Pharmacokinetics, Pharmacodynamics, and Dose Optimization of Fludarabine in Nonmyeloablative Hematopoietic Stem Cell Transplantation

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Kyle Thomas Baron

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

Doctor of Philosophy

October, 2010

Acknowledgements

I would like to take this opportunity to thank my advisor, Richard Brundage, for his mentorship, guidance, and friendship over the past seven years. His investment in me, both professionally and personally, will make a lasting impact on my life. It will be an honor to be known as *his* student wherever my future professional endeavors lead me. I would like to thank the thesis reviewers and committee members: Dr. Pamela Jacobson, Dr. Angela Birnbaum, Dr. Cavan Reilly, and Dr. Richard Brundage. Their careful reading of the manuscript and helpful suggestions greatly improved this work. This work would not have been possible without the previous work of Dr. Janel Long-Boyle and Dr. Pamela Jacobson in Experimental and Clinical Pharmacology as well as many physicians, nurses, pharmacists, patients and others in Hematology, Oncology, & Transplantation at the University of Minnesota. I would like to especially thank Janel Long-Boyle for teaching me how to do good science in a truly collaborative environment

and for believing in what I could do as a pharmacomatrician for the purpose of advancing knowledge. A big thank you also goes out to University of Minnesota Pharmacometrics Group members, both past and present, who provided friendship and a stimulating learning environment. I would like to especially thank Varun Goel for countless hours spent at the whiteboard in our office working through modeling and statistical problems and occasionally clarifying the original question that was posed. I am grateful to my wife, Vera, for so patiently and lovingly supporting me during my Ph.D. work and especially while I was writing the thesis. I thank my family for always believing in me and encouraging me as I studied in yet *another* year of school. A special thank you goes to Matt and Jean Baron who supported me and adopted me into their family when I needed a home base.

Dedication

To my wife, Vera.

Contents

List of Tables

List of Figures

Chapter 1

Thesis Introduction

In this thesis, a quantitative approach is used to learn about the clinical pharmacology of fludarabine in nonmyeloablative hematopoietic stem cell transplantation (HCT). The knowledge that is obtained from a quantitative approach is contained in models that identify sources of variability in fludarabine exposure and establishes relationships between fludarabine exposure and clinical outcomes after HCT. Once these models are built and qualified, they may be leveraged to develop clinically-relevant dose-individualization guidance and to inform future clinical study protocol development. Model-based, quantitative approaches that focus on learning have been the subject of intense research and development and represent a potentially fruitful and bright new direction for clinical pharmacology.

This work is focused on data analysis and the end-products are only models

and data simulated from these models. The information available in a model and the usefulness of the simulated data critically depends on the data set used to estimate model parameters. Clearly, an identifiable model cannot provide insight beyond the information that is already contained in the data. The data set used for the modeling work presented here was collected and published [1, 2] before the current work was initiated. In 2007, Brunstein et. al. [1] published outcomes in 110 patients undergoing nonmyeloablative HCT at the University of Minnesota Medical Center. This study helped to establish the efficacy and safety of transplantation using umbilical cord blood as a stem cell source. In 2010, Long-Boyle et. al. [2] published an exposure-response analysis for 87 of those patients, establishing a link between increased F-ara-A exposure and increased risk of treatment-related mortality. The current work employs advanced, modeling and simulation methodology to maximize information from this data set and thus to build upon and extend these original published results.

Chapter 2 of this thesis provides an overview of HCT and the clinical pharmacology of fludarabine. The goal of this introductory chapter is to define terms and familiarize the reader with the basics of HCT.

Chapter 3 reports the development of a population pharmacokinetic model for F-ara-A[1] in HCT. A key goal in pharmacokinetic modeling is to understand

[1] F-ara-A is the active, circulating form of the pro-drug fludarabine.

relationships between patient-specific covariates and pharmacokinetic model parameters like clearance, so that doses may be individualized to achieve a target exposure.

In chapter 4, pharmacodynamic models characterizing F-ara-A exposure-response relationships for several HCT outcomes are presented. Pharmacodynamic models are estimated using a Bayesian approach that acknowledges uncertainty in model parameters and the models are evaluated using posterior predictive simulations.

Chapter 5 reports the results of simulations from the pharmacokinetic and pharmacodynamic models with the goal of optimizing fludarabine doses in different patient populations. Specifically, data-driven guidance is presented for adjusting doses for patients with different levels of renal function. A utility index is used to simultaneously optimize fludarabine doses across multiple endpoints.

Finally, chapter 6 provides some concluding remarks as well as potential future directions for the work that was started here.

Chapter 2

Hematopoietic Stem Cell Transplantation and Fludarabine

2.1 Hematopoietic Stem Cell Transplantation

2.1.1 Overview

Hematopoietic stem cell transplantation (HCT) is a treatment for various hematologic disorders in which the transplant recipient's own bone marrow is improperly making cells in the blood (hematopoiesis). After the preparative (or conditioning) regimen either totally or partially eradicates the diseased blood cells and

marrow, donor stem cells from a suitable source are infused and establish normal hematopoiesis (engraftment). HCT is a viable treatment option for a wide variety of malignancies including cancer (leukemias and lymphomas), hemoglobin disorders (beta-thalasemia and sickle cell disease), problems with red blood cells (severe aplastic and Fanconi anemia), rarer metabolic disease and other disorders [3]. However, most transplants are performed to treat hematologic cancers [3] and the introductory remarks presented here will be made within this context.

Several factors are important to consider in HCT for all indications, including stem cell donor sources and the type of conditioning regimen that is used prior to transplant. These factors have important influences on short and long-term outcomes (both positive and negative) after HCT.

2.1.2 Stem Cell Donor Sources

Stem cell donor sources for transplant may be classified according to their relationship to the recipient.

In *allogeneic* transplant, the donor and recipient are genetically and immunologically distinct. Donor-recipient pairs may either be unrelated or related to each other. Transplantation of tissue from one individual to another involves the potential for substantial immunological incompatibility or mismatch between the recipient (host) and donor cells (graft). If the host's own immune cells remain

after the preparative regimen, the mismatch will result in rejection where host immune cells identify graft cells as "foreign" and attack and destroy the graft. If the preparative regimen leads to destruction of the host immune system, then mismatched immune cells of graft origin would identify the host as "foreign" and the host will be attacked, a condition known as "graft-versus-host disease" (GVHD), one of the most serious complications of HCT [4].

To avoid these mismatch reactions, the donor and recipient are matched at several human leukocyte antigen (HLA) genes. Five HLA loci are considered in establishing a matched donor-recipient pair: HLA-A, -B, -C for class I loci and HLA-DRB1 and -DQB1 for class II loci [5]. The goal is to match at as many HLA loci as possible. Even though rejection or GVHD are still possible in case of a perfect match, positive outcomes after HCT can still be achieved after partial HLA mismatches from unrelated stem cell donors [6]. Immunosuppressive drug therapy both before and after transplant can also help prevent rejection and GVHD (see below).

In *autologous* HCT, the patient serves as both donor and recipient of the transplanted cells. Here, the patient's own stem cells are harvested and re-introduced after ablative therapy to destroy the diseased bone marrow. Harvested stem cells may be purged of tumor cells before re-introduction in order to decrease the chances of reintroduction of malignant cells in the transplant procedure. Because

the graft is perfectly matched to the recipient, GVHD reactions after transplant are absent, however the effectiveness of the procedure may be reduced due to lack of graft-versus-tumor effect [3] (see below).

In the early days of HCT, stem cells for allogenic or autologous transplantation were frequently harvested by aspiration directly from the donor's bone marrow. For this reason, HCT used to be commonly referred to as "bone marrow transplant" or BMT. This terminology was dropped with the use of stem cell sources other than the bone marrow. Today, peripheral blood stem cells (PBSCs) are commonly used in HCT, especially in autologous transplants [7]. Advantages to using PBSCs in HCT include faster engraftment by 2 to 3 days [8, 9] as well as a harvest process that is more tolerable to the donor.

In cases where an appropriately matched peripheral blood or marrow sources are not and autologous transplantation is not a viable option, hematopoietic stem cells may be harvested from umbilical cord blood (UCB) [1]. Recipient and donor UCB cells are still required to be at least partially HLA matched. The advantage to UCB is that partial HLA mismatch in UCB grafts can result in lower acute GVHD rates than matched or partially mis-matched bone marrow grafts, with similar (but delayed) engraftment rates [10]. Also, UCB is able to be banked and quickly available for transplantation.

2.1.3 HCT Preparative Regimen

The goals of the preparative regimen prior to HCT for hematological malignancies are two-fold [3]. First, the preparative regimen must eliminate the cancer cells and diseased bone marrow in the transplant recipient. Second, in the case of allogeneic HCT, the preparative regimen should provide suppression of the recipient's immune cells so that engraftment may take place. Preparative regimens are highly variable across institutions, but commonly involve total-body irradiation (TBI) as well as high doses of chemotherapeutic agents such as cyclophosphamide and busulfan [11]. Agents with immunosuppressive activity such as anti-thymocyte globulin (ATG) and fludarabine (FLU) may also be included in the regimen [11].

Ablative and Nonmyeloablative Preparative Regimen

The intensity or type of preparative regimen used defines two broad categories of HCT.

An *myeloablative* HCT preparative regimen relies on high-intensity TBI and chemotherapeutic agents to eradicate malignant cells but unfortunately also destroys. Recovery of the patient's native hematopoietic system (autologous recovery) is generally not possible after myeloablative therapy and this preparative regimen would be lethal if no stem cells are reintroduced to re-establish hematopoiesis. Thus, the primary role of transplanted cells in myeloablative HCT is as "rescue

agent" [12]. The benefit of a myeloablative transplant is that there is a high likelihood of complete eradication of malignancy. However, due to the high-intensity of the preparative regimen, treatment-related complications, including death, can be severe. Older patients and patients with significant comorbid conditions do not generally tolerate these toxic consequences well and thus often are not offered myeloablative HCT [13].

A *nonmyeloablative* preparative regimen before HCT involves a reduced-intensity preparative regimen usually including TBI and chemotherapeutic and immunosuppressive drugs. Compared to the ablative preparative regimen, the focus in nonmyeloablative conditioning is on immunosuppression rather than myelosuppression [14]. TBI doses are often greatly reduced compared to a myeloablative conditioning regimen and immunosuppressive agents such as cyclosporine and mycophenolate may be added post-transplant to augment immunosuppression from the pre-transplant regimen. Criteria that define a nonmyeloablative regimen have been proposed [15]. Included in the definition are the requirements that recipient hematopoiesis is not destroyed by the preparative regimen and that autologous recovery should occur in case of graft rejection. The reduced intensity treatment is more tolerable for older patients and those with comorbid conditions, opening up the possibility of HCT to a much broader population [15].

Graft-Versus-Tumor Effect

Instead of relying on irradiation and drug therapy to provide the myelotoxic activity to eradicate the cancer, a nonmyeloablative HCT regimen relies on the graft to exert antitumor activity. A nonmyeloablative preparative regimen is mild enough that cancer cells remain at the time of HCT. Once the donor cells are introduced into the recipient, the graft recognizes residual cancer cells as "foreign" and attack, thus working to eliminate the malignancy. This phenomenon is known as a "graft-versus-leukemia" or "graft-versus-tumor" effect.

Several lines of evidence support the existence of a graft-versus-tumor effect in allogeneic HCT. Sullivan et. al. [16] retrospectively assessed the influence of GVHD on relapse and survival rates in patients undergoing allogeneic HCT for acute and chronic leukemias. Patients with acute leukemia in relapse or chronic leukemia in accelerated or blast crisis generally had better survival and lower relapse rates if they experienced some GVHD post-transplant. For example, five-year relapse probabilities in the patients who survived at least 150 days were 65 to 81% if there was no acute or chronic GVHD, 20 to 43% if there was acute and chronic GVHD, and 31 to 40% if there was acute but not chronic GVHD. Also, survival among these patients was much better in case of either acute or chronic GVHD (42 to 65%) compared to patients with neither acute nor chronic GVHD

(22 to 31%). Howowitz et. al. [17] inferred a graft-versus-tumor effect by studying patients receiving HCT using T-cell-depleted or non-depleted bone marrow grafts. In this retrospective study, recipients of stem cell grafts from identical twin or T-cell depleted donor sources had 3-year relapse rates of 41 to 46%. Relapse rates in patients receiving non-T-cell depleted grafts were much lower: 25% if the patient did not experience GVHD and only 7% if the patient experienced both acute and chronic GVHD. Presumably, the lower relapse rate in patients receiving non-T-cell depleted allogeneic transplants and those experiencing GVHD was due to more effective anti-tumor activity of T-cells in the graft.

