245)
Baseline Height (In) Without 9 missing values		64 (4.45)	64 (57-72)
Sertraline dose administered (mg)		92.47 (18.62)	100 (25-100)
Sertraline concentrations (ng mL <sup>-1</sup> ) Without 5 BQL values		62.94 (47.62)	49 (9-229)

### 3.2 Population Pharmacokinetic Modeling

A one-compartment model with first order absorption and elimination and an additive residual error model best described the sertraline data. The population parameter estimates of CL/F and Vd/F in the base model were 83.1 L/h and 6,620 L, respectively. Inter-individual variability (IIV) was estimated only for CL/F because a significant correlation was found between CL/F and Vd/F. Patients at site C has CL/F approximately 49% lower than that seen in patients other at other clinical sites ( $\chi^2 = 5.576$  df=1,  $p < 0.05$ ). The final covariate model was implemented using the categorical covariate model structure described in the methods section (equation 7). No other significant covariate relationships were found for CL/F or Vd/F. The population PK parameter estimates and goodness-of-fit plots for the final model are listed in Table 3.2 and Figure 3.1, respectively.

Table 3.2. Population pharmacokinetic parameter estimates of sertraline in the final model

Parameters	Population estimate (% SE)	Inter-individual variability (% SE)
CL/F <sub>site=C</sub> , L/h	43.8(34)	59.33% (31.3)
CL/F <sub>site=others</sub> , L/h	89.1(12.2)	59.33% (31.3)
Vd/F, L	6470(70.5)	
Ka, 1/h (fixed)	0.5	
Residual variability	19.6 ng/mL (11.6)	

CL/F<sub>site=C</sub> ; clearance adjusted for bioavailability from site C; CL/F<sub>site=others</sub>; clearance adjusted for bioavailability from other clinical sites; Vd/F, Volume of distribution adjusted for bioavailability; Ka, rate of absorption; SE, standard error

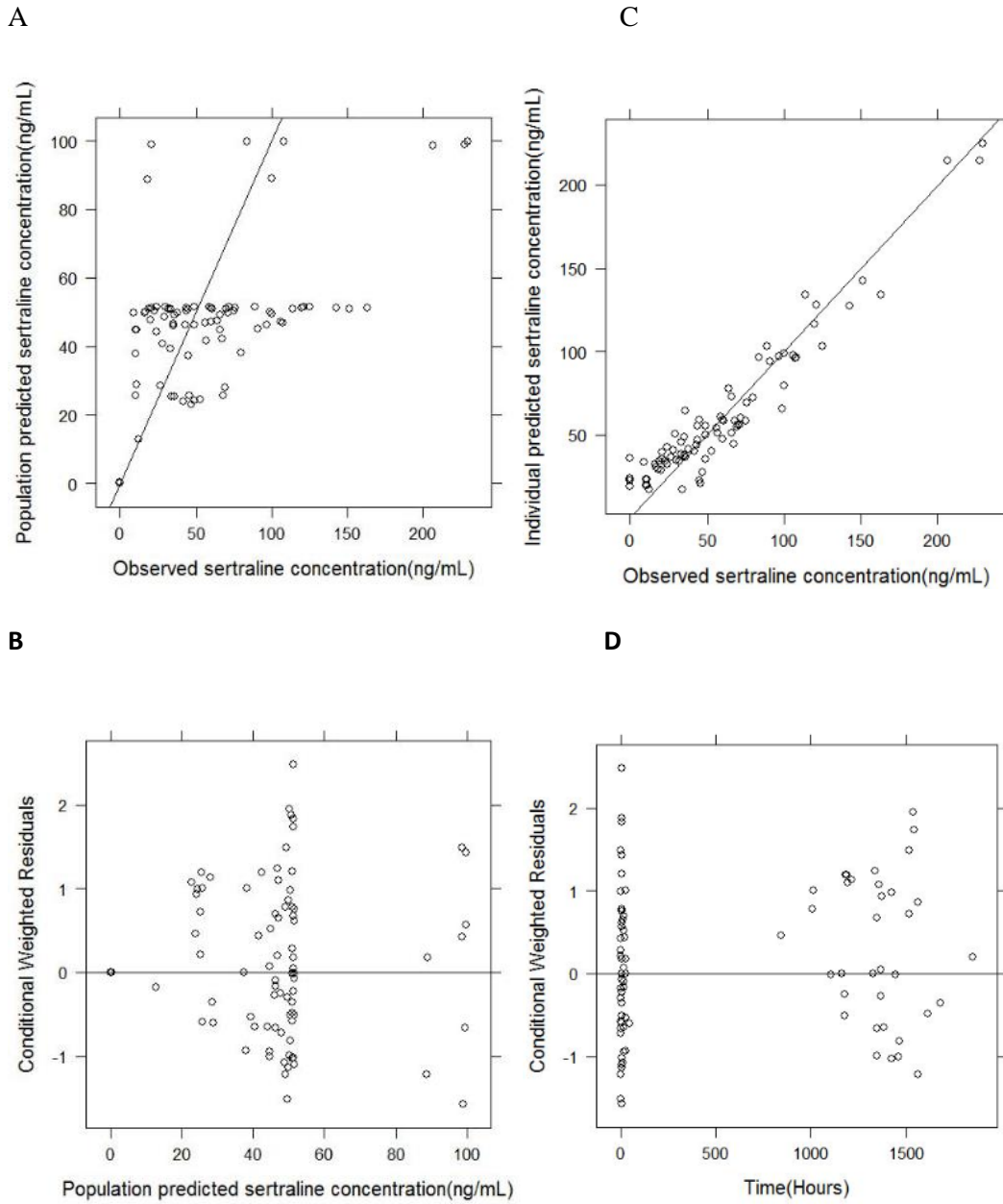


Figure 3.1. **Goodness-of-fit plots for the final sertraline model.**

Plot of population and individual predicted versus observed sertraline concentrations and plot of conditional weighted residuals versus population predicted concentrations and time (hours).

#### 4. Discussion

Nineteen percent of the variance in pharmacokinetic distribution of sertraline in these depressed AD patients was accounted for using a 1-compartment model. The population mean CL/F and Vd/F were 83.1 L/h and 6,620 L, respectively. As most patients in DIADS-2 were elderly (mean age of 75 years; range: 53-89), the mean elimination rate constant of 0.013 1/h is consistent with other literature reported values (Ronfeld et al., 1997). Compared to younger patients, the elderly have a longer sertraline half-life, but we found no difference between elderly healthy volunteers in the literature and these AD patients.

In our analysis, CL/F estimates for the four patients with sertraline observations below the quantification limits were in the top 24% of all CL/F estimates. Indeed, one patient with both observations below the quantification limit had a CL/F value approximately three times higher than the typical population value. The cause of these high CL/F values is not clear. Many possible factors might be considered, such as fasting, poor adherence, or genetic variance.

We also examined the effect of sex, race, age, weight, height and site on the variability of the PK parameter estimates. Previous publications have suggested that the average half-life is 1.5 times longer in women than in men (Ronfeld et al., 1997). We expected to detect an effect of gender on sertraline PK parameters; however, inclusion of sex as a covariate in our model yielded a statistically non-significant association. Possible differences between males and females were likely undetectable because of the small study population. In addition, a commonly reported correlation (Mahmood et al., 2007) between plasma clearance and body weight was not detected in this analysis.

Unexpectedly, the only covariate examined that affected PK variability was related to a single clinical site. Patients at site C had much lower CL/F than other sites. This might be explained by the 3 high sertraline concentrations found in 2 patients out of a total of 9 observations from site C. Nineteen percent of the total variance in inter-individual on CL/F was explained by incorporating the site as a covariate in this analysis. The model CL/F estimates we present that exclude patients at the site C are likely more generalizable to the typical AD population. In conclusion, clinical site was a significant covariate contributing to the CL/F change. The clinical implications for subjects at the outlying site would be important if this represented a true bias in subject selection. In this case diminished clearance could lead to a greater risk of adverse effects such as dizziness, extrapyramidal effects and hyponatremia. Nonetheless, the subjects at this site were found not to differ in demographic nor clinical attributes from the other four sites. This, as well as the failure to detect differing incidences of side effects, leads to the more likely explanation, that there may have been variations in study procedures such as sampling. Site variability in trial procedures is critical to the validity of multi-site trials, and population pharmacokinetics may prove helpful in this assessment. If drug administration and sample collection procedures were found not to be atypical than this may suggest meaningful subject differences in the patient sample at the outlying site e.g., drug metabolism or body size as we have found previously (Jin et al., 2010). Therefore, the dosage regimen and administration should be closely monitored in a multicenter study in order to avoid unnecessary exposures or incomplete treatments.

## **CHAPTER IV: *Data quality constrains utility of computational modeling of tumor burden in non-small cell lung cancer clinical trials***

### **1. *Introduction***

Cancer drug development would be more effective if new therapeutics could be evaluated with readily available technologies, in fewer patients, observed on treatment for shorter periods of time. One set of strategies to achieve these goals has been computational modeling of the longitudinal growth of non-small cell lung cancer (NSCLC) in populations of patients and *in silico* simulation of clinical trials. (Claret et al., 2012; Houk et al., 2009; Wang et al., 2009; Fridlyand et al., 2011; Moertel et al., 1976). The ultimate goal of these efforts is to improve the efficiency of cancer drug clinical development (Barrett et al., 2007; Bruno et al., 2009; Bruno et al., 2010).

In NSCLC, clinical trial simulations have employed a “two-step” joint model (Ibrahim et al., 2010) of overall survival in metastatic disease that was based on a longitudinal tumor growth model. These models were developed with data from nearly 3,400 patients submitted to the FDA. From 4 randomized phase III clinical trials for regulatory approval of bevacizumab, docetaxel, erlotinib, and pemetrexed (Wang et al., 2009) the longitudinal growth model was derived from the sum of the longest dimension measurements of tumors by computed tomography (CT) imaging as recorded in study case report forms. Interpolations of the change in tumor size from baseline to 8 weeks of treatment (the tumor size ratio or TS) proved an important predictor of overall survival. Modeling and simulation with these data could be an efficient means to support decision making at the Phase 2 to Phase 3 transition in drug development (Claret et al., 2012).



Another potential benefit of quantitative analysis of NSCLC tumor burden would be to redesign phase II trials to randomize fewer patients and have shorter observation periods than required for determining progression-free survival (Adjei et al., 2009; Dhani et al., 2009; Maitland et al., 2011; Maitland et al., 2011; Mandrekar et al., 2010; Stein et al., 2011; Yap et al., 2010). A simple strategy in NSCLC would be to measure the median change in tumor size at 8 weeks for randomly-assigned treatments among the study arms (Bruno et al., 2009; Lavin et al., 1981; Karrison et al., 2007). A more complex strategy that might ultimately require fewer resources would be to collect all tumor measurements from all patients over time and have the cumulative data from patients enrolled early in the trial continuously inform a calculated parameter of drug effect. These strategies have had limited testing and require validation. For example, in studies of colorectal cancer therapy and survival outcomes some have found advantages to continuous tumor measurement metrics while others have not (An et al., 2011; Kaiser et al., 2013; Claret et al., 2013).

The purpose of this study was to assess and refine the published FDA longitudinal tumor size model for NSCLC using archived tumor measurement data so that modeling and simulation might lead to smaller, quicker early phase trials for testing new treatments for NSCLC. We obtained archived case report forms from three randomized clinical trials by the Cancer and Leukemia Group B (CALGB) sponsored by the U.S. National Cancer Institute (NCI). We intended to evaluate the power of smaller clinical trials with novel endpoints to detect evidence of anti-cancer drug treatment effects. Instead, we found elements of 3 data sets from multi-center clinical trials that could bias comparisons between continuous measurement and categorical strategies for improving treatment

evaluations. These findings are likely to be common to historical and current solid tumor trial data sets. Here we identify variance in measurement and recording of CT imaging assessments of tumor burden as a modifiable factor that constrains the successful development and validation of novel tumor growth assessment metrics.

## *2. Materials and Methods*

### **2.1 Patients**

Archived case report forms were available from 857 patients enrolled in 3 NCI-supported studies conducted by CALGB listed in Table 4.1. These were front-line trials in metastatic NSCLC: CALGB 9730 (Lilenbaum et al., 2005) was a phase III randomized trial that compared single-agent paclitaxel with combination carboplatin/paclitaxel, CALGB 30203 (Edelman et al., 2008) was a randomized phase II trial that evaluated eicosanoid modulation and examined cyclooxygenase-2 expression as a positive predictive factor for the inclusion of celecoxib in the standard first-line cytotoxic therapy regimens, and CALGB 30303 (Miller et al., 2008) was a phase II randomized study of dose-dense docetaxel and cisplatin administered every two weeks with pegfilgrastim and darbopoetin-alfa with or without the chemoprotectant BNP7787. The inclusion and exclusion criteria of the trials were previously published (Edelman et al., 2008; Lilenbaum et al., 2005; Miller et al., 2008).

**Table 4.1. Three U.S. National Cancer Institute (NCI)-sponsored studies conducted by the Cancer and Leukemia Group B CALGB**

<b>CALGB study</b>	<b>Treatment</b>	<b># of patients enrolled</b>	<b># of patients treated &amp; eligible</b>	<b>Dates of accrual</b>
<b>9730</b>	<b>paclitaxel vs. paclitaxel/carboplatin</b>	<b>561</b>	<b>561</b>	<b>10/1997 – 12/2000</b>
<b>30203</b>	<b>carboplatin/gemcitabine + zileuton/celecoxib/both</b>	<b>140</b>	<b>134</b>	<b>12/2003 – 9/2004</b>
<b>30303</b>	<b>docetaxel/cisplatin +/- BNP7787</b>	<b>160</b>	<b>151</b>	<b>8/2004 – 3/2006</b>

## 2.2 Original Clinical Trial Data Collection

Data relevant to original reporting of the clinical trial results were captured on case report forms and entered into the CALGB digital databases for each of the clinical trials. The coded, patient-level data were stored at the Core Statistical Facility for CALGB (Durham, NC, USA). Treatment response assessments were conducted according to the study protocols. The CALGB 9730 trial incorporated standard World Health Organization response criteria (Miller et al., 1981) based on imaging studies conducted every 2 cycles (6 weeks). Bi-dimensional measurements of radiographically identified-lesions were typically hand-written into the same data-field on a standard spreadsheet at each evaluation time-point. Treatment was discontinued: when it became intolerable; when new lesions were identified; when the product of two perpendicular diameters of any measured lesion increased by at least 25%; or when 6 cycles of therapy were completed. Patients were subsequently followed every 3 months for 2 years. For CALGB studies 30203 and 30303, the Response Evaluation Criteria in Solid Tumors was employed, and categorical responses were based on the sum of the longest unidimensional measurements of criteria-defined “target lesions” (Therasse et al., 2000). Thoracic CT imaging evaluations were conducted in all patients pre-treatment, and at 6 and 12 weeks after treatment. Patients were removed from the studies for unacceptable toxicity or progression of disease. Patients who completed all study therapy were followed at minimum every 12 weeks thereafter.

For studies 30203 and 30303, tumor measurements were recorded on the “CALGB Solid Tumor Evaluation Form” – C-660. The form designates separate rows for each target lesion and columns for each evaluation time-point. The longest unidimensional

measurement was entered for each designated target lesion at baseline in a separate row. Non-target lesions were identified as present in a separate section of the form. During the studies the original case report forms were transmitted by the sites to the CALGB Statistical Facility where the forms were reviewed and verified by the data management team and the RECIST categorical evaluations were entered into digital databases. The sum of the longest dimensions of target lesions and the target lesion measurements were not captured in the study database. After publication of study results, the case report forms and associated source document CT reports were stored off-site in an archive facility.

### **2.3 Tumor Measurement Collection**

To obtain the original target lesion measurements, the archived paper case report forms were obtained from storage, scanned, and saved as portable document format (pdf) files. Tumor measurements from the pdf files for CALGB 30203 and 30303 were manually extracted by a research assistant and entered into a tracking file and into the study databases simultaneously. The transcriptions were independently reviewed by one of the study authors (SK and CL) and inconsistencies were manually corrected. Additionally, individual patient tumor growth plots were inspected for atypical growth and response patterns. All aberrant plots were cross-verified with the original case report form pdf and any additional data entry errors captured by this review were corrected before modeling analyses were performed. This process (Figure 4.1) resulted in 103 patients with evaluable data from CALGB 30203 and 124 from CALGB 30303. A sample of case report forms from CALGB 9730 (Figure 4. 2), consistently revealed insufficient documentation of quantitative measurements of tumors to be useful for this

modeling exercise, and therefore all patients from this trial were excluded from our planned evaluation.

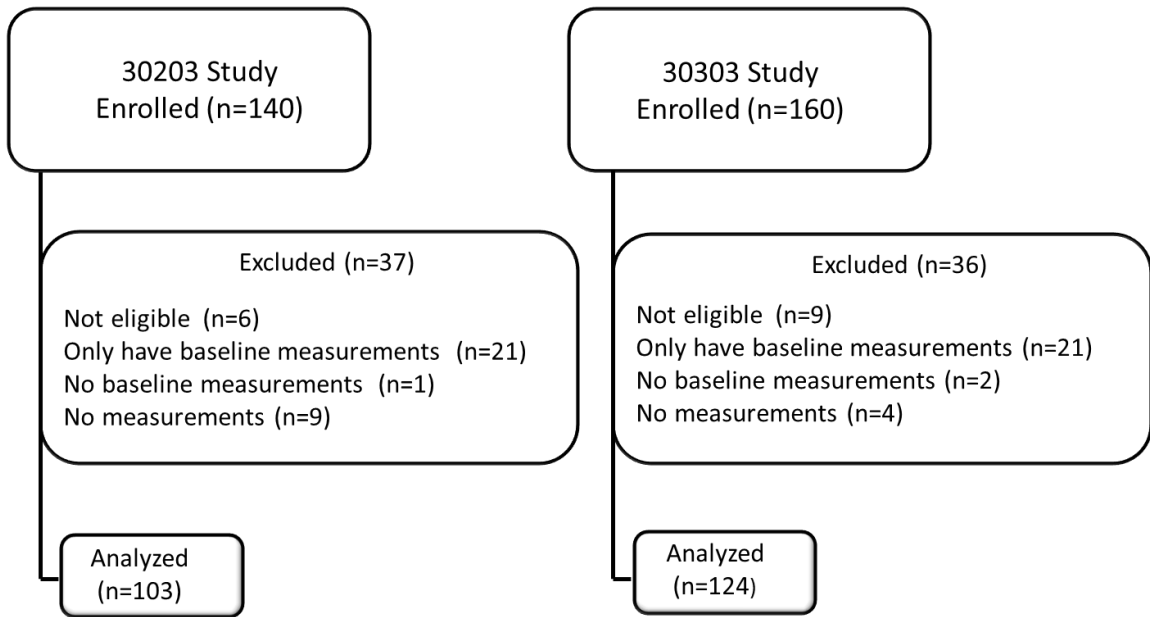


Figure 4.1. Selection of patients contributing images from CALGB 30203 and 30303.

LIST ALL MEASURABLE AND EVALUABLE PARAMETERS FOR RESPONSE.\*

Date of Observation (m/d/y)		cycle 2					13	BMC 12/24/99 (Bone Scan)
		4-27-99	6-17-99	7-27-99	9-21-99	12- <del>25</del> -99	12-6-99	
Response (CR, PR, REGression, SD, PD)		Pre-Study	SD	SD	SD	PD	PR/CR	
71	anterior RUL nodule <sup>pleural based</sup>	1x1	nc	nc	nc	nc	n/a	
2	anteromedial RML nodule	1.5x1.5	nc	nc	nc	nc	n/a	
3	LLL tissue mass	present	nc	nc	nc	nc	n/a	
4	tiny pulmonary nodules					present		
5	uptake proximal R 10th rib						present	
6								

Figure 4.2. An example of a case report form from CALGB 9730 trial.

## 2.4 Tumor Size Modeling

Longitudinal tumor size trajectories (sum of longest tumor diameter) were analyzed with nonlinear mixed effect modeling software, NONMEM, Version VII (GloboMax\_LLC, Ellicott City, MD, USA) using Wings for NONMEM, Version 7 (Holford, 2012) and the published model structure (Wang et al., 2009). This model employed a combination of a linear growth function and an exponential shrinkage function to describe the tumor change respect to baseline size (Eq.1).

$$TS_i(t) = BASE_i e^{-SR_i t} + PR_i t$$

Where  $TS_i(t)$  is the tumor size at time  $t$  for the  $i^{th}$  individual,  $Base_i$  is the baseline tumor size,  $SR_i(t)$  is the exponent tumor shrinkage rate constant, and  $PR_i(t)$  is the linear tumor growth rate constant. The exponential shrinkage function constrains the tumor size to be greater than zero (to avoid negative tumor sizes being predicted) and reflects the drug effect on tumor (whether or not there is actual tumor shrinkage). Tumor size changes were modeled using the first-order conditional estimation (FOCE) and stochastic approximation expectation maximization (SAEM) method with interaction. Between subject variability was assumed to be log-normally distributed and evaluated on baseline tumor size, tumor shrinkage rate and tumor progression rate using an exponential model  $P_i = P_{TV} x e^{\eta_p}$  where  $P_i$  is the parameter estimate for the  $i^{th}$  individual and  $P_{TV}$  is the typical value for the parameter at the population level. The variability between  $i^{th}$  individual and population parameter values was described by  $\eta_p$  which was identically distributed with a mean of 0 and a variance of  $\omega_\eta^2$  (Feng et al., 2006). Residual variability was also estimated using a proportional residual error model ( $y_{ij} = \hat{y}_{ij} (1 + \varepsilon_{ij})$ ) where  $y_{ij}$  and  $\hat{y}_{ij}$  represents the  $j^{th}$  observed tumor trajectory, and its corresponding



model predicted tumor size. The difference between observed and predicted values was described by  $\varepsilon_{ij}$  which was assumed to be normally distributed with a mean of 0 and a variance of  $\sigma^2$ .

A likelihood ratio test was applied based on the change in the objective function value (OFV) in NONMEM to evaluate the random effect on parameters. This change in OFV is approximately equal to  $-2 \times \log$  likelihood. The difference in  $-2 \times \log$  likelihood between two models that are nested follows a  $\chi^2$  distribution. An alpha level for significance of 0.05 was set a priori (3.84 OFV for 1 degree of freedom). The final model was further examined using goodness-of-fit plots generating using R (Version 2.13) based on the conditional weighted residuals distribution and the predicted versus observed tumor size measurements at both the population and individual levels. The tumor size model was developed to evaluate data from both treatment arms individually as well as simultaneously on the combined dataset.

## **2.5 Modeling Matched Cases from ECOG 4599 and CALGB 30203**

The CALGB trial based parameter estimates for the linear growth rate and the treatment-related shrinkage rate differed from those originally published from the large FDA sample. To determine whether the deviation of the parameter estimates were specific to the CALGB data collection we extracted longitudinal tumor measurement data from a clinical trial conducted by the Eastern Cooperative Oncology Group (ECOG), trial 4599. That study compared front-line therapy with carboplatin and paclitaxel with or without the addition of bevacizumab (Sandler et al., 2006). The ECOG4599 data constituted more than one fourth the total sample used to generate the FDA model. One hundred three individual cases were selected from the carboplatin, paclitaxel, and placebo

arm of ECOG 4599 by matching to the CALGB 30203 sample on patient visit time, number of lesions, and sum of the longest diameter of the target lesions at baseline.

## **2.6 Calculation of time to tumor growth**

In addition to change in tumor size at 8 weeks an alternative metric of treatment effects on serial tumor measurements, time to tumor growth (TTG) was evaluated (Claret et al., 2013) (Figure 4.3). TTG is expressed in time units (week) and depends on all the parameters of the empirical tumor model including baseline (Base, cm), tumor shrinkage rate constant (SR, 1/week) and tumor progression rate constant (PR, cm/week) (Eq 2).

$$\frac{dTSi(TTG)}{dt} = Base \cdot e^{-SR(TTG)} \cdot (-SR) + PR = 0$$

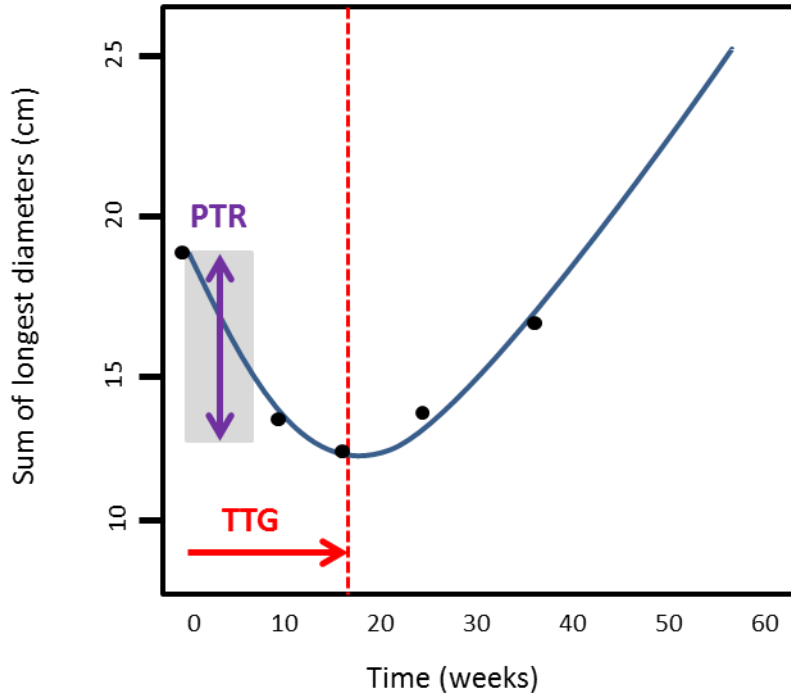


Figure 4.3. **A graphical representation of tumor matrices.**

TTG: time-to-tumor growth estimated using the FDA model; PTR: percentage tumor reduction from baseline.

## **2.7 Survival Analysis**

In order to determine the relationship between tumor size change and overall survival, the tumor size model was used to predict tumor size change at early time points (4, 6 and 8 weeks). In addition to the predicted tumor size changes at 4, 6 and 8 weeks, time to tumor growth and other clinical risk factors (Table 4.2) were tested in Cox model for overall survival using pre-selection with a stepwise significance at 0.05. For the purpose of evaluating the prediction of time to death using measures of tumor dynamics, a parametric survival was adapted to estimate this relationship. Parametric survival functions including: exponential; Weibull; and log normal were tested, and the final survival distribution selected based on likelihood ratio test and Akaike information criterion.

	<b>CALGB30203 (n=103)</b>	<b>CALGB30303(n=124)</b>
ECOG PS	0 (n=40) 1 (n=58) 2 (n=5)	0 (n=61) 1 (n=63)
SEX	Male (n=63) Female (n=40)	Male (n=82) Female (n=42)
AGE	Median=61 (Range:25-83)	Median=61 (Range:29-78)
Height	Median=171 (Range:148-200)	Median=171 (Range:150-198)
Weight	Median=75 (Range:45-138)	Median=75 (Range:43-128)
Histology	Non-small cell (n=3) Adenocarcinomas (n=50) Squamous (n=24) Undifferentiated Large Cell (n=5) Bronchoalveolar (n=4) Undifferentiated non-small cell (n=17)	Non-small cell (n=7) Adenocarcinomas (n=57) Squamous (n=29) Undifferentiated Large Cell (n=2) Bronchoalveolar (n=2) Undifferentiated non-small cell (n=27)
Ethnicity	Hispanic (n=1) Non-Hispanic (n=89) Unknown (n=13)	Hispanic (n=0) Non-Hispanic (n=115) Unknown (n=9)
Race	White (n=85) Black (n=13) Asian (n=2) Native Hawaiian or Pacific Islander (n=1) American Indian or Alaska Native (n=1) Unknown (n=1)	White (n=105) Black (n=18) Asian (n=0) Native Hawaiian or Pacific Islander (n=1) American Indian or Alaska Native (n=0) Unknown (n=0)
Prior_Chemo	No (n=84) Yes (n=17) Missing (n=2)	No (n=124) Yes (n=0) Missing (n=0)
Prior_RT	No (n=102) Yes (n=0) Missing (n=1)	No (n=109) Yes (n=15) Missing (n=0)
Disease stage	IIIB (n=9) IV (n=88) Recurrent (n=6)	IIIB (n=12) IV (n=112) Recurrent (n=0)

**Table 4.2. Demographic and clinical characteristics of patients in CALGB 30203 and 30303 trial.**

Chemo: Chemotherapy; RT: Radiation.

## **2.8 Script to “Clean CRF Data” to Match RECIST 1.1 Consistently**

When considered a conservative evaluation of longest diameter of tumor size, RECIST criteria 1.1 has been proposed to apply for trajectory selection. The tumor dynamic data was further manipulated by a script based program implemented using R (version 2.13) to extract and calculate the sum of longest diameter of the tumor size at each measurement using RECIST 1.1 criteria. The majority of tumors were measured by computed tomography scan (93.3%) and the rest of 6.62% were measured in different methods. Besides removing lesions with missing information and deleting individuals with baseline measurements only, the script based extraction was further applied to comprise the following steps: 1) exclude the lesions with inconsistent follow-up, 2) select up to a total of five lesions with a maximum of two lesions per organ and exclude lymph nodes if the maximum number of target lesions are exceeded.

R script code

```
##### import data #####

library(MIfuns)

library(lattice)

library(reshape)

d<-

read.csv(file="C:/Claire/NSCLC1/NSCLC30303TEST2.csv",skip=0,header=TRUE,sep=
",",as.is=TRUE)

names(d) <-

c("ID","OID","VISIT","TREATMENT","LESION","LOCATION","SIZE","OLOCATIO
N")

d$SIZE <- as.numeric(as.character(d$SIZE))

origi.dat <- d

write.table(d,file="C:/Claire/NSCLC1/30303TEST1.csv",
sep=" ",append=F,quote=F,col.names=T,row.names=F,na=".")

##### calculate sum of baseline tumor sizes #####

d1 <- d[d$VISIT==0,]

reapply <- function(x,INDEX,FUN,...){

  y <- tapply(x,INDEX)

  z <- tapply(x,INDEX,FUN,...)

  z[y]

}
```

```

d1 <- transform(d1,basesum=reapply(SIZE,INDEX=ID,FUN=sum))
d1 <- d1[,c("ID","basesum")]
d2 <-d1[!duplicated(d1$ID),]
d3 <- merge(d,d2,by.x="ID",by.y="ID",all.x=T)
write.table(d,file="C:/Claire/NSCLC1/30303TEST2.csv",
sep="," ,append=F,quote=F,col.names=T,row.names=F,na=".")
##### remove lesion and size which are missing #####
d <- d3
d <- d[!is.na(d$LESION) & !is.na(d$SIZE),]
head(d)
write.table(d,file="C:/Claire/NSCLC1/30303TEST3.csv",
sep="," ,append=F,quote=F,col.names=T,row.names=F,na=".")
##### determine which lesions are present for each individual at every visit #####
#####how many follow up+baseline in each patient
ni <- length(id <- unique(d$ID))
keep.lesion <- NULL
temp <- NULL
for (i in 1:ni){
  visit <- unique(d$VISIT[d$ID==id[i]])
  nvisit <- length(visit)
  s.lesion <- d$LESION[d$ID==id[i]]
  u.lesion <- unique(s.lesion)
  n.sl <- length(s.lesion)

```



```

n.ul <- length(u.lesion)

l <- data.frame(u.lesion,rep(0,n.ul))

names(l) <- c("lesion","times")

for (j in 1:n.ul) {
  for (k in 1:n.sl) {
    if (u.lesion[j]==s.lesion[k])
      l$times[j] <- l$times[j]+1
  }
}

l <- l[l$times==nvisit,]

temp <- data.frame(rep(id[i],length(l$lesion)),l$lesion,rep(nvisit,length(l$lesion))

keep.lesion <- rbind.data.frame(keep.lesion,temp)
}

names(keep.lesion) <- c("ID","LESION","NVISIT")

write.table(keep.lesion,file="C:/Claire/NSCLC1/30303TEST4.csv",
sep="," ,append=F,quote=F,col.names=T,row.names=F,na=".")

##### delete IDs with baseline visit only #####

dd <- keep.lesion

dd <- dd[dd$NVISIT >= 2,]

##### calculate number of lesions that appear at all of the visits #####

ni <- length(dd$ID)

dd$flag <- rep(0,ni)

```

```

n <- ni-1

for (i in 1:n) {

  if (dd$ID[i]==dd$ID[i+1])

    dd$flag[i] <- 1

  else

    dd$flag[i]<- 0

}

dd <- transform( dd, NLESION=reapply(flag,INDEX=ID,FUN=sum)+1)

dd$flag <- NULL

keep.lesion <- dd

##### extract data with only the lesions that appear at all of the visits

#####

a <- merge(d,keep.lesion,by=c("ID", "LESION"),all.y=T)

a <- a[order(a$ID,a$VISIT,a$LESION),]

a <-

a[,c("ID", "VISIT", "LESION", "SIZE", "basesum", "NLESION", "TREATMENT", "LOCA
TION", "NVISIT")]

all.lesions <- a

write.table(all.lesions,file="C:/Claire/NSCLC1/30303TEST5.csv",

sep=",",append=F,quote=F,col.names=T,row.names=F,na=".")

### add flag for lesions which are labeled as "LYMPH NODE" #####

flag <- rep(0,length(a$ID))

for (i in grep("LYMPH NODE",a$LOCATION)){

```

```

    flag[i] <- 1
  }
a$LYM <- flag
lym <- a[a$LYM==1,c("ID","LESION","LYM")]
lym <- unique(lym)
write.table(lym,file="C:/Claire/NSCLC1/30303TEST6.csv",
sep="," ,append=F,quote=F,col.names=T,row.names=F,na=".")
a$LYM <- NULL
a <- merge(a,lym,by=c("ID","LESION"),all.x=T)
a <- a[order(a$ID,a$VISIT,a$LESION),]
a$LYM[is.na(a$LYM)] <- 0

write.table(a,file="C:/Claire/NSCLC1/30303TEST7.csv",
sep="," ,append=F,quote=F,col.names=T,row.names=F,na=".")
##### calculate sum of lesions size at each visit for IDs with <= 3 lesions present at
each visit #####
a1 <- a[a$NLESION <=3,]
aa1 <-
aggregate(x=list(visitsum=a1$SIZE),by=list(ID=a1$ID,VISIT=a1$VISIT),FUN=sum)
aa1 <- aa1[order(aa1$ID,aa1$VISIT),]
a1 <- merge(a1,aa1,by=c("ID","VISIT"),all.x=T)
a1 <- a1[order(a1$ID,a1$VISIT,a1$LYM,-a1$SIZE),]

```

```

write.table(a1,file="C:/Claire/NSCLC1/30303TEST8.csv",
sep="," ,append=F,quote=F,col.names=T,row.names=F,na=".")

##### IDs lesion # >3 #####

a2 <- a[a$NLESION >3,]

a2 <- a2[order(a2$ID,a2$VISIT,a2$LYM,-a2$SIZE),]

write.table(a2,file="C:/Claire/NSCLC1/30303TEST9.csv",
sep="," ,append=F,quote=F,col.names=T,row.names=F,na=".")

##### identify three lesion numbers with biggest size at baseline and exclude "lymph
node" for #lesion >=4 #####

d <- a2[a2$VISIT==0,c("ID","LESION")]

d1 <- NULL

temp <- NULL

ni <- length(unique(d$ID))

for (i in 1:ni){

  lesion <- unique(d$LESION[d$ID==id[i]])

  nlesion <- length(lesion)

  temp <- data.frame(rep(id[i],3),lesion[1:3])

  d1 <- rbind.data.frame(d1,temp)

}

names(d1) <- c("ID","LESION")

d1 <- d1[order(d1$ID,d1$LESION),]

v <- d1

##### calculated sum of three largest lesion sizes at each visit #####

```

```

v1 <- merge(v,a2,by=c("ID","LESION"),all.x=T)

vv1 <-

aggregate(x=list(visitsum=v1$SIZE),by=list(ID=v1$ID,VISIT=v1$VISIT),FUN=sum)

v1 <- merge(v1,vv1,by=c("ID","VISIT"),all.x=T)

a2 <- v1

a2 <- a2[order(a2$ID,a2$VISIT,a2$LYM,-a2$SIZE),]

a2$NLESION <- 3

##### combine two subset together #####

a <- rbind.data.frame(a1,a2)

###a$TREAT <- ifelse(a$TREATMENT=="PLACEBO",0,1)###

a <- a[,c("ID","VISIT","LESION","SIZE","visitsum","basesum","NLESION"
,"LYM","TREATMENT","NVISIT")]

names(a) <- c("ID","VISIT","LESION","SIZE","VSUM","BSUM","NLESION"
,"LYM","TREATMENT","NVISIT")

a <- a[order(a$ID,a$VISIT,a$LYM,-a$SIZE),]

##### export dataset with all ids #####

write.table(a,file="C:/Claire/NSCLC1/NSCLC30303FINAL.csv",
sep="," ,append=F,quote=F,col.names=T,row.names=F,na=".")

```

## **2.9 Blinded Re-Evaluation of Imaging Data**

To identify sources of variance between patient outcomes and the modeled change in tumor size over time, we obtained the original sets of images from patients enrolled at the University of Chicago in studies 30203 and 30303. One radiologist, blinded to the original case report forms and radiology reports (co-author AO) reviewed all of the baseline images and identified and measured all target lesions. Unblinded to the lesions, but still blinded to the original measurements, the radiologist performed serial measurements on the identified target lesions on the University of Chicago patients.

### *3.Results*

#### **3.1 Data Quality Control**

We discovered that 9730 study case report forms (CRFs) frequently included no change (nc) or not available (na) rather than tumor size measurements on subsequent CT scans as in Figure 4.2. Consequently the entire trial data set had too much missing data to be useful for validating the longitudinal tumor growth model and data from all 561 subjects was excluded.

For the CALGB trials, we subjected the patients' data to the same standard for inclusion as in the FDA model (at least a baseline measurement and measurements recorded at some subsequent time-point). Figure 4.1 describes the attrition for 140 original cases in the CALGB 30203 trial and 160 in CALGB 30303. This resulted in a total of 227 patients available for the analyses.

#### **3.2 Longitudinal Modeling of Tumor Size**

Parameter estimates for sum of longest tumor dimensions at baseline (M\_BASE),

the treatment-effect/shrinkage rate (M\_SR), and the linear tumor growth rate (M\_PR) were determined and compared to the results of similar study arms from the original study as shown in Table 4.3. The individual predictions vs. observations of sum of the longest diameter were compared in each individual between FOCE and SAEM methods. The final parameters were estimated using SAEM based on individual fitting.

Study	Treatment	N	M_BASE (cm)	M_SR (1/week)	M_PR (cm/week)	
FDA trial treatment arms (E4599& TAX 326)	PCB	434	9.1(0.33)	0.06(0.004)	0.13(0.02)	
	PC	444	8.0(0.30)	0.038(0.01)	0.14(0.04)	
	DC	408	8.7(0.31)	0.052(0.01)	0.16(0.02)	
	DCb	406	9.2(0.38)	0.047(0.005)	0.16(0.02)	
	VC	404	8.5(0.28)	0.063(0.01)	0.17(0.02)	
NCI trial treatment arms (CALGB30203)	GCb+/- Zi or Ce	103	8.14(0.53)	0.014(0.005)	0.041(0.016)	
	(CALGB30303)	DC+/-BNP	124	8.66(0.48)	0.037(0.005)	0.079(0.011)
	(CALGB30203+CALGB30303)	GCb+/-Zi or Ce and DC+/-BNP	227	8.44(0.35)	0.027(0.003)	0.066(0.008)

PCB, Paclitaxel, Carboplatin and Bevacizumab; PC, Paclitaxel and Carboplatin; DC, Docetaxel and Cisplatin; DCb, Docetaxel and Carboplatin; VC, Vinorelbine and Cisplatin; GCb, Carboplatin and Gemcitabine; Zi, Zileuton; Ce, Celecoxib; BNP, BNP7787

**Table 4.3. Tumor model parameter estimates and precision standard error (SE) of baseline (M\_BASE), shrinkage rate (M\_SR) and progression rate (M\_PR) for FDA registration trials and CALGB 30203 and 30303 trial.**



Variance in parameter estimates increased as sample size was reduced from typical phase III to typical phase II size study arms. With a combination of both 30203 and 30303 trials, the model estimates of baseline tumor size, shrinkage rate and progression rate were 8.44cm, 0.027 1/week and 0.066 cm/week, respectively. For example a patient with an average baseline tumor size at 8.1cm will after 1 month, have the typical tumor burden decrease to  $8.44 \times e^{(-0.027 \times 4)} + 0.066 \times 4 = 7.84$ cm. This 7.1% decrease reflects the drug effect on tumor size. Compared to the average parameter estimates from five treatment arms of FDA data, the CALGB results underestimated M\_BASE, M\_SR, M\_PR in 3%, 48% and 57%, respectively.