Further evidence that the graft itself is at least partially responsible for eradicating cancer in allogeneic HCT comes from the ability of infused donor lymphocytes (DLI) to induce remission in certain patients who relapsed after HCT. Complete remission was re-established in 60 to 70% of patients with relapsed chronic myelogenous leukemia after DLI [18, 19]. Also, the ability to achieve complete remission after DLI also appears to be associated with development of GVHD. One study found that 89 to 93% of patients with complete response to DLI had acute or chronic GVHD, while complete remission was only established in 13% of patients who developed neither acute nor chronic GVHD [18].

2.1.4 HCT Complications

Complications after HCT can involve a variety of organ systems, may occur either early or late after transplant, and the incidence and severity of the complications depends on the type of HCT performed. Early complications include mucositis in the mouth and in the intestine, and hepatic injury known as sinusoidal obstruction syndrome [3]. Both are less common after nonmyeloablative HCT [20]. Lung injury may also occur in the acute phase after HCT. Until engraftment of transplanted cells, HCT patients are also at high risk of bacterial, viral, and fungal infection. Granulocyte colony stimulating factor may be given to speed white cell recovery antimicrobial agents are commonly given to treat or prevent infection. Complications that may occur in the later phases after HCT include endocrine problems, problems with the digestive system, secondary cancers, and complications of over-immunosuppression and infection [3].

Graft-Versus-Host Disease

One of the most common and serious complication after HCT is GVHD [4]. As noted above, GVHD occurs only in allogeneic transplant, when the donor immune cells attack recipient tissues after identifying them as "foreign" antigens. GVHD may occur either in the acute or chronic phase after transplant. Acute GVHD (aGVHD) typically occurs in the first 100 days after transplant and involves skin

rash, cholestatic liver dysfunction and often-severe gastrointestinal tract dysfunction including abdominal pain, diarrhea, anorexia, nausea, and vomiting [21]. Symptoms unique to chronic GVHD (cGVHD), generally occurring after the first 100 days, include xerostomia, ulceration and sclerosis in the mouth, sclerosis and eruptions on the skin, dry eyes and conjunctivitis, jaundice, restricitve or obstructive pulmonary disease, muscle pain and arthralgia [4, 21]. Gastrointestinal and dermatologic symptoms of aGVHD may also occur in an "overlap syndrome" with cGVHD [22, 21]. The most important predictor of cGVHD is the occurrence of previous aGVHD [4] and anti-aGVHD measures may also help prevent cGVHD [21]. Systemic and topical corticosteroids are the primary choices for treating GVHD. Prevention of GVHD is important and agents such as ATG, calcineurin inhibitors, mycophenolates, and steroids are used prophylactically [4, 21].

2.2 Fludarabine

2.2.1 Clinical Pharmacology

Fludarabine (9-β-D-arabinofuranosyl-2-fluoroadenine 5'-monophosphate, F-ara-AMP) is a purine analog anti-metabolite drug with antitumor and immunosuppressive activity [23, 24]. Fludarabine is a pro-drug of F-ara-A (9-β-D-arabinofuranosyl-2-fluoroadenine); the phosphate group attached at the 5'-hydroxyl

position of F-ara-A enhances solubility of the drug and allows parenteral administration [25, 23]. Upon intravenous infusion, the 5'-phosphate group of fludarabine is rapidly and completely cleaved by esterases in the blood yielding the circulating form of fludarabine, F-ara-A. [23, 26, 27, 28, 29]. The fluorine at the 2-position of the adenine ring renders F-ara-A resistant to deactivation through deamination [23].

Because fludarabine is active at intracellular targets in the 5'-triphosphate form (F-ara-ATP), the transport and metabolism of F-ara-A are likely to be important determinants of its activity. Due to overall negative charge at physiological pH, F-ara-A does not readily cross cell membranes on its own, but can be transported into cells through both active (hCNT) and passive (hENT) nucleotide transporters [30]. Once inside the cell, F-ara-ATP is generated by phosphorylation reactions catalyzed by intracellular kinases. The rate-lmiting step in bioactivation of F-ara-A is its phosphorylation back to the F-ara-AMP by deoxycytidine kinase [25]. F-ara-AMP is phosphorylated by adenylic kinase to create the 5'-diphosphate, which is subsequently phosphorylated by nucleoside diphosphate kinase to make the active 5'-triphosphate [25]. Gandi et. al. showed that, when 30 mg/m^2 fludarabine is given as a 30 min intravenous infusion to patients with chronic lymphocytic leukemia, median peak intracellular F-ara-ATP occurs at about four hours post-dose and high intracellular concentrations can be maintained up to 24

hours [31].

The major intracellular action of F-ara-ATP is to inhibit DNA synthesis [23, 24, 25]. Because of its structural similarity to deoxyadenosine triphosphate (dATP), F-ara-ATP competes with dATP in DNA chain-elongation reactions catalyzed by DNA-polymerase and become inserted into the nascent DNA chain. The hydroxyl group at the 2' position of F-ara-A makes it a poor substrate for subsequent 3'-5' phosphodiester bond formation, and DNA chain elongation is terminated. F-ara-ATP also inhibits ribounucleotide reductase (RNR) [23, 25], the enzyme responsible for the intracellular production of 2'-deoxyribonucleotides, including dATP, thus potentially creating a competitive advantage for F-ara-A at the DNA-polymerase active site. Once DNA synthesis is inhibited by F-ara-ATP, affected cells undergo apoptotic cell death [23].

2.2.2 Clinical Use of Fludarabine in CLL and HCT

Fludarabine is FDA-approved for use in adults with relapsed or refractory chronic lymphocytic leukemia (CLL). Approved intravenous dosing for CLL treatment is 25 mg/m^2 as a 30 minute infusion daily for five days, repeated in a 28 day cycle [32]. Oral doses for CLL are 40 mg/m^2 daily for 5 days. The manufacturer recommends a 20% dose decrease in case of moderate renal impairment (CLCR

[1] of 30 to 70 ml/min/1.73 m^2) and intravenous doses are contraindicated when CLCR is less than 30 ml/min/1.73 m^2.

In 1998, Slavin et. al. provided some of the earliest results for allogeneic HCT after a nonmyeloablative conditioning regimen containing fludarabine [12]. ATG, busulfan, and fludarabine at 30 mg/m^2 daliy for six days prior to transplant along with cyclosporine for post-transplant immunosuppression was given to 28 patients. They reported high engraftment rates and low incidence and severity of treatment-related complications. Now, after many years of intense study of the nonmyeloablative HCT approach, fludarabine is included in most nonmyeloablative preparative regimens and is thought to provide an important role in promoting engraftment and avoiding early graft rejection [33, 11]. Depending on institutional protocols, intravenous fludarabine is typically given at doses of 25-40 mg/m^2 daily for five days or 30 mg/m^2 daily for three to six days and is usually combined with cyclophosphamide or busulfan and TBI. Given the relatively recent introduction of nonmyeloablative allogeneic HCT, optimal use of fludarabine in the preparative regimen is of great interest. Recent results from Long-Boyle et. al. [2] indicate that excessive fludarabine exposure can lead to increased treatment-related mortality, suggesting optimization or individualization of fludarabine doses is necessary.

[1] creatinine clearance

Chapter 3

Population pharmacokinetics of fludarabine in HCT

3.1 Introduction

F-ara-A pharmacokinetics (PK) have been previously described after intravenous, oral, and subcutaneous fludarabine (FLU) doses in a variety of patient populations, including several studies in patients undergoing non-myeloablative HCT.

After intravenous FLU administration, FLU is rapidly and completely converted to F-ara-A in the plasma [23, 26, 27, 28, 29]. F-ara-A concentrations decline in a multiphasic manner with a terminal half-life of 9-12 hours [29, 34, 35] in patients with normal renal function with little F-ara-A accumulating over a

5-day treatment period [27]. F-ara-A exposures change approximately linearly with changes in FLU doses across a wide range of intravenous [34, 36, 35] and oral [36] doses. The fraction of an orally-administered dose of FLU appearing in the systemic circulation as F-ara-A is about 0.5 - 0.6 [36, 26]. Kuo et. al. [37] report that 105% of a subcutaneously administered dose of FLU appears in the plasma as F-ara-A.

A significant fraction of the administered FLU dose is recovered in the urine as F-ara-A (i.e. with no further metabolism beyond the rapid, initial dephosphorylation from F-ara-AMP to F-ara-A) and the fraction of this primary metabolite increases with increasing renal function. Lichtman et. al. [38] used non-compartmental analysis methods to measure renal clearance of F-ara-A in patients with normal (creatinine clearance[1] range: 71 - 152 ml/min/1.73 m^2), moderately impaired (47.4 - 66.6 ml/min/1.73 m^2), and impaired (18.5 - 25.1 ml/min/1.73 m^2) renal function. As a fraction of total clearance, renal clearance of F-ara-A was 64%, 56%, and 35% in the normal, moderately impaired, and impaired renal function group, respectively. In another study [27] of seven patients with mean CLCR of 63 ml/min (range: 37 - 77 ml/min), the fraction of the FLU dose excreted unchanged as F-ara-A was 0.25 when averaged over the 5-day study period. In four studies that didn't not report CLCR estimates for study subjects,

[1] Creatinine clearance was calculated from 24-hour urine collection data and serum creatinine measurement prior to start of the study.

the apparent fraction excreted unchanged after intravenous or subcutaneous FLU administration varied from 0.37 to 0.41 [35, 37, 29, 26]. Across all of the published data, the apparent fraction excreted unchanged varied from 0.25 to 0.6, indicating significant clearance through renal elimination.

F-ara-A pharmacokinetic parameters evaluated through both modeling-based and non-compartmental analysis methods have been reported. Most compartment-based modeling of F-ara-A PK assumed a 2-compartmental structure, although one report describes the fitting of a 3-compartment model. Hersch et. al. [27] describe a phase-I pharmacokinetic study of FLU 18-25 mg/m^2 given as a 30 minute infusion in seven adult patients with solid tumors. The study population had a mean mean body surface area (BSA) of 1.8 m^2 and mean CLCR of 63 ml/min. A 2-compartment model was fit to the F-ara-A concentration time data, with mean clearance (CL) of 9.1 L/hr/m^2 and steady-state volume of distribution (Vss) of 96.2 L/m^2. Across the 5-day treatment regimen, approximately 25% of the administered daily dose was recovered unchanged in the urine. Knebel et. al. [34] fit a 2-compartment model to F-ara-A concentration-time data in 26 rheumatoid arthritis patients receiving FLU 20-30 mg/m^2 as an intravenous infusion daily for three days. PK modeling results showed a mean CL of 13.7 L/hr and Vss of 170 L. Mean BSA in the study population was 1.8 m^2. CLCR was not reported in this study.

Two studies in the literature [35, 2] have addressed F-ara-A PK as a part of a reduced-intensity preparative regimen prior to HCT through non-compartmental analysis methods. Both studies report F-ara-A PK parameters after the first FLU dose. Bonin et. al. [35] report F-ara-A PK in 16 adult subjects receiving FLU 30 mg/m^2 daily for four days. In this study, the mean total clearance and terminal half-life were 5.35 $L/hr/m^2$ (10.2 L/hr at BSA of 2 m^2) and 8.9 hours, respectively. The reported mean Vss was 53 L, a value that is difficult to reconcile with other published values [2, 34, 27]. Also, this study reported F-ara-A renal clearance approaching 100% of total clearance (5.03 $L/hr/m^2$), but the mean fraction of the dose in the urine was only 0.41.

Long-Boyle et. al. [2] report on F-ara-A PK in 87 adults receiving FLU 40 mg/m^2 daily in a 5 day regimen. In the Long-Boyle study, mean total clearance and terminal half-life (in the standard dose subgroup, N=78) were 16 L/hr and 8.53 hr, respectively. The steady-state volume of distribution was 1.9 L/kg (or 161 L at the median weight). Total F-ara-A clearance was reduced (11.5 L/hr) in nine patients with reduced doses due to renal impairment [2].

Salinger et. al. [39] reported the population pharmacokinetics of F-ara-A in 42 adult patients undergoing HCT. Fludarabine 30 mg/m^2 daily for 4 days or 50 mg/m^2 daily for 5 days was given as a 30 minute intravenous infusion. Median BSA was 2 m^2 and median CLCR was 89 ml/min. In this study, a 2-compartment

structural model was assumed and all pharmacokinetic parameters were assumed to scale directly with BSA through expression of doses on a mg/m^2 basis. The typical value of CL was 5.7 L/hr/m^2 and the typical Vss was 72 L/m^2. After BSA was in the model (through dosing records), the only other covariate to explain variability in clearance was a modest effect due to height. The estimated height effect was 0.0334 L/hr/m^2 change in TVCL with a 1-cm change in height on a TVCL of 5.7 L/hr/m^2 at the reference height of 170.5 cm [39]. Interestingly, CLCR (range: 55-148 ml/min) was not selected as a covariate in the final model for mean F-ara-A clearance.

Despite the previously-reported work describing F-ara-A pharmacokinetics, a comprehensive, model-based characterization of F-ara-A population PK is lacking, particularly in the HCT population. Since over half the total F-ara-A clearance is due to excretion in a population with normal renal function, an understanding of how F-ara-A clearance changes with changes in the patient's renal function would be extremely useful when making clinical dosing decisions. It is unclear why previous modeling work [39] did not detect a relationship between F-ara-A clearance and CLCR. The limited sample size used to estimate the previous population model may have contributed to the difficulty. Also, the modeling methods (assumption of direct relationship between all PK parameters and body surface area) may have obscured the signal in the data. Clearly, more study

is needed to better understand sources of variability in F-ara-A PK in an HCT population.

In the work presented here, a model-based approach is used to understand F-ara-A PK in an HCT population. The dataset published by Long-Boyle et. al. [2] is analyzed under a population PK framework using NONMEM. The goals of the work are to estimate both individual and population F-ara-A PK, identify sources of variability in F-ara-A PK, including fixed covariate effects as well as random variability, and to validate the model for use in simulation.

3.2 Methods

3.2.1 Patients

The FLU pharmacokinetic dataset was collected as part of a larger trial evaluating outcomes after stem cell transplantation at the University of Minnesota Medical Center [1, 2]. Data were collected under protocols UMN-2000LS039 and UMN-2005LS036 (http://www.cancer.gov). Fludarabine plasma concentrations were collected from 87 patients undergoing allogeneic hematopoietic stem cell transplantation. Fludarabine was part of a nonmyeloablative conditioning regimen that also included cyclophosphamide and total-body irradiation (TBI). Post-transplant immunosuppression was provided with cyclosporine A (CSA) and mycophenolate

mofetil (MMF). Detailed inclusion criteria can be found in the study protocol and previously published analyses [1, 2]. All patients provided informed written consent to participate in the study. Table 3.1 shows a summary of the study population characteristics.

3.2.2 HCT Treatment Regimen

The pre-transplant preparative regimen [1, 2] included the following (with transplant occurring on day 0): FLU 40 mg/m^2 IV administered as a one-hour infusion daily on days -6 through -2, cyclophosphamide 50 mg/kg IV on day -6, and TBI 200 cGY on day -1. A subset of patients received anti-thymocyte globulin (ATG) 15 mg/kg IV every 12 hours given over four to six hours on days -3 through -1. Patients receiving ATG also received methylprednisolone 1 mg/kg IV every 12 hours prior to ATG infusion.

After the transplantation, patients were treated with granulocyte colony stimulating factor (G-CSF, 5 μg/kg daily), immunosuppressive agents (MMF 1 gram i.v. or p.o. twice daily and CSA twice daily to target trough 200-400 ng/mL), and antimicrobial agents to treat or prevent viral, bacterial, or fungal infection. Prophylaxis for fungal infection (fluconazole or voriconazole for 100 days after transplant) and pneumocystis carinii (trimethoprim-sulfamethoxazole for 12 months after transplant) was given for all patients. Patients at risk for herpes simplex

virus recurrence or those who were CMV seropositive were given prophylactic acyclovir. Documented CMV infection was treated with ganciclovir plus intravenous immunoglobulin.

3.2.3 Pharmacokinetic Data

F-ara-A pharmacokinetics were studied in 87 patients participating in the UMN-2005LS36 study protocol. Intensive PK samples (N=8) were collected on the first dose and sparse PK samples (N=4) were collected on the fifth (final) dose of the treatment course, with trough concentrations drawn before doses 2, 3, 4, and 5 (N=4). Blood samples (5 mL) were collected at the following times: 1, 2.6, 3, 4, 5, 7, 8, and 12 hours after starting the infusion on the day of the first dose; 4, 8, 24, and 48 hours after starting the infusion on the fifth dose; and pre-dose levels (24, 48, 72, and 96 hours after starting the infusion on the day of the first dose). In total, there were 1389 F-ara-A observations included in the analysis. Blood samples were centrifuged and plasma stored at -80^{o}C until bioanalysis . F-ara-A concentrations were determined by UV-HPLC analysis and the linear range of the analytical method was 10-3000 mcg/L [2].

3.2.4 Software and Estimation

Population pharmacokinetic models were built using Nonlinear Mixed Effects Modeling software (NONMEM; version VI, level 2.0). NONMEM was complied and models estimated using Compaq Visual Fortran Optimizing Compiler (version 6.6, update C). All models were run using NONMEM first order conditional estimation (FOCE) method with interaction option. Model parameters were estimated using a log-transform both sides method [40]. Scripts written for the R statistical software package were used to assemble datasets as well as post-process run results.

3.2.5 Model Building

Standard population pharmacokinetic model-building techniques were employed in the current analysis. Modeling began with base-model exploration to determine the compartmental structure, interindividual variability in PK parameters as well as the correlation between PK parameters within a population, and residual and inter-occasion variability. Compartmental model selection was guided by exploratory data analysis, NONMEM objective function changes, as well as model diagnostic plots. Pharmacokinetic parameters were assumed to be log-normally distributed in the population and modeled as:

$$P_i = TVP \cdot e^{\eta_{P,i}} \tag{3.1}$$

with

$$\eta \sim N(0, \Omega) \tag{3.2}$$

where P_i is the value of parameter P in the ith subject, TVP is the typical value of the parameter in the population, $\eta_{P,i}$ is the subject-specific random effect for parameter P, and Ω is the variance-covariance matrix describing variances (ω_P^2) and covariances (ω_{P_x,P_y}) of the different η in the model. Inter-occasion variability (IOV) [41] was incorporated into the model as:

$$P = TVP \cdot e^{\eta_P + \kappa_{OCC_k}} \tag{3.3}$$

where κ_{OCC_k} is a random effect for the nth occasion with

$$\kappa \sim N(0, \pi^2) \tag{3.4}$$

Two pharmacokinetic "occasions" were derived from the six-day observation period: days 1-2 and days 3-6. Residual error was modeled to have proportional and additive components:

$$Y = \hat{Y} \cdot (1 + \epsilon_1) + \epsilon_2 \tag{3.5}$$

where

$$\epsilon_x \sim N(0, \sigma_x^2) \tag{3.6}$$

Shrinkage was calculated for etas and epsilons [42] . Eta-shrinkage on a parameter (P) was calculated as:

$$sh_P = 1 - \frac{SD(\eta_P)}{\omega_P} \tag{3.7}$$

where η_P is the vector of all η_i in the study population and ω_P is the modeled standard deviation of the random effect distribution for parameter P. Epsilon shrinkage was calculated as:

$$sh_{eps} = 1 - SD(IWRES) \tag{3.8}$$

where IWRES is the vector of the 1389 individual-weighted residuals.

3.2.6 Covariate Modeling and Selection

Covariate models were explored to help describe the relationship between values of known, patient-specific attributes and pharmacokinetic parameters. The following covariates were considered as candidates for incorporation into covariate models for F-ara-A pharmacokinetic parameters: various body size measures, including

actual, lean [43], and ideal body weight (kg), body mass index (kg/m^2), and body surface area (m^2); female sex (no=0, yes=1); age (years); creatinine clearance (ml/min), calculated from the equation of Cockcroft and Gault [44] using actual body weight (kg); diagnosis for which the HCT was indicated (unordered categorical values with six disease categories; see table 3.1); total bilirubin (mg/dL); use of ATG in the preparative regimen (no=0, yes=1); and comorbidity score (ordered categorical data, values of 0 to 7 were observed) [45]. Creatinine clearance was a time-varying covariate, with weight fixed at the day -6 value and serum creatinine measurements recorded daily for each of the five FLU doses. Missing creatinine clearance estimates (due to missing serum creatinine data, 6 measurements in 4 patients) were imputed as the mean of the remaining creatinine clearance estimates for that patient.

Covariate model building was guided by principles of allometric scaling, previous knowledge of F-ara-A pharmacokinetics, diagnostic plots, objective function value (OFV) changes, and clinical interest in the covariate. In general, an allometric relationship between PK parameters and weight was assumed where allometric exponents for clearances were fixed to 0.75 and exponents for volumes of distribution were fixed to 1. Prior to starting the analysis, it was known that a large fraction of the total clearance was through the kidney as unchanged drug [32, 27], so CLCR was a natural focus of the covariate modeling for F-ara-A elimination

clearance.

After an initial base run with actual body weight included as a covariate on clearances and volumes of distribution, exploratory plots based on individual Bayesian post-hoc parameters (P_i) and typical values of the parameters (TVP) were examined to identify additional covariates that could explain variability in PK parameters. So-called "delta-plots" were created where residual unexplained variability in PK parameters (delta = TVP-P_i) was plotted versus individual covariate values. Scatter plots were created for continuous covariates and box and whisker plots were created for categorical covariates. Covariates that showed systematic variation with delta for a given PK parameter were selected for testing in the model for that parameter. Covariates selected from this step were incorporated into the model according to the form suggested in the delta plot (e.g. step change in the parameter value or linear or power relationship). This delta-plot methodology was repeated until $delta_P$ for each PK parameter had no detectable systematic relationship to any remaining covariate in the data set.

Once incorporated in the model, the suitability of that covariate was evaluated based on objective function change relative to a reduced model without the covariate effect, the estimated effect magnitude, and the precision of the estimated effect [46, 47]. Differences in OFV across two nested models are approximately chi-squared distributed with degrees of freedom equal to the difference in the number

of estimable parameters in the two models. Larger values of χ^2 increased the likelihood of retaining the covariate in the final model and no covariate was retained when $\chi^2_n \leq$ the 95th percentile of a χ^2 distribution with n degrees of freedom. In the interest of parsimony, small-magnitude covariate effects that explained only a small fraction of the total variability in the parameter were left out of the final model even in the face of large OFV changes. Similarly, covariate effects of large magnitude or inherent scientific interest were considered for the final model even if limitations in information content precluded efficient estimation of the effect (i.e. large error in estimation of the parameter or small χ^2 value).

Several lines of covariate model building for F-ara-A clearance were initially explored including linear slope-intercept type models [48] as well as a full model parameterization approach [46, 47]. The final model structure chosen in the end was of a previously-published form that partitions total clearance into renal and non-renal clearance components and assumes an allometric relationship with actual body weight [49, 50]. The final structural form for typical value of CL in the population was:

$$TVCL = \theta_{COMOR} \cdot \left[\theta_{CLnr} + \theta_{CLr} \cdot \frac{RF}{RF_{std}}\right] \cdot \left[\frac{WGT}{WGT_{std}}\right]^{\theta^{WGT}} \tag{3.9}$$

where RF is CLCR normalized to a 70 kg individual:

$$RF = \frac{CLCR \cdot 70kg}{WGT} \tag{3.10}$$

WGT_{std} is the standardized weight of 70 kg and RF_{std} is the standardized renal function of 100 ml/min in a patient at the standard weight. CLCR is the Cockcroft-Gault-estimated creatinine clearance using actual body weight [44], θ_{CLnr} is the clearance that changes independently of RF, θ_{CLr} is the clearance that changes with RF and θ_{COMOR} describes the average fractional change in CL when the patient's comorbidity score is ≥ 2. Models were investigated where θ^{WGT} was estimated as well as fixed to its theoretical value of 0.75 [51, 52, 53]. When interpreting model parameters[2] , θ_{CLr} is considered to be the renal clearance and θ_{CLnr} is considered to be the non-renal clearance, so total clearance is taken to be the sum of θ_{CLr} and θ_{CLnr}.