These differences in tumor size and growth metrics between the smaller CALGB dataset and the FDA dataset were associated with tumor size at 8 weeks no longer being a statistically significant predictor of survival in NSCLC (Table 4.4). This inconsistency in the prediction of survival by tumor size with data collected by an experienced cooperative clinical trials group undermined plans to replace RECIST with continuous measures of tumor burden in smaller, prospective, randomized phase II trials. We conducted additional studies to identify factors in acquisition of these data that might have led to this unexpected finding.

Furthermore, no significant changes in the tumor parameter estimates were observed when applying additional script-based filtration by adding the restriction of number of the lesions, consistence of follow-up and number of lymph nodes.

Study	Treatment	N	M_BASE (cm)	M_SR (1/week)	M_PR (cm/week)
FDA trial treatment arms (E4599& TAX 326)	PC	444	8.0	0.038	0.14
	PC subgroup	103	9.26	0.0138	0.0346
NCI trial treatment arm (CALGB30203)	Gcb+/- Zi or Ce	103	8.1	0.0139	0.0407

PC, Paclitaxel and Carboplatin; GCb, Carboplatin and Gemcitabine; Zi, Zileuton; Ce, Celecoxib

**Table 4.4. Comparison of tumor parameter estimates between CALGB30203 trial and a subset of E4599 trial which matched to CALGB30203 tumor information.**

### **3.3 Evaluation of deviations in parameter estimates and survival prediction**

First we hypothesized that these lower estimates might reflect lower quality of data from small cooperative group trials compared to perhaps more meticulously accurate data submitted to the FDA for review. We therefore matched 103 patients from the ECOG 4599 study carboplatin/paclitaxel arm, part of the dataset used to generate the FDA model, with the 103 CALGB 30203 cases (who received carboplatin/gemcitabine). For this subset of matched cases in ECOG 4599 the mean baseline tumor size was larger, but the parameter estimates for M\_SR and M\_PR were now more similar to CALGB 30203 than to the results for the entire 444 subject carboplatin/paclitaxel arm of ECOG 4599 (Table 4.4). This implied that the deviation of parameter estimates between the CALGB data and the similar, but larger, treatment arms in the FDA dataset were unlikely to be due to significant differences in data quality between trials conducted by different groups, and instead reflected effects of decreasing the size of the analyzed subject pool.

Some recent studies suggested that TTG might be more robust than tumor size at early fixed time-points for colorectal cancer (Claret et al., 2013) and NSCLC (Claret et al., 2014). We therefore tested whether TTG might be a significant predictor of survival in the CALGB 30203 and the matched E4599 smaller dataset. In the Cox proportional hazard model for CALGB 30203 patients, the baseline tumor size (centered at 8.5cm) ( $P < 0.001$ ) and TTG ( $P < 0.001$ ) were significant predictors of time to death when incorporated as independent factors. The p-value of each coefficient was no longer significant when both factors were tested in the survival model simultaneously. The TTG and baseline tumor size were correlated ( $r = 0.75$ ) ( $P < 0.001$ ). Therefore, the final survival model included the baseline tumor size and ECOG performance status (0/1/2) (Eq 3).

$$\log(T)=\beta_0 + \beta_1 \times(\text{baseline}-8.5) +\beta_2 \times\text{ECOG} +\varepsilon$$

where T is time to tumor death (day) in a log normal distribution,  $\beta_0$  is the intercept when no covariate was added,  $\beta_1$  and  $\beta_2$  are the coefficients for baseline-8.5 and ECOG, respectively, and  $\varepsilon$  represents the residual error. In the matched subset of E4599 patients, TTG remained a statistically significant predictor in this model. However, when tested in the remaining 232 subjects of the carboplatin/paclitaxel arm, TTG was no longer a significant predictor of survival. We concluded that this loss of tumor size as a significant predictor of survival in smaller datasets was not an effect of the specific parameter of continuous measurement of tumor burden.

A difficult-to-test hypothesis is that the model accurately reflects that the CALGB 30203 and E4599 subsets of patients are genuinely different from the larger population of patients on which the FDA model was based. Our experience with the multi-step process of CT-imaging measurement and transmission of measurements into clinical trial databases offers an alternative hypothesis: the routine variance in tumor burden measurements introduced by our current RECIST-oriented clinical trial methods contributes significantly to this loss of power because of the inherent variance introduced with this approach. This variance further confounds computational models of continuous tumor growth reducing their ability to accurately predict survival in patient populations typical of phase 2 trials.

We therefore explored specific modifiable factors in the collection and reporting of tumor measurements might contribute to the altered shrinkage rate and growth rate parameter estimates in the longitudinal growth model when the size of the population was decreased. To evaluate the reproducibility of the tumor measurements on the case report

forms, an independent radiologist in blinded fashion re-determined and measured the baseline target lesions from the original CT images collected from 15 patients enrolled in CALGB 30203 and 30303 at one institution (Figure 4.4A). For 4 of the 15 patients, at least one additional target lesion was identified. Once unblinded to the determination of target lesions, the radiologist performed blinded serial measurements of the designated lesions (Figure 4.4B). Although displayed for consistency, 3 of the 15 subjects did not have an on-treatment assessment and therefore were not included in subsequent modeling analyses. For the 12 cases with serial measurements, 4 (subjects 7, 8, 9, 12) had trajectories of the measured sums of longest dimensions that were nearly superimposable between the case report forms (CRF) and the blinded evaluator re-assessment. Four (subjects 1, 3, 4, and 5) had obvious divergence between the CRF and blinded evaluations in terms of the magnitude of change in tumor burden and time-points at which these changes are registered. The remaining 4 had differences of unclear significance (2, 6, 10, 11).

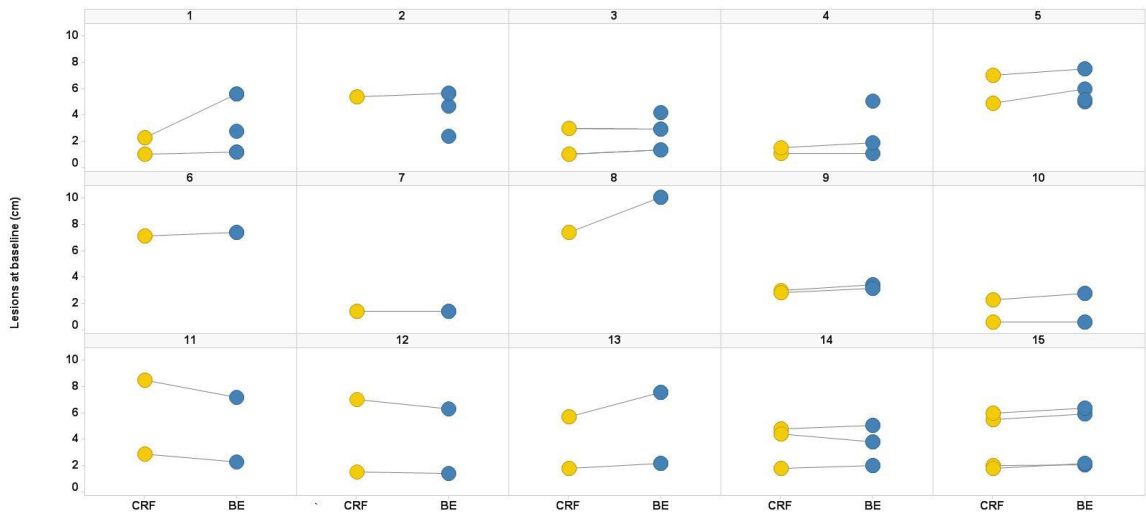
### **3.5 Estimated impact of continuous measurement variance on modeled endpoints**

RECIST was developed to be robust to inter-rater variance in measurements by setting categories for tumor size changes (Progressive Disease, Partial Response, and Complete Response) based on thresholds for magnitudes of change that would be unlikely to be due to the greatest degree of inter-rater variance (Therasse et al., 2000). A patient's category of response would then likely only be due to a significant effect of treatment (Moertel et al., 1976; Miller et al., 1981). It is therefore not surprising that in settings where inter-rater variance is not actively controlled, assessments of continuous

measurements of tumor growth will not improve upon our current RECIST-based categorical and time-to-event strategies

We hypothesized that this inter-rater variance in tumor burden assessments would have a significant effect on a continuous-measurement-based strategy such as TTG with less effect on a RECIST-based time to event endpoint such as progression-free survival (PFS). For the 12 subjects with serial CRF and blinded radiologist measurements (Table 4.5), we identified an average 25% delay in TTG calculated from the re-evaluated scans compared with the CRF data, but no differences in PFS assessments. Despite differences in target lesion assessment and measurement, subjects met criteria for progressive disease at the same imaging session in both datasets.

A)



B)

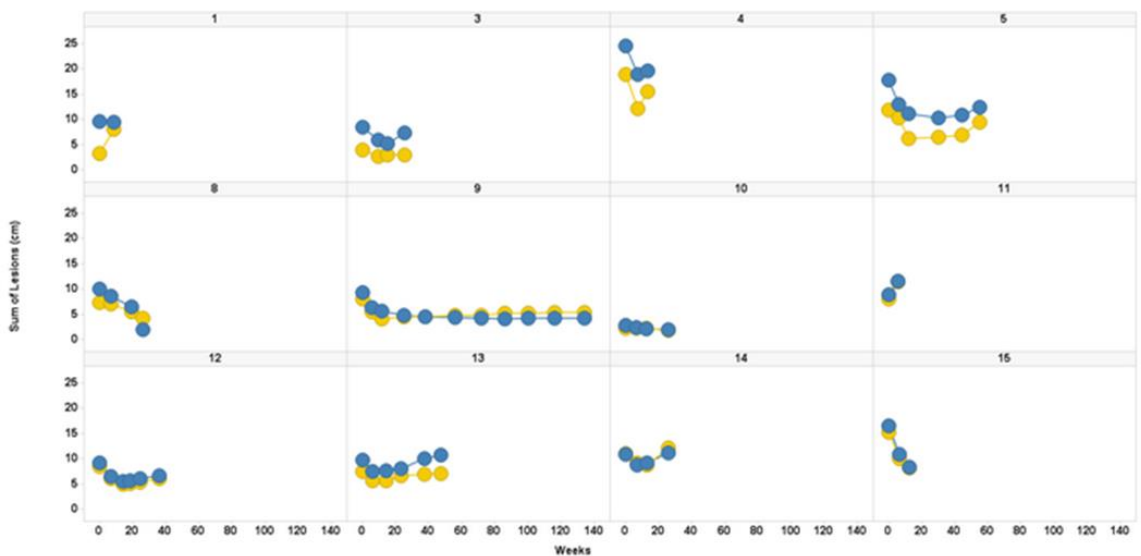


Figure 4.4. Comparisons of identified target lesion measurements from original case report forms and by blinded radiologist reviewer.

A) Target lesion measurements from CALGB 30203 and 30303. X-axis; Yellow circles; CRFs (case report forms), Blue circles; BE (Blinded re-evaluation of imaging data by radiologist from the University of Chicago). Y-axis; Each circle depicts one lesion at baseline (BL) visit in centimeters (cm). Gray line between yellow and blue circle connects the same lesion. B) Sum of the longest dimensions (cm) at each assessment time point for subjects in A over the course of the trial.

	Patient ID	1	2	3	4	5	6	7	8	9	10	11	12
PFS (week)	CRF	8	26	4	48	28	128	40	12	36	48	24	18
	BE	8	26	4	48	28	128	40	12	36	48	24	18
TTG (week)	CRF	1	20	74	27	43	53	21	43	23	20	13	46
	BE	51	31	94	36	52	73	22	52	24	13	26	47

**Table 4.5. Comparisons of PFS and calculated TTG from the target lesion measurements by blinded radiologist evaluation (BE) and case report form (CRF) data**



#### *4. Discussion*

This examination of a NSCLC longitudinal tumor growth model and endpoints in published CALGB studies revealed current limitations to employing continuous measurements of tumor burden in phase II clinical trials. Modeling of typical phase III clinical trial-size samples has reproducibly demonstrated tumor burden metrics as predictors of survival (Claret et al., 2012; Stein et al., 2011; Claret et al., 2013; Stein et al., 2012). Contrary to our expectations, we showed that decreasing the sample size to that of a typical phase II trial results in complete loss of the predictive power of these tumor metrics. Even though TS at 8 weeks is no longer a statistically significant predictor of survival in NSCLC with a typical phase II sample size, the direction of this predictor to survival remains the same as observed in the large phase III trials. As Bonate illustrated in a simulation study based on the FDA model structure, the parameter bias did not translate to prediction bias in tumor size as the measurement error with the expected bounds for tumor size assessment (Bonate et al., 2013). This suggests that we cannot rule out the clinical utility of TS at early time points in phase II survival prediction because of the low predictive power in a small size trial. In fact, other sources of variance such as phase II survival data in contributing to the model predictability may also need to be evaluated. Closer scrutiny of the original images and the recorded data revealed variance in the process by which tumor burden is assessed and recorded to meet RECIST standards. This variance has no apparent effect on RECIST categories or time-to-event endpoints, but does affect tumor burden metrics.

There is no superior alternative approach to RECIST for the standardized assessment of anatomical tumor burden and its change over time (Eisenhauer et al., 2009;

Fojo et al., 2012). This categorical system applied to quantitative assessments of tumor burden provides low inter-rater variance (progressive disease will be determined with high uniformity across sites in a multi-center trial and among trials) at the expense of efficiency (requires more patients to be observed over long periods of time). Our findings are consistent with investigators collecting and curating the quantitative tumor burden data with sufficient precision to support use of RECIST but not to support more computationally intensive methods of evaluating effects of treatments in small clinical trials. It is no surprise that many investigators have found no significant advantages to use of quantitative methods (such as tumor size ratio) over more qualitative time-to-event strategies (such as progression-free-survival) for predicting impact on overall survival (Fridlyand et al., 2011; An et al., 2011; Kaiser et al., 2013).

This study had a limited data sample for analysis, but it required significant effort to obtain these data. The qualitative documentation of tumor response data in earlier trials was not suitable for further quantitative analysis. For the more recent trials, the primary databases maintained the RECIST-based categories in data fields, but obtaining the quantitative tumor measurements required manual retrieval and processing of archived paper forms. The small cohort of patients for whom images were available and reviewed might have been a biased sample, but this site had been a major contributor to enrollment across thoracic oncology trials in CALGB with the stringent audit and quality control processes applied for member sites. The data are therefore likely representative of the overall quality of data in the larger clinical trials. Furthermore, an independent trial, conducted by a completely different set of institutions (ECOG) yielded similar results. We cannot exclude the possibility that this particular subset of patients matched between

the CALGB and ECOG datasets represents a unique group of NSCLC patients whose tumor growth patterns are distinct from the typical patient population. Therefore our findings will require confirmation in other datasets.

The study demonstrates a potential major flaw in assumptions on the use of quantitative measures of tumor burden and computational analysis. The process of measuring, transmitting, analyzing, and interpreting CT imaging-based measures of tumor contributes significant, but potentially modifiable variance. Centralized collection and measurement of CT images with semi-automated and digitally enhanced procedures should significantly reduce this variance. Advances in computing and digital data management in the past several years have reduced costs and make paperless systems with fewer opportunities for manual error possible. Our findings suggest that establishing methods with less inter-rater variance should be a worthwhile investment in the future of cancer therapeutics assessment.

**CHAPTER V: *Clinical Trial Simulation to Evaluate Study Design: Capturing Abiraterone and Nilotinib Exposures***

*1. Introduction*

In the past twenty years, oral anticancer drug therapy has become more prevalent (Singh et al., 2004). Unlike intravenous drug administration, orally administered agents undergo absorption from the gastrointestinal lumen through the intestinal epithelium into the portal vein and then into the systemic circulation. The fraction of an orally administered drug dose that reaches the systemic circulation is the drug's oral bioavailability and significant differences or changes in bioavailability will lead to a significant variation in drug exposure (Martinez et al., 2002) and thus modified therapeutic and or toxic effects.

One of most well-known factors contributing to variability in oral bioavailability is food intake (Singh et al., 2004; Winstanley et al., 1989; Gu et al., 2007). The effect of food on oral bioavailability has the potential to cause a clinically significant impact by causing variable systemic drug exposures that can lead to drug toxicity or therapeutic failure (Koch et al., 2009; Kang et al., 2010). Furthermore, the alternation of pharmacokinetic profile of several standard and investigational anticancer drugs such as busulfan, topotecan and fluorouracil caused by food-drug interactions have been previously reported (Singh et al., 2004). Quantifying the effect of food on drug exposure is important when designing a clinical trial for oral anti-cancer drugs. In addition to food related variability, variability in drug pharmacokinetic parameters between individuals as well as between occasions within an individual also contributes to variability in drug exposure. In order to assess the contribution of different magnitudes of between-

individual and between-occasion variability on drug exposure, this study used a population PK modeling approach and Monte Carlo methods to simulate a virtual clinical trial of patients who took drug in both the fasted and fed states.

Such a clinical trial simulation framework provides insight as to whether a particular study design, with a random sampling schema to reflect a typical clinical practice setting, permits the retrieval of pharmacokinetic (PK) parameters and their variability under different prandial conditions. In addition, simulation of a clinical trial can assess how patient specific covariates, between-individual, and between-occasion variability affect the ability to accurately capture individual level exposures. The aim of this study was to evaluate whether a pre-defined oncology trial with a sparse drug concentration sampling schedule can adequately capture individual level drug exposures, random variability and the food effects of (1) abiraterone (92% decrease in oral clearance), and (2) nilotinib, (18% decrease in oral clearance). To achieve these objectives, a virtual cancer study population was simulated with assigned pharmacokinetic characteristics, food intake and between-individual as well as between-occasion variability. A population pharmacokinetic (Pop PK) approach was utilized to retrieve the Pop PK parameters under these conditions and examine whether or not these parameters could be adequately retrieved on estimation at both the individual and population level.

## *2. Materials and Methods*

### **2.1 Simulation of Patient Pharmacokinetic Characteristics**

Hypothetical patients were created with fasting state Pop PK characteristics mimicking those found in oncology patients. (See Table 5.1 for abiraterone (Ryan et al.,

2010) and Table 5.2 for nilotinib (ZYTIGA, Tanaka et al., 2010) patient characteristics. The Pop PK parameters were assumed to be log- normally distributed and the parameter values for simulation were obtained or calculated based on literature values from a non-compartmental analyses (Ryan et al., 2010; ZYTIGA, Tanaka et al., 2010). A one compartment model with first order absorption and elimination was assumed as the model structure. Between-individual variability of oral clearance (CL/F) and volume of distribution (Vd/F) were based on the mean and standard deviation of pharmacokinetic parameters from published studies. Abiraterone between-individual variability values for oral clearance and volume of distribution were 174% and 73%, respectively, in the fasted state and 42.2% and 36%, respectively, in the fed state [7]. The between-individual variability of the absorption rate constant was assumed to be 20% and a proportional residual error with 30% CV was assumed based on literature values for abiraterone (Ryan et al., 2010). For nilotinib, the between-individual variability of oral clearance and volume of distribution values were set to 55% and 37% CV, respectively, for both fed and fasting states (Tanaka et al., 2010). A between-individual variability for the nilotinib absorption rate constant of 20% and a proportional residual error with 10% CV based on the lower limit of quantification of the nilotinib concentration assay were assumed (Tanaka et al., 2010).

Values of the Pop PK parameters for each hypothetical patient were randomly chosen from their respective distributions with a correlation of 0.6 applied between clearance and volume.

**Table 5.1. Population pharmacokinetic parameters for abiraterone patients**

Abiraterone		
Parameter	Fasting status	Fed status
	Mean (SD)	Mean (SD)
Apparent Clearance / bioavailability [CL/F] (L/h)	2650 (4617)	231 (97.7)
Apparent Volume of distribution / bioavailability [Vd/F] (L)	25494 (18670)	4069 (1462)
Absorption rate (1/h)	1.65 (0.33)	1.65 (0.33)

F = oral bioavailability

**Table 5.2. Population pharmacokinetic parameters for nilotinib patients**

Nilotinib		
Parameter	Fasting status	Fed status
	Mean (SD)	Mean (SD)
Apparent Clearance / bioavailability [CL/F](L/h)	33 (18)	27(15)
Apparent Volume of distribution / bioavailability [Vd/F] (L)	720 (267)	604(181)
Absorption rate (1/h)	0.74 (0.15)	0.74 (0.15)

F = oral bioavailability



## **2.2 Food Effects**

For both abiraterone and nilotinib, it was assumed that 50% of doses were taken with food and that the values of the oral clearance and volume of distribution depended upon whether a particular dose was taken with food or without food.

To mimic the actual drug administration protocol for food intake, we considered a different food effect in the abiraterone versus nilotinib trials. This was done to test whether the model would be able to capture two extreme food effects, the between-individual and between-occasion variability, and the individual level exposure. For abiraterone, patients in the fed prandial state were assumed to have had a high fat meal given the availability of Pop PK parameter values in the reference for simulation from this state; therefore, the oral clearance and volume of distribution were reduced by 92% and 85%, respectively, if the dose was taken with food (Ryan et al.,2010). In contrast to the abiraterone simulations, patients who were simulated as receiving nilotinib were assumed to have a light fat meal and possibly just a glass of grape juice 2 hours before or 1 hour after nilotinib was taken. An 18% reduction in apparent clearance was introduced to reflect this food effect (Tanaka et al., 2010; Yin et al., 2010).

## **2.3 Study Design**

Patients in the simulated abiraterone trials were randomly assigned to take 1,000 mg once daily and patients in the simulated nilotinib trials were assigned to take a 300 or 400 mg tablet twice daily in 1:1 fashion. Virtual pharmacokinetic samples for each patient were assumed to occur at week 1, week 4, and month 2 and month 3 based on a previously established trial sampling schedule (see Figure 5.1) after the start of the trial so each individual had a total of 4 measurements (one sample obtained on each occasion).

This sampling schedule was designed to fit in with the standard clinical follow up of the patient. The samples were assumed to occur at random times during clinic hours. The visit day for each pharmacokinetic sample of each hypothetical patient was randomly selected from a discrete distribution with 50% of virtual patients having a visit sample drawn on the recommended day, 20%  $\pm$  1 day, 20% at  $\pm$  2 days and 10% at  $\pm$  3 days from the scheduled visit date. Each clinic visit time was randomly selected from a uniform distribution assuming regular office hours from 7am to 6pm. To mimic food intake when patients take their medications, a random food covariate was generated for each hypothetical patient and sampled for the four different clinic visit occasions within each patient.

The hypothetical patients were assigned into trial sizes of 20, 50 or 70 individuals, and each trial was replicated 100 times. The same study design was tested using three different between-occasion variability levels: 10, 25, and 40% CV on drug clearance.

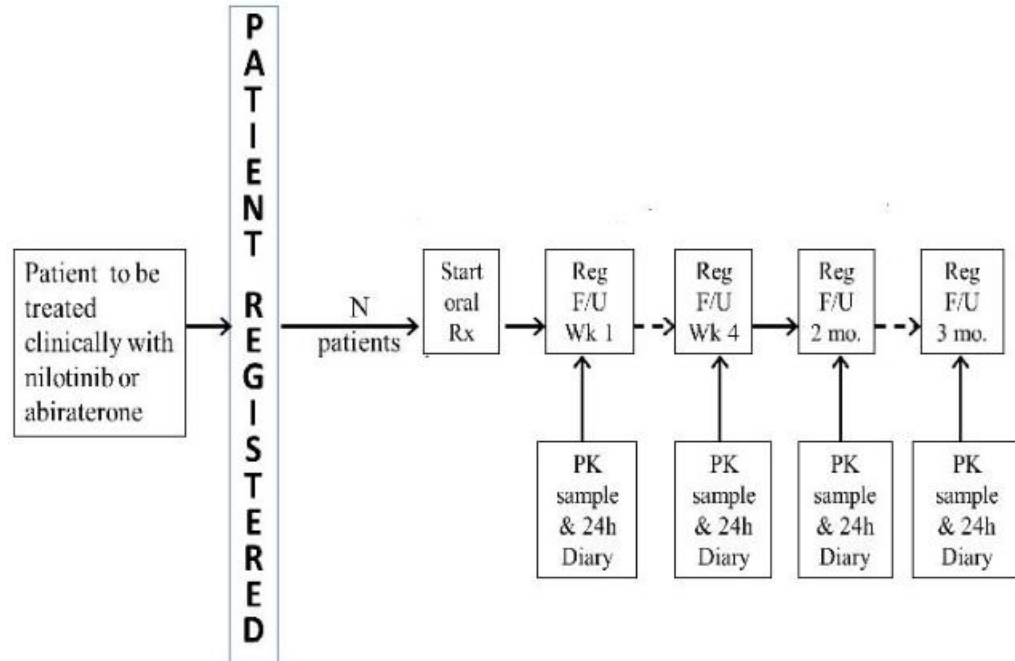


Figure 5.1. Study schema of abiraterone and nilotinib clinical trials.

## 2.4 Population Pharmacokinetic Model Estimation

The population pharmacokinetics of the hypothetical patients in each trial were analyzed using nonlinear mixed effect modeling software, NONMEM, Version VII (GloboMax\_LL, Ellicott City, MD, USA) using Wings for NONMEM, Version 7. One-compartment structural model with a proportional residual error model was tested using the first-order conditional estimation (FOCE) method. Between-individual variability was tested on oral clearance and volume of distribution, and between-occasion variability was tested on oral clearance only. The between-individual and between-occasion variability for Pop PK parameters were assumed to be log-normally distributed and evaluated using an exponential model  $P_{i,k} = P_{TV} \times e^{\eta_{p,i}} \times e^{\eta_{p,k}}$  where  $P_{i,k}$  is the parameter estimate for the  $i^{th}$  individual after the  $k^{th}$  dose administered and  $P_{TV}$  is the typical value for the parameter at the population level. The variability between the  $i^{th}$  individual and the population parameter value was described by  $\eta_{p,i}$  which was identically distributed with a mean of 0 and a variance of  $\omega_{\eta_{p,i}}^2$  (Feng et al., 2006). The variability between the  $k^{th}$  occasion of dose administration and the population parameter value was described by  $\eta_{p,k}$  which was identically distributed with a mean of 0 and a variance of  $\omega_{\eta_{p,k}}^2$ . In addition to the between-individual and between-occasion variability, residual variability was described by a proportional error model. That is,  $y_{ij} = \hat{y}_{ij}(1 + \varepsilon_{ij})$  where  $y_{ij}$  and  $\hat{y}_{ij}$  are the  $j^{th}$  observed nilotinib or abiraterone concentration and its corresponding model predicted concentration, respectively, with the difference described by  $\varepsilon_{ij}$  which is assumed to be normally distributed with a mean of 0 and a variance of  $\sigma^2$ . In order to estimate between-occasion variability in the model, the residual error was fixed in the estimation step.

## 2.5 Evaluation of Clinical Trial Designs

The food effect in each trial was evaluated by comparing the fit of a base model with no food effects with a model that includes a food effect covariate on oral clearance. The model comparison was based on a likelihood ratio test using the objective function value (OFV) from NONMEM. The change in the OFV returned by NONMEM is approximately equal to  $-2 \times \log$  likelihood. The difference in  $-2 \times \log$  likelihood between two models that are nested follows a  $\chi^2$  distribution. The significance level for identifying the food effect corresponded to a decrease in the OFV of greater than 6.63 (p-value  $\leq$  0.01, df=1). The power to detect a food effect is the percent of the trials where the population PK analysis demonstrated significant OFV change among the 100 trials in both abiraterone and nilotinib and retrieving a food effect value within 20% of the true food effect value.

In addition, to determining whether or not a food effect was detected by the trial design, the accuracy and precision of the model for retrieving parameter values of clearance, between-individual and between-occasion variability was assessed using two statistical standard criteria: percent bias and percent precision (Huang et al., 2007).

At the population level, the bias of each parameter represents the difference between the estimated values from the simulated (true) value of the population, and the percent bias is the mean predicted error normalized by the simulated value taken from the literature. The bias is calculated using the equation shown below:

$$\text{Population \% bias} = \frac{1}{m} \sum_{j=1}^m \frac{(Y_{p,j} - Y_e)}{Y_e} \times 100\%$$

where  $Y_{p,j}$  is the predicted value for the  $j^{th}$  trial in total  $m$  trials with the same given between-occasion variability level and the same sample size, and  $Y_e$  is the literature value we simulated from. The precision was estimated by calculating the root mean square error, reflecting the distribution of variance, and the percent precision is the root mean square error normalized by the simulated mean. This included the bias and precision of the estimation of the food effect.

$$\text{Population \% precision} = \sqrt{\frac{\sum_{j=1}^m \left(\frac{Y_{p,j} - Y_e}{Y_e}\right)^2}{m}} \times 100\%$$

Parameter estimates at the individual level were evaluated using percent bias and percent precision.

$$\text{Individual \% bias} = \frac{1}{n} \sum_{i=1}^n \frac{(Y_{p,i} - Y_{e,i})}{Y_{e,i}} \times 100\%$$

where  $Y_{p,i}$  and  $Y_{e,i}$  are the predicted and simulated values for the  $i^{th}$  patient, respectively.  $n$  represents the number of patients in the trial.

$$\text{Individual \% precision} = \sqrt{\frac{\sum_{i=1}^n \left(\frac{Y_{p,i} - Y_{e,i}}{Y_{e,i}}\right)^2}{n}} \times 100\%$$

Then, an average of individual percent bias and precision of the  $m$  number of trials was calculated.

### 3. Results

The goodness-of-fit plots for the final model were shown in Figure 5.2 and 5.3 for abiraterone and nilotinib, respectively. The percent bias and precision for the population

oral clearance for abiraterone ranged from -37% to -31% and 36.2% to 38.1%, respectively (Table 5.3). The average percent bias and precision are noticeably smaller in magnitude for individual clearance estimates (range of 1.2% to 4.4% and 14.5% to 19.9%, respectively). This indicates that the model estimates for oral clearance are more accurate and less variable at the individual level compared to the population level. Across three between-occasion variability levels (10, 25, and 40%), 21%, 10% and 11% of the abiraterone trials, respectively, retrieved population clearance values within 20% of the true value. The ranges of percent bias and precision for between-individual variability were -45.1% to -42.1% and 42.8% to 45.9%, respectively, with minimal variations with number of patients in the trial or between-occasion variability. The ranges of percent bias and precision of between-occasion variability were -30% to -1.1% and 11.6% to 40.7%, respectively. There was a decrease in both the between-occasion variability bias and precision as the number of patients and between-occasion variability increased. The known food effect on oral clearance for abiraterone was identified in 100% of simulated trials with 20, 50 and 70 patients for the 10%, 25%, and 40% between-occasion variability levels. The ranges of percent bias and precision of food effect were 2.01% to 4.42% and 6.81% to 14.4%, respectively.

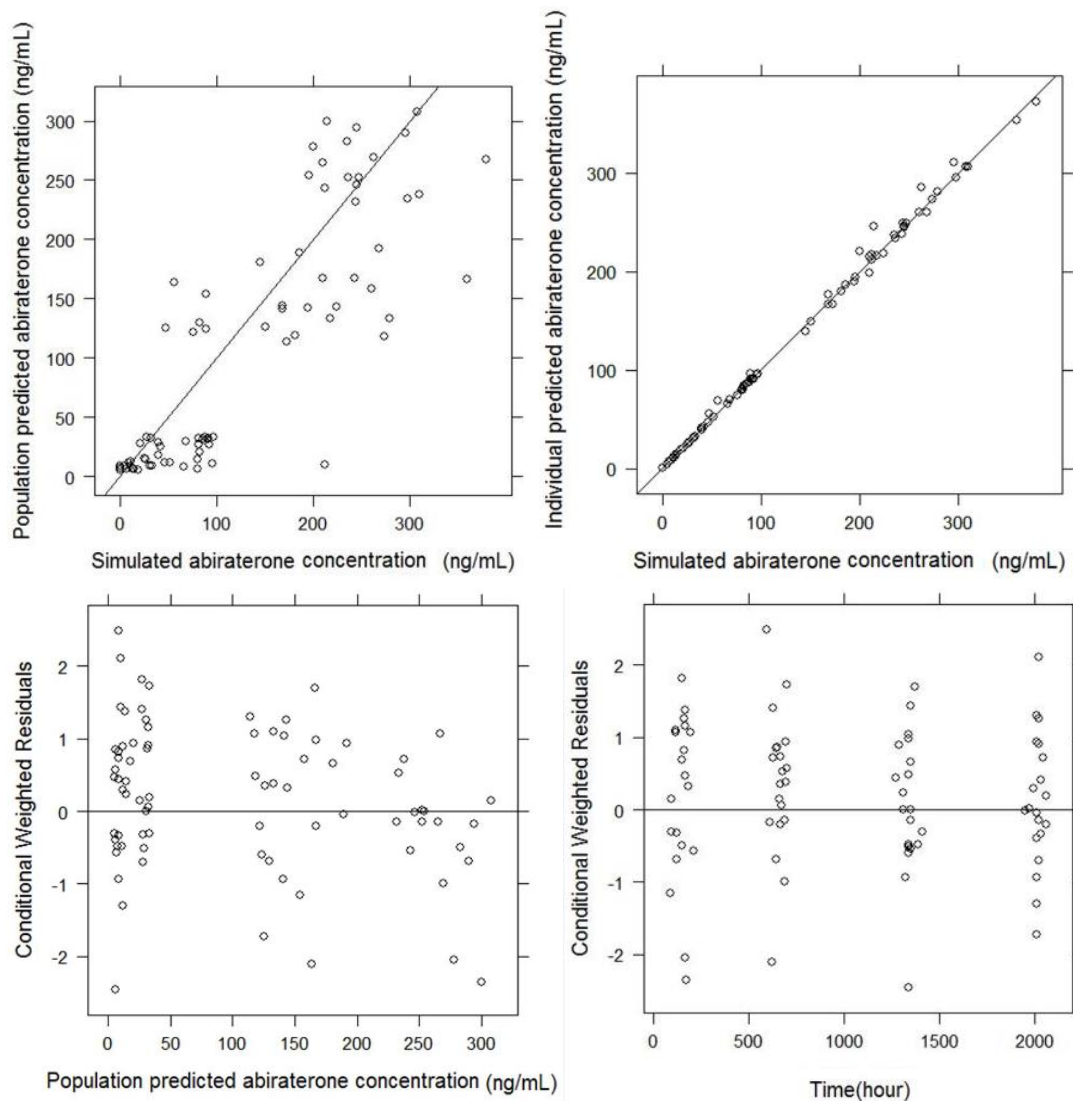


Figure 5.2. **Goodness-of-fit plots for the final model of abiraterone.**

Plot of population and individual predicted versus observed concentrations and plot of conditional weighted residuals versus population predicted concentrations and time (hours).



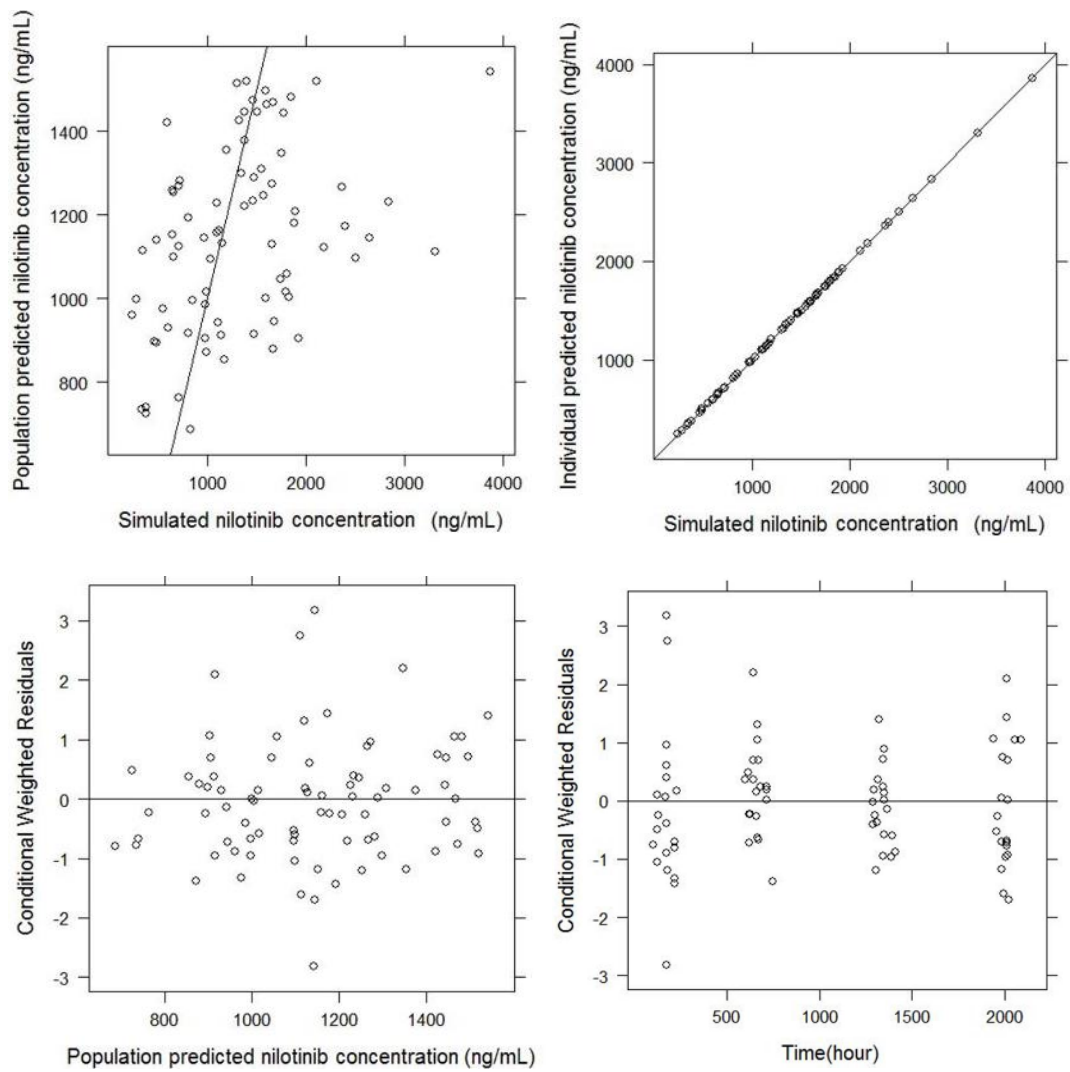


Figure 5.3. **Goodness-of-fit plots for the final model of nilotinib.**

Plot of population and individual predicted versus observed concentrations and plot of conditional weighted residuals versus population predicted concentrations and time (hours).