Intercompartmental clearances were assumed to scale with actual body weight according to the allometric relationship:

$$TVQ_x = \theta_{Q_x} \cdot \left(\frac{WGT}{WGT_{std}}\right) \tag{3.11}$$

where TVQ_x is the typical value of a generic intercompartmental clearance in a

[2] The interpretation as presented here is for a 70 kg patient with CLCR of 100 ml/min.

patient with a given WGT and θ_{Q_x} is an intercompartmental clearance in a patient at the standard weight (WGT_{std}) of 70 kg.

Base covariate models for volumes of distribution initially assumed that Vd varied directly with actual body weight [53]:

$$TVVx = \theta_{V_x} \cdot \left(\frac{WGT}{WGT_{std}} \right) \tag{3.12}$$

where TVV_x is the typical value of a generic volume of distribution given the patient's WGT and θ_{V_x} is the volume of distribution for a compartment in a patient at the standard weight of 70 kg. Other body size metrics (BSA, lean body weight, ideal body weight) were also explored as covariates for Vd in a form analogous to equation 3.12 (see equation 3.14 below).

3.2.7 Model Qualification

The final pharmacokinetic model was qualified through two predictive checks [54, 55]. The principle underlying predictive checks is that, when a model has an acceptable fit to the observed data, data simulated from the model should be similar to the observed data that were used to inform model parameter estimates [56]. The degree of similarity (or discrepancy) is quantified either numerically

through a test statistic or through visual comparison of observed and simulated data. Significant discrepancies between the predictive distribution of the data and the observed data can indicate model mis-fit and may be grounds for rejecting a model.

When visually evaluating discrepancy (VPC [55]), observed data is plotted (as a function of time) on top of various quantiles of the distribution of simulated data at each observation time and discrepancy is assessed through graphical comparison. In the VPC, new F-ara-A concentration time data sets were simulated (N=1000) with the same design and covariate values as that in the observed data set. At each nominal observation time after the dose, the $5th$, $50th$, and $95th$ percentile of the simulated F-ara-Concentration was calculated and plotted versus time after dose with the observed data overlaid on top of this prediction interval. In order to retain the model, the median and 90% prediction intervals across time should be consistent with the median and prediction interval of the observed data, and approximately 10% of the observed data should lie outside the 90% prediction interval.

When numerically evaluating discrepancies [54], a test statistic is defined and calculated in both the observed data ($T(y)$) and each of the simulated data sets ($T(y^{rep})$, for $y^{rep} = 1, ..., n^{rep}$). Then $T(y)$ is compared with the distribution of $T(y^{rep})$. A p-value (p^{PPC}) relating the degree of discrepancy is:

$$p^{PPC} = Pr(T(y^{rep}) > T(y)) \tag{3.13}$$

The p^{PPC} value is easily calculated as the number of times $T(y^{rep})$ is greater than $T(y)$ divided by the number of datasets simulated in the predictive check (n^{rep}). It is expected that there may be discrepancies between $T(y)$ and $T(y^{rep})$ by chance, even if the model is correct. Extreme values of p^{PPC} signify that such a discrepancy is unlikely to happen by chance and indicate a poor model fit. Like the VPC, F-ara-A concentrations are repeatedly simulated in the numerical check for each observed concentration (no extra smoothing observation are needed). The test statistics $T(y)$ and $T(y^{rep})$ used in this analysis are the quantiles (10 to 90 %) of F-ara-A concentrations calculated for observed and simulated data, respectively.

3.3 Results

F-ara-A concentration-time profiles are shown in figure 3.1. Exploratory data analyses clearly indicated multi-compartmental structure of the data. Therefore, 2- and 3-compartment models were considered for fitting the F-araA PK data

using NONMEM ADVAN3 and ADVAN11, respectively. All models were parameterized in terms of clearances and volumes (TRANS4). The objective function for the 2-compartment model was -2481 while the objective function for the 3-compartment model was -2616. The change in objective function between 2- and 3-compartment models was -134 on 2 degrees of freedom, showing a strong signal in the data in favor of the 3-compartment model ($p < 0.0001$). Figure 3.2 shows plots of weighted residuals (WRES) versus time for the 2- and 3-compartment model fits. The 3-compartment model fit shows less misspecification compared with the 2-compartment fit especially at very early times (0-3 hours after the first dose) and at very late times (48 hours after dose 5). Therefore, the 3-compartment model was retained moving forward in the model development process.

Parameter estimates along with relative asymptotic standard estimation errors from the final model are summarized in table 3.2. The NMTRAN control stream to generate this final model is included in the code appendix.

Total clearance was 12.9 L/hr in a 70 kg patient with CLCR=100 ml/min and a 0/1 comorbidity score. Renal clearance was 7.51 (RSE=12%) L/hr and the fraction excreted unchanged (fe) was 0.58. Removing CLCR as a covariate on clearance increased the OFV by 114 units (df=1). The estimate for θ^{WGT} was 0.64 with 95% CI [0.5 , 0.79], which included the theoretical value of 0.75. Therefore, this factor was fixed to 0.75 in the final model. Figure 3.3 illustrates the modeled

relationships between weight, CLCR, and TVCL.

Exploratory plots indicated that patients with the very lowest comorbidity scores had slightly higher clearance than the rest of the population, but it wasn't totally clear from the plots how to divide the covariate into low and high values. Comorbidity scores ranged from 0-7 and were tested in the PK model according to several different groupings. On average, patients with non-zero comorbidity score (N=78, 90%) had 14% (RSE=3.3%) lower clearance than patients with zero comorbidity score. Patients with comorbidity score greater than one (N=64, 74%) had a 9.1% (3.5%) reduction in clearance and patients with comorbidity score greater than two (N=46, 53%) had a 7.6% (3.3%) decrease. The magnitude of the effect was modest for the three ways of expressing comorbidity score as a categorical covariate and unlikely to significantly alter clinical dosing decisions when pursuing a therapeutic target exposure. Allowing a 9.1% decrease in TVCL when comorbidity score is greater than one seemed to appropriately represent the effect magnitude with at least 25% of the population in the smallest group and was retained in the final model.

Volumes of distribution were initially assumed to vary directly with actual body weight according to principles of allometry (i.e. allometric exponent fixed to 1) [52]. After running this model, plots of remaining unexplained variability in V (as DELTA=TVV-Vi) by patient sex showed persistent differences in volume

of distribution between male and female patients after weight was accounted for in the model. When this sex-difference was modeled as a proportional change in each of the three volumes of distribution, the objective function dropped by 16.7 units (df=3, p=0.00081) and estimates ranged from 30% [55% , 87%] decrease in second peripheral volume to a 13% [0% , 27%] increase in first peripheral volume in female patients compared to male subjects. After accounting for these sex differences in volumes of distribution, there was no detectable relationship between unexplained interindividual variability in Vd and any other covariate in the data set. In an attempt to create a more parsimonious model, other body size metrics were used to model volumes of distribution. Delta plots from model fits where Vd was expressed as a function of lean body weight [43] showed no persistent sex-related differences in unexplained variability of volumes of distribution. This covariate model using lean body weight:

$$Vd = \theta_{Vx} \cdot \left(\frac{LBW}{LBW_{std}} \right) \tag{3.14}$$

was retained in the final population PK model.

Variance component estimates describing inter-individual variability (IIV, ω_P^2) and residual unexplained variability (RUV, σ_x^2) in PK parameters are shown in table 3.2. IIV was estimable on CL and all volumes of distribution and correlations in random effects were estimable between CL/V1, Cl/V3, and V1/V3. In the final

model (with all retained covariate effects), IIV in CL was low (CV=21%). IIV for volumes of distribution ranged from CV=19 to 46%. The proportional RUV component estimate was 14.3% and additive error variance was 8.5 mcg/L.

Shrinkage estimates are shown in table 3.3. Shrinkage on clearance was very low (1.3%). Shrinkage values for V1 and V3 were also low (< 7%), but shrinkage on V2 was considerable (48%). Epsilon shrinkage was 7%.

IOV for clearance was 4.5% (RSE=100%) and incorporation of this parameter in the model reduced the objective function by 2.77 (df=1, p=0.096). IOV for the volume of the central compartment was driven to zero by NONMEM ($\omega^2_{IOV-V1} = 1.42^{-9}$) and incorporating this parameter did not change the objective function. Hence, variance components describing IOV on any PK parameter were not included in the final model.

Diagnostic plots for the final model are shown in figures 3.4 to 3.8. Figure 3.4 shows population- and individual-predictions of F-ara-A concentrations under the model versus observed F-ara-A concentration. Predictions are generally unbiased, with an exception at very high concentrations, when the model tends to under-predict F-ara-A concentration. Figure 3.5 shows observed and modeled population-predictions plotted versus time. Mis-specification in the 4- and 8-hour observations on day 5 are also noted here as well as a tendency to over-predict trough concentrations on days two through five. Figure 3.6 shows a plot

of conditional weighted residuals [57, 58] (CWRES) versus population-predicted F-ara-A concentrations. Figure 3.7 shows a plot of CWRES versus time for the final model. This plot shows slight model-misspecification in the 4- and 8-hour observations. Figure 3.8 shows population- and individual-predicted and observed F-ara-A concentrations versus time for six patients in the dataset. Overall, the misspecification indicated in the diagnostic plots is small and is not sufficient to reject the model.

A visual predictive model check is shown in figure 3.9. The final population PK model was used to simulate F-ara-A concentrations. Across all nominal time points, 13% of simulated F-ara-A concentrations were outside of the 90% prediction interval. Figure 3.10 shows histograms summarizing a predictive check comparing quantiles (5 and 95%) of observed F-ara-A concentrations with quantiles of concentrations from 1000 datasets simulated under the final model. p^{ppc} values ranged from 0.067 to 0.938. Neither predictive check reveals mis-specification that would warrant rejection of the model.

3.4 Discussion

The population pharmacokinetic model presented here describes the concentration-time relationship for F-ara-A after intravenous FLU in 87 patients undergoing non-myeloablative HCT. F-ara-A pharmacokinetics have been published previously [35, 28, 27, 34], including one published population analysis of F-ara-A PK in patients undergoing HCT [39]. The modeling work presented here adds to the knowledge base in significant ways, including a robust covariate model for clearance and the expanded sample size in terms of both numbers of patients (N=87) and amount of data per patient (1389 F-ara-A concentrations).

F-ara-A clearance was modeled as a function of actual body weight, creatinine clearance, and comorbidity score (equation 3.9). The model estimated that patients with a comorbidity score greater than one had on approximately 10% lower total clearance compared with those with 0/1 comorbidity score and this effect was well estimated (RSE=3.5%). However, the modest effect size is unlikely to affect clinical dosing decisions to achieve a therapeutic target. This issue will be addressed in chapter 5.

The renal clearance of F-ara-A was taken as that portion of the total clearance that varied with the estimated CLCR normalized to a patient weight of 70 kg. The estimated fraction excreted unchanged (at typical body size and renal function)

was 0.58, consistent with previous studies reporting fe of 25-60%. Under the model, total F-ara-A clearance in a 70 kg patient with comorbidity score ≥ 2 decreased by 0.68 L/hr for every 10 ml/min decrease in CLCR (equation 3.9) across the observed CLCR range of 46 - 218 ml/min. Current guidelines for adjusting FLU doses in patients with renal impairment suggest a 25% decrease in dose when CLCR is less than 70 ml/min. The modeled relationship presented in this work may be used as the basis for a more refined method for adjusting FLU doses when the patient has compromised renal function. Furthermore, it is possible that FLU doses may need to be increased in the case of a patient with exceptionally efficient renal function (see chapter 5), depending on the therapeutic exposure target. A previous F-ara-A population PK model [39] failed to identify F-ara-A as a significant covariate for CL after PK parameters in the model were normalized to body surface area. It is unclear why CLCR was such a prominent covariate in the current model (ΔOFV=114 when taken out of the final model) but not in the previous work [39] when study populations were so similar.

Estimating the independent effect glomerular filtration capacity on F-ara-A clearance was an important challenge in the population PK modeling. While, within a given population, individuals with increased body size are also likely to have increased glomerular filtration capacity (and vice versa), the toxicity of cancer treatments may alter renal function in an individual independently of changes

in that individual's body size. Therefore, it is critical that the model properly quantifies the unique contribution of each covariate to the value of TVCL. The difficulty in modeling these effects lies in the use of the Cockcroft-Gault (CG) equation [44] as an estimate of CLCR, a surrogate marker for glomerular filtration rate. In Cockcroft-Gault, creatinine excretion rate is modeled as a function of age, sex, and actual body weight[3] and creatinine clearance is obtained after dividing by the observed serum creatinine. Thus, when CG is used in a covariate model for clearance, TVCL becomes explicitly a function of ABW. This can be problematic if actual body weight is already in the model for TVCL in order to describe changes in clearance related to body size: information from one covariate (actual body weight) could influence estimates of multiple parameters.