Table 5.3. **Bias and precision of abiraterone oral clearance at the population level and individual level, between-individual variability, and between-occasion variability**

Between-occasion variability	Number of patients in trial	Oral Clearance – Population level		Oral Clearance – Individual level		Between-individual variability		Between-occasion variability		Food effect	
		Bias	Precision	Bias	Precision	Bias	Precision	Bias	Precision	Bias	Precision
10%	20	-32.3%	37.1%	1.5%	14.5%	-44%	45.6%	-30%	40.7%	2.65%	9.95%
	50	-36.6%	37.8%	1.3%	14.6%	-43.6%	44.5%	-25.8%	30.8%	4.42%	7.42%
	70	-36.8%	37.8%	1.4%	14.9%	-43.3%	43.5%	-15.3%	23.3%	3.84%	6.81%
25%	20	-31.7%	36.6%	3.3%	19.6%	-45.1%	45.9%	-5.85%	18.8%	2.36%	11.8%
	50	-36.3%	37.6%	1.2%	18.9%	-42.4%	42.8%	-5.51%	14.6%	4.16%	8%
	70	-36.9%	37.9%	3.2%	19.9%	-43 %	43%	-3.08%	14.8%	3.61%	7.22%
40%	20	-31%	36.2%	1.4%	19.7%	-44.9%	46.1%	-1.34%	17.4%	2.01%	14.4%
	50	-35.7%	37.1%	1.6%	19.9%	-42.1%	42.8%	-1.44%	13.4%	2.52%	11.9%
	70	-37%	38.1%	4.4%	19.6%	-42.6%	43%	-1.11%	11.6%	2.64%	8.6%

The percent bias and precision on the population oral clearance for nilotinib ranged from -13.3% to -11.8% and 14.2% to 17.0%, respectively (Table 5.4) and were consistent across between-occasion variability levels (10%, 25% and 40%). In contrast to abiraterone, the average individual nilotinib oral clearance estimates were significantly more accurate and precise than population estimates with the percent bias and precision ranging from -1.9% to -0.5% and 4.2% to 8.6%, respectively. Across three between-occasion variability levels (10, 25 and 40%), 86, 84 and 83% of the nilotinib trials, respectively, retrieved the population oral clearance within 20% of the true value. The ranges of percent bias and precision of between-individual variability were -9.9% to -7.9% and 11.3% to 19.4%, respectively and ranged from -3.9% to -0.4% and 4.9% to 11.0% for percent bias and precision on between-occasion variability. Retrieval of the known food effect in this system was observed in 100% of the simulated nilotinib trials with 10% between-occasion variability with trial sizes of 25, 50, and 70 patients. For nilotinib trials simulated with 25% between-occasion variability, significant food effects on oral clearance were retrieved in 80% of 20 patient trials, 99% of 50 patient trials, and 100% of 70 patient trials. Nilotinib trials simulated with 40% between-occasion variability resulted in significant food effects on clearance being retrieved in 50% of 20 patient trials, 78% of 50 patient trials, and 88% of 70 patient trials. The ranges of percent bias and precision of food effect were -2.16% to 11.8% and 6.3% to 38.8%, respectively

**Table 5.4. Bias and precision of nilotinib oral clearance at the population level and individual level, between-individual variability, and between-occasion variability**

Between-occasion variability	Number of patients in trial	Oral Clearance – Population level		Oral Clearance – Individual level		Between-individual variability		Between-occasion variability		Food effect	
		Bias	Precision	Bias	Precision	Bias	Precision	Bias	Precision	Bias	Precision
10%	20	-11.8%	16.0%	-0.6%	4.6%	-9.0%	18.1%	-3.9%	11.0%	-0.12%	12.3%
	50	-12.0%	16.3%	-0.5%	4.3%	-9.9%	13.4%	-2.1%	6.9%	-0.05%	12.3%
	70	-12.8%	14.2%	-0.5%	4.2%	-8.0%	11.3%	-2.3%	5.7%	-0.04%	6.3%
25%	20	-12.0%	16.2%	-1.5%	6.6%	-8.8%	18.5%	-1.9%	8.9%	-0.4%	30%
	50	-12.4%	16.7%	-1.0%	6%	-9.1%	13.1%	-1.2%	5.2%	0.9%	20.9%
	70	-13.3%	14.7%	-0.8%	5.7%	-7.9%	11.7%	-0.9%	4.9%	-1.39%	15.5%
40%	20	-12%	16.4%	-1.9%	8.6%	-8.5%	19.4%	-1.8%	8.7%	11.8%	38.8%
	50	-12.7%	17.0%	-1.0%	7.2%	-9.2%	13.9%	-0.4%	7.7%	2.32%	30%
	70	-13.3%	14.7%	-0.9%	7.1%	-7.9%	12.3%	-0.7%	4.9%	-2.16%	24.3%

#### *4. Discussion*

Virtual clinical trials of abiraterone and nilotinib using sparse concentration measurements and a population PK sampling time design were simulated to test whether the drug exposure of each simulated patient and its variability under different prandial conditions on oral clearance for two recently approved drugs would be retrieved accurately and precisely. The trial design is a typical phase II design in the NCI cooperative system and this assessment provides an evaluation of whether there is value in drawing sparse samples to utilize in population pharmacokinetic analysis for the determination of individual drug exposure. It was important to assess this particular study design as it is widely used and insight on the value of drug concentration sampling under these conditions was unclear. Population as well as individual pharmacokinetic parameters for abiraterone, with a large food effect, and nilotinib, with a smaller food effect, were well estimated from the virtual trial results. This evaluation of whether between-individual and between-occasion variability can be well captured with a significant covariate effect on oral clearance (the underlying food effect) at different levels of variability on anticancer drug exposure is a novel observation. As the prior knowledge of more than 100% between-individual variability was introduced to oral clearance in the abiraterone trial, the model estimation of the population oral clearance parameter and between-individual variability have relatively poor percent bias and precision compared to the nilotinib trial. This is reflected in the power calculation showing that only 10 to 20% of the trials across three between-occasion variability levels (10, 25, 40%) have population oral clearance estimates within 20% of the true value. This finding indicates that retrieving population level effects with large between-individual

variability may need a much larger trial size or more intense sampling schedule or a combination of both.

Between-occasion variability estimates were reasonable considering both percent bias and precision were less than 30% in most trial simulations. However, the percent bias and precision were relatively poor for the simulated abiraterone trials with 20 patients and 10% within-individual variability. One possible explanation of this is that the system is less able to capture between-individual variability if the actual variability is small, but the true mechanism contributing to this poor bias at 10% between-individual variability is as yet unclear.

This model was unable to estimate both residual error and between-occasion variability accurately simultaneously. In order to accurately capture the between-individual variability, the residual error estimation was fixed to published values. This suggests that more than one sample per occasion will be required to distinguish these hierarchies of variability simultaneously.

The effects of two important features of such clinical trials were quantified using a NONMEM based simulation analysis. These were the power to detect differences in oral drug clearance related to the prandial state of the study participants and the degree of between-individual variability and sample size. When the between-individual variability was set to 25% for a trial with 20 patients, 80% of the simulated nilotinib trials resulted in the detection of a statistically significant reduction in oral clearance caused by the food effect. The power to detect this food effect on oral drug clearance increased to 100% when the number of patients per trial was increased from 20 to 70. At 40% between-individual variability, the food effect signal was observed in only 50% of trials with 20

patients, increasing to 88% when the number of patients increased to 50. The percent precision also indicated that the food effect as a covariate was captured less precisely when between-occasion variability increased, and the trial sample size was reduced. In contrast, abiraterone with a much more substantial food effect (92% reduction in oral clearance) resulted in a power to detect the food effect of 100% in the smallest trial evaluated (20 patients). This virtual trial also had 40% between-individual variability and resulted percent bias and precision on the Pop PK parameters that were all less than 20%. Using a modeling approach, we were able to simulate a complex clinical oncology population pharmacokinetic trial setting and capture the between-occasion variability and the magnitude of individual drug exposure in the presence of a large food effect for two recently approved oral anti-cancer agents. This simulated approach facilitated an early evaluation of the proposed trial design. However, clinical trial simulations are generated based on many trial assumptions, and these assumptions may include uncertainty. In fact, with different underlying assumptions, the simulated outcomes can differ, so multiple scenarios as well as assumptions must be tested in order to fully interpret the relevance of the results.

There are limitations to this analysis. First, a dropout model was not incorporated in this study design which can potentially contribute to censor events in a 3 month study. The simulations also assumed parameter distributions based on available food effect assessment in previous non-compartmental analysis. Because of the established sampling schedule, both drugs were better estimated with a simplified model structure. Despite this, the model provides an approximation of the actual behavior and can capture the trend of the variability within the population. In fact, this simplified modeling approach

was previously proposed to assess the pharmacokinetics of some drug entities in the trials with a relative sparse sampling schedule (Scerwin et al., 2012; van Erp et al., 2011). The sampling schedule could also be optimized for the identification of food effects and model parameters, but the objective of this work was to evaluate trial designs given a commonly used sampling schedule. Second, non-compliance was not considered in this study design as we assumed that the compliance rate should be reasonably controlled in the clinical trial although this will result in a higher residual error and between-individual variability than compliance accounted for using electronically monitoring (Vrijens et al., 2005). This study emphasizes the importance of addressing trial designs where intensive sampling cannot be obtained and yet there is a need to understand the drug exposure characteristics to what are otherwise medications with highly variable pharmacokinetic disposition.



**CHAPTER VI:** *Genetic determinants of vincristine neuropathy in pediatric acute lymphoblastic leukemia patients*

*1. Introduction*

Vincristine is one of the most commonly prescribed anticancer agents, and it has been used to treat a wide variety of malignancies. Even though the drug is commonly used, the dose optimization of vincristine and potential causes of severe side effects and lack of efficacy still remain unclear. In fact, vincristine plays a role in many cellular mechanisms. It not only binds irreversibly to microtubules and spindle proteins and interferes with the formation of the mitotic spindle, but also depolymerizes microtubules and may also interfere with amino acid, cyclic AMP, and glutathione metabolism. In addition, it is associated with calmodulin-dependent  $Ca^{++}$ -transport ATPase activity, cellular respiration, and nucleic acid and lipid biosynthesis (National Cancer Institute, 2011).

Similar to other chemotherapy agents such as taxanes and platinum, vincristine is associated with highly variable cumulative dose-dependent neurotoxicity, secondary to its binding to tubulin dimers, which can inhibit the assembly of microtubule and arrest mitosis in metaphase (Park et al., 2013; Howlader et al., 2012; Verstappen et al., 2005; Postma et al., 1993; Haim et al., 1994). In severe cases, neurotoxicity may lead to dose reduction, which may negatively impact the drug efficacy. In addition, substantial variability in vincristine pharmacokinetics with up to a 40-fold interpatient variation has been reported in the literature (Crom et al., 1994; de Graaf et al., 1995), and this pharmacokinetic variability may be associated with severity of neurotoxicity. Published data also indicate that vincristine pharmacokinetics may in fact be associated with long-

term outcomes in children of acute lymphoblastic leukemia with rapid clearance being associated with a greater risk of relapse. (Lonnerholm et al., 2008).

The variability of vincristine pharmacokinetics might be potentially explained in part by genetic differences in cytochrome P450 3A family enzymes, which are important for vincristine metabolism (Dennison, et al., 2006).

As previous study shown, Caucasian children with ALL more commonly suffer from vincristine-induced peripheral neuropathy (VIPN) than do children of other ethnicities due to CYP3A5 genotype (Dennison et al., 2006). In fact, another study suggests that polymorphisms of multi-drug resistance (MDR1) gene which code as an efflux transporter are associated with peripheral neuropathy (Te Loo et al., 2010). This might also contribute to the variability of vincristine PK, and the genetic impact on drug toxicity. In addition, study published by Hertz et al. suggested that CYP2C8 polymorphisms can increase the risk of neuropathy in breast cancer patients treated with paclitaxel (Hertz et al., 2013). This indicated that in the case of paclitaxel drug toxicity may be associated with pharmacokinetic exposure through metabolic pathways. However, those findings are controversial across studies with limited validation in the literature (Hertz et al., 2014). In a study of 329 subjects by Broyl et al., an early-onset vincristine-induced neuropathy was characterized by the up-regulation of genes involved in cell cycle and proliferation, including AURKA and MKI67, and also by the presence of single-nucleotide polymorphisms (SNPs) in genes involved in these processes, such as GLI1 (rs2228224 and rs2242578) (Broyl et al., 2010). Similar to the vincristine association analyses, many studies have aimed to identify potential genetic risks in predicting neuropathy induced by platinum agents or Taxanes. In those genomic studies,

few candidate genes (FGD4, TUBB2A, ARHGEG and PRX) were found to be associated with paclitaxel-induced peripheral neurotoxicity. These results indicated that drug-induced peripheral neurotoxicity may share genetic roots as hereditary neuropathies (Baldwin et al., 2012; Leandro-Garcia et al., 2012; Beutler et al., 2014; Travis et al., 2014). Furthermore, McWhinney- Glass's group further explored genetic risk of four SNPs (SOX10, BCL2, OPRM1 and TRPV1) in an association analysis of platinum/taxane-induced neurotoxicity. Interestingly, a cumulative association of these genetic variations with neurotoxicity was observed in this population of 404 patients with ovarian disease (McWhinney-Glass et al., 2013).

In order to better understand VIPN and predict risk for toxicity to ultimately aid in vincristine dose optimization and to further explore the findings from the previous studies, a genome-wide association analysis was proposed to take an unbiased approach in evaluating the association between germline variants and VIPN in pediatric ALL patients. Indeed, as the genomics technologies improve, genetic biomarkers provide a unique tool for prediction of disease outcomes of interest. The objective of this study is to build a candidate gene list including polymorphisms associated with VIPN from the GWAS study as well as other association analysis approaches. The rationale for the proposed research is that once we understand the impact of specific pharmacogenetic changes on vincristine toxicity, we will be able to use this information to optimize dosing of this important drug for individual patients.

## *2. Materials and Methods*

### **2.1 ALL Patient Population and Study Design**

A total of 2154 children with precursor B cell acute lymphoblastic leukemia from the Pediatric Oncology Group (POG) 9904, 9905 and 9906 trials were enrolled. In conjunction with 9905, the objective of the 9904 trial was to determine if a delayed multi-drug intensification, administered in the context of intensive anti-metabolite therapy improved outcomes and compare the different durations of methotrexate infusion. 9906 was aimed at specifically evaluating patients at high risk for treatment failure. Those who had Down's syndrome, Charcot Marie Tooth disease, baseline peripheral neuropathy, or a history of liver disease with chronic elevation in liver function tests to greater than 5-times the upper limit of normal based on normal values for age were excluded from this GWAS. Genotyping of germline samples from 1888 patients on the 9900 trials has been completed using the Affymetrix GeneChip Human Mapping 6 and 500K arrays for 9904/9905 and 9906, respectively. The number of vincristine doses per patient over the course of treatment ranged from 18-23, depending on specific protocol and treatment arm. Standard vincristine dosing of  $1.5\text{mg}/\text{m}^2$ /dose was used. The complete treatment duration is 2.5 years from the date of diagnosis for females and 3.5 years for males. The treatment plan is summarized in Figure 6.1.

Day	1	8	15	22	29	36	43	50	57	64	71	78	85	92	99	106	113	120	127	134	141	148	155	162	169	176	183	190	197	204	211	218	225	232	239	246	253	260	267	274							
Week	INDUCTION				CONTINUATION																DELAYED INTENSIFICATION																										
9904 GroupA	V	V	V	V				V	V								V	V							VV																						
9904 GroupB	V	V	V	V				V	V								V	V							VV																						
9904 GroupC	V	V	V	V				V	V								V	V							VV																						
9904 GroupD	V	V	V	V				V	V								V	V							VV																						
Day	281	288	295	302	309	316	323	330	337	344	351	358	365	372	379	386	393	400	407	414	421																										
Week	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61																										
9904 GroupA	VV																																														
9904 GroupB	VV																																														
9904 GroupC	VV																																														
9904 GroupD	VV																																														
Day	561	568	575	582	589	596	603	610	617	624	631	638	645	652	659	666	673	680	687	694	701																										
Week	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101																										
9904 GroupA																																															
9904 GroupB																																															
9904 GroupC																																															
9904 GroupD																																															

Vincristine dose regimen scheme in POG 9904 trial

Day	1	8	15	22	29	36	43	50	57	64	71	78	85	92	99	106	113	120	127	134	141	148	155	162	169	176	183	190	197	204	211	218	225	232	239	246	253	260	267	274								
Week	INDUCTION				CONTINUATION																INTENSE CONTINUATION																											
9905 GroupA	V	V	V	V				V	V								V	V																														
9905 GroupB	V	V	V	V				V	V								V	V																														
9905 GroupC	V	V	V	V				V	V								V	V																														
9905 GroupD	V	V	V	V				V	V								V	V																														
Day	281	288	295	302	309	316	323	330	337	344	351	358	365	372	379	386	393	400	407	414	421																											
Week	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61																											
9905 GroupA																																																
9905 GroupB																																																
9905 GroupC																																																
9905 GroupD																																																
Day	561	568	575	582	589	596	603	610	617	624	631	638	645	652	659	666	673	680	687	694	701																											
Week	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101																											
9905 GroupA																																																
9905 GroupB																																																
9905 GroupC																																																
9905 GroupD																																																
Day	841	848	855	862	869	876	883	890	897	904																																						
Week	121	122	123	124	125	126	127	128	129	130																																						
9905 GroupA																																																
9905 GroupB																																																
9905 GroupC																																																
9905 GroupD																																																

Vincristine dose regimen scheme in POG 9905 trial



## **2.2 Genotype Data and Quality Control**

DNA(500ng) was digested with restriction enzymes, amplified, labeled, and hybridized to the Affymetric GeneChip Human Mapping 6 set for P9004 and 9005 and 500K for 9006. The genotypic data included raw data files, genotypes, quality scores, intensity values, SNPs and sample summary tables. QC was performed to remove both samples and markers which were unreliable in the following fashion. Samples with >5% missing rate were excluded. Furthermore, SNPs with a study-wide missing data rate of >5% and/or evidence of Hardy-Weinberg disequilibrium ( $p \leq 0.0001$ ) were discarded. SNPs with minor allele frequency (MAF) of <0.05 were also removed from the analysis because previous studies have shown that these SNPs have little power to detect association and are more prone to genotypic errors resulting in false positive evidence of association. In addition, MAF < 0.10 cutoff was applied for the recessive model in the analysis. PLINK was proposed to manipulate the data by generating both per sample and per SNP metrics to assess the quality of the genotypic data (Purcell et al., 2007).

## **2.3 Population Stratification**

A principal components approach (Price et al., 2006) was applied to correct for any possible stratification errors. A subset of the genome-wide SNP data was used to identify components that reflect population structure. Scores for each individual from the initial principal component assessment were calculated and included as covariates in the subsequent association analyses.

## **2.4 Phenotype Identification**

Vincristine is associated with highly variable cumulative dose-dependent peripheral neurotoxicity that often necessitates chemotherapy dose reductions, thereby

compromising efficacy. To do this association analysis, the first step was to define our phenotype of interest. National Cancer Institute Common Terminology Criteria for Adverse Events, version 3.0 (CTCAE v3.0) was used to capture and grade adverse events in this study. We defined neuropathy for the purposes of this analysis as any sensory or motor neuropathy  $\geq$  grade 2. Neuropathic pain was also captured in this study. For the purpose of this analysis, we designate neuropathic pain as secondary neuropathy as any sensory neuropathy, motor neuropathy or neuropathic pain  $\geq$  grade 2.

## **2.5 Genome-Wide Association Study**

The analysis started with association analysis of genotype frequencies with primary neuropathy (CTCAT V3.0 sensory or motor neuropathy). Neuropathy was treated as a binary variable; ie, 1 for having neuropathy and 0 for no neuropathy. Clinical, demographic, and population stratification variables were tested as covariates in the regression analyses. Genotype frequencies were tested in dominant, recessive and additive (gene-dose) model. This strategy was also used to test the association between SNPs and secondary neuropathy. These analyses were performed in R coxph package (Therneau et al., 2014). The corresponding p-value of each SNP was summarized across chromosomes in Manhattan and QQ plots.

## **2.6 Gene Enrichment Analysis**

The gene enrichment analysis was done by focusing on enrichment of specific target genes (Shamir, 2010). In this case, the target genes are genes previously identified as being related to neuropathy or pain. The initial step of the analysis was to obtain SNP rsIDs along with their genomic positions and p-value from the SNP6.0 Affymetrix COG data which was previously analyzed in the survival analysis with neuropathy phenotype.



SNPs were annotated with their respective genes using their positions and with the annotation file from UCSC genome browser. This was done by matching the SNP chromosome number with the gene chromosome number and then looking for SNPs that within the gene base pair co-ordinates. Upstream or downstream regions of the genes were not considered. Using the significance frequency table, a Fisher's exact test was run to calculate the odds ratio based on equation 1 for each gene.

	P-value <0.05	P-value >0.05
Candidate gene	N11	N12
Other genes	N21	N22

$$\text{Odds Ratio} = N11 * N22 / N12 * N21 \text{ (Equation 1)}$$

N is the number of SNPs in each category. The p-value was estimated in Fisher's exact test. Based on the criteria of p-value <0.05 from Fisher's exact test and calculated odds ratio >1, genes which passed the criteria were selected.

## 2.7 Pathway Analysis

Significant pathways in biological processes were identified in an interactive pathway analysis from Ingenuity systems (Ingenuity systems analysis). The following steps were used in the analysis: "(1) Genes identified as significant from the experimental data sets were overlaid onto the interactome. Focus genes we re-identified as the subset having direct interaction(s) with other genes in the database. (2) The specificity of connections for each focus gene was calculated by the percentage of its connections to other significant genes. The initiation and growth of pathways proceeded from genes with the highest specificity of connections. Each pathway had a maximum of 35 genes. (3)

Pathways of highly interconnected genes were identified by statistical likelihood using the following equation:

$$\text{Score} = -\log_{10}\left(1 - \sum_{i=0}^{f-1} \frac{C(G,i)C(N-G,s-i)}{C(N,s)}\right)$$

where N is the number of genes in the genomic network, of which G are focus genes, for a pathway of genes, f of which are focus genes, and C (n,k) is the binomial coefficient. (4)

Pathways with a score greater than 4 (P < 0.0001) were combined to form a composite network representing the underlying biology of the process.”

## 2.8 Gene Expression Analysis

Gene expression omnibus(GEO) is a public functional genomics data repository supporting MIAME-compliant data submissions (Lehnhardt et al.,2005). To explore whether the gene expression level was changed after treating with vincristine, a keyword search of vincristine in GEO database was performed. SNPs in COG data were annotated with their respective genes using their positions and with the annotation file from UCSC genome browser. These genes were compared with the genes in GEO. Those overlapped genes were further filtered by the gene expression level with a threshold of greater than 50% up-regulation or down-regulation of the vincristine treated replicates compared to the control. For those genes which passed the threshold, the SNPs with their p-values from the GWAS study were evaluated and visualized in the QQ plot.

## 2.9 eQTL(GTEEx) Analysis

In addition to the GEO database, tissue data from GTEx (Genotype-Tissue Expression) project (Gibbs et al., 2010) was explored to determine the correlations between genotype and the tissue-specific gene expression level. This could help to identify the expression quantitative trait locus (eQTL) that might regulate the expression level of mRNAs or proteins. Based on the available data from the published tissue bank,

four of the brain tissues including cerebellum, frontal cortex, temporal cortex and pons were considered in the correlation analysis with neuropathy.

### *3. Results*

#### **3.1 ALL Patient Characteristics**

Within 2154 ALL patients enrolled in POG 9904, 9905 and 9906 trials, genotyping of 1888 individuals were eligible. To further insure the genotyping quality, only patients from POG 9904 and 9905 who were genotyped using Affymetrix 6.0 were included in the initial analysis. Furthermore, a principle component approach was applied to correct the population stratification error by using Hapmap references of Han Chinese in Beijing, China (CHB), Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) and African ancestry in Southwest USA (ASW). Based on the result of the two principal component analyses (PC1 and PC2) in Figure 6.2, Caucasians were better overlapped with the reference and represented the majority of the population. In addition, data which did not pass the additional quality control including samples or SNP with >5% missing data rate and/or evidence of Hardy-Weinberg disequilibrium ( $p \leq 0.0001$ ) and/or MAF < 0.05 for dominant and additive model and/or MAF < 0.10 for recessive model were excluded, and a total of 587,014 SNPs of 1068 individuals remained in the analysis (Figure 6.3).

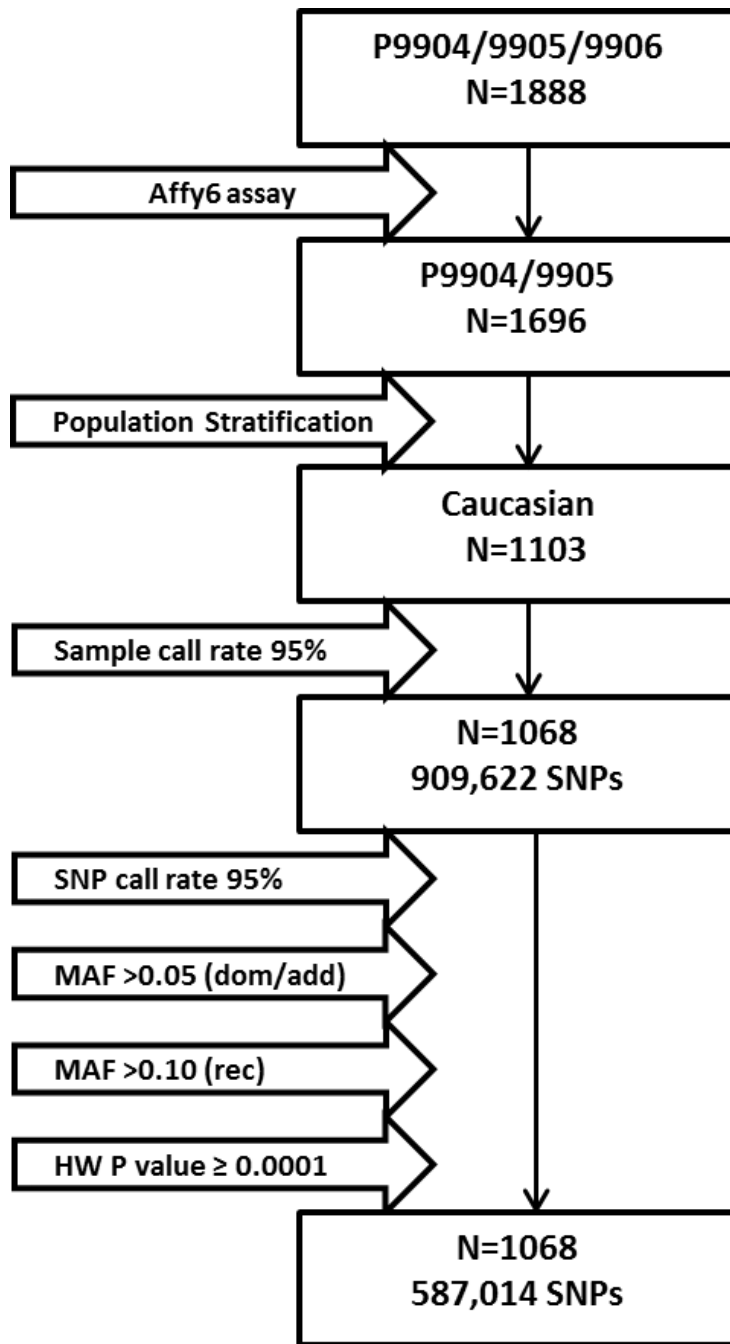


Figure 6.2. **Flowchart of ALL patient selection in the final GWAS analysis.**

MAF: minor allele frequency; dom: dominant model; add: additive model; rec: recessive model; HW: Hardy Weinberg equilibrium.

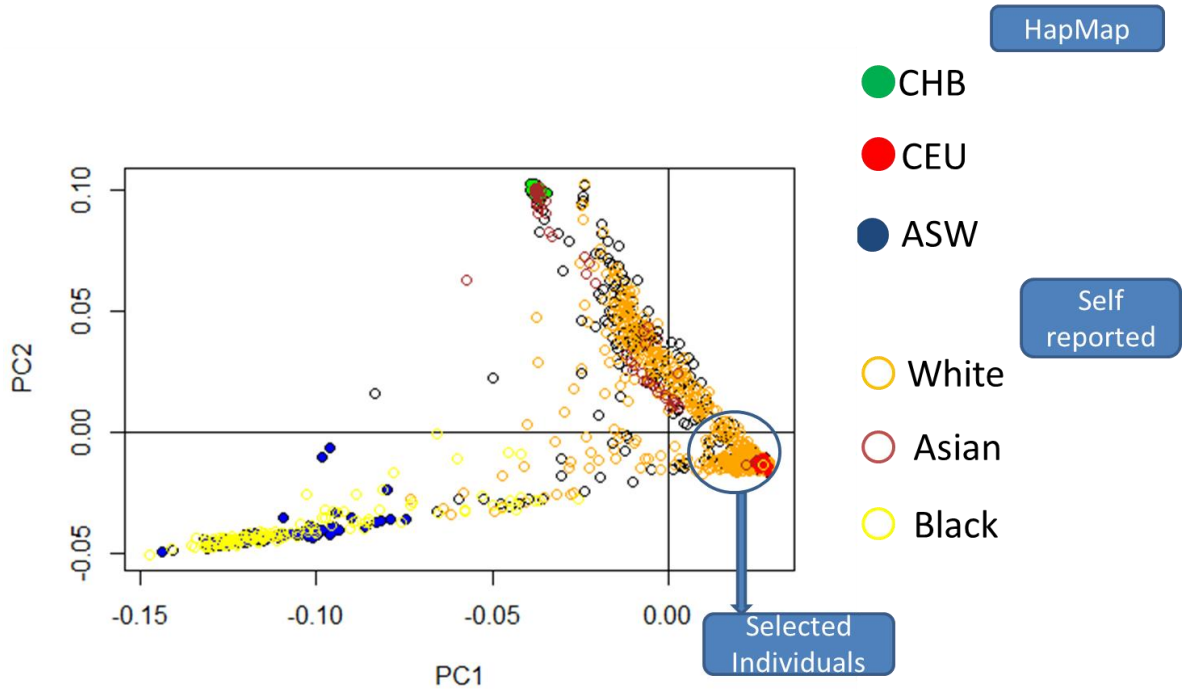


Figure 6.3. **Population stratification using principal component analysis.**

HapMap reference was used to compare the data. CHB: Han Chinese in Beijing, China. CEU:Utah residents with Northern and Western European ancestry from the CEPH. ASW: African ancestry in Southwest USA.

### **3.2 Identification of Genomic Loci Associated with ALL Neuropathy Risk in GWAS**

#### **Analysis**

The initial step of the analysis coded homozygous major alleles, heterozygous, and homozygous minor alleles as 0, 1 and 2, respectively using PLINK. After the data was manipulated, a Cox proportional hazard model was performed to test association between markers and the primary and the secondary neuropathy in three genetic models. The p-values in  $-\log_{10}$  were shown across chromosomes in Manhattan plot and QQ plot for primary neuropathy and secondary neuropathy in Figure 6.4 A, B, C, D, E and F and Figure 6.5A, B, C, D, E and F, respectively.

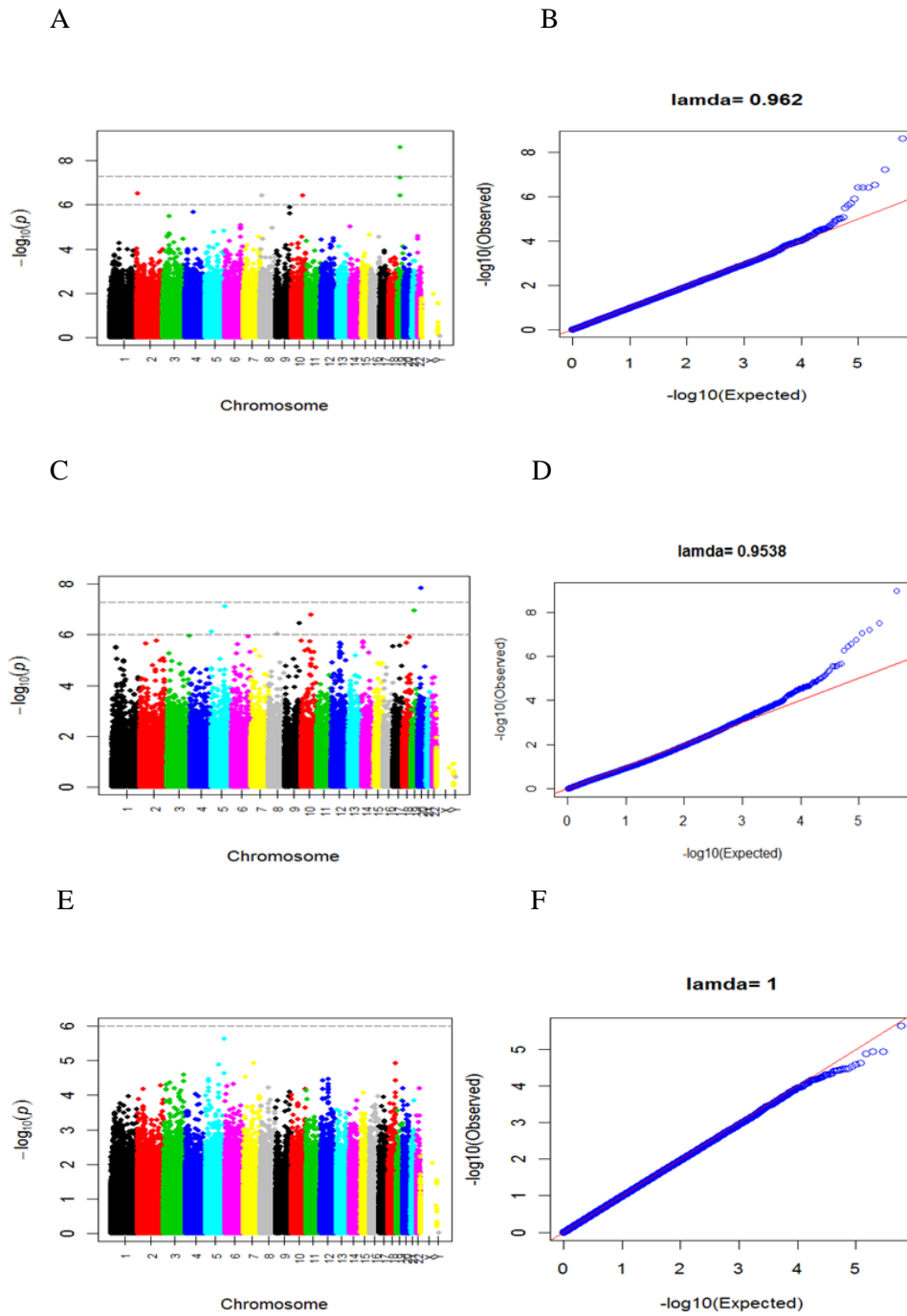


Figure 6.4. **Manhattan plot and QQ plot of GWAS association analysis with primary neuropathy.**

A and B: dominant model; C and D: recessive model; E and F: additive model.

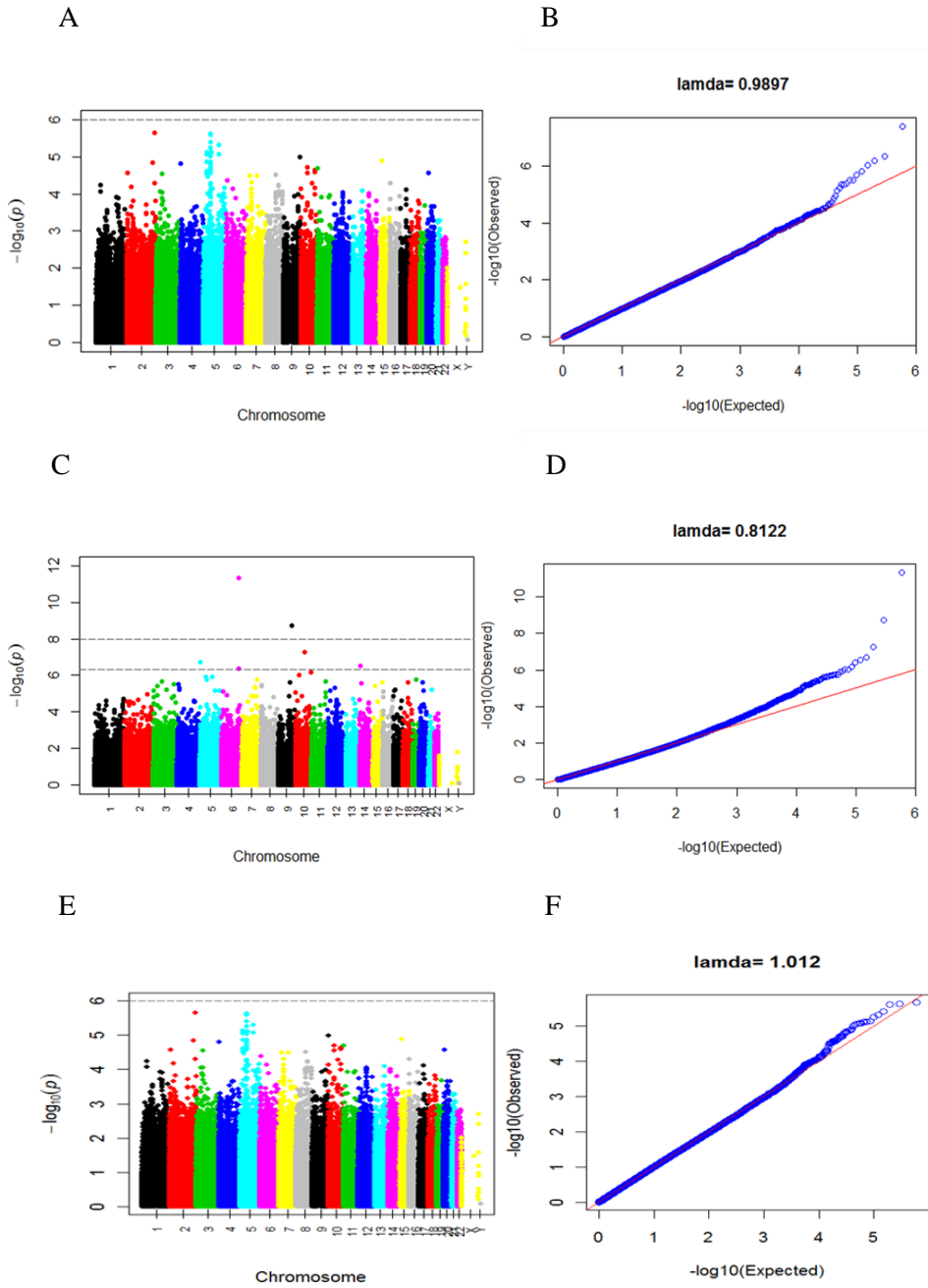


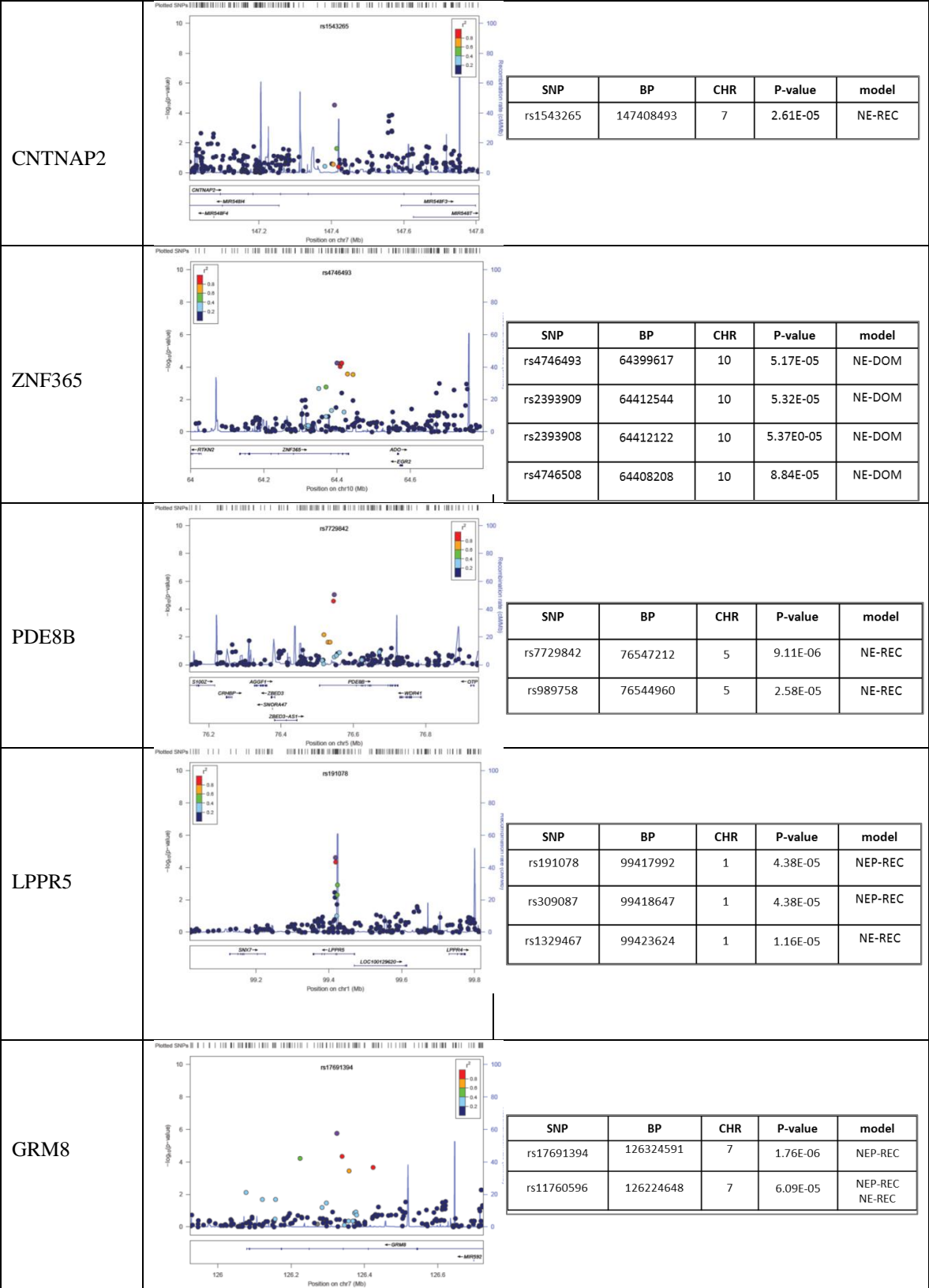
Figure 6.5. **Manhattan plot and QQ plot of GWAS association analysis with secondary neuropathy.**

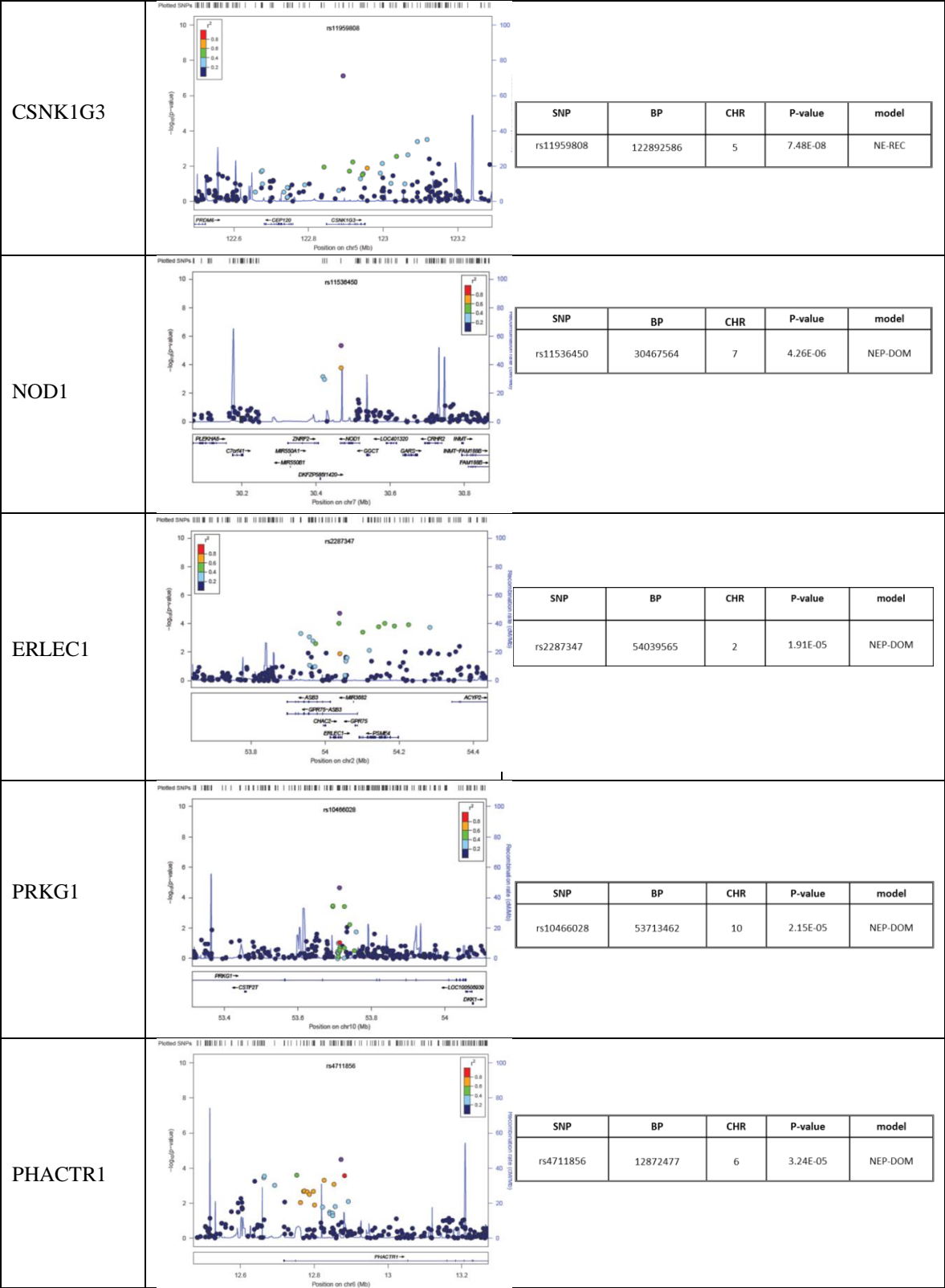
A and B: dominant model; C and D: recessive model; E and F: additive model.

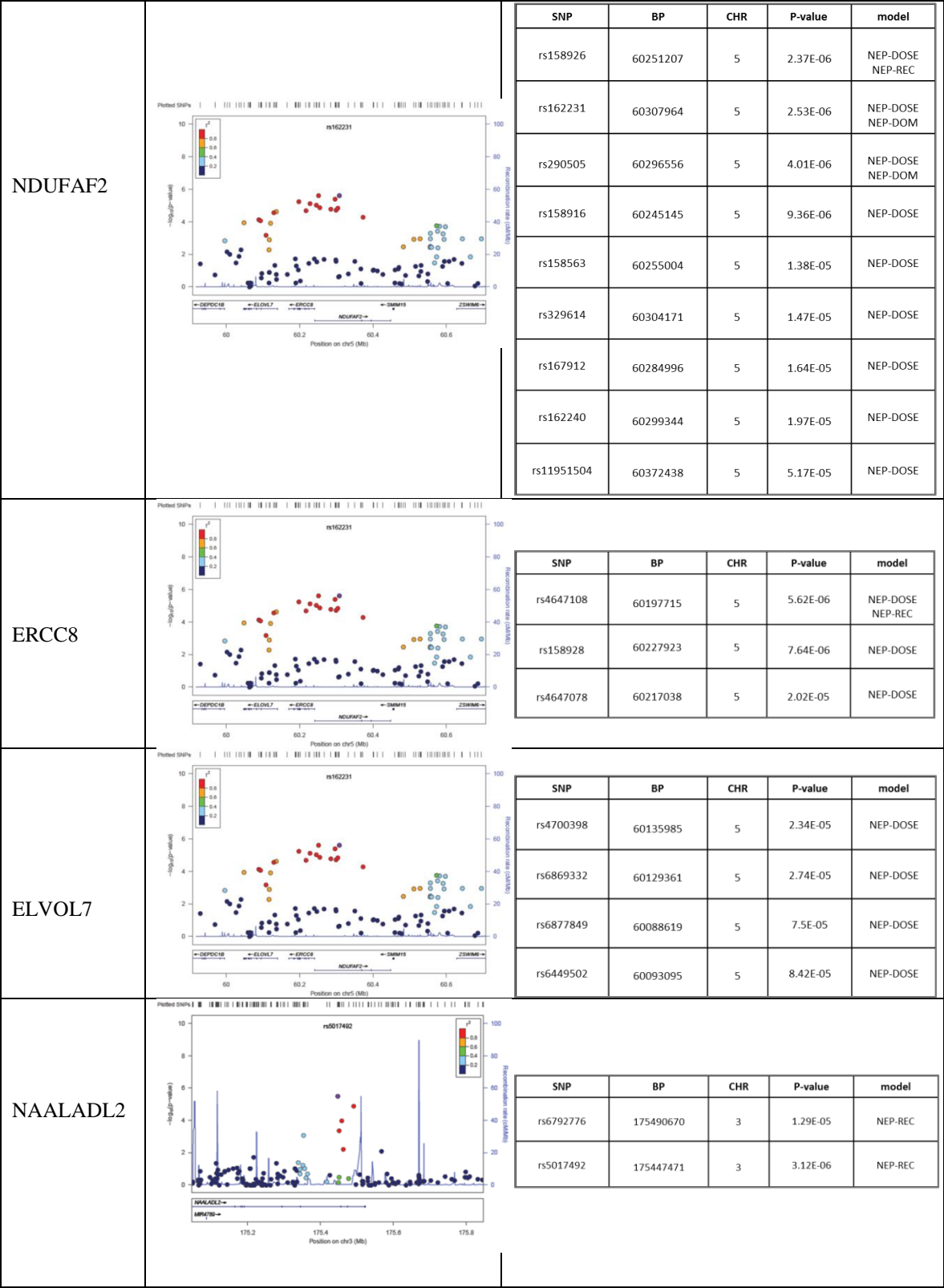


In order to determine the candidate SNPs for validation, a threshold of p-value of  $1 \times 10^{-4}$  was selected. Those candidate SNPs with p-value  $< 1 \times 10^{-5}$  were further investigated and identified the location of gene region and visualized closely in zoom-in plot (Figure 6.6). A total of 23 genes were identified across both neuropathy phenotypes and three genetic model structures. The rsID, position, minor allele, MAF and hazard ratio of those candidate SNPs as well as its corresponding gene were summarized in Table 6.1.

Gene	Zoom-in plot	Corresponding SNPs																														
NEK6		<table border="1"> <thead> <tr> <th>SNP</th> <th>BP</th> <th>CHR</th> <th>P-value</th> <th>model</th> </tr> </thead> <tbody> <tr> <td>rs1722478</td> <td>127065836</td> <td>9</td> <td>1.28E-06</td> <td>NE-DOM</td> </tr> <tr> <td>rs4838140</td> <td>127010114</td> <td>9</td> <td>2.34E-06</td> <td>(not on NEK6)</td> </tr> <tr> <td>rs16927327</td> <td>127047450</td> <td>9</td> <td>3.57E-07</td> <td>NE-REC</td> </tr> <tr> <td>rs4838158</td> <td>127068176</td> <td>9</td> <td>2.16E-05</td> <td>NE-REC</td> </tr> </tbody> </table>	SNP	BP	CHR	P-value	model	rs1722478	127065836	9	1.28E-06	NE-DOM	rs4838140	127010114	9	2.34E-06	(not on NEK6)	rs16927327	127047450	9	3.57E-07	NE-REC	rs4838158	127068176	9	2.16E-05	NE-REC					
SNP	BP	CHR	P-value	model																												
rs1722478	127065836	9	1.28E-06	NE-DOM																												
rs4838140	127010114	9	2.34E-06	(not on NEK6)																												
rs16927327	127047450	9	3.57E-07	NE-REC																												
rs4838158	127068176	9	2.16E-05	NE-REC																												
ZFPM2		<table border="1"> <thead> <tr> <th>SNP</th> <th>BP</th> <th>CHR</th> <th>P-value</th> <th>model</th> </tr> </thead> <tbody> <tr> <td>rs4734869</td> <td>106512828</td> <td>8</td> <td>8.40E-06</td> <td>NE-DOM</td> </tr> </tbody> </table>	SNP	BP	CHR	P-value	model	rs4734869	106512828	8	8.40E-06	NE-DOM																				
SNP	BP	CHR	P-value	model																												
rs4734869	106512828	8	8.40E-06	NE-DOM																												
FHIT		<table border="1"> <thead> <tr> <th>SNP</th> <th>BP</th> <th>CHR</th> <th>P-value</th> <th>model</th> </tr> </thead> <tbody> <tr> <td>rs241704</td> <td>60512889</td> <td>3</td> <td>2.53E-06</td> <td>NE-DOM</td> </tr> <tr> <td>rs10451993</td> <td>60566492</td> <td>3</td> <td>1.52E-05</td> <td>NE-DOM</td> </tr> <tr> <td>rs2734365</td> <td>60565903</td> <td>3</td> <td>1.97E-05</td> <td>NE-DOM</td> </tr> <tr> <td>rs6805296</td> <td>60544132</td> <td>3</td> <td>6.05E-05</td> <td>NE-DOM</td> </tr> <tr> <td>rs7651468</td> <td>60572073</td> <td>3</td> <td>9.21E-05</td> <td>NE-DOM</td> </tr> </tbody> </table>	SNP	BP	CHR	P-value	model	rs241704	60512889	3	2.53E-06	NE-DOM	rs10451993	60566492	3	1.52E-05	NE-DOM	rs2734365	60565903	3	1.97E-05	NE-DOM	rs6805296	60544132	3	6.05E-05	NE-DOM	rs7651468	60572073	3	9.21E-05	NE-DOM
SNP	BP	CHR	P-value	model																												
rs241704	60512889	3	2.53E-06	NE-DOM																												
rs10451993	60566492	3	1.52E-05	NE-DOM																												
rs2734365	60565903	3	1.97E-05	NE-DOM																												
rs6805296	60544132	3	6.05E-05	NE-DOM																												
rs7651468	60572073	3	9.21E-05	NE-DOM																												
MICAL3		<table border="1"> <thead> <tr> <th>SNP</th> <th>BP</th> <th>CHR</th> <th>P-value</th> <th>model</th> </tr> </thead> <tbody> <tr> <td>rs2075455</td> <td>18373885</td> <td>22</td> <td>2.36E-05</td> <td>NE-DOM</td> </tr> <tr> <td>rs2165971</td> <td>18364579</td> <td>22</td> <td>3.49E-05</td> <td>NE-DOM</td> </tr> </tbody> </table>	SNP	BP	CHR	P-value	model	rs2075455	18373885	22	2.36E-05	NE-DOM	rs2165971	18364579	22	3.49E-05	NE-DOM															
SNP	BP	CHR	P-value	model																												
rs2075455	18373885	22	2.36E-05	NE-DOM																												
rs2165971	18364579	22	3.49E-05	NE-DOM																												







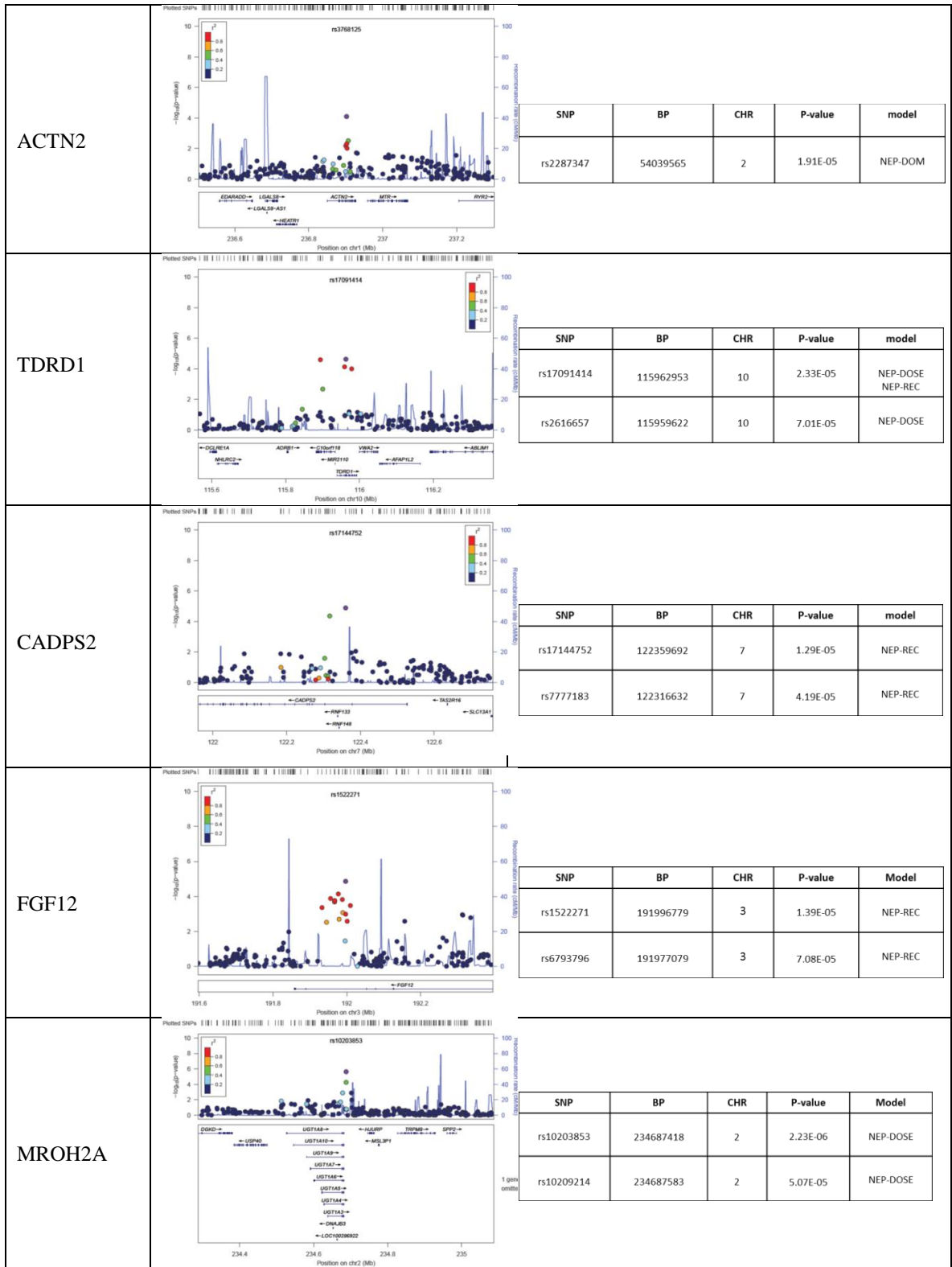


Figure 6.6. Zoom-in plots of each candidate gene from GWAS study and their corresponding SNPs. NE:neuropathy( primary neuropathy); NEP:neuropathy and neuropathic

pain (secondary neuropathy); DOM: dominant model; REC: recessive model; DOSE: additive model.

**Table 6.1. Candidate genes selected from GWAS study.**

Top SNPs from GWAS to vincristine induced primary neuropathy								
SNP	Chr	Position	Gene	Genetic model	Minor Allele	MAF	HR	P
rs17222478	9	127065836	NEK6	Dominant	T	0.11	0.129	1.28E-06
rs16927327	9	127047450	NEK6	Recessive	G	0.15	6.008	3.57E-07
rs4838158	9	127068176	NEK6	Recessive	T	0.11	6.949	2.16E-05
rs4734869	8	106512828	ZFPM2	Dominant	T	0.25	0.287	8.4E-06
rs241704	3	69512778	FHIT	Dominant	T	0.38	0.344	2.53E-06
rs10451993	3	60566492	FHIT	Dominant	T	0.45	0.400	1.52E-05
rs2734365	3	60565903	FHIT	Dominant	T	0.46	0.404	1.97E-05
rs6805296	3	60544132	FHIT	Dominant	C	0.27	0.321	6.05E-05
rs7651468	3	60572073	FHIT	Dominant	T	0.1	0.101	9.21E-05
rs2075455	22	18373885	MICAL3	Dominant	T	0.35	0.360	2.36E-05
rs2165971	22	18364579	MICAL3	Dominant	G	0.37	0.376	3.49E-05
rs1543265	7	147408493	CNTNAP2	Recessive	T	0.36	0.819	2.61E-05
rs4746493	10	64399617	ZNF365	Dominant	T	0.11	0.796	5.17E-05
rs2393909	10	64412544	ZNF365	Dominant	G	0.15	2.376	5.32E-05
rs2393908	10	64412122	ZNF365	Dominant	T	0.15	2.367	5.37E-05
rs4746508	10	64408208	ZNF365	Dominant	G	0.15	2.331	8.84E-05
rs7729842	5	76547212	PDE8B	Recessive	G	0.38	2.835	9.11E-06
rs989758	5	76544960	PDE8B	Recessive	G	0.38	2.686	2.58E-05
rs191078	1	99417992	LPPR5	Recessive	T	0.16	4.772	4.38E-05
rs11760596	7	126224648	GRM8	Recessive	G	0.22	3.791	6.09E-05
rs11959808	5	122892586	CSNK1G3	Recessive	G	0.30	0.635	7.48E-08



Top SNPs from GWAS to vincristine induced secondary neuropathy								
SNP	Chr	Position	Gene	Genetic model	Minor Allele	MAF	HR	P
rs309087	1	99418647	LPPR5	Recessive	G	0.14	3.833	4.38E-05
rs1329467	1	99423624	LPPR5	Recessive	G	0.31	2.574	1.15E-05
rs17691394	7	126324591	GRM8	Recessive	T	0.13	4.501	1.76E-06
rs11760596	7	126224648	GRM8	Recessive	G	0.41	3.105	6.09E-05
rs11536450	7	30467564	NOD1	Dominant	G	0.13	1.011	4.26E-06
rs2287347	2	54039565	ERLEC1	Dominant	G	0.20	2.226	1.91E-05
rs10466028	10	53713462	PRKG1	Dominant	G	0.36	2.083	2.15E-05
rs4711856	6	12872477	PHACTR1	Dominant	G	0.06	2.418	3.24E-05
rs158926	5	60251207	NDUFAF2	Additive	C	0.19	1.947	2.37E-06
rs290505	5	60307964	NDUFAF2	Additive	T	0.20	0.532	2.53E-06
rs162231	5	60296556	NDUFAF2	Additive	T	0.20	1.887	4.01E-06
rs158916	5	60245145	NDUFAF2	Additive	G	0.20	0.545	9.36E-06
rs158563	5	60255004	NDUFAF2	Additive	T	0.34	1.796	1.38E-05
rs329614	5	60304171	NDUFAF2	Additive	T	0.17	1.802	1.47E-05
rs167912	5	60284996	NDUFAF2	Additive	T	0.20	0.558	1.64E-06
rs162240	5	60299344	NDUFAF2	Additive	G	0.35	1.796	1.97E-05
rs11951504	5	60372438	NDUFAF2	Additive	G	0.17	1.745	5.17E-05
rs4647108	5	60197715	ERCC8	Additive	G	0.21	1.849	5.62E-06
rs158928	5	60227923	ERCC8	Additive	T	0.21	1.837	7.64E-06
rs4647078	5	60217038	ERCC8	Additive	G	0.20	1.788	2.02E-05
rs4700398	5	60135985	ELOVL7	Additive	G	0.26	1.723	2.34E-05
rs6869332	5	60129361	ELOVL7	Additive	T	0.20	0.567	2.74E-05
rs6877849	5	60088619	ELOVL7	Additive	G	0.21	0.582	7.5E-05
rs6449502	5	60093095	ELOVL7	Additive	T	0.21	0.581	8.42E-05
rs5017492	3	175447471	NAALADL2	Recessive	T	0.27	3.176	3.12E-06
rs6792776	3	175490670	NAALADL2	Recessive	G	0.25	2.944	1.29E-05
rs3768125	1	236901818	ACTN2	Dominant	T	0.42	0.502	7.48E-05
rs17091414	10	115962953	TDRD1	Additive	G	0.23	1.801	2.33E-05
rs2616657	10	115969622	TDRD1	Additive	G	0.17	1.724	7.01E-05
rs17144752	7	122359692	CADPS2	Recessive	G	0.12	1.38	1.29E-05
rs7777183	7	122316632	CADPS2	Recessive	C	0.30	2.897	4.19E-05
rs1522271	3	191996779	FGF12	Recessive	G	0.43	0.865	1.39E-05
rs6793796	3	191977079	FGF12	Recessive	G	0.41	0.831	7.08E-05
rs10203853	2	234687418	MROH2A	Additive	T	0.24	0.943	2.23E-06
rs10209214	2	234687583	MROH2A	Additive	C	0.34	0.606	5.07E-05

### **3.3 Identification of Genomic Loci Associated with ALL Neuropathy Risk in Gene-Enrichment Analysis**

A gene-enrichment analysis based on a bottom-up approach was conducted to explore genes previously identified as drug targets for treating neuropathy or pain. A total of 101 genes were identified for evaluation in this analysis. Among these 101 target genes, the SNPs were enriched in the gene regions were tested using Fisher's exact test. The threshold of significance was set to be p-value at 0.05. A total of eight genes including CACNA1D, SLC29A4, CACNA1C, GRIK1, SCN8A, CACNB1, GRIN3A and SLC22A1 with odds ratio greater than 1 were identified and listed in Table 6.2. The zoom-in plots of those genes with their p-values were summarized in Figure 6.7.

<b>Gene</b>	<b>Length (bp)</b>	<b>Location</b>
CACNA1D	317,810	53,528,683-53,846,492
SLC29A4	29,705	5,314,000-5,343,704
CACNA1C	727,164	2,079,952-2,807,115
GRIK1	403,029	30,909,254-31,312,282
SCN8A	218,258	51,984,050-52,202,307
CACNB1	24,248	37,329,709-37,353,956
GRIN3A	169,229	104,331,634-104,500,862
SLC22A1	36,930	160,542,821-160,579,750

Table 6.2. **Candidate genes selected from gene-enrichment analysis.**

Gene	P-value	Odds ratio	Zoom-in plot
CACNAID	2.96E-10	4.405549	
SLC29A4	0.000133	24.02168	
CACNA1C	0.004477	1.894156	
GRIK1	0.008791	2.109875	

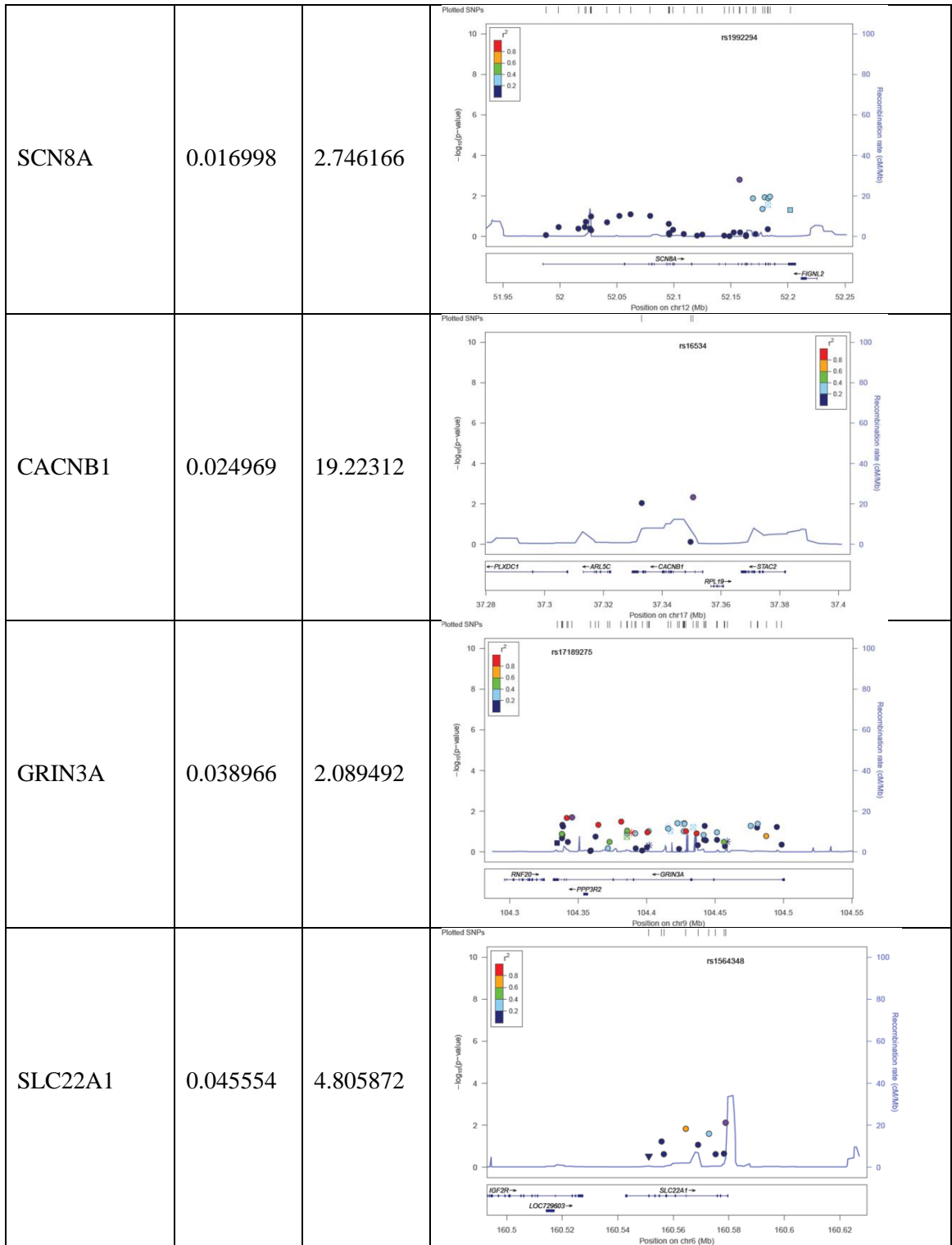


Figure 6.7. Zoom-in plots of each candidate gene from Gene enrichment analysis study and their corresponding P-value and odds ratio.

### **3.4 Pathway Identification of Genomic Loci Associated with ALL Neuropathy Risk**

The candidate genes selected from the association analysis were further evaluated in an interactive pathway analysis. Only one pathway, pregnenolone biosynthesis, passed the p-value threshold. In addition, two networks were selected based on the score that the system calculated. The first network had a score of 35 with 14 focus moles including ACTN2, CNTNAP2, CSNK1G, ERCC8, ERLEC1, FGF12, GRM8, MICAL3, NAALADL2, NDUFAF2, NEK6, NOD1, PHACTR1 and TDRD1; it is cell-to cell signaling and interaction, infectious disease, inflammatory (Figure 6.8). The second network with a score of 15 with 7 focus moles including CADPS2, ELOVL7, FHIT, LPPR5, PDE8B, PRKG1 and ZFPM2 is tissue development, tumor morphology, cancer (Figure 6.9).

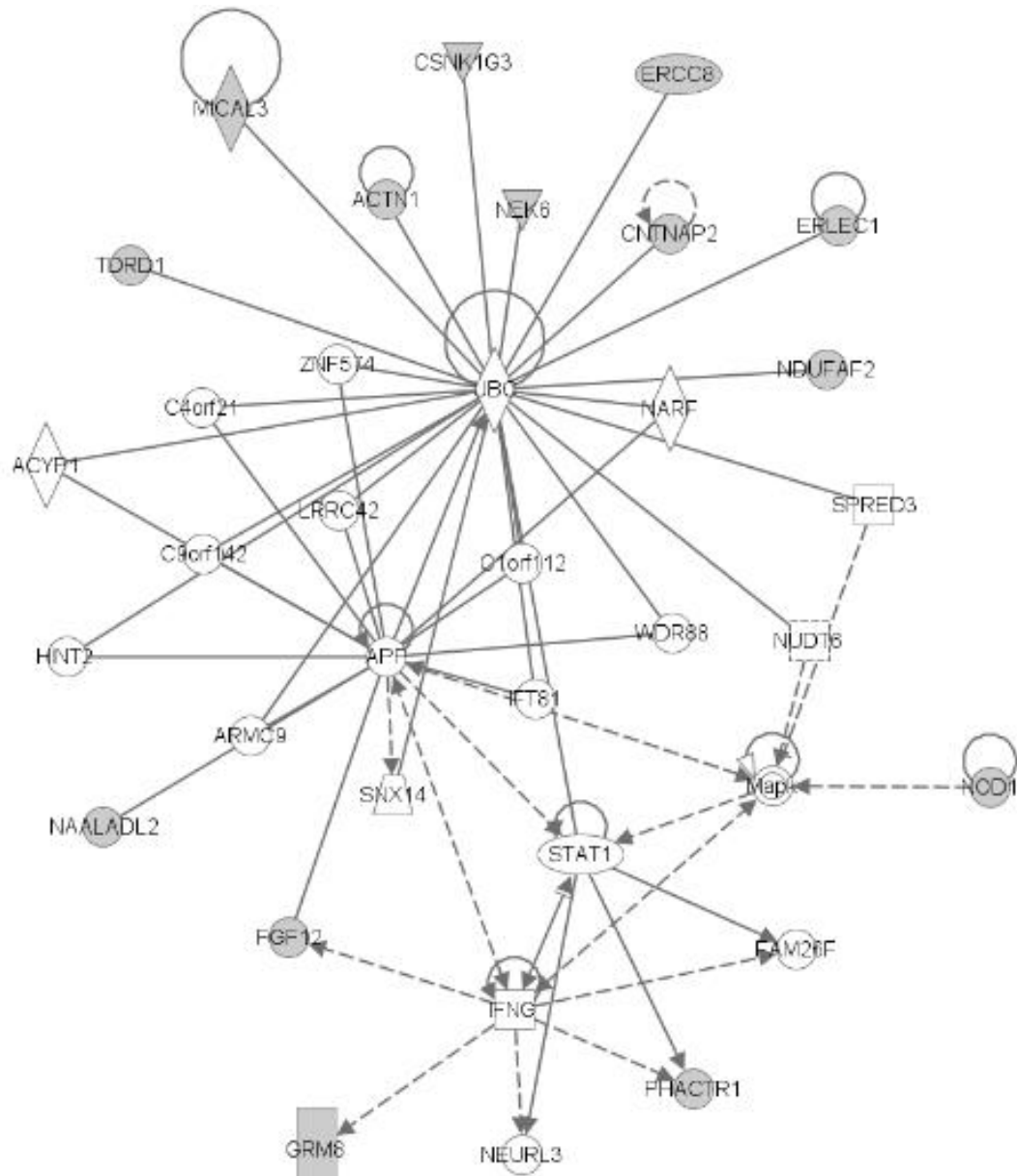


Figure 6.8. **The network of cell-to cell signaling and interaction, infectious disease and inflammatory.** 14 genes included ACTN2, CNTNAP2, CSNK1G, ERCC8, ERLEC1, FGF12, GRM8, MICAL3, NAALADL2, NDUFAF2, NEK6, NOD1, PHACTR1 and TDRD1 were identified in this network.

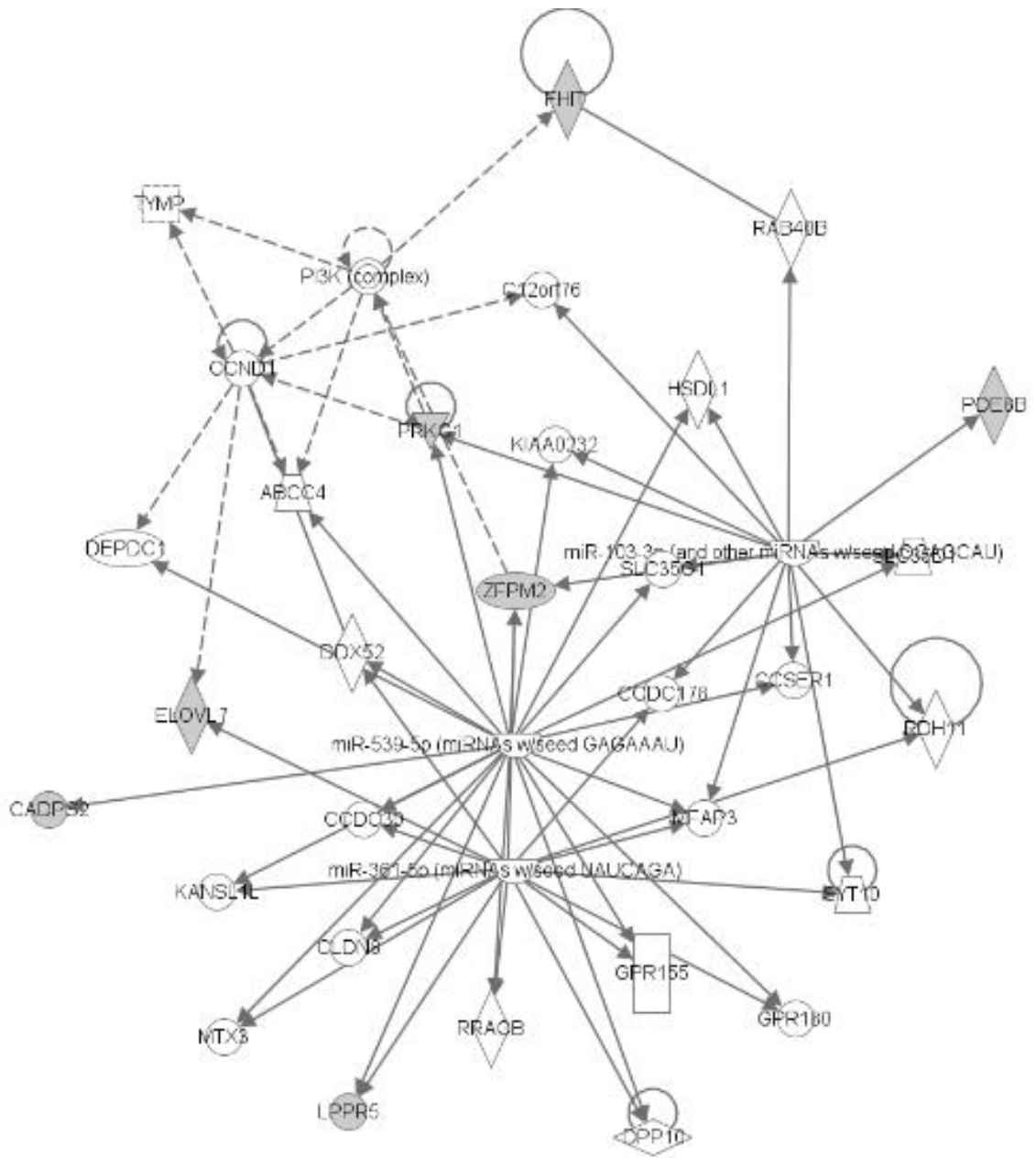


Figure 6.9. **The network of tissue development, tumor morphology and cancer.**

7 genes including CADPS2, ELOVL7, FHD, LPPR5, PDE8B, PRKG1 and ZFPM2 were identified in this network.



### **3.5 The Change in Gene Expression of Genomic Loci Associated with ALL Neuropathy Risk**

The correlation of genetic information and toxicity was further evaluated at the gene expression level to consider whether the genes of interest are functional and expressed. A study of a fibrosarcoma cell line response to various cytostatic drugs was utilized in which the control cell line was compared to that (Figure 6.10). The gene expression level of the sample treated with vincristine was normalized by the control. A total of 4619 and 16893 SNPs reached the threshold of greater than 50% down-regulation and up-regulation, respectively. Based on the QQ plots, most of the p-values of SNPs were shown to approximately lie on the line  $y = x$  meaning that the p-values of the SNPs are likely to be randomly distributed. Few SNPs were deviated from the line  $y=x$ . These are summarized in Figures 6.11 and 6.12 with their corresponding genes.