To address these issues a model that was reported previously [49, 50] was used to assess covariate effects on TVCL. Instead of modeling CG-CLCR (which is explicitly a function of actual body weight), a renal function (RF) measure is derived as the CG-CLCR normalized to a 70 kg subject (equation 3.10). When the

[3] It is emphasized here that the patient's actual body weight (ABW), not the ideal body weight, was used in estimating creatinine clearances with the Cockcroft-Gault equation. Parameter estimates in the published regression equation [44] were obtained from using ABW-normalized creatinine excretion rate as the dependent variable for the regression. Cockcroft and Gault speculate in that publication that using IBW in place of ABW in obese patients may provide less-biased estimates. Although this IBW substitution has become common practice in clinical practice (e.g. [2]), predictive performance evaluations have shown this substitution to result in severe under prediction of CLCR, especially in obese patients [59]. Furthermore, it is noted that U.S. Food and Drug Administration guidelines on developing renal-dosing nomograms recommend using the Cockcroft-Gault equation as originally formulated - using ABW.

influence of ABW is removed from the renal function measure, differences in RF are more likely to be reflect changes in filtration capacity, rather than differences in creatinine production rate [49]. Note that in equation 3.9, renal clearance is still allowed to scale with ABW, but the scaling assumes the theoretical allometric relationship (see below).

Weight was included in the model for clearance through an allometric relationship [52, 53]. There is considerable debate in the pharmacometrics community about fixing the allometric exponent to it's theoretical value of 0.75 or allowing it to be estimated. In the current study, the exponent for weight on clearance was fixed to 0.75. Doing this allows the model to be used for simulation to aid in designing a future trial of FLU PK in a pediatric population. Furthermore, exploratory models were created where the exponent was estimated based only on the data in this study. The NONMEM estimate of 0.64 was very close to the theoretical value of 0.75 and an asymptotic 95% confidence interval around the estimate (0.5 to 0.79) included the theoretical value.

One of the goals of the population PK modeling was to obtain individual estimates of F-ara-A exposure for use in pharmacodynamic modeling (chapter 4) and dose optimizing exercises (chapter 5). If the data are insufficiently informative with respect to a parameter, these individual post-hoc predictions will "shrink" to the typical value of that parameter and the predictions will be unreliable [60, 42, **?**].

In this study, shrinkage on clearance was extremely small (∼1%) presumably due to an abundance of informative data, especially pre-dose observations on days 2-5 and the terminal observation 48 hours after dose 5. Because of the low shrinkage on CL, the individual post-hoc parameter estimates are suitable for use to calculate area-under-the-curve ($AUC_{0\rightarrow\infty}$; see chapter 4).

In this analysis, volumes of distribution were modeled to scale directly with lean body weight (LBW) instead of ABW as is usually done in allometric approaches [51]. Using ABW alone was insufficient to explain sex-based differences in volumes of distribution. When included in the model, these sex-differences were able to be precisely estimated and estimates for some volumes were clinically significant. For example, after accounting for ABW, V3 was on average 30% lower in female patients compared to male patients. In other words not only did female patients tend to have lower weights and lower V3 than male subjects, but female patients tended to have lower V3 even compared to male patients at the same ABW. The equation for estimating LBW published by Janmahastian et. al. [43] is a complicated function of height, weight, and sex. Using LBW in the model as the covariate for volumes of distribution left no detectable residual sex-differences in the parameters and led to a more parsimonious model. One serious drawback to scaling volumes to LBW comes when using the model to simulate in a pediatric population. Principles of allometry hold that volumes of distribution scale with

weight according to an allometric exponent of 1. It has not been established if the linear relationship between Vd and LBW as modeled here remains linear (i.e. allometric exponent of 1) from an adult population into a pediatric population. Even if the relationship could be assumed to be linear into a pediatric population, methods for accurate estimation of LBW in pediatric patients are not immediately available.

In summary, the 3-compartment model with ABW, CLCR, and comorbidity score as covariates on CL adequately described the data and the parameters were precisely estimated. The modest model misspecification was not sufficient to reject the model. Individual post-hoc CL estimates are expected to be suitable for use in calculating individual F-ara-A AUC.

Table 3.1: Study Patient Characteristics

Patient Characteristic	Median (range) / Percent
N	87
Age (years)	55 (20-69)
FLU Dose (mg/m^2)	40 (30-40)
Creatinine Clearance (ml/min)	106 (46-218)
Serum Creatinine (mg%)	0.9 (0.4-1.5)
Weight (kg)	82.6 (41.5-139.5)
BSA (m^2)	1.9 (1.3-2.5)
Female (%)	64
ATG (%)	46
Cord blood as stem cell source (%)	74
HLA-matched Related Donor (%)	25
Unrelated Donor (%)	75
Acute lymphocytic leukemia (%)	7
Acute myelogenous leukemia (%)	30
Chronic myelogenous leukemia (%)	1
Other Leukemia (%)	7
Myelodysplastic syndrome (%)	16
Non-Hodgkins Lymphoma (%)	20
Hodgkins Lymphoma (%)	9
Other (%)	10
Comorbidity Score = 0 (%)	10
Comorbidity Score = 1 (%)	16
Comorbidity Score = 2 (%)	21
Comorbidity Score $\geq$ 3 (%)	54

Table 3.2: Pharmacokinetic Parameter Estimates

Parameter	Unit	Estimate	RSE (%)	IIV (%)	RSE (%)
CL_{TOTAL}	L/hr	12.9		21	17
θ_{CLnr}	L/hr	5.35	15		
θ_{CLr}	L/hr	7.51	12		
θ_{Q2}	L/hr	4.16	13		
θ_{Q3}	L/hr	17.4	22		
θ_{V1}	L	66.8	6.7	25	25
θ_{V2}	L	69.4	8.1	19	83
θ_{V3}	L	37.9	8.3	46	23
θ_{COMOR}	% decrease	9.1	3.5		
$\rho_{CL,V1}$	-	0.789			
$\rho_{Cl,V3}$	-	0.599			
$\rho_{V1,V3}$	-	0.848			
σ_1^2	%	14.3	8.8		
σ_2^2	mcg/ml	8.5	40		

CLnr estimate is given for 70kg subject

CLr estimate is given for 70kg subject with CLCR of 100 ml/min

Q_x estimates are given for 70kg subject

θ_{Vx} are given for subject with lean body weight of 63 kg

Table 3.3: Shrinkage

Parameter	Shrinkage (%)
η-CL	1.3
η-V1	5.5
η-V2	48
η-V3	6.7
ϵ	7

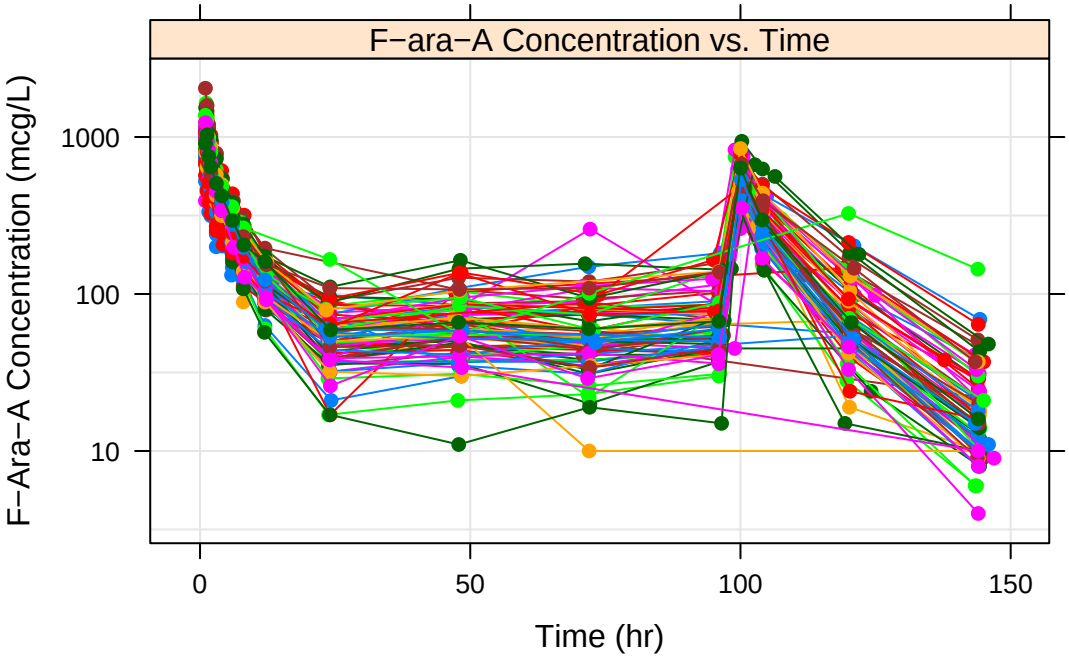

Figure 3.1: Observed plasma F-ara-A concentrations versus time for 87 patients in the study.

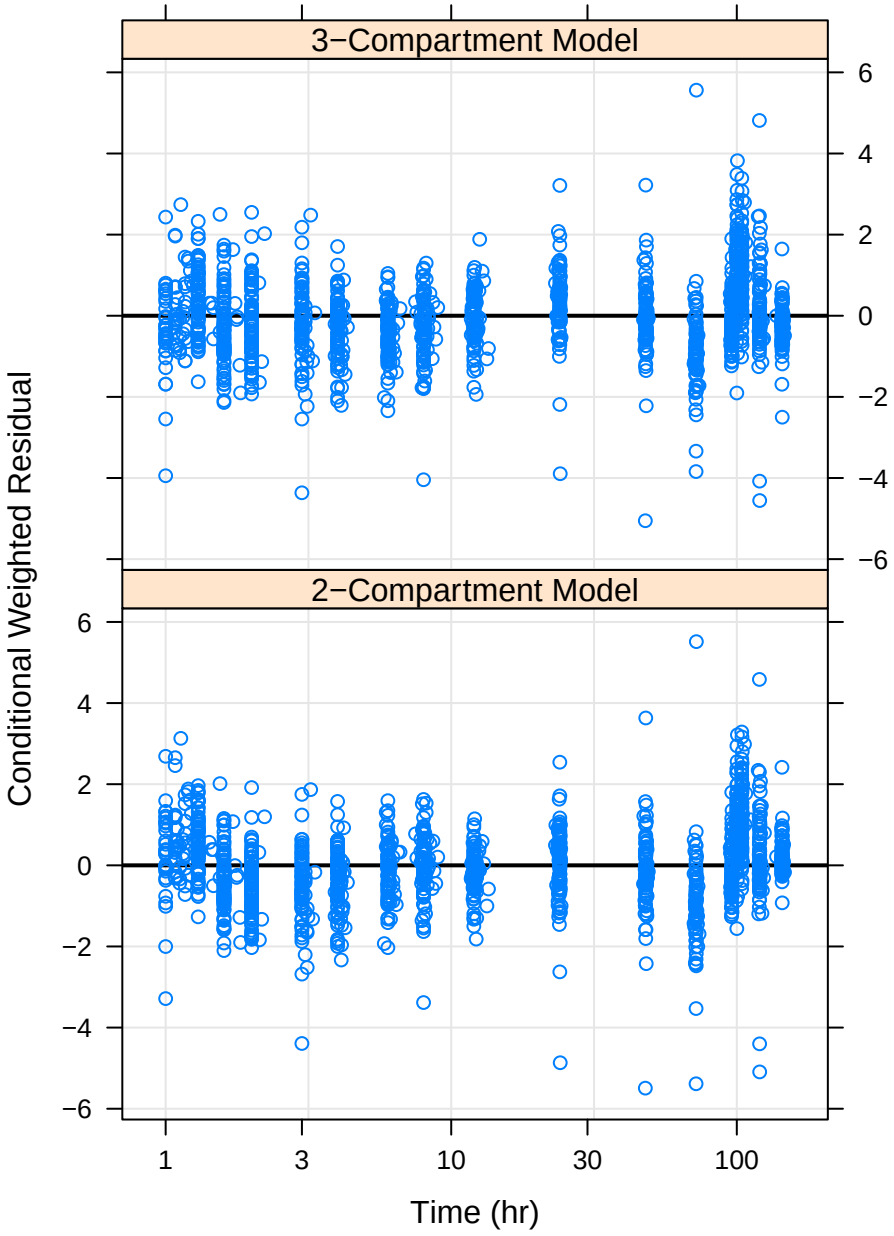

Figure 3.2: Diagnostic plots used in base model selection. Two- and three-compartment models were fit to the FLU pharmacokinetic data and conditional weighted residuals (CWRES) were plotted versus time. The top and bottom panels show CWRES from three- and two-compartment fits, respectively. The abscissa is shown on the log scale.