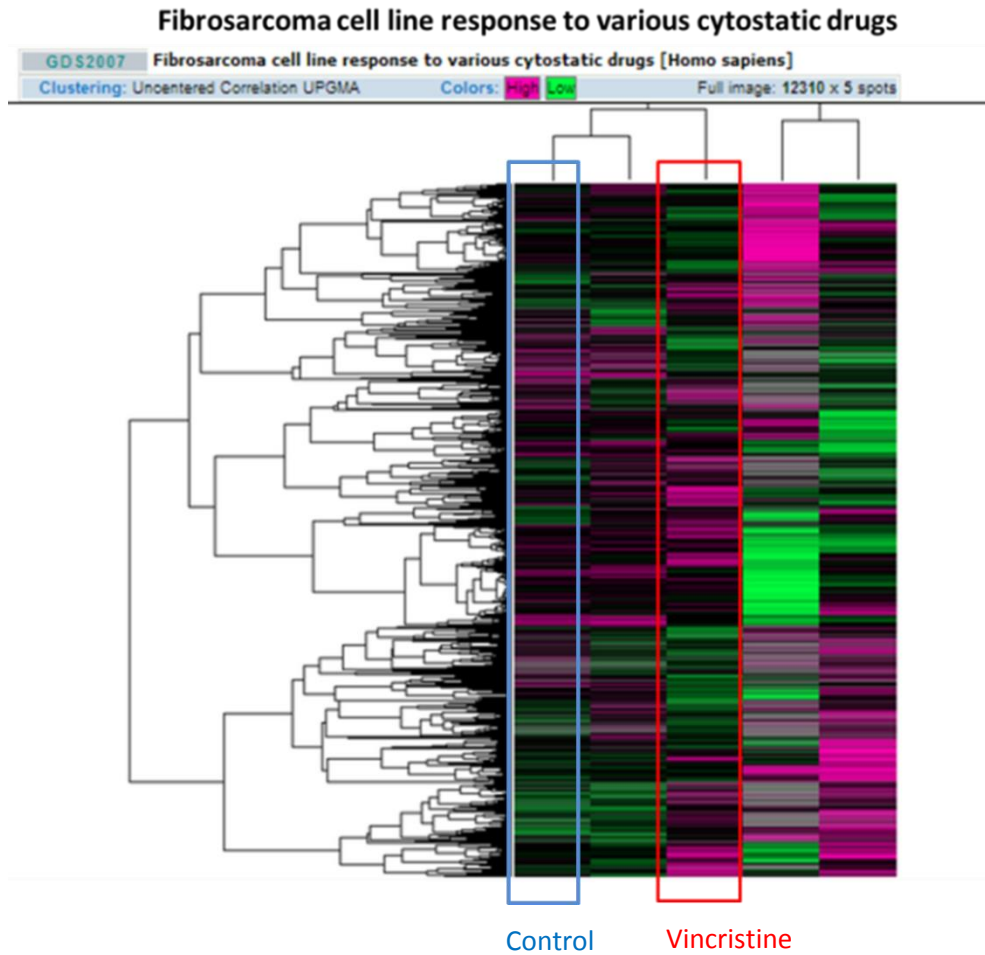


Figure 6.10. **Fibrosarcoma cell line response to various cytostatic drugs in the GEO database.** The cell line was treated with Vincristine, doxorubicin or actinomycin.

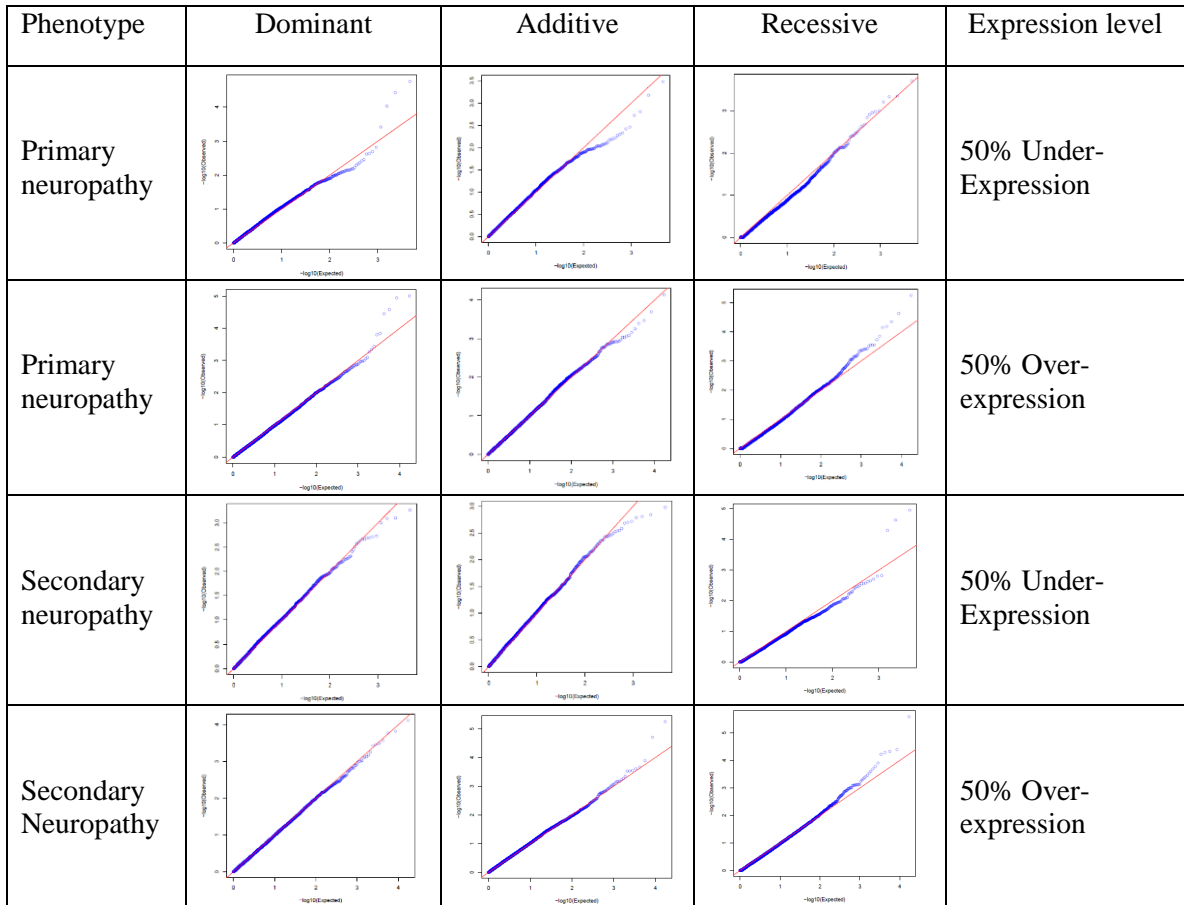


Figure 6.11. QQ-plots of genomic loci associated with ALL neuropathy risk in fibrosarcoma cell line.

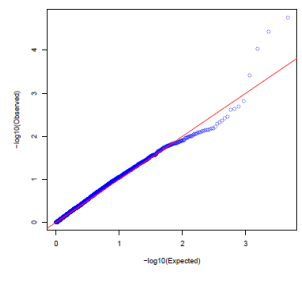
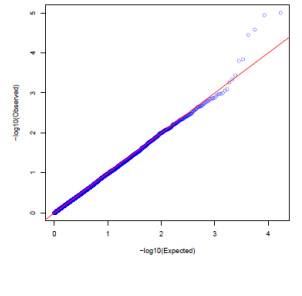
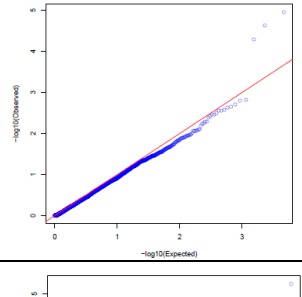
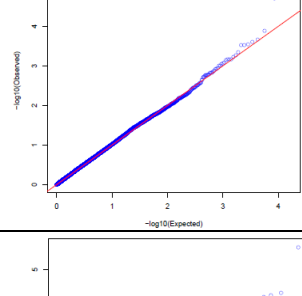
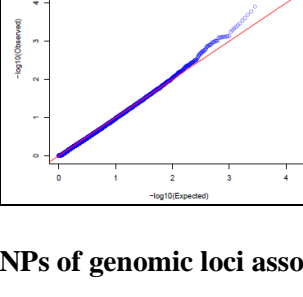
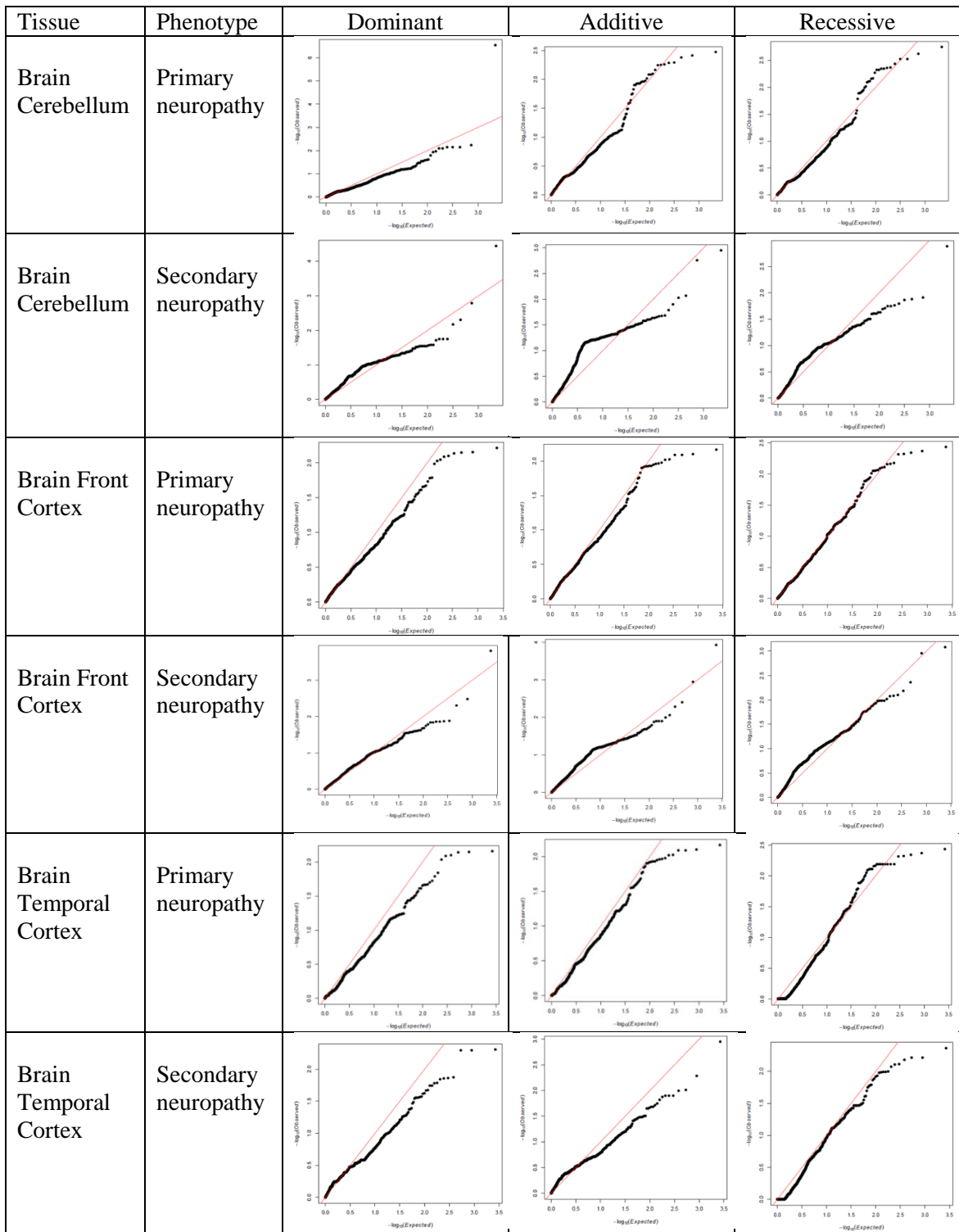
Phenotype	Expression level	Genetic model	QQ-plot	SNP										
Primary neuropathy	50% Under-Expression	Dominant		<table border="1"> <thead> <tr> <th>SNP</th> <th>Gene</th> </tr> </thead> <tbody> <tr> <td>rs2238034</td> <td>CACNA1C</td> </tr> <tr> <td>rs6882294</td> <td>ATG10</td> </tr> <tr> <td>rs7696226</td> <td>OTUD4</td> </tr> <tr> <td>rs7712216</td> <td>ATG10</td> </tr> </tbody> </table>	SNP	Gene	rs2238034	CACNA1C	rs6882294	ATG10	rs7696226	OTUD4	rs7712216	ATG10
SNP	Gene													
rs2238034	CACNA1C													
rs6882294	ATG10													
rs7696226	OTUD4													
rs7712216	ATG10													
Primary neuropathy	50% Over-Expression	Dominant		<table border="1"> <thead> <tr> <th>SNP</th> <th>Gene</th> </tr> </thead> <tbody> <tr> <td>rs1867353</td> <td>MICAL3</td> </tr> <tr> <td>rs2075455</td> <td>MICAL3</td> </tr> <tr> <td>rs2165971</td> <td>MICAL3</td> </tr> <tr> <td>rs2167143</td> <td>KCNAB1</td> </tr> </tbody> </table>	SNP	Gene	rs1867353	MICAL3	rs2075455	MICAL3	rs2165971	MICAL3	rs2167143	KCNAB1
SNP	Gene													
rs1867353	MICAL3													
rs2075455	MICAL3													
rs2165971	MICAL3													
rs2167143	KCNAB1													
Secondary neuropathy	50% Under-Expression	Recessive		<table border="1"> <thead> <tr> <th>SNP</th> <th>Gene</th> </tr> </thead> <tbody> <tr> <td>rs4543885</td> <td>NEBL</td> </tr> <tr> <td>rs6751734</td> <td>DNAH7</td> </tr> <tr> <td>rs8119756</td> <td>ADA</td> </tr> </tbody> </table>	SNP	Gene	rs4543885	NEBL	rs6751734	DNAH7	rs8119756	ADA		
SNP	Gene													
rs4543885	NEBL													
rs6751734	DNAH7													
rs8119756	ADA													
Secondary neuropathy	50% Over-expression	Additive		<table border="1"> <thead> <tr> <th>SNP</th> <th>Gene</th> </tr> </thead> <tbody> <tr> <td>rs4647078</td> <td>ERCC8</td> </tr> <tr> <td>rs4647108</td> <td>ERCC8</td> </tr> </tbody> </table>	SNP	Gene	rs4647078	ERCC8	rs4647108	ERCC8				
SNP	Gene													
rs4647078	ERCC8													
rs4647108	ERCC8													
Secondary neuropathy	50% Over-expression	Recessive		<table border="1"> <thead> <tr> <th>SNP</th> <th>Gene</th> </tr> </thead> <tbody> <tr> <td>rs2717425</td> <td>PTPRB</td> </tr> </tbody> </table>	SNP	Gene	rs2717425	PTPRB						
SNP	Gene													
rs2717425	PTPRB													

Figure 6.12. Genes and corresponding SNPs of genomic loci associated with ALL neuropathy risk in fibrosarcoma cell line

### **3.6 Genomic Loci Associated with ALL Neuropathy Risk in Genotype-Tissue Expression eQTL**

In addition to the GEO database, the correlation was also tested in the GTEx database. In this database, mRNA expressions were estimated in the different tissues. Based on current available tissues in the database, this study was carried out to analyze data from four brain tissues which are potentially related to our phenotype of interest. After matching the SNPs between our data and the GTEx database, p-values of these matched SNPs were further visualized in QQ plots across three genetic models and two phenotypes in Figure 6.13. Similar to GEO strategy, the SNPs with deviated  $-\log_{10}$  p-values from the line  $y = x$  were selected. eQTLs can be described as being cis, where the genotyped marker is within 2 MB of the expressed gene, or trans, in which the genotyped marker is far away from the expressed gene or even on another chromosome (Gibbs et al., 2010). The corresponding genes of the selected SNPs were based on the probe position where the gene expression level was measured in the original eQTL study. As Figure 6.14 showed, in the brain cerebellum tissue, rs10153783 was found to correlate to the gene expression of collectin sub-family member 11 (COLEC11) in the dominant model for both primary and secondary neuropathies. Another SNP, rs359436, is associated with hydroxysteroid (17-beta) dehydrogenase 4 (HSD17B4) gene expression in both dominant and additive models for the secondary neuropathy.



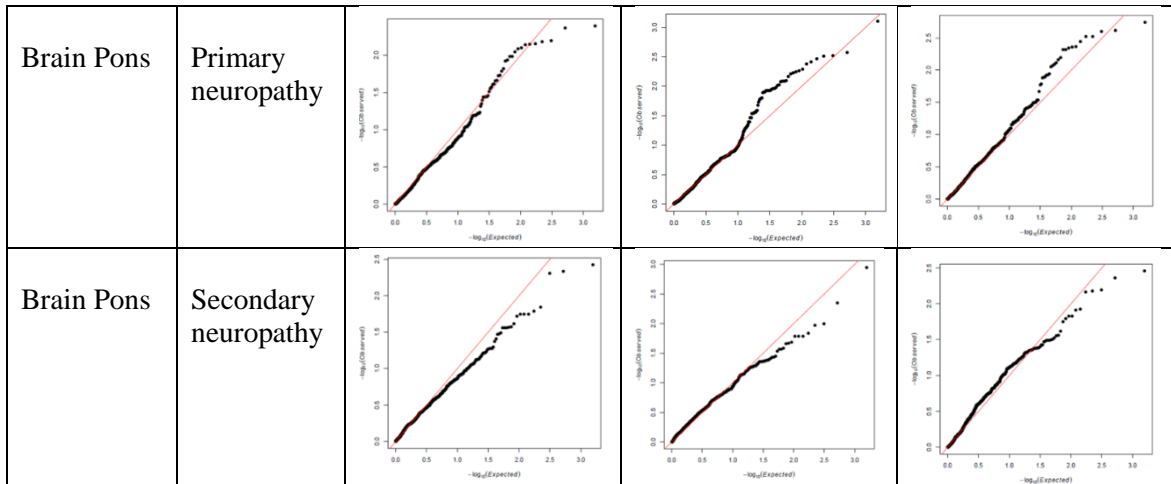


Figure 6.13. QQ-plots of genomic loci associated with ALL neuropathy risk in eQTL data of four brain tissues.

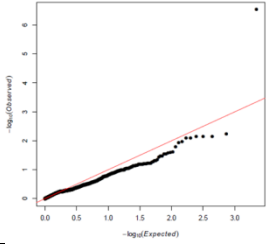
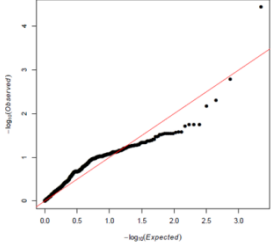
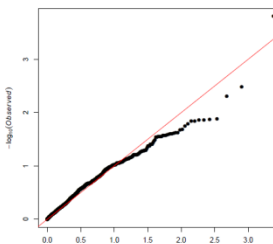
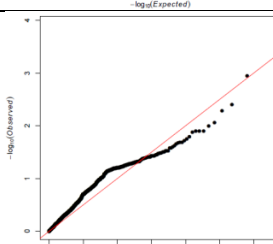
Tissue	Phenotype	Genetic model	QQ-plot	eQTL probe gene
Brain Cerebellum	Primary neuropathy	Dominant		COLEC11
Brain Cerebellum	Secondary neuropathy	Dominant		COLEC11
Brain Frontal Cortex	Secondary neuropathy	Dominant		HSD17B4
Brain Frontal Cortex	Secondary neuropathy	Additive		HSD17B4

Figure 6.14. **Indirect association of ALL neuropathy risk with eQTL expression.**



#### *4. Discussion*

Based on a previous study from Dennison and colleagues, they concluded that CYP3A5 expressers experienced less vincristine-induced peripheral neuropathy and had lower metabolic ratios compared to CYP3A5 non-expressers (Dennison, et al., 2006). This finding stimulates our interests in identifying potential genetic explanations for the difference in the pharmacokinetic exposures in a large population and identifying associations between genetic markers and drug-induced neuropathy. In fact, a subgroup of patients had to discontinue the treatment because of the severity of the neuropathy and its major impact on their quality of life. Even though this study focused on initial discovery of association, we believe that this might be useful to suggest an alternative therapy for those patients before administration of vincristine. In this genome-wide association study, we identified several genetic loci associated with an incidence of VIPN  $\geq$  grade 2. Based on the top-ranked SNPs in Table 6.1, 23 genes were identified in the association with either the primary or secondary neuropathy.

Among these 23 genes, one example is CSNK1G3. CSNK1G3 (Casein kinase 1, gamma 3) is an isoform of a monomeric serine-threonine protein kinase. Casein kinase has been implicated in a wide range of signaling activities such as cell differentiation, proliferation and apoptosis. Sakurai et al. previously found that the percentage of CK1 $\epsilon$ -positive neurons and the expression level of CK1 $\epsilon$  protein were increased in dorsal root ganglion and the spinal cord of the neuropathic mice (Sakurai et al., 2009). Another example is Nek6 (NIMA-related kinase 6). As RNA interference depletion studies reported by Yin et al., Nek6 and family isoforms are required for cell cycle progression through mitosis. Nek6 together with Nek7 and 9 form a mitotically activated module,

which plays a role in mitotic progression and more specifically spindle organization. Loss of Nek6 can lead to failure of centrosome separation in prophase and formation of weak mitotic spindles with reducing microtubule density (Yin et al., 2003). This modification of microtubule dynamic through the change of Nek6 expression can potentially lead to an association with neuropathy. In addition, MICAL3 is microtubule associated monooxygenase, calponin and LIM domain containing, which is an important regulator for microtubule polymerization (Terman et al., 2002; Fischer et al., 2005). Furthermore, ERCC8 is excision-repair-cross-complementing group 8. A study has shown that mutations in ERCC8 are associated with Cockayne syndrome, which is characterized by some neurological abnormalities including dys- or demyelinating neuropathy (Ohnishi et al., 1987; Smits et al., 1982).

Additional eight genes were discovered in an association with neuropathy from the gene enrichment analysis. In these eight genes, three (CACNAID, CACNA1C, CACNB1) are calcium channel, voltage-dependent subunits, and one (SCN8A) is a sodium channel, voltage gated subunit (Bock et al., 2011; Tan et al., 2011; Samak et al., 2011; Hori et al., 2012; Kloiber et al., 2012; Yang et al., 2012; Soeiro-de-Souza et al., 2012; Burgess et al., 1995). In particular, the phenotypes of mice with a mutation in SCN8A gene were found to be consistent with a defect in a neuronal sodium channel expressed in motor neuron (Burgess et al., 1995). Another two genes are in the transporter category. One (SLC29A4) is an equilibrative nucleoside transporter 4, which catalyzes the reuptake of monoamines into presynaptic neurons (Duan et al., 2013; Barnes et al., 2006). Another (SLC22A1) is a solute carrier family 22 member 1 which is a polyspecific organic cation transporter in many organs for elimination of endogenous small organic cations and toxins (Koehler et

al., 1997). In addition, two glutamate receptor, ionotropic subunits (GRIK1, GRIN3A) were also identified in the analysis. These two receptors have a critical role in excitatory synaptic transmission and plasticity in the CNS (Braga et al., 2009; Kaminski et al., 2004; Gryder et al., 2003; Rogawski et al., 2003). They govern a range of physiological conditions including neurological disorders caused by excitotoxic neuronal injury, psychiatric disorders and neuropathic pain syndromes (Petroff et al., 2002). These can be potential targets for further understanding of VIPN mechanism.

In the gene expression analyses, the expression levels were changed over 50% in eleven genes when treated with vincristine in the fibrosarcoma cell line. Interestingly, two (MICAL3 and ERCC8) and one (CANA1C) of these eleven genes were previously identified in the association and gene enrichment analyses, respectively. These findings confirmed that these genetic variants altered the gene expression, and the changes in expression may lead to toxicity. Different to the GEO analysis, two genes (COLEC11 and HSD18B4) found in the eQTL data were in a trans- association with the probe genes, and this indicated an indirect association with lipid biosynthesis. In fact, similar to what we observed in the pathway analysis, an association between VIPN and lipid biosynthesis mechanism was shown.

With the design of the analyses by using different association approaches, some direct or indirect associations between genomic variations and VIPN in children with ALL were identified. In the discovery phase of potential targets, we came across some similar correlations observed from different approaches which provided an additional confirmation of the findings. These findings could be clinical relevance given the widespread use of vincristine in treating childhood cancers. To validate these

observations, candidate genes will be sequenced, and the association between genomic variations and VIPN will be further examined in the sequencing data.

**CHAPTERVII:** *Genetic signature to predict vincristine neuropathy and relapse in children with acute lymphoblastic leukemia*

*1.Introduction*

A genome-wide association study (GWAS) is an examination of many common genetic variants in different individuals to determine if any variant is associated with a trait, and a way for scientists to identify genes involved in human disease. This method searches the genome for small variations that occur more frequently in people with a particular disease than in people without the disease. Each study can look at hundreds or thousands of single-nucleotide polymorphisms (SNPs) at the same time. Researchers use data from this type of study to pinpoint genes that may contribute to a person's risk of developing a certain disease. Typically, GWAS focuses only on associations between SNPs and traits. There has been an increasing interest in relating genetic profiles to survival phenotypes such as time to event. Because of high dimensionality of SNP data, there is a limitation of collinearity in fitting a prediction model such as the Cox proportional hazards model (Sohn et al., 2009). To avoid the collinearity problem, several methods based on penalized Cox proportional hazards models have been proposed. One of the approaches that we used is a LASSO penalized Cox proportional hazard model (Tibshirani, 1997).

The benefit of using a LASSO approach, instead of other penalized regression Smodels, is that this approach allows shrinking coefficients of insignificant predictors to exact zero which can simplify the model building in selecting a complex signature. In addition, this approach may potentially improve the predictability of genetic biomarkers to drug-induced toxicity as well as drug efficacy and provide a comprehensive and

unique signature. The objective of the study is to apply an alternative modeling approach to evaluate genetic biomarkers in predicting vincristine-induced neuropathy and relapse in childhood acute lymphoblastic leukemia patients. With this signature approach, a risk assessment of patients treated with vincristine can be further applied to distinguish the clinical outcome in a population based on the accuracy of the model performance and to provide clinically useful information to avoid unnecessary exposure to patients who are likely to experience severe toxicity but not benefit from the treatment.

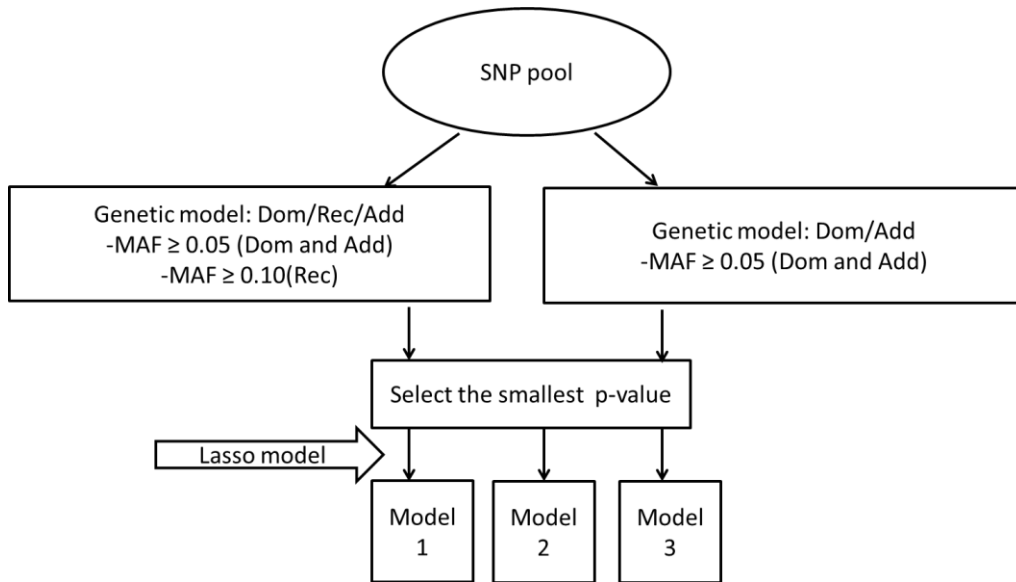


Figure 7.1. **Model strategies using the LASSO penalized regression approach.**

Model 1: demographic and clinical variables were tested in the LASSO model; Model 2: demographic, clinical variables and SNPs with  $p\text{-value} < 1 \times 10^{-5}$  from the previous univariate analysis were tested in the LASSO model; Model 3: demographic, clinical variables and SNPs with  $p\text{-value} < 1 \times 10^{-4}$  from the previous univariate analysis were tested in the LASSO model. Dom: dominant model; Rec: recessive model; Add: additive model.

## *2. Materials and Methods*

### **2.1 Acute Lymphoblastic Leukemia (ALL) Patient Population and Study Design**

Children with precursor B cell acute lymphoblastic leukemia from Pediatric Oncology Group (POG) 9904 and 9905 trials treated with 1.5mg/m<sup>2</sup>/dose vincristine doses and other chemotherapy treatments were included in the analysis as a training data set. The number of vincristine doses per patient over the course of treatment ranged from 18-23, depending on specific protocol and treatment arm. In this population, genotyping of germline samples from a total of 1068 subjects using an Affymatrix 6 assay remained in the analysis after a sequential data quality control including population stratification, sample and SNP missing call rate, Hardy-Weinberg disequilibrium and minor allele frequency. The details of data quality control were described in **CHAPTER VI** (Fareed et al., 2013). In addition, genotyping and clinical data of 122 ALL subjects from POG 9906 trial were used as a validation data set.

### **2.2 Phenotype Identification**

As an extension to the association studies between genomic variations and vincristine-induced neurotoxicity in Chapter VI, this study continually evaluated and focused on both primary and secondary neuropathy as toxicity endpoints. The definition of the primary neuropathy based on National Cancer Institute Common Terminology Criteria for Adverse Events, version 3.0, is when patients experience sensor or motor neuropathy (grade  $\geq 2$ ). Time to event in this case means that the time to the first neuropathy event.

A secondary neuropathy which is considered as any sensory neuropathy, motor neuropathy or neuropathic pain  $\geq$  grade 2 was also tested in the model. In addition to the



toxicity endpoints, an efficacy endpoint, the time to first relapse, was also taken into evaluation. By the definition of relapse of ALL patients, it included four types of relapse: an isolated bone marrow relapse: the presence of  $\geq 25\%$  lymphoblasts in a bone marrow aspirate following the first complete remission (CR); CNS relapse: Positive cytomorphology and  $> 5$  WBC/ $\mu\text{L}$  or positive cytomorphology with cerebrospinal fluid (CSF) WBC 0-4/ $\mu\text{L}$  on two successive occasions; testicular relapse: the histological evidence of lymphoblastic infiltration in one or both testes; and combined ALL relapse ( protocol 9004; protocol 9905).

### **2.3 Genome-Wide Association Study**

An association analysis of genotype frequencies with each phenotype was first carried out for the biomarker pre-selection process. Each phenotype was treated as a binary variable; ie, 1 for the event and 0 for no event. Clinical, demographic, and population stratification variables were also tested as covariates in the regression analyses. Genotype frequencies were evaluated in dominant, recessive and additive (gene-dose) models, and these analyses were performed in R coxph package (Therneau et al., 2014).

### **2.4 Genotype Imputation**

To infer missing genotypes in the data, an imputation software called MACH (Li et al., 2009; Li et al., 2010) was applied to impute sporadic missingness of typed markers using a phasing approach with the following command: `mach1 -d sample.dat -p sample.ped --states 200 -r 50 --dosage --prob --geno --phase` After the imputation, the genotypes were coded as 0, 1 and 2 for minor allele copies.

## 2.5 LASSO Penalized Regression Model

After the association analysis in the GWAS study, each SNP was extracted with a statistical p-value and a hazard ratio (Therneau et al., 2014). Then, the SNPs were selected based on a pre-selection process which was defined in the modeling strategies (Figure 7.1). First, the smallest p-value across three genetic models was identified for each SNP. After ranking the p-value of each SNP, three following signature models were conducted. In model 1, only demographic and clinical variables were tested in the association. In model 2, in addition to variables tested in model 1, SNPs with a p-value  $< 1 \times 10^{-5}$  were included in the association analysis. Furthermore, an extended SNP pool with p-value  $< 1 \times 10^{-4}$  were tested in model 3. The same modeling sequence was also examined with SNPs selected from only the dominant and additive models.

A particular feature of GWAS studies specifies that the dimension of the predictor space (number of genes) is typically larger than the number of samples (Sohn et al., 2009; Tibshirani et al., 1997; Benner et al., 2010; Fan et al., 2002). A penalized log partial likelihood in the LASSO method was utilized in conjunction with the proportional hazards model. The LASSO solution proposed by Tibshirani (Tibshirani et al., 1997) and Fan and Li (Fan et al., 2002) allows for parameter estimation by minimizing the log partial likelihood. The log partial likelihood was calculated using the following equation:  $l(\beta) = \sum_{i:C_i=1} (\beta X_i - \log \sum_{j:Y_j \geq Y_i} \exp(\beta X_j))$  subject to  $\sum |\beta_j| \leq s$  where  $s$  is a tuning parameter. If  $\sum |\beta_j| > s$ , the regression coefficients are shrunken toward zero. In this context, imputed SNPs as well as clinical and demographic variables were analyzed in relation to the phenotype using the LASSO penalized Cox model in R package glmnet

(Friedman et al., 2010). A cross-validation was implemented in the program to optimize lambda for model selection.

## 2.6 Model Evaluation

The predictive accuracy of a survival model was summarized using extensions of the proportion of variation explained by the model. The time-dependent sensitivity and specificity, and time-dependent receiver operating characteristic (ROC) curves were applied to evaluate the survival regression model by focusing on the correct classification rates. For survival data, there are several potential extensions of cross-sectional sensitivity and specificity. In this study, a cumulative/dynamic approach was performed to define the population who is classified as either a case or a control on the basis of vital status at time  $t$  at any fixed time  $t$ . Each individual serves as a control for time  $t < T_i$ , but as a case for time  $t \geq T_i$  (Heagerty et al., 2000; Heagerty et al., 2005). At any time point  $t$ , the sensitivity and specificity for a given threshold  $c$  are defined as

Sensitivity:  $P(M_i > c \mid T_i \leq t)$

Specificity:  $P(M_i \leq c \mid T_i > t)$

Where  $T_i$  is either time to event or censoring for the subject  $i$ ,  $M_i$  is a risk factor for the subject  $i$ . Furthermore, the nearest neighbor estimation (NNE) approach was applied to create a monotonic ROC curve. The sensitivity and specificity are based on joint and marginal distributions for time to event and risk factors, and the estimators of sensitivity and specificity were defined below

$$F_M(c) = \frac{\sum 1(M_i \leq c)}{n} \text{ where } 1(M_i \leq c) \text{ is the indicator function.}$$

$$S_{\lambda_n}(c, t) = \frac{1}{n} \sum_i S_{\lambda_n}(t \mid M = M_i) 1(M_i > c)$$

$$\hat{S}_{\lambda_n}(t|M = M_i) = \prod_{s \in \tau_n, s \leq t} \left\{ 1 - \frac{\sum_j 1(|\hat{F}_M(M_j) - \hat{F}_M(M_i)| < \lambda_n) 1(Z_j = s) \delta_j}{\sum_j 1(|\hat{F}_M(M_j) - \hat{F}_M(M_i)| < \lambda_n) 1(Z_j \geq s)} \right\} \text{ with a chosen}$$

$$\lambda_n = n^{-1/3}$$

Where  $\hat{F}_M(c)$  is an estimator of the risk factor distribution;  $\hat{S}_{\lambda_n}(c, t)$  is an estimator of the joint survival function;  $\hat{S}_{\lambda_n}(t|M = M_i)$  is an estimator of the conditional survival function with a specific risk factor.

Then, the sensitivity and specificity were defined as

$$\text{Sensitivity} = \frac{1 - F_M(c) - S_{\lambda_n}(c, t)}{1 - S_{\lambda_n}(-\infty, t)}$$

$$\text{Specificity} = 1 - \frac{S_{\lambda_n}(c, t)}{S_{\lambda_n}(-\infty, t)}$$

## 2.7 Signature Models Validation

A signature for each phenotype constructed in the training data (POG 9904 and 9905 trials) was further validated in a completely independent validation dataset (POG 9906 trial) with a total of 122 ALL patients. To do so, SNPs which were identified in both training and validation data were first mapped. Among these selected SNPs and demographic variables, the estimated coefficient of each covariate from the univariate analysis in the training data were multiplied by the value of covariate in the validation data to generate a predictive risk score for each ALL patient in the validation data.

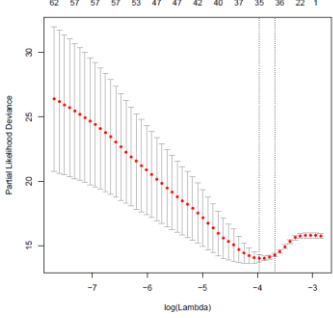
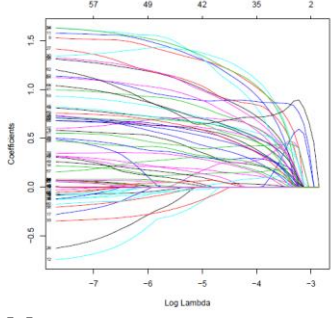
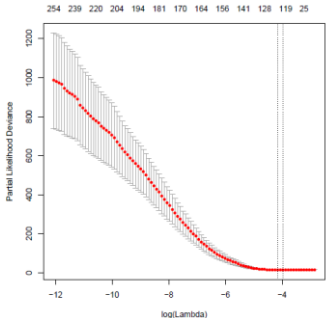
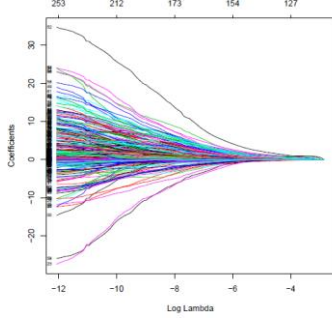
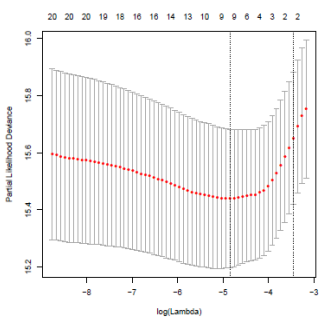
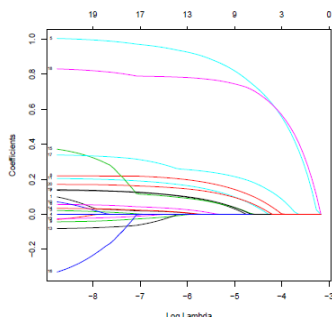
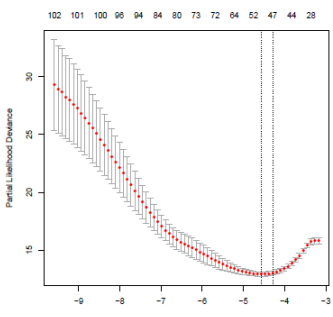
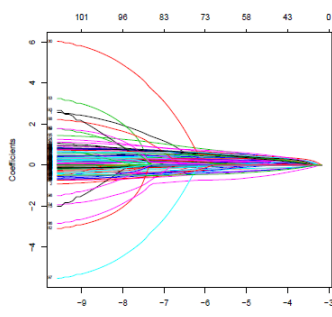
## 2.8 Clinical Application

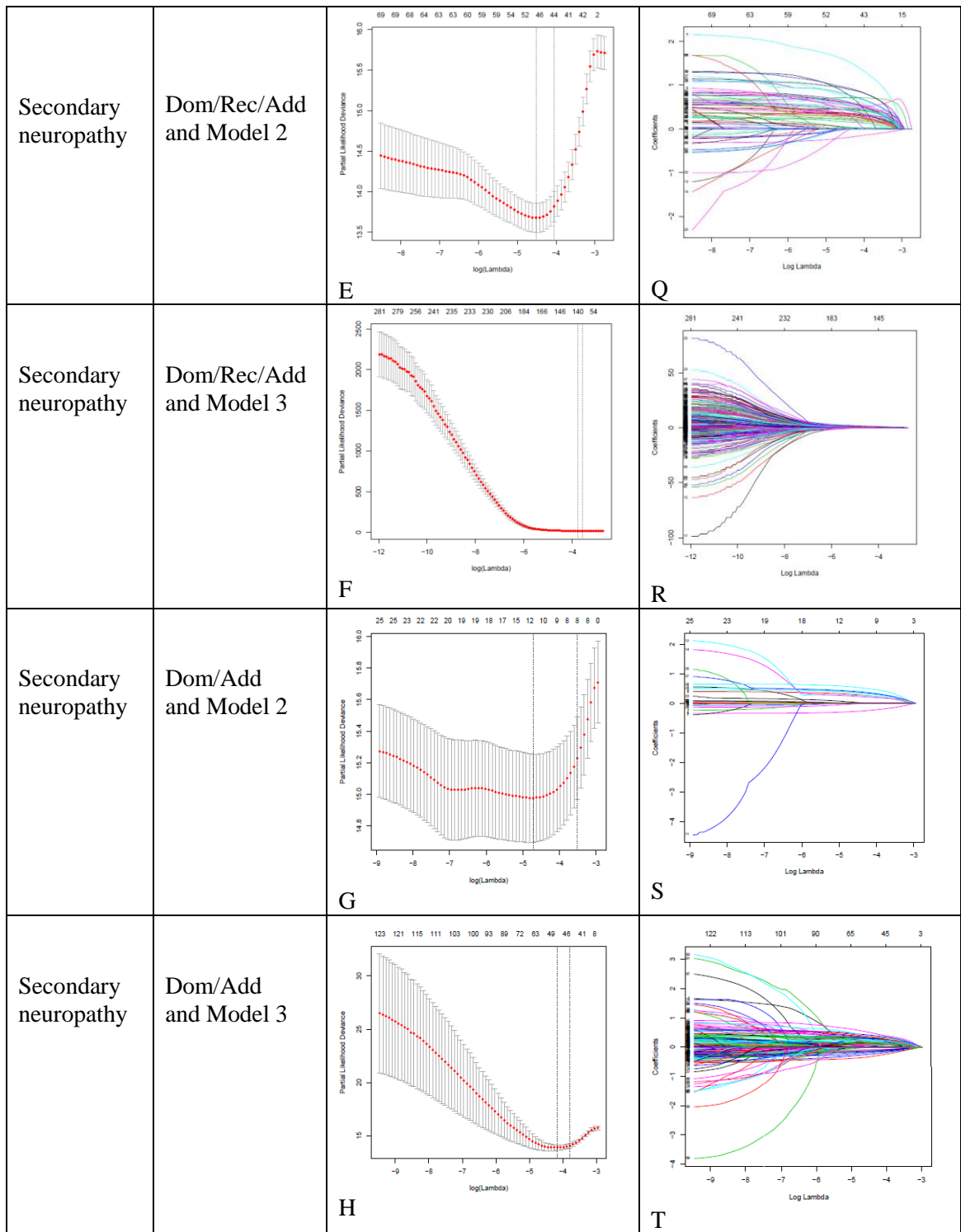
To consider the clinical application, sensitivity and specificity calculated from ROC curves were evaluated to determine the cutoff for the risk score. Patients with a risk score less than the threshold were considered to be a low-risk group. In contrast, those who have a score above the threshold were assigned to be a high-risk group.