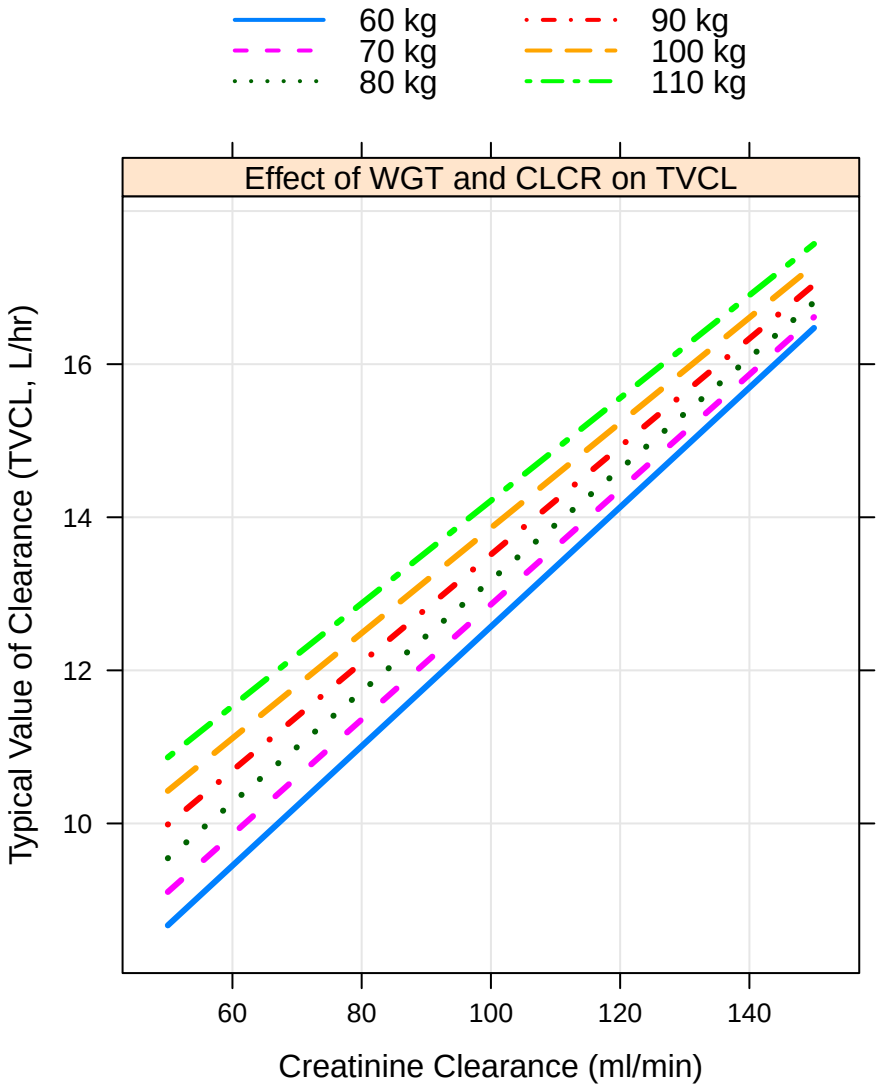

Figure 3.3: TVCL was calculated at various weights and CLCR to show the modeled relationships between the fixed effects and clearance. All simulations were done for a subject with 0/1 comorbidity score.

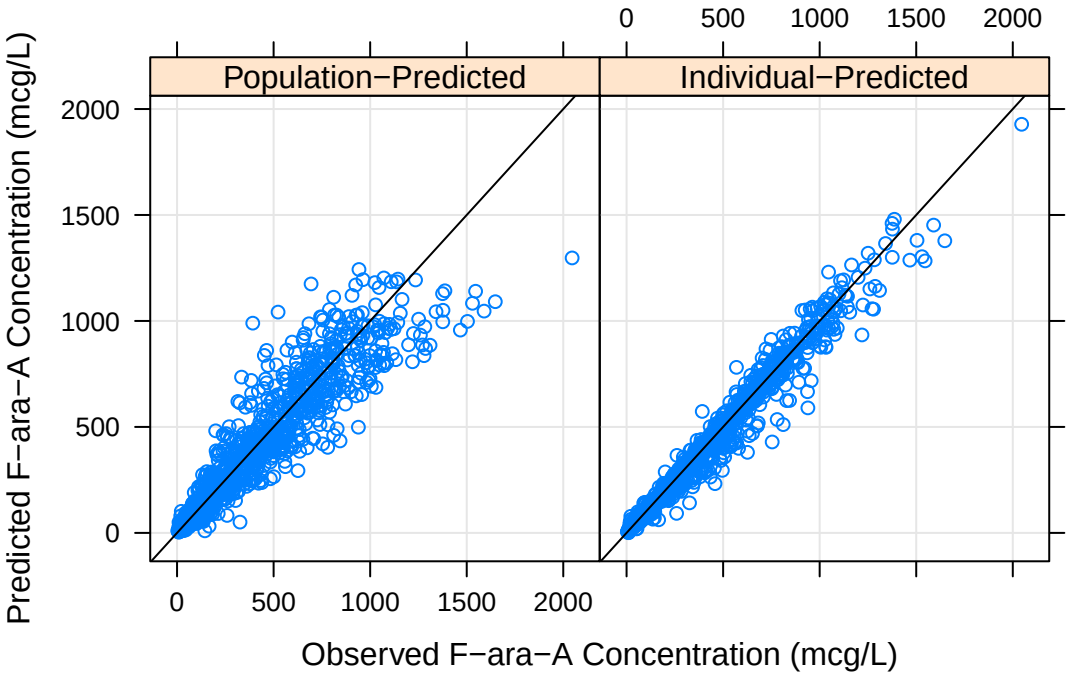

Figure 3.4: Population-predicted (left panel) and individual-predicted (right panel) F-ara-A concentrations are plotted versus observed concentrations. In each panel, the line of identity is shown.

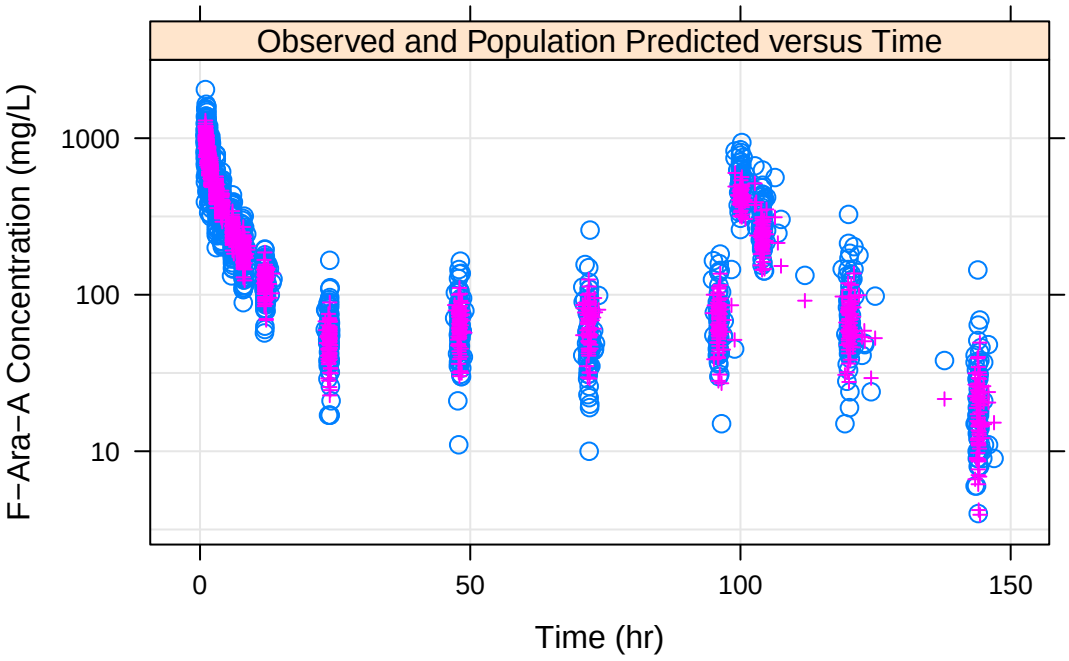

Figure 3.5: Population-predicted (crosses) and observed (open circles) F-ara-A concentrations plotted versus time.

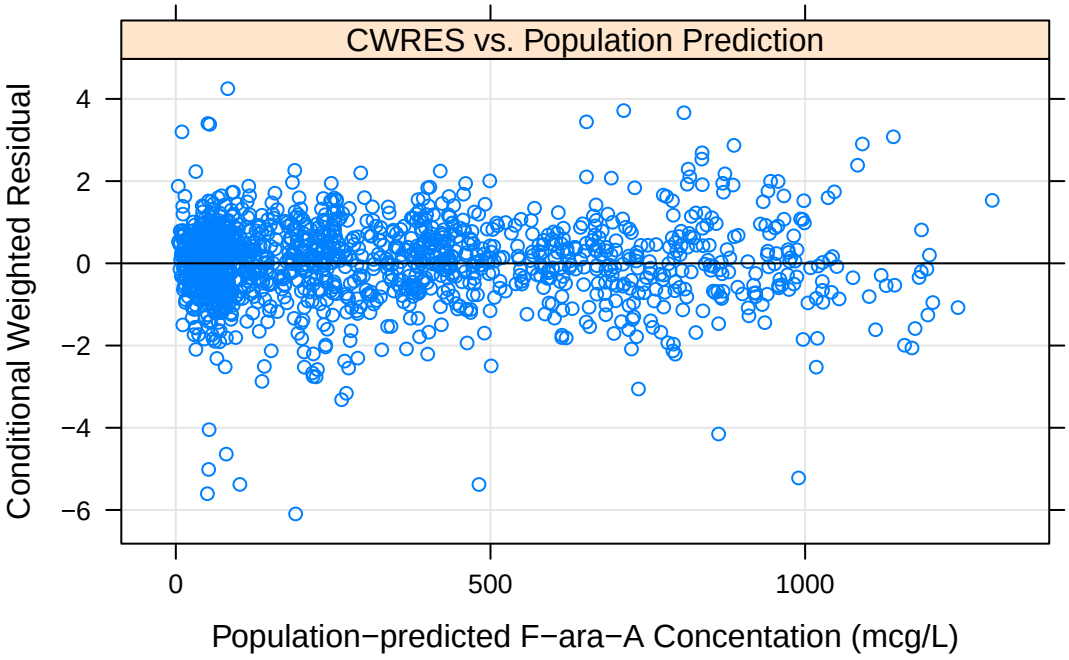

Figure 3.6: Conditional weighted residuals are plotted versus population-predicted F-ara-A concentrations from final model fit.