### 3. Results

To avoid the collinearity problem and high-dimensionality, a LASSO penalized Cox proportional hazard model was applied to the analysis. Different to a typical Cox proportional hazard mode, the LASSO approach cannot handle missing genotypes which were observed commonly in the high-dimensional genome wide association study. In order to overcome this limitation, an imputation of typed SNPs was performed by MACH based on the haplotype of the dataset. At the initial stage of the model development, a signature model was introduced independently to test SNPs selected from each genetic model. As we moved forward to design a more comprehensive signature model, the strategy was modified and represented in Figure 7.1. For each phenotype, a minimum of p-value of each SNP was selected across either strategy 1, which included all three genetic models, or strategy 2, which only included the dominant and additive models. As Figures 7.2 A to L showed, the LASSO regression model in glmnet package allowed conducting a 10-fold cross-validation to optimize lambda which is a parameter to determine which coefficient was shrunken toward 0. The y-axis represented the partial likelihood error. The line on the right was drawn at the minimum error, and the other was drawn at the maximum value of lambda within 1 standard error of the minimum. At the maximum lambda, the LASSO coefficient paths were shown in Figures 7.2 M to X. The model was further evaluated by the ROC curves. This ROC analysis provided tools to select the optimal model, and the accuracy of the model was measured by the area under the ROC curve (AUC). The accuracy of the model represents how well the model separates the group being tested into those with and without the phenotype in question. For the primary neuropathy, no biomarker was selected in model 1 when only the

demographic and clinical factors were considered. The AUC is 0.7069419 in model 2, and the AUC increases to 0.9472682 in model 3. When the model was tested under the strategy of considering all three genetic models, the numbers of SNPs were increased in the test. As a result, the calculated AUCs improve to 0.908435 in model 2 and 0.9710504 in model 3, respectively. For the secondary neuropathy, the AUC is 0.5412067 without considering any genetic information. Furthermore, the AUCs increase to 0.7668192 and 0.9254674 in models 2 and 3, respectively. When the SNPs from the recessive model were also tested, the AUCs are 0.8876722 and 0.9691573 in models 2 and 3, respectively. The numbers of the selected predictors in each model were listed in Figures 7.3 E to F. For relapse, the accuracy of the model was evaluated in a time dependent manner. As the duration of clinical follow-up lasted almost 10 years for some patients, the sensitivity and specificity of the model at four time points (1, 2, 4 and 8 years) were evaluated. In Figures 7.4 E to F, the AUCs are between 0.687354 and 0.712445 across four time points in model 1. Under the strategy 1, the AUCs are between 0.731131 and 0.756252 in model 2 and between 0.925097 and 0.933578 in the model 3 across time points. When additional SNPs selected from the recessive model were tested, the AUCs are between 0.837579 and 0.861887 in model 2 and between 0.926813 and 0.969303 in model 3. Overall, signature models with model 3 approach predicted both neuropathy and relapse more accurately than the models 1 and 2 based on the ROC curve evaluation.

Phenotype	Model Strategy	Cross-validation	Lasso coefficient path
Primary neuropathy	Dom/Rec/Dose and Model 2	 <p data-bbox="683 625 711 653">A</p>	 <p data-bbox="1057 625 1084 653">M</p>
Primary neuropathy	Dom/Rec/Dose and Model 3	 <p data-bbox="683 976 711 1010">B</p>	 <p data-bbox="1057 976 1084 1010">N</p>
Primary neuropathy	Dom/Dose and Model 2	 <p data-bbox="683 1333 711 1373">C</p>	 <p data-bbox="1057 1333 1084 1373">O</p>
Primary neuropathy	Dom/Add and Model 3	 <p data-bbox="683 1696 711 1736">D</p>	 <p data-bbox="1057 1696 1084 1736">P</p>





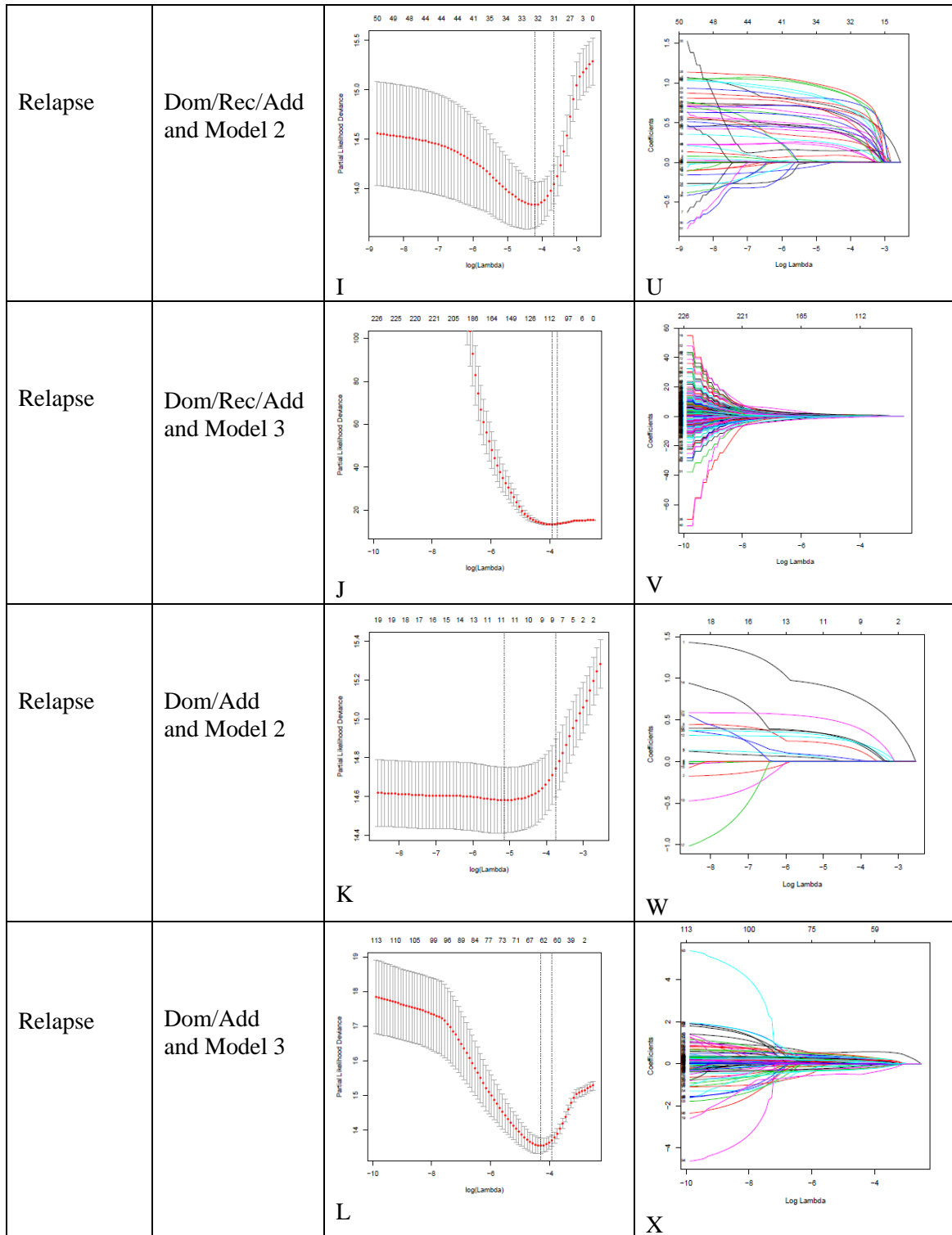


Figure 7.2. **Cross-validation and LASSO coefficient path plots for primary neuropathy, secondary neuropathy and relapse.**

A-L represents the 10-fold cross-validation result with Model 2 and 3 as well as Dom/Rec/Add and Dom/Add genetic model strategies.

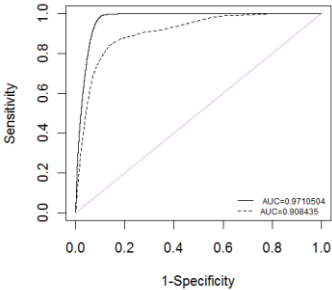
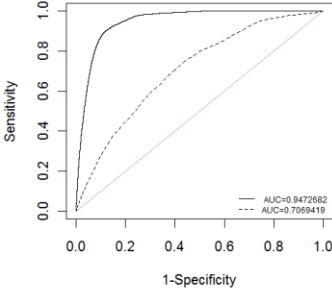
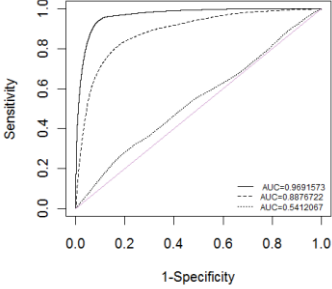
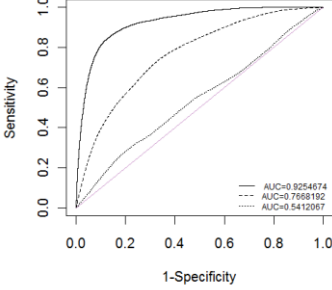
M-X represents the LASSO coefficient path result with Model 2 and 3 as well as Dom/Rec/Add and Dom/Add genetic model strategies. With the selected lambda, the coefficient of marker shrunk toward 0 on the right.

Dom: dominant model; Rec: recessive model; Add: additive model.

Furthermore, model 3 was used to identify patients who are at high risk of having an event based on selected sensitivity and specificity. A threshold of sensitivity at 0.90308144 and specificity at 0.901608868 were selected for primary neuropathy. The survival curves of high risk vs. low risk groups were plotted in Figures 7.5A and B. For the secondary neuropathy, sensitivity at 0.901325 and specificity at 0.900154 were selected to distinguish the survival of high risk vs. low risk groups in Figures 7.6A and B. Survival curves of high risk vs. low risk groups based on sensitivity at 0.900651992 and specificity at 0.9000682719 for relapse were shown in Figures 7.7A and B.

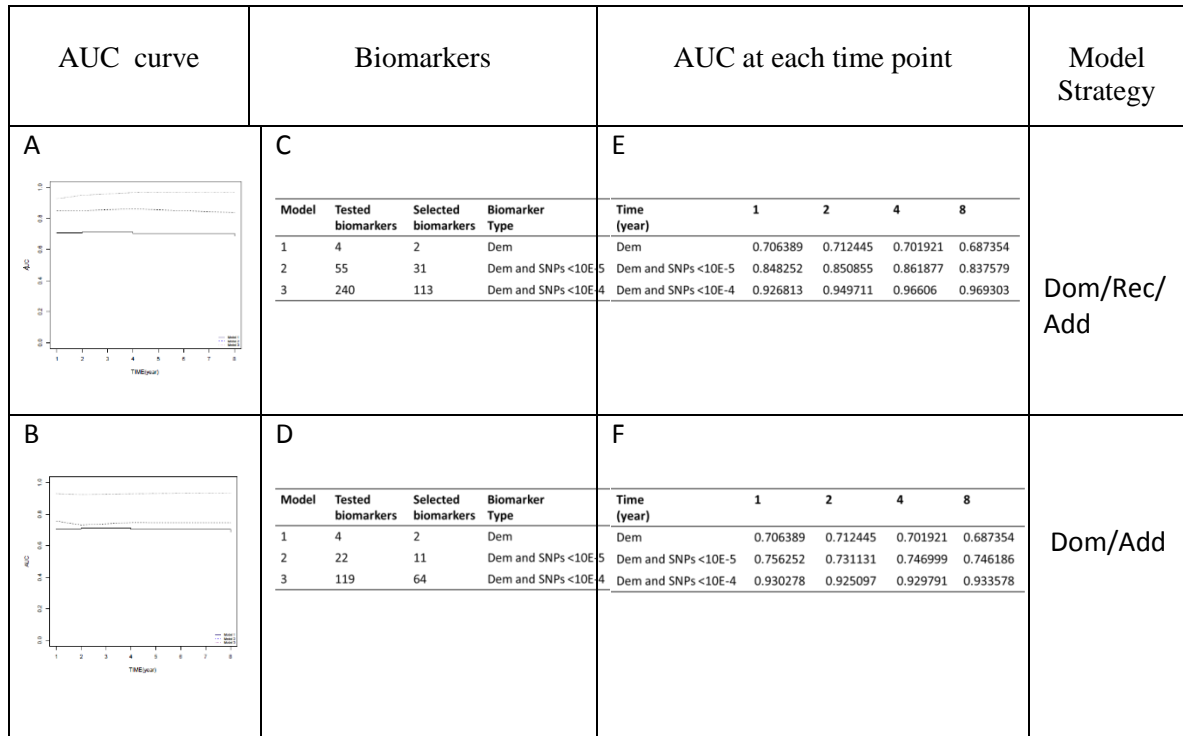
In order to evaluate the model predictability, the signature of each phenotype was further validated by predicting the risk score in a new population. There are 265,126 SNPs found in both training and validation sets. Among these SNPs, SNPs with p-value  $<1 \times 10^{-3}$  in the training set and p-value  $<1 \times 10^{-2}$  in the validation set were then evaluated. By taking the median risk score across the population, the time to primary neuropathy was compared between group 1 ( $>$ median risk score) and group 2 ( $<$ median risk score). A p-value of 0.00066 was given based on  $\chi^2$  distribution in the log rank test. Similarly, a significant p-value of 0.00791 in  $\chi^2$  distribution was observed when comparing the time to secondary neuropathy between group 1 ( $>$ median risk score) and group 2 ( $<$ median risk score). For the efficacy endpoint, the difference of time to relapse between the two groups, based on a cutoff of the median predictive risk score from the validation data, was with a p-value of 0.000656. All these findings indicated that the signature of each phenotype was well validated in an independent population based on the proposed validation scenario.

In addition, to apply the clinical utility of the signature model by identifying patients who are less likely to be benefited from the treatment, patients at high risk of relapse were selected by a risk score greater than 222.009734 based on specificity at 0.9000682719. Furthermore, a subgroup of those patients who are likely to experience primary neuropathy was identified based on a risk score greater than 1 at sensitivity of 0.820777. A total of 48 individuals were identified in this category (Figure 7.8). A sub-population who is most likely to receive the benefit of the treatment was also classified in Figure 7.8. Similar to the primary neuropathy, patients were divided into four sub-groups based on the risk assessment of both relapse and secondary neuropathy in Figure 7.9.

Phenotype	ROC curve	Model strategy	Biomarkers																
Primary neuropathy	<p><b>A</b></p> 	Dom/Rec/Add	<p><b>E</b></p> <table border="1"> <thead> <tr> <th>Model</th> <th>Tested biomarkers</th> <th>Selected biomarkers</th> <th>Biomarker Type</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>4</td> <td>0</td> <td>Dem</td> </tr> <tr> <td>2</td> <td>70</td> <td>35</td> <td>Dem and SNPs &lt;10E-5</td> </tr> <tr> <td>3</td> <td>289</td> <td>132</td> <td>Dem and SNPs &lt;10E-4</td> </tr> </tbody> </table>	Model	Tested biomarkers	Selected biomarkers	Biomarker Type	1	4	0	Dem	2	70	35	Dem and SNPs <10E-5	3	289	132	Dem and SNPs <10E-4
Model	Tested biomarkers	Selected biomarkers	Biomarker Type																
1	4	0	Dem																
2	70	35	Dem and SNPs <10E-5																
3	289	132	Dem and SNPs <10E-4																
Primary neuropathy	<p><b>B</b></p> 	Dom/Add	<p><b>F</b></p> <table border="1"> <thead> <tr> <th>Model</th> <th>Tested biomarkers</th> <th>Selected biomarkers</th> <th>Biomarker Type</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>4</td> <td>0</td> <td>Dem</td> </tr> <tr> <td>2</td> <td>20</td> <td>8</td> <td>Dem and SNPs &lt;10E-5</td> </tr> <tr> <td>3</td> <td>107</td> <td>51</td> <td>Dem and SNPs &lt;10E-4</td> </tr> </tbody> </table>	Model	Tested biomarkers	Selected biomarkers	Biomarker Type	1	4	0	Dem	2	20	8	Dem and SNPs <10E-5	3	107	51	Dem and SNPs <10E-4
Model	Tested biomarkers	Selected biomarkers	Biomarker Type																
1	4	0	Dem																
2	20	8	Dem and SNPs <10E-5																
3	107	51	Dem and SNPs <10E-4																
Secondary Neuropathy	<p><b>C</b></p> 	Dom/Rec/Add	<p><b>G</b></p> <table border="1"> <thead> <tr> <th>Model</th> <th>Tested biomarkers</th> <th>Selected biomarkers</th> <th>Biomarker Type</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>4</td> <td>1</td> <td>Dem</td> </tr> <tr> <td>2</td> <td>74</td> <td>48</td> <td>Dem and SNPs &lt;10E-5</td> </tr> <tr> <td>3</td> <td>291</td> <td>140</td> <td>Dem and SNPs &lt;10E-4</td> </tr> </tbody> </table>	Model	Tested biomarkers	Selected biomarkers	Biomarker Type	1	4	1	Dem	2	74	48	Dem and SNPs <10E-5	3	291	140	Dem and SNPs <10E-4
Model	Tested biomarkers	Selected biomarkers	Biomarker Type																
1	4	1	Dem																
2	74	48	Dem and SNPs <10E-5																
3	291	140	Dem and SNPs <10E-4																
Secondary Neuropathy	<p><b>D</b></p> 	Dom/Add	<p><b>H</b></p> <table border="1"> <thead> <tr> <th>Model</th> <th>Tested biomarkers</th> <th>Selected biomarkers</th> <th>Biomarker Type</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>4</td> <td>1</td> <td>Dem</td> </tr> <tr> <td>2</td> <td>30</td> <td>12</td> <td>Dem and SNPs &lt;10E-5</td> </tr> <tr> <td>3</td> <td>126</td> <td>50</td> <td>Dem and SNPs &lt;10E-4</td> </tr> </tbody> </table>	Model	Tested biomarkers	Selected biomarkers	Biomarker Type	1	4	1	Dem	2	30	12	Dem and SNPs <10E-5	3	126	50	Dem and SNPs <10E-4
Model	Tested biomarkers	Selected biomarkers	Biomarker Type																
1	4	1	Dem																
2	30	12	Dem and SNPs <10E-5																
3	126	50	Dem and SNPs <10E-4																

**Figure 7.3. ROC curves of primary and secondary neuropathy in two genetic model strategies and the numbers of tested and selected biomarkers in the models.**

Dom: dominant model; Rec: recessive model; Add; additive model; Dem: demographic and clinical factors.



**Figure 7.4. Time dependent ROC of relapse in two genetic models.**

A-B: ROC curves from 1 to 8 years

C-D: The number of tested and selected biomarkers in model 1 to 3.

E-F: Area under the ROC curve (AUC) of each model at time 1, 2, 4 and 8 years.

Dom: dominant model; Rec : recessive model; Add: additive model.

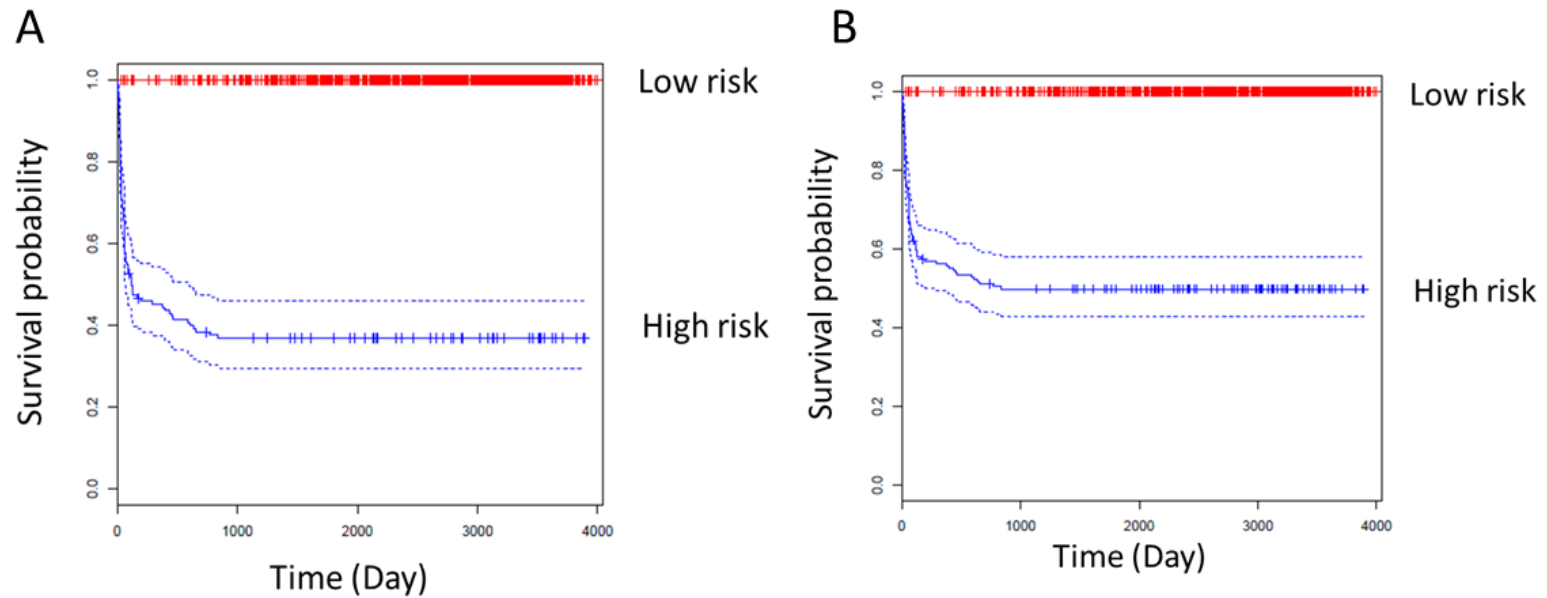


Figure 7.5. **Survival plot of high risk vs. low risk group of primary neuropathy with (A) sensitivity=0.903081 and (B) specificity=0.901608.** (A) Patients have risk score greater than 9.15 considered as high risk group; (B) Patients have risk score greater than 7.26 considered as high risk group.



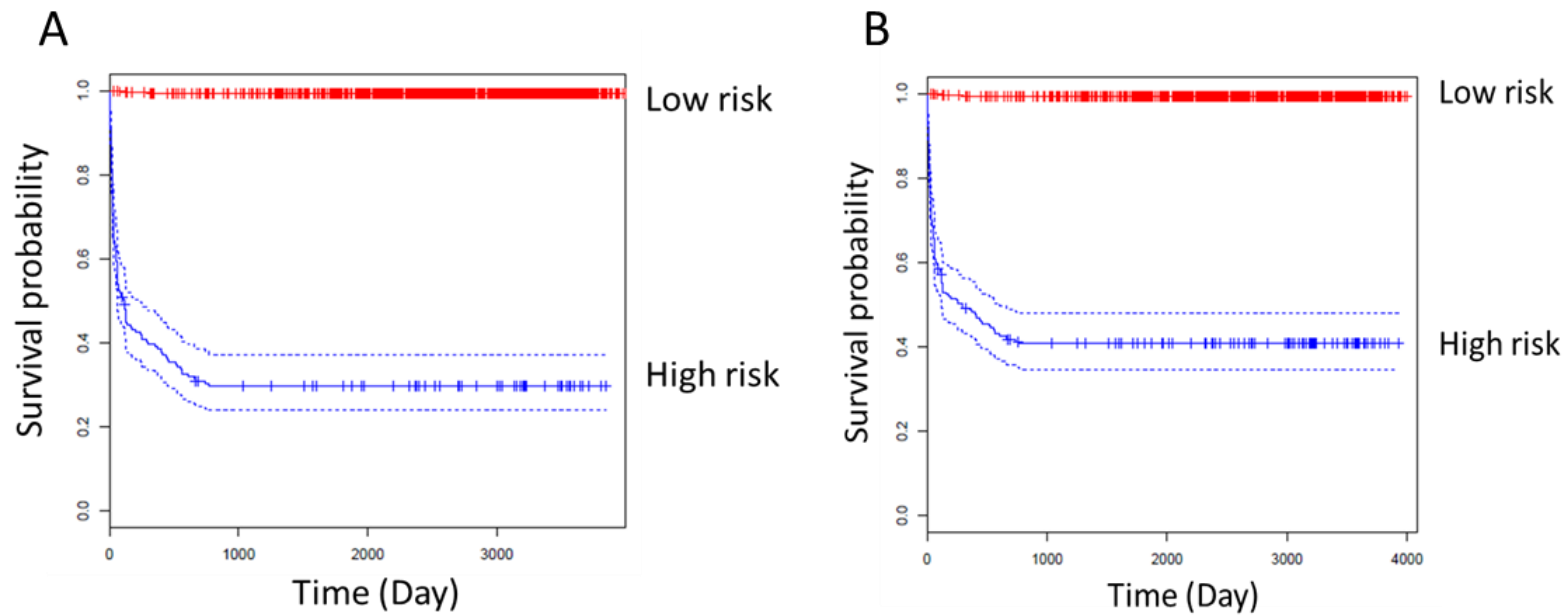


Figure 7.6. **Survival plot of high risk vs. low risk group of secondary neuropathy with (A) sensitivity=0.901325 and (B) specificity=0.900154.**

(A) Patients have risk score greater than 8.01 considered as high risk group; (B) Patients have risk score greater than 6.34 considered as high risk group.

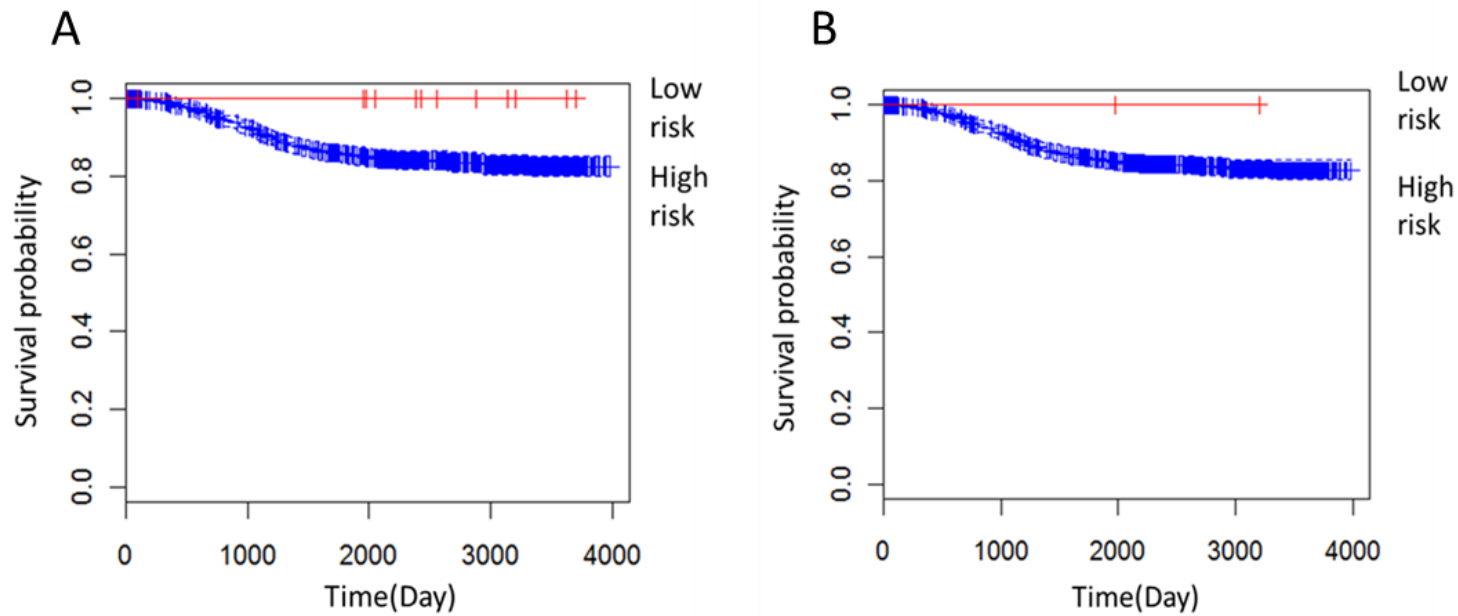


Figure 7.7. **Survival plot of high risk vs. low risk group of relapse with (A) sensitivity=0.900651 and (B) specificity=0.900068 .**  
(A) Patients have risk score greater than 204.74 considered as high risk group; (B) Patients have risk score greater than 222.01 considered as high risk group.

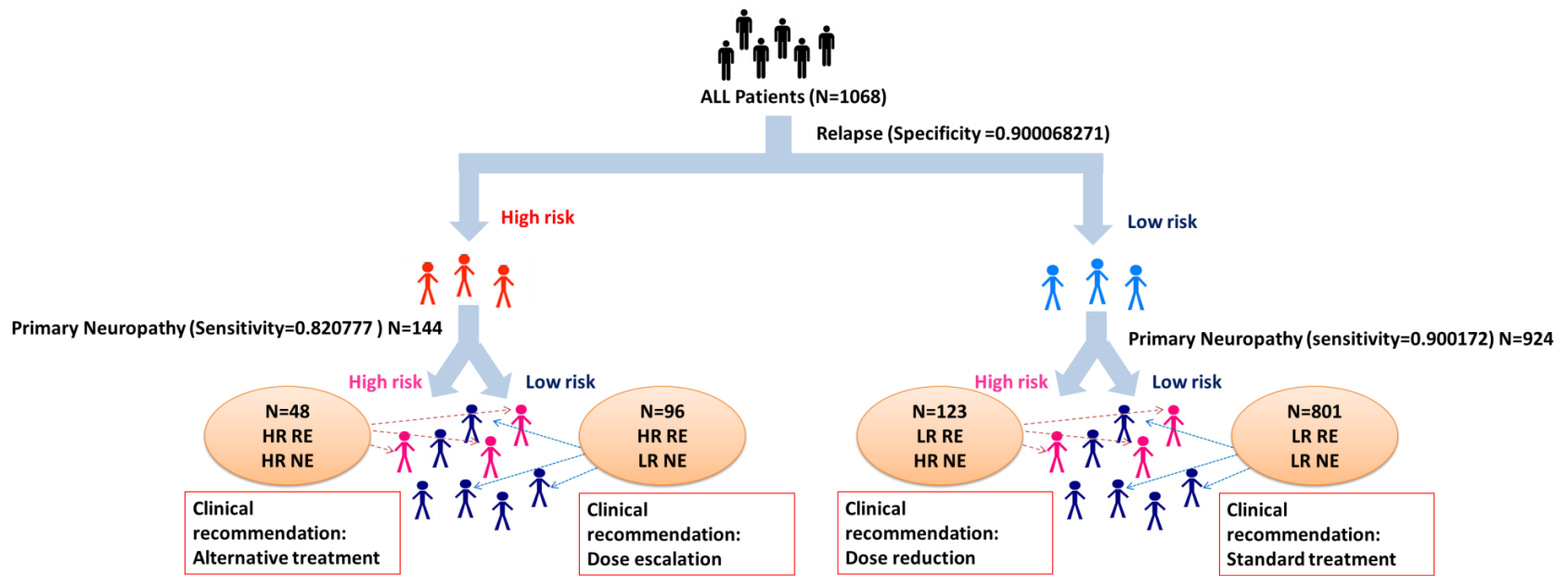


Figure 7.8. A graphical representation of the clinical application using the signature model to identify ALL patient in four subgroups (HR RE/HR NE; HR RE/ LR NE; LR RE/HR NE; LR RE/ LR NE) based on relapse and primary neuropathy endpoints and their clinical recommendation.

NE: primary neuropathy; RE: relapse; HR: high risk; LR: low risk.

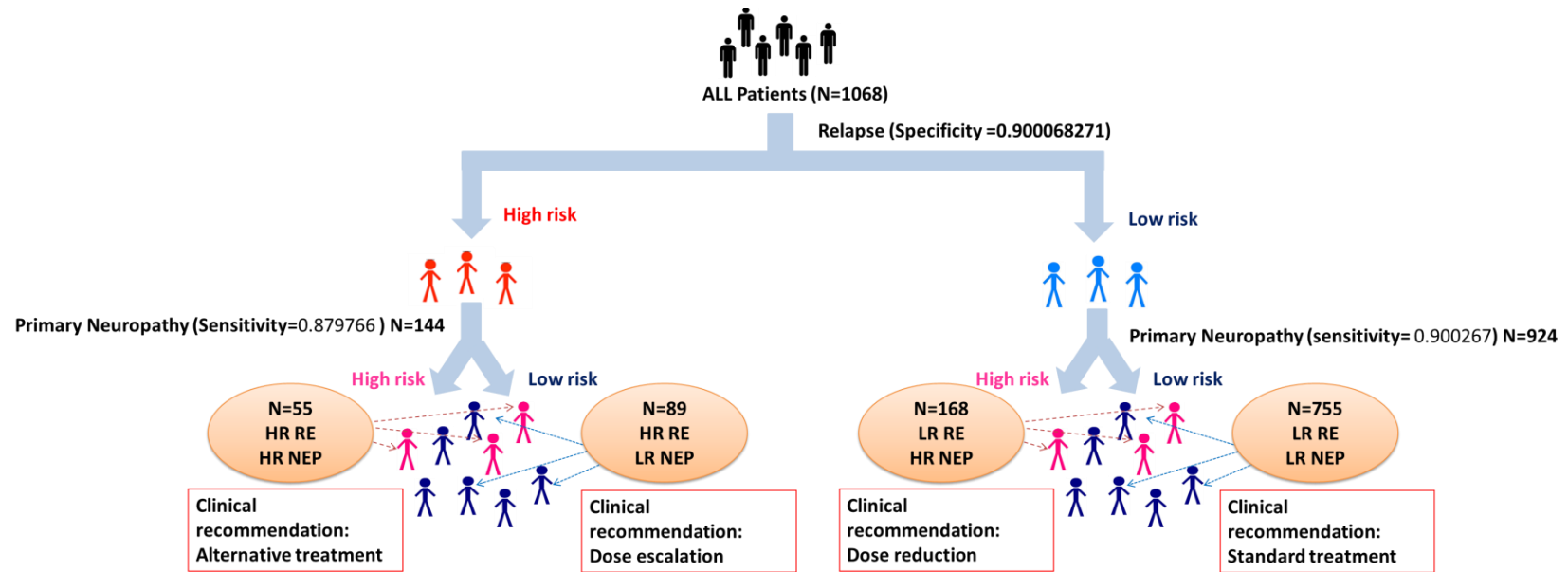


Figure 7.9. A graphical representation of the clinical application using the signature model to identify ALL patient in four subgroups (HR RE/HR NEP; HR RE/ LR NEP; LR RE/HR NEP; LR RE/ LR NEP based on relapse and secondary neuropathy endpoints and clinical recommendation.

NEP: secondary neuropathy; RE: relapse; HR: high risk; LR: low risk

#### *4. Discussion*

Our analysis of single nucleotide polymorphism data demonstrated that the LASSO penalized regression is capable of identifying pertinent predictors. The LASSO technique for variable selection in the Cox model seemed to be a worthy competitor to stepwise selection. It is less variable than the stepwise approach and still yields interpretable models. The computational time is relatively quick compared to the traditional GWAS study. Initially, we expected that the model would not only be capable of predicting time to event by a genetic pattern, but also would provide a reasonable prediction in a relative small association data set with no statistical significant association identified in the GWAS study. Based on our study design, we constructed three models to compare how much information would be enough in the test to generate a reasonable prediction to the same phenotype. As the model performed 10-fold cross-validation, it allowed the function to run 10fold+1 times to get the lambda sequence and compute the fit with each of the folds omitted. The optimal lambda was selected based on the mean of accumulated cross-validated error. This process was done repeatedly in each model and generated a unique signature for each phenotype.

Based on ROC curve evaluation, the area under the curve of ROCs in model 3, which contained demographic and SNPs with p-value  $<1 \times 10^{-4}$  across both neuropathy and relapse, generally exceeded 0.9. This indicated that the model well predicted patients with and without event. All important findings are subject to replication. This attitude allows us to define the most informative SNPs rather than only declaring their global significance. Our approach to data analysis is motivated by this consideration. In addition, the signature of each phenotype was further validated in the validation data set. In this

backward validation, the overlapped SNPs with a predefined p-value threshold were first identified; then, the predicted risk scores were calculated for each individual in the validation data using estimated coefficients from univariate analysis in the training data. As the result indicated a statistically significant separation of patients' time to neuropathy or relapse based on the median of predicated risk score using the log-rank test, this validation gave a preliminary overlook of how well the model predicts based on the best scenario. However, the current validation data set, which mainly has a population at high risk for treatment failure, may not be the best candidate to validate the signature model. Therefore, further analysis needs to be done, not only to validate the model by taking all the tested SNPs into consideration which may be more precise in evaluating the signature applicability, but also to evaluate the signature in a standard risk disease population similar to POG 9904 and 9905 trials.

In addition to validating the generalizability of the model, the clinical application of the signature model is also a major focus in this study. Clinically, as ALL relapse still carries a very poor prognosis and the majority of relapse continues to occur in patients without apparent "high-risk" features (Stanulla et al., 2009; Schrappe et al., 2000a; Schrappe et al, 2000b), the utility of the signature model becomes meaningful when a subgroup of patients who have no benefit from the treatment because of the high risk of relapse and neurotoxicity can be predicted and alternative treatments can be recommended. Similarly, for those who are either at high risk of relapse or high risk of neuropathy, the clinicians can evaluate the cost and effectiveness to recommend dose modification. Those who are less likely to experience toxicity or relapse can be the best targets for the treatment and for understanding drug exposure variability. With all these

factors taken into account, our signature modeling approach not only provides an innovated assessment to high dimensional genomic data for identifying potential associations with toxicity and efficacy, but also integrates a statistical model with clinical useful application and interpretation.

## **CHAPTER VIII: *Conclusions and future directions***

### *1. Conclusions*

The major foci of this dissertation were to explore the impact of modeling and simulation in drug development and patient care: 1) where a population approach in pharmacometrics can be used to understand pharmacokinetics of drugs in humans from a translational in vivo model, the covariate effect on exposure variability in a special population, the pharmacodynamics of biomarkers for predicting the clinical outcome of a treatment and the evaluation of a trial design based on a simulated virtual population; and 2) how a statistical modeling approach can be utilized to evaluate the association between pharmacogenomic variations, drug-induced toxicity and efficacy to establish a signature pattern that determines patients' risk of experiencing these effects. The analyses from this dissertation address limitations of each modeling approach presented. These limitations are tied back to the underlying assumptions in the modeling approach as well as the potential sources of variability that can contribute to trial failure and should be accurately measured. This dissertation also points out an integration between different modeling approaches which potentially allows more sophisticated understanding of the biomarker application and helps to capture disease status and treatment effect across a spectrum of modeling approaches and data types.