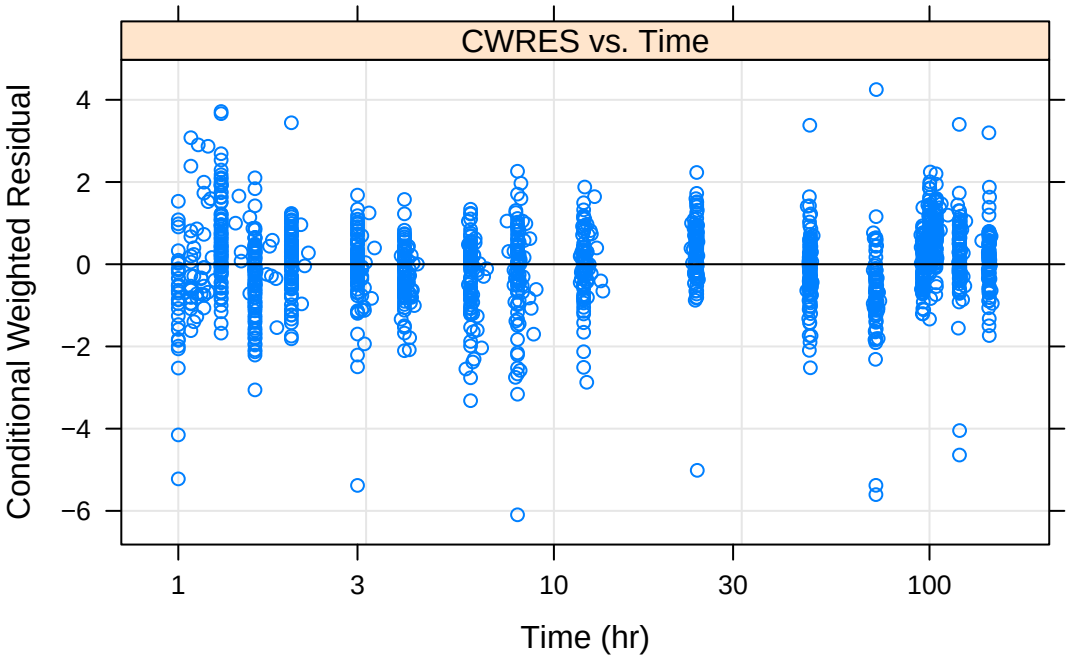

Figure 3.7: Conditional weighted residuals plotted versus time. The abscissa is shown on the log scale.

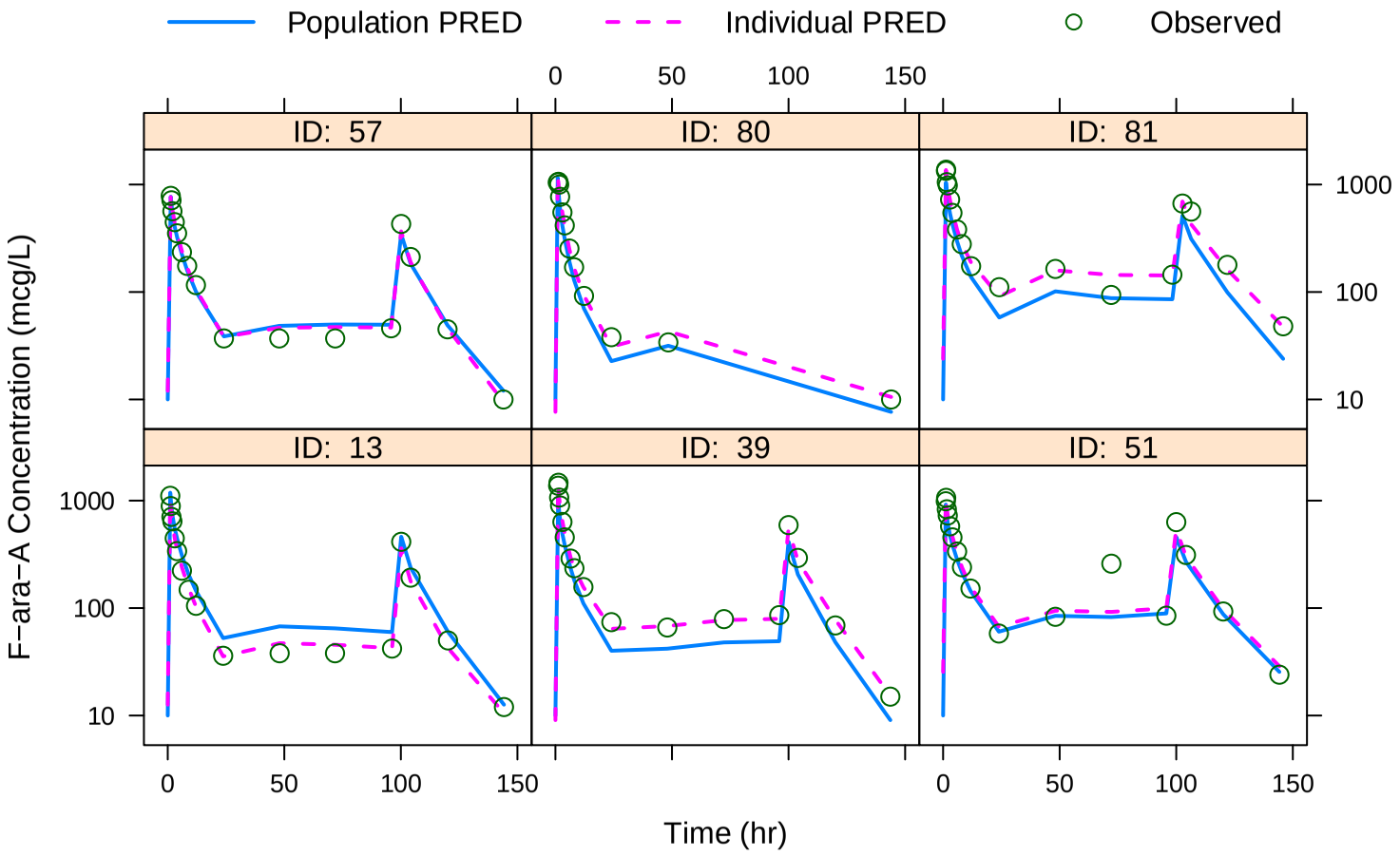

Figure 3.8: Population predictions (solid line), individual predictions (dashed line) and observed (open circles) F-ara-A concentrations are plotted versus time for six individuals.

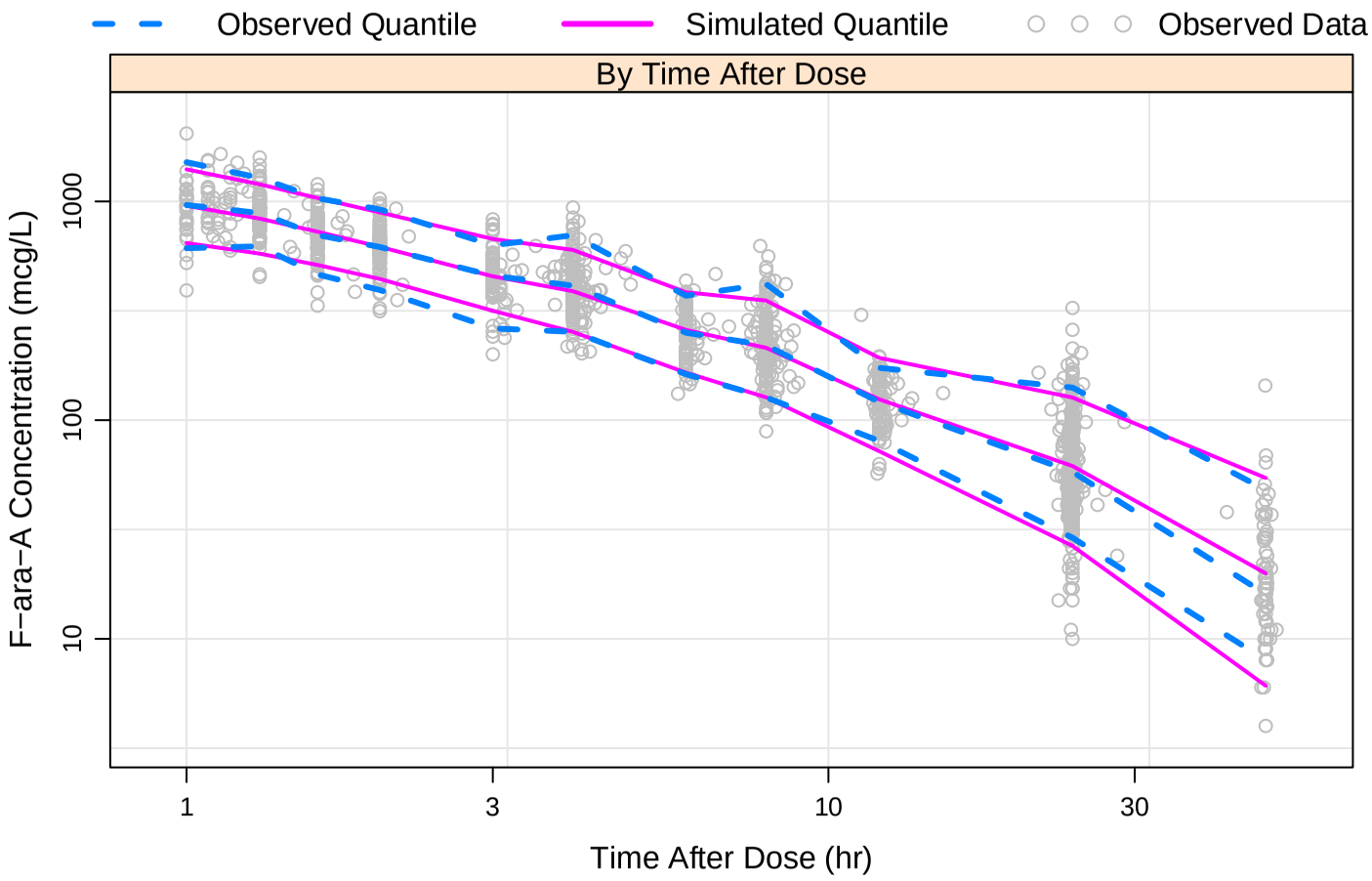

Figure 3.9: Observed F-ara-A concentrations (open circles) are plotted versus time after dose. Median and 90% intervals for observed and model-predicted concentrations are indicated by solid or broken lines as indicated in figure legend. The abscissa is shown on log-scale.

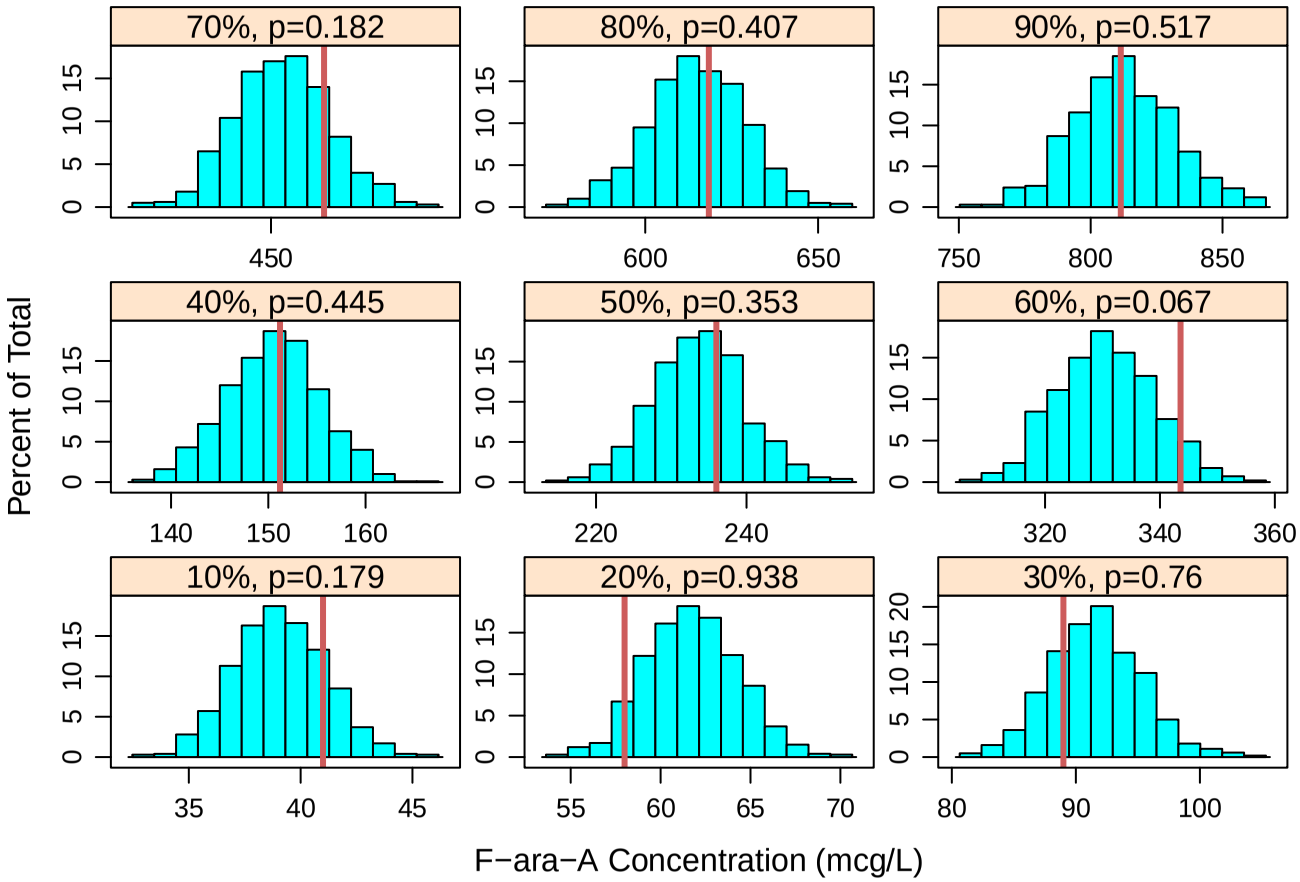

Figure 3.10: Numerical Predictive check. F-ara-A concentrations were simulated from the final model under the same design and covariate values as the original study data set. Test statistics are quantiles of F-ara-A concentration across entire data set. Each panel shows a different quantile (10 - 90%). Histograms show the distribution of test statistics for simulated data ($T(y^{rep})$). Vertical spikes in each panel shows the test statistics for observed data ($T(y)$).