Based on the major findings outlined above, the overarching conclusion of each modeling and simulation application can be drawn concerning the body of work presented in this dissertation. These are: 1) a translational PK modeling approach that was able to predict clozapine and norclozapine CNS exposures in humans relating these exposures to receptor binding kinetics at multiple receptors; 2) a population



pharmacokinetic analysis of a study of sertraline in depressed elderly patients with Alzheimer's disease (DIADS-2) identified site specific differences in drug exposure contributing to the overall variability in sertraline exposure; 3) the utility of a longitudinal tumor dynamic model developed by the Food and Drug Administration for predicting survival in NSCLC patients drawn from NCI funded studies, including an exploration of the limitations of this approach; 4) a Monte Carlo clinical trial simulation approach that was used to evaluate a pre-defined oncology trial with a sparse drug concentration sampling schedule with the aim to quantify how well individual drug exposures, random variability, and the food effects of abiraterone and nilotinib were determined under these conditions; 5) a time to event analysis that facilitated the identification of candidate genes including polymorphisms associated with vincristine-induced neuropathy from several association analyses in childhood ALL patients; and 6) a LASSO penalized regression model that predicted vincristine-induced neuropathy and relapse in ALL patients and provided the basis for a risk assessment of the population. Each of these conclusions is summarized followed by further directions in the section which follows.

*A. A translational PK modeling approach that predicted clozapine and norclozapine CNS exposures in humans relating these exposures to receptor binding kinetics at multiple receptors*

From a clinical perspective, it is important to understand the broad receptor coverage of clozapine that is thought to account for clozapine's unique superiority in treatment resistant schizophrenia. With the innovation and improvement of measurement techniques, clozapine concentrations now can be measured at the target site using microdialysis so that the concentration can be more precisely related to response at the

effect site (Derendorf et al., 2000; Chefer et al., 2009; Jacobson et al., 2013). As part of this dissertation (**CHAPTER II**), we demonstrated that the proposed PK model can be extrapolated to predict human systemic exposure including clozapine brain concentrations and associated receptor occupancy profiles in humans at clinically relevant doses. In addition, the model simultaneously captured the parent-drug and metabolite in the system. This is relevant as norclozapine also has activity at multiple receptors in the CNS that could contribute to response (Bishara et al., 2008). This modeling approach could form the foundation for the design of the future comprehensive PK-PD models as well as extend our understanding of clozapine's complex behavioral effects in humans. Furthermore, this model framework can serve to support the discovery of a new drug entity that may share the same mechanism of action as clozapine.

*B. A population pharmacokinetic analysis of a study of sertraline in depressed elderly patients with Alzheimer's disease (DIADS-2) identified site specific differences in drug exposure contributing to the overall variability in sertraline exposure*

This is the first population pharmacokinetic (PPK) study of sertraline focusing on AD patients with depression. Compared to younger patients, the elderly have a longer half-life, but there was no difference between elderly healthy volunteers in the literature and these AD patients in PK characteristics. In addition, a covariate analysis was conducted to explore whether the variability of the PK parameter estimates could be explained by patient specific characteristics. Based on the analysis, the clinical site was the only significant covariate explaining variability in the CL/F change. The clinical implications for subjects at the outlying site would be important if this represented a true bias in subject selection. As we discussed in **CHAPTER III**, no significant difference in

demographic or clinical attributes between this study site and the others were found; therefore, a potential source of variation is the implementation of study procedures at this site. In conclusion, site variability in trial procedures is critical to the validity of multi-site trials, and population pharmacokinetics may prove helpful in this assessment. With the application of the population PK model approach, a close monitoring of dosage regimen and administration is recommended in the future multicenter study especially with sparse PK sample collections so that patients would not be exposed to unnecessary harm. This can also minimize what may be artificial systematic differences in exposure, and therefore potentially confound the interpretation of response and/or toxicity.

*C. The utility of a longitudinal tumor dynamic model developed by the Food and Drug Administration for predicting survival in NSCLC patients drawn from NCI funded studies, including an exploration of the limitations of this approach*

Previous studies have shown that modeling and simulation is an efficient means to support decision making at the Phase 2 to Phase 3 transition in drug development (Claret et al., 2012), therefore, it was thought that this longitudinal tumor model could serve as a framework to predict clinical outcomes of cancer patients ideally resulting in smaller trials with shorter observation periods (Adjei et al., 2009; Dhani et al., 2009; Maitland et al., 2011; Maitland et al., 2011; Mandrekar et al., 2010; Stein et al., 2011; Yap et al., 2010). Based on this initial idea, we demonstrated in **CHAPTER IV** that a modeling strategy to assess the tumor load dynamics found that the tumor model parameter estimates of tumor shrinkage and progression rates were both lower than the findings from the previous FDA large registration trial data. However, when the FDA trial was reduced to a similar size as our CALGB30203 study and patient characteristics matched,

the parameter estimates became comparable. The lower parameter values for tumor dynamics from the model using the data from the CALGB trials suggested that the prediction of overall survival using this metric was likely impacted by the sample size and heterogeneity in the tumor presentation. In addition, a simplified time to tumor growth (TTG) parameter was observed to be more robust in predicting overall survival in our small trial as a single factor, but the clinical interpretation and utility need to be further explored. In conclusion, the data quality is critical for a model to give a meaningful contribution to predicting patient outcome. The limitations inherent in tumor measurement can lead to the misrepresentation of the true tumor growth dynamic. This variation can be more problematic when the sample size is small. Further analysis is required to explore the utility of the tumor dynamic model in predicting patients' survival outcomes using alternative measurements such as automated capture of the three-dimensional tumor volumes from raw CT scan data (Schwartz et al., 2000). It appears that this approach may be more accurate in describing the tumor change over time, and potentially eliminates the inter-rater variability from the traditional sum of the longest diameter measurement carried out by radiologists. In addition, a simulation approach could be applicable to evaluate how different level of variance of the tumor measurements can affect the overall tumor model parameter estimates. Until more precise and/or accurate measures are available, it will also be difficult to ascertain the true relationship between tumor loads (especially across tumor types that may be treated with agents that are –static instead of –cidal in their effects) and patient specific outcomes such as progression free survival and overall survival.

*D. A Monte Carlo clinical trial simulation approach was used to evaluate a pre-defined oncology trial with a sparse drug concentration sampling schedule with the aim to quantify how well individual drug exposures, random variability, and the food effects of abiraterone and nilotinib were determined under these conditions*

Food intake is one important contributor to variability in oral bioavailability (Singh et al., 2004; Winstanley et al., 1989; Gu et al., 2007) in particular for some of the newer oral anticancer agents such as abiraterone and nilotinib (Ryan et al., 2010; Yin et al., 2010). This can cause a significant impact on systemic drug exposures that can lead to drug toxicity and/or therapeutic failure (Koch et al., 2009; Kang et al., 2010). We demonstrated in **CHAPTER V** how a clinical trial simulation using a Monte Carlo method under the framework of a Pop PK approach provided understanding of a clinical oncology population pharmacokinetic trial design with a random sampling schema. More specifically, we were able to quantify how well this design and the data collection scheme associated with it and were able to capture the variability and the magnitude of individual drug exposure in the presence of a large food effect in nilotinib and abiraterone trials. This simulated approach facilitated an early evaluation of the proposed trial design. However, clinical trial simulations are generated based on many trial assumptions, and these assumptions may include uncertainty. For future studies, different scenarios with underlying assumptions can be tested to optimize the design. With the limitations of current analysis discussed in **CHAPTER V**, additional components such as a drop-out pattern of the trial and patient compliance, which may vary in the clinical practice setting, should be explored in future analyses.

This study emphasizes the importance of addressing trial designs where intensive sampling cannot be obtained and yet there is a need to understand the drug exposure characteristics to what are otherwise medications with highly variable pharmacokinetic dispositions.

*E. A time to event analysis facilitated the identification of candidate genes including polymorphisms associated with vincristine-induced neuropathy from several association analyses in childhood ALL patients*

Vincristine is among the most commonly used anticancer agents, however, little is known regarding its disposition and optimal dosing. This gap in knowledge can lead to negative clinical outcomes such as toxicity due to drug overdosing or lack of efficacy due to sub-therapeutic dosing. Vincristine is associated with highly variable cumulative dose-dependent peripheral neuropathy. To explore the association between germline variants and VIPN in pediatric ALL patients, a genome-wide association analysis was carried out initially. The analysis was expanded to examine the association using gene enrichment analysis. In both analyses, we identified several genetic loci associated with incidence of  $\geq$  grade 2 VIPN. Based on the top-ranked SNPs, 31 genes were found in an association with either primary or secondary neuropathy. In **CHAPTER VI**, a discussion was carried out to identify those candidate genes and their corresponding pathways. In this stage of discovery, we found that there are many genes which are indirectly associated with the mechanisms of leading vincristine-induced neuropathy, but we believe these findings can be potential focuses for the future *in vitro* studies after association is confirmed by sequencing an independent data set from a multi-center trial. Furthermore, the candidate genes selected from the association analysis were evaluated in the

interactive pathway analysis and gene expression analysis. Based on both the GEO and eQTL databases, few additional gene candidates appeared in association with neuropathy. In conclusion, we came across some similar correlations observed from different associations and data approaches that provided additional confirmation of the findings. These findings have significant potential clinical relevance given the widespread use of vincristine in treating childhood cancers.

*F. A LASSO penalized regression model predicted vincristine-induced neuropathy and relapse in ALL patients and provided the basis for a risk assessment of the population*

As the previous section (**CHAPTER VI**) discovered the association between germline variants and VIPN using multiple association analyses, this chapter focused on an alternative modeling approach to identify a genetic signature pattern which allows predicting the traits and avoid the collinearity in fitting which is a common limitation found in typical Cox proportional hazards model (Sohn et al., 2009). By performing a LASSO penalized regression model, a signature was developed not only for neuropathy but also for relapse independently as previous study showed no correlation between two phenotypes. Under different model strategies, we demonstrated that a unique genetic signature was capable to predict neuropathy or relapse with reasonable precision based on ROC evaluations. In order to indicate the clinical utility of the signature model, a risk assessment was carried out under the structure of genetic signature. In sum, the whole population was divided into four major groups based on the statistical threshold for both relapse and neuropathy. This identification would provide clinicians an initial guidance for dose recommendations which might result in a more efficient treatment plan and avoid unnecessary drug exposure to patients. The utility of the signature model seems

promising in predicting time to event, but a more comprehensive validation plan will need to further explore and test the model predictability. In fact, instead of building a signature only based on typed SNPs from the genotyping data, an additional approach was suggested to expand the signature by examining the untyped SNPs from the imputation. This approach might provide a more representative pattern of predictive signature across different populations and data, but the quality of imputation data needs to be carefully evaluated as we found to be a challenge initially when we imputed our training data. In terms of clinical utility of the signature, a challenge we are facing right now is if there is a surrogate endpoint which can help us to understand the disease and treatment effect as early as possible and enhance the application of the modeling approach. To consider this, minimal residual disease which was suggested by some studies (Campana 2010; Borowitz MJ, et al., 2008) as an important predictor of ALL relapse may be valuable to explore in this case.

Overall, this dissertation demonstrated the utility of modeling and simulation in various study design scenarios and its application and contribution for the future studies in drug development and patient care. Although different population and disease were focused in each chapter to explore a specific underlying modeling approach, the whole dissertation provided a projection of how each approach can be integrated in drug development process. An integration of multiple modeling approaches is also suggested by this dissertation to overcome the challenges and to examine the proposed solutions for the limitations of the current modeling approach. As recalled the drug development process in Figure 1.1, if there is a new compound targeted at early drug discovery phase, a combination of preclinical and translational modeling approach (CHAPTER II) can



capture its exposure-response relationship *in vivo* system, extrapolate a first to human dose and further predict some pharmacological relationships which have not been assessable in the human system. As this new compound is moved forward to the clinical phase, PPK model (CHAPTER III) can serve as a tool to continuously evaluate its exposure-response relationship, quantify the level of variability and identify covariate effect. This approach can be further expanded to predict PK in a special population such as elderly or another disease population. In addition to understand the PK profile of this new compound, assessment of the dynamics of a biomarker (CHAPTER IV, VI and VII) can provide insight into the disease progression and capture a therapeutic drug effect of this compound which might be useful for early go/no go decision or personalized medicine later on before the clinical endpoints are available. Meanwhile, a clinical trial simulation (CHAPTER V) can explore various dosage regimens of the compound and optimize the trial design which might allow reducing the numbers of the studies and be more efficient and informative in bridging studies from one phase to another. As a demonstration of how these approaches emphasized in the dissertation can be integrated in the drug development process, we also need to be aware that there are many more assessments could be done by modeling and simulation in the process.

In summary, modeling and simulation is a tool that enhances our understanding of pharmacokinetics and pharmacodynamics for specific drug entities and provides critical recommendations for dose selection and treatment plans. Leveraging the findings from biological studies, it will continue to be an essential element in drug development and patient care.

## REFERENCES

9904 Protocol:ALinC #17 TREATMENT FOR PATIENTS WITH LOW RISK ACUTE LYMPHOBLASTIC LEUKEMIA: A PHASE III STUDY

9905 Protocol"ALinC 17: PROTOCOL FOR PATIENTS WITH NEWLY DIAGNOSED STANDARD RISK

Adjei AA, Christian M, and Ivy P(2009) Novel designs and end points for phase II clinical trials. *Clin Cancer Res* **15**:1866-1872.

Ahn J E, Karlsson M O, Dunne A, and Ludden TM (2008) Likelihood based approaches to handling data below the quantification limit using NONMEM VI. *J Pharmacokinet Pharmacodyn* **35**:401-421.

An MW, Mandrekar SJ, Branda ME, Hillman SL, Adjei AA, Pitot HC, Goldberg RM, and Sargent DJ (2011) Comparison of Continuous versus categorical Tumor Measurement-Based Metrics to Predict Overall Survival in Cancer Treatment Trials. *Clin Cancer Res* **17**(20):6592-6599.

Axelsson DA, Perel JM, Birmasrewhar B, Rudolph GR, Nuss S, Bridge J, and Brent DA (2002) Sertraline pharmacokinetics and dynamics in adolescents. *J Am Acad Child Adolesc Psychiatry* **41**:9.

Baldwin RM, Owzar K, Zembutsu H, Chhibber A, Kubo M, Jiang C, Watson D, Eclow RJ, Mefford J, McLeod HL, Friedman PN, Hudis CA, Winer EP, Jorgenson EM, Witte JS, Shulman LN, Nakamura Y, Ratain MJ, and Kroetz DL (2012) A genome-wide association study identifies novel loci for paclitaxel-induced sensory peripheral neuropathy in CALGB 40101. *Clin Cancer Res* **18**(18):5099-5109.

Barnes K, Dobrzynski H, Foppolo S, Beal PR, Ismat F, Scullion ER, Sun L, Tellez J, Ritzel MW, Claycomb WC, Cass CE, Young JD, Billeter-Clark R, Boyett MR, and Baldwin SA (2006) Distribution and functional characterization of equilibrative nucleoside transporter-4, a novel cardiac adenosine transporter activated at acidic pH. *Circ Res* **99**(5):510-509.

Barrett JS, Gupta M, and Mondick JT(2007) Model-based Drug Development for Oncology Agents. *Expert Opinion on Drug Discovery* **2**:185-209.

Benner A, Zucknick M, Hielscher T, Ittrich C, and Mansmann U (2010) High-dimensional Cox models: the choice of penalty as part of the model building process. *Biometrical Journal* **52**(1):50-69.

Beutler AS, Kulkarni A, Kanwar R, Qin R, Chunningham JM, Therneau TM, and Loprinzi CL. Sequencing symptom control: results from the alliance N08C1 and N08CA genetics of chemotherapy neuropathy trials. In: 104th Annual Meeting of the American

Association for Cancer Research; Washington, DC; 2013. Accessed April 27, 2014. Available from:<http://www.abstractsonline.com/Plan/ViewAbstract.aspx?sKey=d4332bf8-005b-4cb2-a378-921ae4782081&cKey=52268d6b-7335-4aca-9338-653d6c27ffdb&mKey=%7b9B2D28E7-24A0-466F-A3C9-07C21F6E9BC9%7d>

Bishara D, and Taylor D (2008) Upcoming agents for the treatment of schizophrenia: mechanism of action, efficacy and tolerability. *Drugs* **68**(16):2269-2292.

Bonate PL (2011) Pharmacokinetic-pharmacodynamic modeling and simulation. New York, N.Y.: Spring.

Bock G, Gebhart M, Scharinger A, Jangsangthong W, Busquet P, Poggiani C, Sartori S, Mangoni ME, Sinnegger-Brauns MJ, Herzig S, Striessnig J, and Koschak A (2011) Functional properties of a newly identified C-terminal splice variant of Cav1.3 L-typeCa<sup>2+</sup> channels. *J Biol Chem* **286**(49):42736-42748.

Bonate PL and Suttle B (2013) Effect of censoring due to progressive disease on tumor size kinetic parameter estimates. *AAPS J* **15**(3):832-839.

Bonate PL (2005) Recommended reading in population pharmacokinetic pharmacodynamics. *AAPS J* **7**(2):E363-E373.

Borowitz MJ, Devidas M, Hunger SP, Bowman WP, Carroll AJ, Carroll WL, Linda S, Martin PL, Pullen DJ, Viswanatha D, Willman CL, Winick N and Camitta BM for the Children's Oncology Group (2008) Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood* **111**(12):5477-5485.

Braga MF, Aroniadou-Anderjaska V, Li H, and Rogawski MA (2009) Topiramate reduces excitability in the basolateral amygdala by selectively inhibiting GluK1 (GluR5) kainate receptors on interneurons and positively modulating GABAA receptors on principal neurons. *J Pharmacol Exp Ther* **330**(2):558-566.

Broyl A, Corthals SL, Jongen JL, van der Holt B, Kuiper R, de Knecht Y, van Duin M, el Jarari L, Bertsch U, Lokhorst HM, Durie BG, Goldschmidt H, and Sonneveld P (2010) Mechanisms of peripheral neuropathy associated with bortezomib and vincristine in patients with newly diagnosed multiple myeloma: a prospective analysis of data from the HOVON-65/GMMG-HD4 trial. *Lancet Oncol* **11**:1057-1065.

Bruno R, and Claret L (2009) On the use of change in tumor size to predict survival in clinical oncology studies: toward a new paradigm to design and evaluate phase II studies. *Clin Pharmacol Ther* **86**(2):136-138.

Bruno R, Lu JF, Sun YN, and Claret L (2010) A modeling and simulation framework to support early clinical drug development decisions in oncology. *J Clin Pharmacol* **51**:6-8.

Burgess DL, Kohrman DC, Galt J, Plummer NW, Jones JM, Spear B, and Meisler MH (1995) Mutation of a new sodium channel gene, Scn8a, in the mouse mutant 'motor endplate disease'. *Nat Genet* **10**(4):461-465.

Bymaster FP, Calligaro DO, Falcone JF, Marsh RD, Moore NA, Tye NC, Seeman P, and Wong DT (1996) Radioreceptor binding profile of the atypical antipsychotic olanzapine. *Neuropsychopharmacology* **14**:87-96.

Campana D (2010) Minimal residual disease in acute lymphoblastic leukemia. *Hematology* **2010**(1):7-12.

Centorrino F, Baldessarri RJ, Kando JC, Frankenburg FR, Volpicelli SA, and Flood JG (1994) Clozapine and metabolites: concentrations in serum and clinical findings during treatment of chronically psychotic patients. *J Clin Psychopharmacol* **14**:119-125.

Chefer VI, Thompson AC, Zapata A, and Shippenberg TS (2009). Over view of brain microdialysis. *Curr Protoc Neurosci* **47**(7.1):1-7.

Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, Watanabe N, Nakagawa A, Omon IM, McGuire H, Tansella M, and Barbui C (2009) Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* **373**(9665):746-758.

Claret L, Lu J, Bruno R, Hsu CO, Hei YJ, and Sun YN (2012) Simulations using a drug-disease modeling framework and phase II data predict phase III survival outcome in first-line non-small-cell lung cancer. *Clin Pharmacol Ther* **92**(5):631-634.

Claret L, Gupta M, Han K, Joshi A, Sarapa N, He J, Powell B, and Bruno R (2013) Evaluation of tumor-size response metrics to predict overall survival in western and Chinese patients with first-line metastatic colorectal cancer. *J Clin Oncol* **31**(17):2110-2114.

Clozapine Product Insert, In: *Clinical Pharmacology Database*, 2013, Elsevier/Gold Standard, Tampa, FL.

Connolly C (2001) Price tag for a new drug: \$802 million. Findings of Tufts University Study are disputed by several watch dog groups. *Washington post*.

Couchman L, Morgan PE, Spencer EP, and Flanagan RJ (2010) Plasma clozapine, norclozapine, and the clozapine: norclozapine ratio in relation to prescribed dose and other factors: Data from a therapeutic drug monitoring service, 1993-2007. *Ther Drug Monit* **32**:438-447.

Cox DR (1972) Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B* **34**:187-220.

Creemers TIFH, Flik G, Hofland C, and Stratford RE Jr (2012) Microdialysis evaluation of clozapine and n-desmethylclozapine pharmacokinetics in rat brain. *Drug Metab Dispos* **40**:1909-1916.

Crom WR, de Graaf SS, Synold T, Uges DR, Bloemhof H, Rivera G, Christensen ML, Mahmoud H, and Evans WE (1994) Pharmacokinetics of vincristine in children and adolescents with acute lymphoblastic leukemia. *J Pediatr* **125**: 642-649.

Dawson SJ, Tsui DWY, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forsheew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, and Rosenfeld N (2013) Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med* **368**:1199-1209.

De Bono JS, Scher HI, Montgomery RB, Parker C, Miller MC, Tissing H, Doyle GV, Terstappen LWW, Pienta KJ, and Raghavan D (2008) Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin Cancer Res* **14**(19):6302–6309.

de Graaf SS, Bloemhof H, Vendrig DE, and Uges DR (1995) Vincristine disposition in children with acute lymphoblastic leukemia. *Med Pediatr Oncol* **24**:235-240.

de Lange ECM, Ravenstijn PGM, Groenendaal D, and van Steeg TS (2005) Towards the prediction of CNS drug effect profiles in physiological and pathological conditions using microdialysis and mechanism-based pharmacokinetic-pharmacodynamic modeling. *AAPS J* **7**(3): E532-E543.

de Lange ECM (2013) The mastermind approach to CNS drug therapy: Translational prediction of human brain distribution, target site kinetics and therapeutic effects. *Fluids Barriers CNS* **10**:12-28.

Dennison JB, Kulanthaivel P, Barbuch RJ, Renbarger JL, Ehlhardt WJ, and Hall SD (2006) Selective metabolism of vincristine in vitro by CYP3A5. *Drug Metab Dispos* **34**(8):1317-1327.

Department of health and human service, Guidance for Industry: Population Pharmacokinetics (1999) US Food and Drug Administration. Rockville, MD.

Derendorf H, Lesko LJ, Chaikin P, Colburn WA, Lee P, Miller R, Powell R, Rhodes G, Stanski D, and Venitz J (2000) Pharmacokinetic/pharmacodynamics modeling in drug research and development. *J Clin Pharmacol* **40**:1399-1418.

Dhani N, Tu D, Sargent DJ, Seymour L, and Moore MJ (2009) Alternate endpoints for screening phase II studies. *Clin Cancer Res* **15**:1873-1882.

DiMasi JA (2001) Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Ther* **79**:297-301.

DiMasi JA, Feldman L, Seckler A and Wilson A (2010) Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther* **87**(3):272-277.

Doran A, Obach RS, Smith BJ, Hosea NA, Becker S, Callegari E, Chen C, Chen X, Choo E, Cuabriga J, Cox LM, Gibbs JP, Gibbs MA, Hatch H, Hop CE, Kasman IN, Laperle J, Liu J, Liu X, Logman M, Maclin D, Nedza FM, Nelson F, Olson E, Rahematpura S, Raunig D, Rogers S, Schmidt K, Spracklin DK, Szewc M, Troutman M, Tseng E, Tu M, Van Deusen JW, Venkatakrishnan K, Walens G, Wang EQ, Wong D, Yasgar AS, and Zhang C (2005) The impact of P-glycoprotein on the disposition of drugs targeted for indications of the central nervous system: evaluation using the MDR1A/1B knockout mouse model. *Drug Metab Dispos* **33**:165–174.

Drye LT, Martin BK, Frangakis CE, Mintzer JE, Weintraub D, Porsteinsson AP, Schneider LS, Rabins PV, Munro CA, Meinert CL, Lyketsos CG, and DIADS-2 Research Group (2011) Do treatment effects vary among differing baseline depression criteria in depression in Alzheimer's disease study  $\pm 2$  (DIADS-2)? *Geriatric Psychiatry* **26**:573-583.

Duan H, and Wang J (2013) Impaired monoamine and organic cation uptake in choroid plexus in mice with targeted disruption of the plasma membrane monoamine transporter (Slc29a4) gene. *J Biol Chem* **288**(5):3535-3544.

Edelman MJ, Watson D, Wang X, Morrison C, Kratzke RA, Jewell S, Hodgson L, Mauer AM, Gajra A, Masters GA, Bedor M, Vokes EE, and Green MJ (2008) Eicosanoid Modulation in Advanced Lung Cancer: Cyclooxygenase-2 Expression Is a Positive Predictive Factor for Celecoxib+ Chemotherapy- Cancer and Leukemia Group B Trial 30203. *J Clin Oncol* **26**(6):848-855.

Ette EI, and Williams PJ (2007) *Pharmacometrics: the science of quantitative pharmacology*, Hoboken, N.J.: Wiley.

Fan J, and Li R (2002) Variable selection for Cox's proportional hazards model and frailty model. *Ann. Statist* **30**:74–99.

Fareed M, and Afzal M (2013) Single nucleotide polymorphism in genome-wide association of human population: A tool for broad spectrum service. *Egyptian Journal of Medical Human Genetics* **14**:123–134.

Feng Y, Pollock BG, Ferrell RE, Kimak MA, Reynold CF 3<sup>rd</sup>, and Bies RR (2006) Paroxetine: population pharmacokinetic analysis in late-life depression using sparse concentration sampling. *Br J Clin Pharmacol* **61**:558-569.

Ferguson MA, and Vaidya VS (2010) *Biomarkers: an evolutionary perspective*. Hoboken, N.J.: Wiley.

- Fischer J, Weide T, and Barnekow A (2005) The MICAL proteins and rab1: a possible link to the cytoskeleton? *Biochem Biophys Res Commun* **328**(2):415-423.
- Fridén M, Gupta A, Antonsson M, Bredberg U, and Hammarlund-Udenaes M (2007) In vitro methods for estimating unbound drug concentrations in brain interstitial and intracellular fluids. *Drug Metab Dispos* **35**:1711-1719.
- Fridlyand J, Kaiser LD, and Fyfe G(2011) Analysis of tumor burden versus progression-free survival for Phase II decision making. *Contemp Clin Trials* **32**:446-452.
- Friedman J, Hastie T, and Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical software* **33**(1):1-20.
- Gastonguay MR (2011) Full covariate models as an alternative to Methods relying on statistical significance for inferences about covariate effects: a review of methodology and 42 case studies. Population approach group in Europe meeting. Accessed January 2, 2014. Available from: <http://metrumrg.com/assets/pubs/GastonguayPAGE2011.pdf>
- Gastonguay MR, and El-Tahtawy A (2005) Modeling and Simulation Guided Design of a Pediatric Population Pharmacokinetic Trial for Hydromorphone. *The AAPS Journal* **7**:S2: Abstract W5318.
- Gastonguay MR, Gillespie WR, Khoo KC, and PPRU Network (1999) Population Pharmacokinetics in Pediatric Patients Using Bayesian Approaches with Informative Prior Distributions Based on Adults. *Pharm Res* **16**(S).
- Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, Johnson R, Zielke HR, Ferrucci L, Longo DL, Cookson MR, and Singleton AB (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* **6**(5):1-13.
- Grasela TH, and Sheiner LB (1984) Population pharmacokinetics of procainamide from routine clinical data. *Clin Pharmacokinet* **9**:545-554.
- Grasela T, and Donn SM (1985) Neonatal population pharmacokinetics for phenobarbital derived from routine clinical data. *Dev Pharmacol Ther* **8**:374-383.
- Grasela TH, Antal EJ, Ereshefsky L, Well BG, Evans RL, and Smith RB (1987) An evaluation of population pharmacokinetics in therapeutic trials. Part II. Detection of a drug-drug interaction. *Clin Pharmacol Ther* **42**:433-441.
- Gross JL, Rogers J, Polhamus D, Gillespie W, Friedrich C, Gong Y, Monz BU, Patel S, Staab A, and Retlich S (2013) A novel model-based meta-analysis to indirectly estimate the comparative efficacy of two medications: an example using DPP-4 inhibitors, sitagliptin and linagliptin, in treatment of type 2 diabetes mellitus. *Br Med J* **3**(3):1-11.

Gryder DS, and Rogawski MA (2003) Selective antagonism of GluR5 kainate-receptor-mediated synaptic currents by topiramate in rat basolateral amygdala neurons. *J Neurosci* **23**(18):7069-7074.

Gu CH, Li H, Levons J, Lentz K, Gandhi RB, Raghavan K, and Smith RL (2007) Predicting effect of food on extent of drug absorption based on physicochemical properties. *Pharm Res* **24**:1118-1130.

Haim N, Epelbaum R, Ben-Shahar M, Yarnitsky D, Simri W, and Robinson E (1994) Full dose vincristine (without 2-mg dose limit) in the treatment of lymphomas. *Cancer* **73**:2515-2519.

Heagerty PJ, Lumley T, and Pepe MS (2000) Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**:337-344.

Heagerty PJ, and Zheng Y (2005) Survival model predictive accuracy and ROC curves. *Biometrics* **61**:92-105.

Hertz DL, Roy S, Motsinger-Reif AA, Drobish A, Clark LS, MeLeod HL, Carey LA, and Dees EC (2013) CYP2C8\*3 increases risk of neuropathy in breast cancer patients treated with paclitaxel. *Annals of Oncol* **24**:1472-1478.

Hertz DL, Roy S, Jack J, Motsinger-Reif AA, Drobish A, Clark LS, Carey LA, Dees EC, and MeLeod HL (2014) Genetic heterogeneity beyond CYP2C8\*3 dose not explain differential sensitivity to paclitaxel-induced neuropathy. *Breast Cancer Res Treat* **145**:245-254.

Hiemke C, and Härtter S (2000) Pharmacokinetics of selective serotonin reuptake inhibitors. *Pharmacol Ther* **85**:11-28.

Holford NHG (2013) Wings for NONMEM. Available from: <http://wfn.sourceforge.net/>

Holford NHG (2013) Principle of covariate modeling. Accessed December 20, 2013. Available from: <http://holford.fmhs.auckland.ac.nz/docs/principles-of-covariate-modelling.pdf>

Horacek J, Bubenikova-Valesova V, Kopecek M, Palenicek T, Dockery C, Mohr P, and Höschl C (2006) Mechanism of action of atypical antipsychotic drugs and the neurobiology of schizophrenia. *CNS Drugs* **20**:389-409.

Hori H, Yamamoto N, Fujii T, Teraishi T, Sasayama D, Matsuo J, Kawamoto Y, Kinoshita Y, Ota M, Hattori K, Tatsumi M, Arima K, and Kunugi H (2012). Effects of the CACNA1C risk allele on neurocognition in patients with schizophrenia and healthy individuals. *Sci Rep* **2**:634.



Houk B (2009) Predictive capability of a Food and Drug Administration (FDA) overall survival model in non-small cell lung cancer in two phase II trials utilizing different anti-cancer agents. *Clin Pharmacol Ther* **85**.

Howlader N, Noone AM, and Krapcho M (2013) SEER cancer statistics review, 1975-2010. Bethesda, MD: National cancer institute. Accessed January 14, 2014. Available from: [seer.cancer.gov/csr/1975\\_2010/](http://seer.cancer.gov/csr/1975_2010/)

Huang J, Jacobson P, and Brundage R (2007) Prediction of unbound mycophenolic acid concentrations in patients after hematopoietic cell transplantation. *Ther Drug Monit* **29**: 385-390.

Huang SM, Abernethy DR, Wang Y, Zhao P, and Zineh I. Ingenuity pathways analysis. Available from: <https://analysis.ingenuity.com/pa/launch.jsp>

Ibrahim JG, Chu H and Chen LM(2010) Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol* **28**:2796-2801.

Ismail Z, Wessels A, Ucbida H, Ng W, Mamo DC, Rajji TK, Pollock BG, Mulsant BH, and Bies RR (2012) Age and sex impact clozapine plasma concentrations in inpatients and outpatients with schizophrenia. *Am J Geriatr Psychiatry* **20**(1):53-60.

Itman DG (1992) Analysis of Survival times. In: practical statistics for Medical research. London (UK): Chapman and Hall 365–393.

Jacobson GM, Marano G, Harbron H, and Abraham J (2013) Detection of axillary lymph node metastases and extra-axillary metastases with FDG PET/CT in breast cancer patients scheduled for neoadjuvant chemotherapy. *Pract Radiat Oncol* **3**(2Suppl 1):S24-25.

Jenkins M, Flynn A, Smart T, Harbron C, Sabin T, Ratnayake J, Delmar P, Herath A, Jarvis P, Matcham J, and on behalf of the PSI biomarker special interest group (2011) A statistician's perspective on biomarkers in drug development. *Pharmaceut Statist* **10**(6):494-507.

Jin Y, Pollock BG, Frank E, Cassano GB, Rucci P, Müller DJ, Kennedy JL, Forgione RN, Kirshner M, Kepple G, Fagiolini A, Kupfer DJ, and Bies RR (2010) Effect of age, weight and CYP2C19 genotype on escitalopram exposure. *J Clin Pharmacol* **50**:62-72.

Johnson M, Kozielska M, Pilla Reddy V, Vermeulen A, Li C, Grimwood S, de Greef R, Groothuis GM, Danhof M, and Proost JH (2011) Mechanism-based pharmacokinetic-pharmacodynamic modeling of the dopamine D2 receptor occupancy of olanzapine in rats. *Pharm Res* **28**:2490-2504.

- Kaminski RM, Banerjee M, and Rogawski MA (2004) Topiramate selectively protects against seizures induced by ATPA, a GluR5 kainate receptor agonist. *Neuropharmacology* **46**(8):1097-1104.
- Kang SP, and Ratain MJ (2010) Inconsistent labeling of food effect for oral agents across Therapeutic Areas: Differences between Oncology and Non Oncology Products. *Clin Cancer Res* **16**:4446-4451.
- Kaiser LD (2013) Tumor burden modeling versus progression-free survival for phase II decision making. *Clin Cancer Res* **19**:314-319.
- Kaitin KI (2010) Tumor burden modeling versus progression-free survival for phase II decision making. *Clin Cancer Res*. **19**(31):314-319.
- Karrison TG, Maitland ML, Stadler WM, and Ratain MJ (2007) Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non-small-cell lung cancer. *J Natl Cancer Inst* **99**:1455-1461.
- Kielbasa W, and Stratford RE (2012) Exploratory translational modeling approach in drug development to predict human brain pharmacokinetics and pharmacologically relevant clinical doses. *Drug Metab Dispos* **40**:1-12.
- Kloiber S, Czamara D, Karbalai N, Müller-Myhsok B, Hennings J, Holsboer F, and Lucae S (2012) ANK3 and CACNA1C—missing genetic link for bipolar disorder and major depressive disorder in two German case-control samples. *J Psychiatr Res* **46**(8):973-979.
- Kobayashi K, Ishizuka T, Shimada N, Yoshimura Y, Kamijima K, and Chiba K (1999) Sertraline N-Demethylation is catalyzed by multiple isoforms of human cytochrome P-450 in vitro Drug metabolism and disposition. *Drug Metab Dispos* **27**(7):763-766.
- Koch KM, Reddy NJ, Cohen RB, Lewis NL, Whitehead B, Mackay K, Stead A, Beelen AP, and Lewis LD (2009) Effects of food on the relative bioavailability of lapatinib in cancer patients. *J Clin Oncol* **27**:1191-1196.
- Kola I, and Landis J (2004) Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov*. **3**:711-716.
- Koehler MR, Wissinger B, Gorboulev V, Koepsell H, and Schmid M (1997) The two human organic cation transporter genes SLC22A1 and SLC22A2 are located on chromosome 6q26. *Cytogenet Cell Genet* **79**(3-4):198-200.
- Kozlelska M, Johnson M, Pilla Reddy V, Vermeulen A, Li C, Grimwood S, de Greef R, Groothuis GM, Danhof M, and Proost JH (2012) Pharmacokinetic-Pharmacodynamic modeling of the D<sub>2</sub> and 5-HT(2A) receptor occupancy of risperidone and paliperidone in rats. *Pharm Res* **29**(7):1932-1948.

- Kroeze WK, Hufeisen SJ, Popadak BA, Renock SM, Steinberg S, Ernsberger P, Jayathilake K, Meltzer HY, and Roth BL (2003) H<sub>1</sub>-histamine receptor affinity predicts short-term weight gain for typical and atypical antipsychotic drugs. *Neuropsychopharmacology* **28**:519–526.
- Lalonde RL, Kowalski KG, Hutmacher MM, Ewy W, Nichols DJ, Milligan PA, Corrigan BW, Lockwood PA, Marshall SA, Benincosa LJ, Tensfeldt TG, Parivar K, Amantea M, Glue P, Koide H, and Miller R (2007) Model-based drug development. *Clin Pharmacol & Therap* **82**(1):21-32.
- Lavin PT (1981) An alternative model for the evaluation of antitumor activity. *Cancer Clin Trials* **4**(4):379-380.
- Leandro-Garcia LJ, Leskelä S, Jara C, Gréen H, Avall-Lundqvist E, Wheeler HE, Dolan ME, Inglada-Perez L, Maliszewska A, de Cubas AA, Comino-Méndez I, Mancikova V, Cascón A, Robledo M, and Rodríguez-Antona C (2012). Regulatory polymorphisms in B-tubulin IIa are associated with paclitaxel-induced peripheral neuropathy. *Clin Cancer Res* **18**(16):4441-4448.
- Lee HP (1982) On clinical trials and survival analysis. *Singapore Med J* **23**:164–167.
- Lehnhardt M, Klein-Hitpass L, Kuhnen C, Homann HH, Daigeler A, Steinau HU, Roehrs S, Schnoor L, Steintraesser L, and Mueller O (2005) Response rate of fibrosarcoma cells to cytotoxic drugs on the expression level correlates to the therapeutic response rate of fibrosarcomas and is mediated by regulation of apoptotic pathways. *BMC Cancer* **5**(74):1-16.
- Lesko LJ, Rowland M, Peck CC, and Blaschke TF(2000) Optimizing the science of drug development : opportunities for better candidate selection and accelerated evaluation in humans. *Pharm Res* **17**:1335-1344.
- Li Y, Willer CJ, Sanna S, and Abecasis GR (2009) Genotype Imputation. *Annu Rev Genomics Hum Genet* **10**:387-406.
- Li Y, Willer CJ, Ding J, Scheet P, and Abecasis GR (2010) MACH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**:816-834.
- Lidow MS (2000) Neurotransmitter Receptors in Actions of Antipsychotic Medications. Lilenbaum RC, Herndon JE 2nd, List MA, Desch C, Watson DM, Miller AA, Graziano SL, Perry MC, Saville W, Chahinian P, Weeks JC, Holland JC and Green MR (2005) Single-agent versus combination chemotherapy in advanced non-small-cell lung cancer; the cancer and leukemia group B (study 9730). *J Clin Oncol* **23**:190-196.
- Lönnnerholm G, Frost BM, Abrahamsson J, Behrendz M, Castor A, Forestier E, Hevman M, Uges DR, and deGraaf SS (2008) Vincristine pharmacokinetics is related to clinical

outcome in children with standard risk acute lymphoblastic leukemia. *Br J Haematol* **142**(4):616-621.

Mahmood I (2007) Prediction of drug clearance in children: impact of allometric exponents, body weight, and age. *Ther Drug Monit* **29**(3):271-278.

Maitland ML, Bies RR, and Barrett JS (2011) A time to keep and a time to cast away categories of tumor response. *J Clin Oncol* **29**:3109-3111.

Maitland ML, and Schilsky RL (2011) Clinical trials in the era of personalized oncology. *CA Cancer J Clin* **61**:365-381.

Maitland ML, Schwartz LH, and Ratain MJ (2013) Time to tumor growth: a model end point and new metric system for oncology clinical trials. *J Clin Oncol* **31**(17):2070-2072.

Mandrekar SJ, and Sargent DJ (2010) Randomized phase II trials: time for a new era in clinical trial design. *J Thorac Oncol* **5**:932-934.

Martin BK, Frangakis CE, Rosenberg PB, Mintzer JE, Katz IR, Porsteinsson AP, Schneider LS, Rabins PV, Munro CA, Meinert CL, Niederehe G, and Lyketsos CG (2006) Design of Depression in Alzheimer's Disease Study-2. *Am J Geriatr Psychiatry* **14**(11):920-930.

Martinez MN, and Amidon GL (2002) A mechanistic approach to understanding the factors affecting drug absorption: A review of fundamentals. *J Clin Pharmacol* **42**:620-643.

Mathers CD, Lopez AD, and Murray CJL (2006) The burden of disease and mortality by condition: Data, methods and results, Chapter 3. In: Lopez AD, Mathers CD, Ezzati M, Jamison DT and Murray CJL, eds. *Global Burden of Disease and Risk Factors*. New York, NY: The World Bank and Oxford University Press.

McWhinney-Glass S, Winham SJ, Hertz DL, Revollo JY, Paul J, He Y, Brown R, Motsinger-Reif AA, and McLeod HL for the Scottish gynaecological clinical trials group (2013) Cumulative genetic risk predicts platinum/ taxane-induced neurotoxicity. *Clin Cancer Res* **19**(20):5769-5776.

Meibohm B, and Derendorf H (2002) Pharmacokinetic/pharmacodynamic studies in drug product development. *J Pharm Sci* **91**(1):18-31.

Meltzer HY, Matsubara S, and Lee JC (1989) Classification of typical and atypical antipsychotic drugs on the basis of Dopamine D-1, D-2 and Serotonin<sub>2</sub> Pki Values. *J Pharmacol Exp Ther* **251**(1):238-246.

Miller AA, Wang XF, Gu L, Hoffman P, Khatri J, Dunphy F, Edelman MJ, Bolger M, Vokes EE, Green MR and Cancer, and Leukemia Group B (CALGB) (2008) Phase II

randomized study of dose-dense docetaxel and cisplatin every 2 weeks with pegfilgrastim and darbepoetin alfa with and without the chemoprotector BNP7787 in patients with advanced non-small cell lung cancer (CALGB 30303). *J Thorac Oncol* **3**(10):1159-1165.

Miller AB, Hoogstraten B, Staquet M, and Winkler A (1981) Reporting results of cancer treatment. *Cancer* **47**:207-214.

Milligan PA, Brown MJ, Marchant B, Martin SW, van der Graaf PH, Benson N, Nucci G, Nichols DJ, Boyd RA, Mandema JW, Krishnaswami S, Zwillich S, Gruben D, Anziano RJ, and Laonde RL (2013) Model-based drug development: a rational approach to efficiently accelerate drug development. *Clin Pharmacol Ther* **93**(6):502-514.

Moertel CC, and Hanley JA (1976) The effect of measuring error in the results of therapeutic trials in advanced cancers. *Cancer* **38**:388-394.

Moriguchi S, Bies RR, Remington G, Suzuki T, Mamo DC, Watanabe K, Mimura M, Pollock BG, and Uchida H (2013) Estimated dopamine D<sub>2</sub> receptor occupancy and remission in schizophrenia: analysis of the CATIE data. *J Clin Psychopharmacol* **33**(5):682-685.

Muijsers RB, Plosker GL, and Noble S (2002) Sertraline: a review of its use in the management of major depressive disorder in elderly patients. *Drugs & Aging* **19**(5):377-392.

Murray RM, Lappin J, and Di FM (2008) Schizophrenia: from developmental deviance to dopamine dysregulation. *Eur Neuropsychopharmacol* **18**(Suppl 3):S129-S134.

National Cancer Institute (2009) Common terminology criteria for adverse event v4.0 (CTCAE). The National Cancer Institute of the National Institutes of Health (NIH). Accessed February 4, 2013, Accessed February 2, 2012. Available from: [http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE\\_4.03\\_2010-06-14\\_QuickReference\\_5x7.pdf](http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf)

National Cancer Institute (2011) National Cancer Institute of the National Institutes of Health (NIH). Cancer topics, In: Cancer Drug Information. Accessed January 20, 2014, Available from: <http://www.cancer.gov/cancertopics/druginfo/alphabetical>

Nordström AL, Farde L, Nyberg S, Karlsson P, Halldin C, and Sedvall G (1995) D<sub>1</sub>, D<sub>2</sub>, and 5-HT<sub>2</sub> receptor occupancy in relation to clozapine serum concentration: a PET study of schizophrenia. *Am J Psychiatry* **152**:1444-1449.

Ohnishi A, Mitsudome A, and Murai Y (1987) Primary segmental demyelination in the sural nerve in Cockayne's syndrome. *Muscle Nerve* **10**:163-167.

Olsen CK, Brennum LT, and Kreilgaard M (2008) Using pharmacokinetic-pharmacodynamic modelling as a tool for prediction of therapeutic effective plasma levels of antipsychotics. *Eur J Pharmacol* **584**:318-327.

Owen MJ, Morgan WN, Plott SJ, and Nemeroff CB (1997) Neurotransmitter receptor and transporter binding profile of antidepressants and their metabolites. *J Pharmacol Exp Ther* **283**:1305-1322.

Oxnard GR, Zhao B, Sima CS, Ginsberg MS, James LP, Lefkowitz RA, Guo P, Kris MG, Schwartz LH, and Riely GJ (2011) Variability of lung tumor measurements on repeat computed tomography scans taken within 15 minutes. *J Clin Oncol* **29**(23):3114-3119.

Park SB, Goldstein D, Krishnan AV, Lin CSY, Friedlander L, Cassidy J, Koltzenburg M, and Kiernan M (2013) Chemotherapy-induced peripheral neurotoxicity: a critical analysis. *CA Cancer J Clin* **63**:419-437.

Petroff OA (2002) GABA and glutamate in the human brain. *Neuroscientist* **8**(6):562-573.

Purcell S. PLINK. Available from: <http://pngu.mgh.harvard.edu/purcell/plink/>

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, and Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**(8):904-909.

Postma TJ, Benard BA, Huijgens PC, Ossenkuppele GJ, and Heimans JJ (1993) Long-term effects of vincristine on the peripheral nervous system. *J Neurooncol* **15**:23-27.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, and Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* **81**(3):559-575.

Raedler TJ, Hinkelmann K, and Wiedemann K (2008) Variability of the in vitro metabolism of clozapine. *Clin Neuropharmacol* **31**:347-352.

Ribbing J (2007) Covariate model building in nonlinear mixed effects models. *Acta Universitatis Upsalienensis. Digital comprehensive summaries of Uppsala dissertations from the faculty of pharmacy.* 59.

Rogawski MA, Gryder D, Castaneda D, Yonekawa W, Banks MK, and Lia H (2003) GluR5 kainate receptors, seizures, and the amygdala. *Ann N Y Acad Sci* **985**:150-162.

Romero OJ, Garcia GB, Campos JF, and Touijer KA (2013) Prostate cancer biomarkers: an update. *Urol Oncol* 1-9.

Ronfeld RA, Tremaine LM, and Wilner KD (1997) Pharmacokinetics of sertraline and its N-demethyl metabolite in elderly and young male and female volunteers. *Clin Pharmacokinet* **32** Suppl 1:22-30.

Rosenberg PB, Drye LT, Martin BK, Frangakis C, Mintzer JE, Weintraub D, Porsteinsson AP, Schneider LS, Rabins PV, Munro CA, Meinert CL, Lyketsos CG, and

DIADS-2 Research Group (2010) Sertraline for the treatment of depression in Alzheimer's disease. *Am J Geriatr Psychiatry* **18**(2):136-145.

Ryan CJ, Smith MR, Fong L, Rosenberg JE, Kantoff P, Raynaud F, Martins V, Lee G, Kheoh T, Kim J, Molina A, and Small EJ (2010) Phase I clinical trial of the CYP17 inhibitor abiraterone acetate demonstrating clinical activity in patients with castration-resistant prostate cancer who received prior ketoconazole therapy. *J Clin Oncol* **28**:1481-1488.

Sakurai E, Kurihara T, Kouchi K, Saegusa H, Zong S, and Tanabe T (2009) Upregulation of casein kinase 1 $\epsilon$  in dorsal root ganglia and spinal cord after mouse spinal nerve injury contributes to neuropathic pain. *Mol Pain* **5**:1-14.

Samak G, Narayanan D, Jaggar JH, and Rao R (2011) CaV1.3 channels and intracellular calcium mediate osmotic stress-induced N-terminal c-Jun kinase activation and disruption of tight junctions in Caco-2 CELLMONOLAYERS. *J Biol Chem*. **286**(34):30232-30243.

Savic RM, Jonker DM, Kerbusch T, and Karlsson MO (2007) Implementation of a transit compartment model for describing drug absorption in pharmacokinetic studies. *J Pharmacokinet Pharmacodyn* **34**:711-726.

Scerwin CM, Saldaña SN, Bies RR, Aman MG, and Vinks AA (2012) Population pharmacokinetic modeling of risperidone and 9-hydroxyrisperidone to estimate CYP2D6 subpopulations in children and adolescents. *Ther Drug Monit* **34**(5):535-544.

Schmidt C. CA-125: a biomarker put to the test (2011) *J Natl Cancer Inst* **103**(17):1290-1291.

Schneider LS, Nelson JC, Clary CM, Newhouse P, Krishnan KR, Shiovitz T, Weihs K, and Sertraline Elderly Depression Study Group (2003) An 8-Week multicenter, parallel-group, double-blind, placebo-controlled study of sertraline in elderly outpatients with major depression. *Am J Psychiatry* **160**(7):1277-1285.

Schrapppe M, Reiter A, Ludwig WD, Harbott J, Zimmermann M, Hiddemann W, Niemeyer C, Henze G, Feldges A, Zintl F, Kornhuber B, Ritter J, Welte K, Gadner H, and Riehm H (2000) Improved outcome in childhood acute lymphoblastic leukemia despite reduced use of anthracyclines and cranial radiotherapy: results of trial ALL-BFM90. German-Austrian-Swiss ALL-BFM Study Group. *Blood* **95**(11):3310-3322.

Schrapppe M, Reiter A, Zimmermann M, Harbott J, Ludwig WD, Henze G, Gadner H, Odenwald E, and Riehm H (2000) Long-term results of four consecutive trials in childhood ALL performed by the ALL-BFM study group from 1981 to 1995. Berlin-Frankfurt-Münster. *Leukemia* **14**(12):2205-2222.

Schwartz LH, Ginsberg MS, DeCorato D, Rothenberg LN, Einstein S, Kijewski P, and Panicek DM (2000). *J Clin Oncol* **18**(10):2179-2184.

Seeman PT (2002) Atypical antipsychotics: mechanism of action. *Can J Psychiatry* **47**:27-38.

Shamir R (2010) Gene enrichment analysis. Available from:  
[http://lectures.molgen.mpg.de/Algorithmische\\_Bioinformatik\\_WS1213/lec14a\\_Shamir\\_GO\\_GSEA.pdf](http://lectures.molgen.mpg.de/Algorithmische_Bioinformatik_WS1213/lec14a_Shamir_GO_GSEA.pdf)

Sharma V, and McNeill JH (2009) To scale or not to scale: the principles of dose extrapolation. *Br J Clin Pharmacol* **157**:907-921.

Sheiner LB, Rosenberg B, and Marathe V (1977) Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *J Pharmacokinet Biopharm* **5**:445-479.

Sheiner LB, and Beal SL (1980) Evaluation of methods for estimating population pharmacokinetics parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. *J Pharmacokinet Biopharm* **8**:553-571.

Sheiner LB, and Beal SL (1981) Evaluation of methods for estimating population pharmacokinetic parameters. II. Biexponential model and experimental pharmacokinetic data. *J Pharmacokinet Biopharm* **9**:635-651.

Sheiner LB, and Beal SL (1982) Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: routine pharmacokinetic data. *J Pharmacokinet Biopharm* **11**:303-319.

Sheiner LB (1997) Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* **61**(3):275-291.

Singh BN, and Mallhotra BK (2004). Effects of food on the clinical pharmacokinetics of anticancer agents. *Clin Pharmacokinet* **43**:1127-1156.

Singh R, and Mukhopadhyay K (2011) Survival analysis in clinical trials: basics and must know areas. *Perspect Clin Res* **2**(4):145-148.

Smits MG, Gabreels FJ, Renier WO, Joosten EM, Gabreels-Festen AA, ter Laak HJ, Pinckers AJ, Hombergen GC, Notermans SL, and Thijssen HO (1982) Peripheral and central myelinopathy in Cockayne's syndrome. Report of 3 siblings. *Neuropediatrics* **13**:161-167.



- Soeiro-de-Souza MG, Otaduy MC, Dias CZ, Bio DS, Machado-Vieira R, and Moreno RA (2012) The impact of the CACNA1C risk allele on limbic structures and facial emotions recognition in bipolar disorder subjects and healthy controls. *J Affect Disord* **141**(1):94-101.
- Sohn I, Kim J, Jung SH, and Park C (2009) Gradient lasso for Cox proportional hazards model. *Bioinformatics* **25**(14):1775-1781.
- Sprouse J, Clarke T, Reynolds L, Heym J, and Rollema H (1996) Comparison of the effects of sertraline and its metabolite demethylsertraline on blockade of central 5-HT reuptake in vivo. *Neuropsychopharmacology* **14**:225-231.
- Stanulla M, and Schrappe M (2009) Treatment of childhood acute lymphoblastic leukemia. *Semin Hematol* **46**(1):52-63.
- Steimer JL, and Gieschke R (2000) Pharmacometrics: modeling and simulation tools to improve decision making in clinical drug development. *Eur J Drug Metab Pharmacokinet.* **25**(1):49-58.
- Stein WD, Gulley JL, Schlom J, Madan RA, Dahut W, Figg WD, Ning YM, Arlen PM, Price D, Bates SE, and Fojo T (2011) Tumor regression and growth rates determined in five intramural NCI prostate cancer trials: the growth rate constant as an indicator of therapeutic efficacy. *Clin Cancer Res* **17**:907-917.
- Steinberg M, Shao H, Zandi P, Lyketsos CG, Welsh-Bohmer KA, Norton MC, Breitner JC, Steffens DC, Tschanz JT, and Cache County Investigators (2008) Point and 5-year period prevalence of neuropsychiatric symptoms in dementia: the Cache County Study. *Int J Geriatr Psychiatry* **23**(2):170-177.
- Takeuchi H, Suzuki T, Bies RR, Remington G, Mamo DC, Pollock BG, Mimura M, and Uchida H (2013) Estimated dopamine D2 receptor occupancy from plasma concentrations of atypical antipsychotics and subjective experience/drug attitude in schizophrenia: an analysis of the CATIE data. *Schizophr Res* **150**(2-3):373-379.
- Tan BZ, Jiang F, Tan MY, Yu D, Huang H, Shen Y, and Soong TW (2011). Functional characterization of alternative splicing in the C terminus of L-type CaV1.3 channels. *J Biol Chem* **286**(49):42725-42735.
- Tanaka C, Yin OQ, Sethuraman V, Smith T, Wang X, Grouss K, Kantarjian H, Giles F, Ottmann OG, Galitz L, and Schran H (2010) Clinical pharmacokinetics of the BCR-ABL tyrosine kinase inhibitor nilotinib. *Clin Pharmacol Ther* **87**:197-203.
- Tardif JC, Heinonen T, Orloff D, and Libby P (2006) Vascular biomarkers and surrogates in cardiovascular disease. *Circulation* **113**:2936-2942.

Te Loo DM, Hagleitner MM, van Schie R, van der Sluis IM, Coenen M.J, and Hoogerbrugge P (2010) Genetic variants in MDR1 and CYP3A5 and vincristine toxicity in children with acute lymphoblastic leukemia. *J Clin Oncol* **28**(15\_suppl):9526.

Teorell T (1937) Kinetics of distribution of substances administered to the body. I. The extra-vascular modes of administration. *Arch Int Pharmacodyn* **57**:205-225.

Teorell T (1937) Kinetics of distribution of substances administered to the body. II. The intra-vascular modes of administration. *Arch Int Pharmacodyn* **57**:226-240.

Terman JR, Mao T, Pasterkamp RJ, Yu HH, and Kolodkin AL (2002) MICALs, a family of conserved flavoprotein oxidoreductases, function in plexin-mediated axonal repulsion. *Cell* **109**(7):887-900.

Tibshirani RJ (1996) Regression shrinkage and selection via the LASSO. *J Roy Statist Soc B* **58**(1):267-288.

Tibshirani RJ (1997) The LASSO method for variable selection in the cox model. *Stat Med* **16**:385-395.

Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, and Gwyther SG (2000) New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* **92**:205-216.

Therneau TM (2014) R package “survival v 2.37-7”. Available from: <http://cran.rproject.org/web/packages/survival/survival.pdf>

Travis LB, Fossa SD, Sesso HD, Frisina RD, Herrmann D++N, Beard CJ, Feldman DR, Pagliaro LC, Miller RC, Vaughn DJ, Einhorn LH, Cox NJ, Dolan ME, and for the platinum study group (2014) Chemotherapy-induced peripheral neurotoxicity and ototoxicity: new paradigms for translational genomics. *J Natl Cancer Inst* **106**(5).

Tsuboi T, Bies RR, Suzuki T, Mamo DC, Pollock BG, Graff-Guerrero A, Mimura M, and Uchida H (2013) Hyperprolactinemia and estimated dopamine D2 receptor occupancy in patients with schizophrenia: analysis of the CATIE data. *Prog Neuropsychopharmacol Biol Psychiatry* **45**:178-182.

van Erp NP, Baker SD, Zandvliet AS, Ploeger BA, den Hollander M, Chen Z, den Hartigh J, König-Quartel JM, Guchelaar HJ, and Gelderblom H (2011) Marginal increase of sunitinib exposure by grapefruit juice. *Cancer Chemother Pharmacol* **67**(3):695–703.

Verstappen CC, Koeppen S, Heimans JJ, Huijgens PC, Scheulen ME, Strumberg D, Kiburg B, and Postma TJ (2005) Dose-related vincristine-induced peripheral neuropathy with unexpected off-therapy worsening. *Neurology* **64**(6):1076-1077.

Vozech S, Katz G, Steiner V, and Follath F (1982) Population pharmacokinetic parameters in patients treated with oral mexiletine. *Eur J Clin Pharmacol* **23**:445-451.

Vrijens B, Tousset E, Rode R, Bertz R, Mayer S, and Urquhart J (2009) Successful projection of the time course of drug concentration in plasma during a 1-year period from electronically compiled dosing-time data used as input to individually parameterized pharmacokinetic models. *J Clin Pharmacol* **45**(4):461-467.

Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E, and Gobburu J (2009) Elucidation of Relationship Between Tumor Size and Survival in Non-Small-Cell Lung Cancer Patients Can Aid Early Decision Making in Clinical Drug Development. *Clin Pharmacol Ther* **86**(2):167-174.

Wanger JG (1973) A modern view of pharmacokinetics. *J Pharmacokinet Biopharm.* **1**(5):363-401.

Warrington SJ (1991) Clinical Implication of the pharmacology of sertraline. *Int Clin Psychopharmacol Suppl* **2**:11-21.

Weintraub D, Rosenberg PB, Drye LT, Martin BK, Frankakis C, Mintzer JE, Porsteinsson AP, Schneider LS, Rabins PV, Munro CA, Meinert CL, Lyketsos CG, and DIADS-2 Research Group (2010) Sertraline for the treatment of depression in Alzheimer's disease: Week-24 outcomes. *Am J Geriatr Psychiatry* **18**(4):332-340.

Widmark E (1933) Verteilung und Umwandlung des Äthylalkohols im Organismus des Hundes. *Biochem Z* **267**:128-134.

Widmark E (1933) Der Einfluss der Nahrungsbestandteile auf den Alkoholgehalt des Blutes. *Biochem Z* **267**:135-142.

Winstanley PA, and Orme M (1989) The effects of food on drug bioavailability. *Br J clin Pharmacol* **28**:621-628.

Yang T, Puckerin A, and Colecraft HM (2012) Distinct RGK GTPases differentially use  $\alpha$ 1- and auxiliary  $\beta$ -binding-dependent mechanisms to inhibit CaV1.2/CaV2.2 channels. *PLoS One* **7**(5):e37079.

Yang Y. Multiple Imputation for Missing Data: Concepts and New Development (Version 9.0) Available from:

<http://support.sas.com/rnd/app/papers/multipleimputation.pdf>

Yap TA, Sandhu SK, Workman P, and de Bono JS (2010) Envisioning the future of early anticancer drug development. *Nat Rev Cancer* **10**:514-523.

Yin MJ, Shao L, Voehringer D, Smeal T, and Jallal B (2003) The serine/threonine kinase Nek6 is required for cell cycle progression through mitosis. *J Biol Chem* **278**(52):52454-52460.

Yin OQ, Gallagher N, Li A, Zhou W, Harrell R, and Schran H (2010) Effect of grapefruit juice on the pharmacokinetics of nilotinib in healthy participants. *J Clin Pharmacol* **50**:188–194.

Zolof® product information New York: Roeg (2001) Division of Pfizer Inc. Accessed May 5, 2012. Available from:  
[http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2005/019839s053S0541bl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2005/019839s053S0541bl.pdf)

ZYTIGA label. Accessed December 12, 2012. Available from:  
[www.accessdata.fda.gov/drugsatfda\\_docs/label/2011/2023791bl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/2023791bl.pdf)

## CURRICULUM VITAE

Claire Li

### EDUCATION TRAINING

**Doctor of Philosophy (Medical and Molecular Genetics)** 2009- 08/2014

Indiana University, Indianapolis, IN

Advisor: Robert R. Bies, PharmD, Ph.D.

Overall GPA:3.95/4.00

**Bachelor of Science in Biochemistry and Molecular Biology** 2005- 2008

The University of Georgia, School of Science, Athens, GA

Overall GPA:3.83/4.00 Major GPA:3.96/4.00

**Disease and Therapeutic modeling program Training** 2010-2013

2010: **R Programming for Pharmacometrics** (Metrum Institute)

2010: PHAR-F 813 **Clinical Pharmacokinetics** (Indiana University)

2011: **Essentials of Population PK-PD Modeling & Simulation** (Metrum Institute)

2011: **Monolix 3.2 workshop** (INRIA, ACOP)

2011: **New and Advanced Features of NONMEM7&PDx-Pop4**  
(ICON Development Solutions, ACOP)

2011: **Uppsala Pharmacometric Summer School**

2011: PHAR-F 836 **Pharmacogenetics and Physiologic Disposition of Drugs** (Indiana University)

2011: **Introduction to Bayesian PK-PD modeling & simulation** (Metrum Institute)

2012: **Pharmacokinetic-pharmacodynamic modeling** (State University of New York at Buffalo)

2012: **2-Day workshop on monoclonal antibody pharmacokinetics & Pharmacodynamics** (State University of New York at Buffalo)

2012: **PKPD-modeling of Continuous and Categorical data in NONMEM** (Uppsala University, PAGE)

2013: **Quantitative System Pharmacology** (ASCPT)

2013: **New & Advanced Methods of NONMEM 7 & PDx-Pop 5.0 including New Features in the Upcoming Release of NONMEM 7.3** (ICON Development Solutions, ACOP)

2013: **Using MATLAB® and SimBiology® for Mechanistic Systems Pharmacology Workflows** (Mathworks,ACOP)

2013: **Pharmacokinetic-pharmacodynamic modeling from Sunny Buffalo** (Genentech)

## PROFESSIONALEXPERIENCES

- June 2013-Sep 2013 **Genentech** [Internship]  
Clinical Pharmacology, Genentech, South San Francisco, CA
- May 2008-Jul 2009 **Volt Temp Services at Eli Lilly and Company** [Biochemist-Contractor]  
In vitro Metabolism and Pharmacogenomics group, Eli Lilly and Company, Indianapolis, IN
- Jan-Dec 2007 **Complex Carbohydrate Research Center** [Research Assistant]  
University of Georgia, Athens, GA
- Jan- May 2008 **Department of Biochemistry and Molecular Biology** [Research Assistant]  
University of Georgia, Athens, GA
- Aug-Dec 2007 **Complex Carbohydrate Research Center** [Technician]  
University of Georgia, Athens, GA
- May-August 2007 **Summer Undergraduate Research Program (SURP)** [Internship]  
University of Georgia, Athens, GA

## RESEARCH EXPERIENCES

- 2013-2014 **Population pharmacokinetic modeling of vincristine and metabolite in children with acute lymphoblastic leukemia from Children's Oncology Group**  
Indiana University School of Medicine, Indianapolis, IN  
Mentor: Robert R. Bies, PharmD, Ph.D.
- 2013 **Pharmacokinetic and pharmacodynamic modeling of antibody-drug conjugate target glycosylated antigen in patients with ovarian cancer**  
Genentech, South San Francisco, CA  
Mentor: Jian Xu, Ph.D.
- 2012-2014 **Develop a genetic signature model to predict efficacy and toxicity using GWAS analysis by Lasso penalized logistic regression in Childhood Acute Lymphoblastic Leukemia patients**  
Indiana University School of Medicine, Indianapolis, IN  
Mentor: Lang Li, Ph.D.
- 2012-2013 **Prediction of brain clozapine and norclozapine concentrations in human using a scaled PK model for rat brain and plasma PK**  
Indiana University School of Medicine, Indianapolis, IN

Mentor: Robert R. Bies, PharmD, Ph.D.

- 2012-2013 **Limitations on longitudinal tumor size modeling of non-small cell lung cancer with data collected from completed cooperative group trials**  
Indiana University School of Medicine, Indianapolis, IN  
Mentor: Robert R. Bies, PharmD, Ph.D.
- 2011-2013 **Population pharmacokinetic modeling of vincristine, IU Eldoret Kenya and the Advance trials**  
Indiana University School of Medicine, Indianapolis, IN  
Mentor: Robert R. Bies, PharmD, Ph.D.
- 2011-2012 **The utility of clinical trial simulation for evaluation of study design with a significant food effect on drug exposure: application to abiraterone and nilotinib**  
Indiana University School of Medicine, Indianapolis, IN  
Mentor: Robert R. Bies, PharmD, Ph.D.
- 2010-2011 **Population pharmacokinetic modeling of sertraline treatment in Alzheimer's disease patients: the DIADS-2 study**  
Indiana University School of Medicine, Indianapolis, IN  
Mentor: Robert R. Bies, PharmD, Ph.D.
- 2009-2010 **Population pharmacokinetic modeling of ziprasidone in patients with schizophrenia from the CATIE study**  
Indiana University School of Medicine, Indianapolis, IN  
Mentor: Robert R. Bies, PharmD, Ph.D.
- 2008-2009 **Pharmacogenomics of gemcitabine metabolism: functional analysis of genetic variants in cytidine deaminase and deoxycytidine kinase**  
In Vitro Metabolism and Pharmacogenomics group, Eli Lilly and Company, Indianapolis, IN  
Mentor: Yingying Guo, Ph.D.
- 2008 **Neo gene construction into an *Escherichia coli* cell.**  
Biochemistry and Molecular Biology Department, University of Georgia, Athens, GA  
Mentor: Shaying, Zhao, Ph.D.
- 2007-2008 **Protein expression of human slit3 related to congenital diaphragmatic hernia(CDH) in *Pichia Pastoris*.**  
Complex Carbohydrate Research Center, University of Georgia, Athens, GA  
Mentor: Lianchun, Wang, Ph.D.

### TEACHING EXPERIENCES (LECTURE)

- 2014 Time to event (survival analysis overview)  
Didactics course (Division of Clinical Pharmacology)  
IU school of Medicine, Indianapolis, IN.
- 2014 Time to event (survival analysis overview)  
Quantitative Aspects Translational Research course (GRAD G668)  
IU school of Medicine, Indianapolis, IN.
- 2013 Time to event (survival analysis overview)  
Quantitative Aspects Translational Research course (GRAD G668)  
IU school of Medicine, Indianapolis, IN.

### TEACHING ASSISTANT EXPERIENCES

- 2014 Quantitative Aspects Translational Research course (GRAD G668)  
IU school of Medicine, Indianapolis, IN.  
Instructor: Robert R. Bies, PharmD, Ph.D.
- 2013 Quantitative Aspects Translational Research course (GRAD G668)  
IU school of Medicine, Indianapolis, IN.  
Instructor: Robert R. Bies, PharmD, Ph.D.
- 2012 Quantitative Aspects Translational Research course (GRAD G668)  
IU school of Medicine, Indianapolis, IN.  
Instructor: Robert R. Bies, PharmD, Ph.D.

### SYMPOSIUM HOSTING EXPERIENCES

- 2013 The Third Indiana Clinical and Translational Sciences Institute (CTSI)  
Symposium on Disease and Therapeutic Response Modeling  
[Assist to organize and coordinate the event]
- 2012 The Second Indiana Clinical and Translational Sciences Institute (CTSI)  
Symposium on Disease and Therapeutic Response Modeling  
[Assist to organize and coordinate the event]
- 2011 The First Indiana Clinical and Translational Sciences Institute (CTSI)  
Symposium on Disease and Therapeutic Response Modeling  
[Assist to organize and coordinate the event]



## **Awards/Honors/Citations**

Hope Scholarship, 2005-2007

CRC Press General Chemistry Award, 2005

James W. Mitchell Scholarship, 2005

## **PUBLICATIONS**

1. **Li CH**, Li L, Zhang P, Renbarger JL. Genetic signature to predict vincristine neuropathy and relapse in children with acute lymphoblastic leukemia. [In preparation]
2. **Li CH**, Li L, Skiles JL, Then, JE, Renbarger JL. Genetic determinants of vincristine neuropathy in pediatric acute lymphoblastic leukemia patients. [In preparation]
3. **Li CH**, Bies RR, Yaning Wang, Sharma MR, Karovic S, Werk L, Jiang C, Owzar K, Edelman MJ, Miller AA, Vokes EE, Oto A, Schwartz LH, Ratain MJ, Maitland ML. Data quality constrains utility of computational modeling of tumor burden in non-small cell lung cancer clinical trials. [In preparation for JCO]
4. Skiles JL, **Li CH**, ChienWei C, Smith EM, Olbara G, Jones DR, Mostert S, Kaspers GJ, Li L, Njuguna F, Renbarger JL. Biomarkers of vincristine-induced peripheral neuropathy in Kenyan children with cancer. [Submitted to JCO]
5. **Li CH**, Stratford RE, Velez de Mendizabal N, Thomas I.F.H. Cremers, Bies RR. Prediction of brain clozapine and norclozapine concentrations in human using a scaled PK model for rat brain and plasma PK. [Accepted for publication in Journal of translational medicine]
6. **Li CH**, Sherer EA, Lewis LD, Bies RR. Clinical trial simulation to evaluate study design: capturing abiraterone and nilotinib exposure. [In preparation for JCP]
7. Baker AJ, Wickremsinhe E, **Li CH**, Oluyedun OA, Dantzig AH, Hall SD, Qian YW, Ring BJ, Wrighton SA, Guo Y. Pharmacogenomics of Gemcitabine Metabolism: Functional Analysis of Genetic Variants in Cytidine Deaminase and Deoxycytidine Kinase. *Drug Metab Dispos*, 2013, 41(3):541-545.
8. **Li CH**, Pollock BG, Lyketsos CG, Vaidya V, Drye LT, Kirshner M, Sorisio D, Bies RR and for the DIADS-2 Research Group. Population Pharmacokinetic Modeling of Sertraline Treatment in Alzheimer's Disease Patients: The DIADS-2 Study. *J Clin Pharmacol*, 2012.
9. Wessels AM, Bies RR, Pollock BG, Schneider LS, Lieberman JA, Stroup S, **Li CH**, Coley K, Kirshner MM, Marder SR. Population Pharmacokinetic Modeling of

Ziprasidone in Patients with Schizophrenia from the CATIE Study. *J Clin Pharmacol*, 2011, 51(11):1587-1591.

#### ABSTRACTS AND SCIENTIFIC PRESENTATIONS

1. Xu J, **Li CH**, Zhang R, Samineni D, Wang Y, Girish S, Jin J. Semi-mechanistic PK/PD modeling for CA125 assessment of drug candidates in ovarian cancer patients. In: American Conference on Pharmacometrics, Las Vegas, CA, 2014.
2. Skiles JL, **Li CH**, Li L, Renbarger JL. Genetic determinants of vincristine neuropathy in pediatric acute lymphoblastic leukemia patients. In: 46<sup>th</sup> Congress of the International Society of Pediatric Oncology, Toronto, Canada, 2014.
3. Renbarger JL, **Li CH**, Skiles JL, Li L. Functional Genome Wide Enrichment Association Analysis of Vincristine Neuropathy. In: 46<sup>th</sup> Congress of the International Society of Pediatric Oncology, Toronto, Canada, 2014.
4. **Li CH**, Stratford Jr RE, Velez de Mendizabal N, Thomas I.F.H. Cremers, Bies RR. Prediction of brain clozapine and norclozapine concentrations in human using a scaled PK model for rat brain and plasma PK. In: The Third Indiana CTSI Symposium on Disease and Therapeutic Response Modeling, Indianapolis, IN, 2013.
5. **Li CH**, Grish S, Xu J. Semi-mechanistic model based assessment of longitudinal cancer antigen 125 in ovarian cancer patients. In: Genentech, South San Francisco, CA, 2013.
6. **Li CH**, Grish S, Xu J. Semi-mechanistic model based assessment of longitudinal cancer antigen 125 in ovarian cancer patients. In: Genentech, South San Francisco, CA, 2013.
7. **Li CH**, Grish S, Xu J. A review of cancer antigen 125 clinical utility in ovarian cancer patients leading towards a quantitative modeling approach. In: Genentech, South San Francisco, CA, 2013.
8. **Li CH**, Stratford Jr RE, Velez de Mendizabal N, Thomas I.F.H. Cremers, Bies RR. Prediction of brain clozapine and norclozapine concentrations in human using scaled PK model for rat brain and plasma PK. In: American Conference on Pharmacometrics, Fort Lauderdale, FL, 2013.
9. Quinney SK, Chang C, **Li CH**, Li L, Bies RR, Egbelakin A, Ho R, Hutchinson R, Smith EL, Wells E, Renbarger JL. Comparison of population pharmacokinetic and physiologic based pharmacokinetic models for vincristine in pediatric acute lymphoblastic leukemia. ASCPT, Indianapolis, IN, 2013.
10. **Li CH**, Sherer EA, Lewis LD, Bies RR. The Utility of Clinical Trial Simulation for Evaluation of Study Design with a significant food effect on Drug Exposure: Application to Abiraterone and Nilotinib. In: The Second Indiana Clinical and Translational Sciences

Institute (CTSI) Symposium on Disease and Therapeutic Response Modeling, Indianapolis, IN, 2012.

11. **Li CH**, Sherer EA, Lewis LD, Bies RR. The Utility of Clinical Trial Simulation for Evaluation of Study Design with a significant food effect on Drug Exposure: Application to Abiraterone and Nilotinib. In: Indiana CTSI's Fourth Annual Meeting, Indianapolis, IN, 2012.

12. **Li CH**, Sherer EA, Lewis LD, Bies RR. The Utility of Clinical Trial Simulation for Evaluation of Study Design with a significant food effect on Drug Exposure: Application to Abiraterone and Nilotinib. In: Population Approach Group Europe (PAGE) meeting, Venice, Italy, 2012.

13. **Li CH**, Sherer EA, Lewis LD, Bies RR. The Utility of Clinical Trial Simulation for Evaluation of Study Design with a significant food effect on Drug Exposure: Application to Abiraterone and Nilotinib. In: Department of Medical and Molecular Genetics Annual poster section, Indiana University School of Medicine, Indianapolis, IN, 2012.

14. **Li, CH**, Pollock BG, Lyketsos CG, Vaidya V, Drye LT, Kirshner M, Sorisio D, Bies RR and for the DIADS-2 Research Group. Population Pharmacokinetic Modeling of Sertraline Treatment in Alzheimer's Disease Patients: The DIADS-2 Study. In: The Indiana University School of Medicine BioMedicalGateway (IBMG) Program recruitment poster section, Indiana University School of Medicine, Indianapolis, IN, 2012.

15. **Li, CH**, Pollock BG, Lyketsos CG, Vaidya V, Drye LT, Kirshner M, Sorisio D, Bies RR and for the DIADS-2 Research Group. Population Pharmacokinetic Modeling of Sertraline Treatment in Alzheimer's Disease Patients: The DIADS-2 Study. In: The First Indiana Clinical and Translational Sciences Institute (CTSI) Symposium on Disease and Therapeutic Response Modeling, Indianapolis, IN, 2011.

16. **Li, CH**, Sherer EA, Bies RR. The Clinical trial simulations of the anticancer agents: Abiraterone and Nilotinib. In: Cancer Leukemia Group B Annual Meeting, Chicago, IL, 2011.

17. **Li, CH**. Application of likelihood-based methods to evaluate the impact of below-quantification-limit data arising from non-adherence in a population pharmacokinetics model of Sertraline in Alzheimer's patients. Uppsala University, Uppsala, Sweden, 2011.

18. **Li, CH**. Study on the allelic variants of Cytidine Deaminase (CDA) and Deoxycytidine Kinase (DCK) on the formation of the metabolite of Gemcitabine and prodrug in vitro pharmacogenomics approaches, Eli Lilly and Company, Indianapolis, IN, 2009.

19. **Li, CH**, Zhao, S. Neo AI gene construction with BAC recombineering using galK gene. Biochemistry Department Research Meeting, University of Georgia, Athens, GA, 2008.

20. **Li, CH**, Wang, LC, Wu, SC. Expression human Slit3 protein related to Congenital Diaphragmatic Hernia(CDH) in *Pichia Pastoris*. Complex Carbohydrate Research Center Meeting, University of Georgia, Athens, GA, 2007.

21. **Li, CH**, Wang, LC, Wu, SC. Expression human Slit3 protein related to Congenital Diaphragmatic Hernia(CDH) in *Pichia Pastoris*. A Poster Presentation Meeting at Georgia Center, Athens, GA, 2007.

22. **Li, CH**, Wang, LC, Wu, SC. Expression human Slit3 protein related to Congenital Diaphragmatic Hernia(CDH) in *Pichia Pastoris*. Summer Undergraduate Research Program Presentation at University of Georgia, Athens, GA, 2007.

#### PROFESSIONAL ASSOCIATION MEMBERSHIPS

2012-Present     **American Association of Pharmaceutical Scientist (AAPS)**

2012-Present     **American Society of Pharmacometrics (AsoP)**

#### SCIENTIFIC MEETINGS

2011           **American Conference on Pharmacometrics (ACOP)** San Diego, CA

2011           **Alliance Annual meeting (Cancer Leukemia Group B)** Chicago, IL

2011           **The First Indiana Clinical and Translational Sciences Institute (CTSI) Symposium on Disease and Therapeutic Response Modeling**  
Indianapolis, IN

2012           **Population Approach Group in Europe (PAGE)** Venice, Italy

2012           **American Association of Pharmaceutical Scientists (AAPS)** Chicago, IL

2012           **The Second Indiana Clinical and Translational Sciences Institute (CTSI) Symposium on Disease and Therapeutic Response Modeling**  
Indianapolis, IN

2013           **American Society for Clinical Pharmacology and Therapeutics (ASCPT)** Indianapolis, IN

2013           **American Conference of Pharmacometrics (ACOP)**  
Fort Lauderdale, FL

#### SOFTWARE SKILLS

NONMEM, PsN, Xpose, WFN, Pirana, R, PLINK, Linux, MACH, IMPUTE2, GTOOL, WinNolin, Vensim