Chapter 4

Fludarabine exposure-response modeling in HCT

4.1 Introduction

While literature reports about the pharmacokinetics of F-ara-A are abundant, there are very few reports on its pharmacodynamics in an HCT population. Long-Boyle et. al. [2] reported results of exposure-response analyses of the same data that are analyzed here. This review of the earlier work focuses on the methodology used and results obtained from the exposure-response analyses and is intended to specify the context and rationale for the current modeling work.

In the work by Long-Boyle et. al. [2], F-ara-A exposures ($AUC_{0\rightarrow\infty}$) for

each individual were determined through non-compartmental analysis of first-dose concentration-time profiles. Pharmacodynamic (PD) endpoints for exposure-response analyses included neutrophil engraftment, treatment-related mortality (TRM) , overall survival (OS), and acute graft-versus-host disease (GVHD). Relationships between PD endpoints and F-ara-A exposures were assessed through time-to-event analyses (Kaplan-Meier analyses and Cox proportional hazards models) as well as the cumulative incidence function [2].

In the analysis, all exposure-response relationships were assumed to be single step functions, where a critical F-ara-A exposure is identified that defines two populations based on their increased or decreased risk of experiencing the PD outcome. Taking the TRM response as an example, the analysis identifies a F-ara-A AUC "cutpoint" where individuals with AUC *greater* than the cutpoint have greater likelihood of experiencing TRM and vice versa for those with F-ara-A AUC *lower* than the cutpoint. These critical exposure values in the PD models were identified through a recursive partitioning process leading to an "optimal cutpoint" (OC) for F-ara-A AUC and other exposure metrics for each PD outcome [2].

Although the details of the process are not reported, OC is a commonly-used statistical method for identifying important predictors of some outcome by repeatedly dividing the dataset in to two groups based on values of some continuous

predictor (e.g. low and high exposure) and modeling differences in the outcome based on this arbitrary split. The range of observed values of the continuous predictor in question is explored and the OC is the value of the continuous predictor that gives the "most significant" difference in the responses between the two hypothesized groups (largest covariate effect or lowest p-value). OC methodology provides simple clinical tools for therapeutic drug monitoring (TDM).

In the Long-Boyle analysis, the OC approach showed a significantly greater TRM and significantly lower OS and grade II-IV GVHD with higher F-ara-A exposures, but no significant relationships for neutrophil engraftment [2]. These results are briefly summarized here. Across all F-ara-A exposures, 6-month TRM cumulative incidence was 21%. OC analysis showed that a 50% cumulative incidence of TRM if F-ara-A AUC was above the 6.5 μg·h/ml cutpoint (N=7 subjects) and only 15% if AUC was below the cutpoint (N=80). Multivariate Cox regression modeling showed a 6-month relative risk of TRM of 4.56 for subjects with AUC above the cutpoint [2]. With respect to the day-100 GVHD endpoint, subjects with F-ara-A AUC above 6.5 μg·h/ml had decreased incidence (14%) of grade II-IV GVHD compared with subjects below the cutpoint (52%). An F-ara-A effect was not detectable when grade III-IV GVHD was considered. Across all observed AUCs, there was a 50% incidence of grade II-IV GVHD and 17% incidence of grade III-IV GVHD. Overall, engraftment was high (86%), but no relationship

with F-ara-A exposure was detectable [2].

This analysis gives important new insights about TRM and GVHD endpoints and F-ara-A exposure, but a robust description of the graded exposure-response relationships have yet to be elucidated. Specifically, the assumption of a step exposure-response model and the implementation of recursive partitioning methodology limit the available signal in the data. Additionally, this approach is not consistent with the principles of human biology and systems pharmacology that support graded PD responses.

The step model hypothesizes that risk is constant above and below the cutpoint (figure 4.1). The range of observed AUC was 2 - 11.5 μg·h/ml by non-compartmental analysis and the OC was 6.5 μg·h/ml. Under the model, the risk of TRM is the same from 2 to 6.5 μg·h/ml cutpoint, and doesn't increase as exposure increases from 6.5 to 11.5 μg·h/ml. These are the assumptions of the model, but it is unlikely that the biology is actually consistent with this step function. Questions left unanswered by this model include: (1) While the model predicts benefit (lower TRM risk) when exposure targets are lowered below the cutpoint (e.g. 6 μg·h/ml compared with 8 μg·h/ml), is there additional benefit to an even lower exposure target (e.g. 4 μg·h/ml compared to 6 μg·h/ml) and (2) Is there an exposure associated with TRM risk less than, say, 10%? What is the exposure? In other words, the OC analysis *confirmed* that there was more TRM and less OS

at highest F-ara-A exposures; there remains much to *learn* about the nature of the exposure-response relationships [61].

Furthermore, while the OC methodology of data-determined conversion of continuous covariate (exposure) measurements into dichotomous covariates is commonly used, the methodology can be problematic [62, 63, 64]. Problems such as data reduction, power loss, incorrect selection of "significant" covariates, bias in estimates of covariate effects, and inappropriately small p-values and confidence intervals have been discussed in the statistics literature and raise concern over the use of the methodology. Although there are ways to adjust the results of the OC procedure, it would be desirable to bypass this methodology altogether and develop PD models that utilize the full information available in a continuous exposure metric.

Therefore, the objective of the work presented in this chapter are to develop predictive models describing the relationships between F-ara-A exposure and probability of neutrophil engraftment, TRM, and GVHD.

4.2 Methods

4.2.1 Pharmacodynamic Data

The HCT treatment regimen, including fludarabine therapy, was detailed in chapter 3. Patients were followed after transplant and evaluated on several outcome measures. Treatment outcomes considered for pharmacodynamic modeling included treatment-related mortality (TRM), development of severe acute graft-versus-host disease (GVHD), and engraftment (ENGRAFT) [1, 2]. All three outcomes were treated as binary random variables, coded 1 if the event was observed in a patient ("success") or 0 if the event was not observed over the duration of that patient's observation period ("failure") [1] . TRM was defined as death with no relapse or disease progression (no TRM = 0, TRM = 1). Acute GVHD was graded on a 0-4 scale [65] by clinicians on the care team. Modeling of GVHD responses focused on development of severe GVHD. A binary random variable was created indicating the development of maximum acute GVHD grade III or IV (Grade 0-II = 0, Grade III or IV = 1). Engraftment was defined as absolute neutrophil count (ANC) recovery to greater that 500 cells per microliter by day

[1] The "success" / "failure" terminology will be used sometimes in this text. It is acknowledged here that neither TRM nor serve GVHD are considered "successful" outcomes from treatment, with the positive connotation that usually comes with using the word "success". Here, "success" is only used to indicate that an outcome (good or bad) has occurred and not as a comment on the desirability of that outcome.

+42 and no auto recovery (engraftment criteria not met = 0, engraftment criteria met = 1). Clinicians involved in the transplant procedure were responsible for determining the patient status for each PD outcome. Time in days to development of the outcome and censoring times were also available for ENGRAFT and TRM, but not the GVHD outcome, which was the maximum score over the observation interval. ENGRAFT and TRM responses for censored patients were assigned values of 0 (event not observed).

4.2.2 Software and Estimation

Parameters for pharmacodynamic generalized linear regression models were estimated within a Bayesian framework using the BRugs interface to OpenBUGS (version 3.0.3). Posterior distributions for model parameters were characterized by drawing 10,000 production samples from these distributions. Production samples were summarized as a point-estimate (50^{th} percentile) and a 95% credible interval (the range from the 2.5^{th} to 97.5^{th} percentile). Bayesian credible intervals allow direct probability statements about the uncertainty of an estimate: there is a 95% probability that the parameter value is within the 95% credible interval [66]. To ensure that the Gibbs sampler converged to the true posterior distribution for the parameter, 10,000 burn-in samples were drawn and discarded before drawing the production samples. Sampler convergence was assessed through several

mechanisms as described in section 4.2.4.

4.2.3 Exposure-response Models

Relationships between F-ara-A exposure and PD endpoints were modeled using generalized linear models (GLM) [67]. The general form of the model applicable to every PD outcome is described here, but it is noted that each outcome was analyzed on it's own, independently of the other endpoints. The analysis assumes that outcomes ($Y = 0$ or 1) have a Bernoulli distribution:

$$Y \sim ber(\pi) \tag{4.1}$$

where π is the probability of success (i.e. drawing a 1) in a given trial. The probability distribution function of a the bernoulli random variable is:

$$f(y|\pi) = \pi^y(1-\pi)^{1-y} \tag{4.2}$$

with expectation $\mathrm{E}[Y] = \pi$ and variance $\mathrm{var}[Y] = \pi(1-\pi)$. Modeling $\mathrm{E}[Y]$ directly is problematic because of difficulties constraining model predictions to only take values of 0 or 1. Instead, a link function is used to transform the expectation of the random variable from [0,1] to [-∞, ∞]. It is this transformed expected response which is subsequently related to explanatory variables (covariates) through estimable regression parameters.

The GLM for any PD response in this analysis has the general form:

$$g(\pi_i) = \beta_0 + \beta_1 \cdot (AUC_i - 6) \tag{4.3}$$

where π_i is the probability of the event happening in the ith subject, AUC_i is the F-ara-A area under the curve ($\text{mg}\cdot\text{hr}\cdot\text{L}^{-1}$) from time zero to infinity[2] [68] and $g(\cdot)$ is the link function transforming the expectation of the response. The reference AUC (6 $\text{mg}\cdot\text{hr}\cdot\text{L}^{-1}$) centers the covariate approximately at the median observed AUC (5.8 $\text{mg}\cdot\text{hr}\cdot\text{L}^{-1}$). Different link functions were explored in the modeling, including logit, probit, and complementary log-log links. The *logit* transformation is the log of the odds:

$$g(\pi) = logit(\pi) = log\left[\frac{\pi}{1-\pi}\right] \tag{4.4}$$

The *logit* link function is the canonical link function. The probit link function uses the inverse cumulative standard normal distribution function (Φ^{-1}):

$$g(\pi) = probit(\pi) = \Phi^{-1}[\pi] \tag{4.5}$$

The complementary log-log link function is:

$$g(\pi) = cloglog(\pi) = log[-log(\pi)] \tag{4.6}$$

[2] The AUC abbreviation will always refer to the area under the concentration-time curve from time zero to time infinity: $AUC = \int_0^\infty C(t)dt$

AUCs were derived from the individual Bayesian post-hoc clearances from the population PK analysis and the dose according to the relationship :

$$AUC_i = \frac{DOSE_i}{CL_i} \tag{4.7}$$

Interpreting model parameters using the logit link function: the parameter β_0 is the logit of the success probability at the reference AUC. The parameter β_1 describes the magnitude of the linear effect of F-ara-A exposure on the log odds of success: a unit change in AUC changes $logit(\pi)$ by β_1 units. Positive values of β_1 indicate an increase in success probability with increasing AUC and negative values of β_1 indicate a decrease in success probability with increase in AUC. Alternatively, FLU effect may be interpreted as an odds ratio where the odds of success change by e^{β_1} with each unit change in F-ara-A AUC. In this interpretation, an odds ratio greater than 1 indicates increasing odds of success with increasing AUC, an odds ratios less than 1 indicate decreasing odds of success with increasing AUC, and an odds ratio equal to 1 indicates no effect of F-ara-A AUC on odds of success. Only linear exposure-response models were considered in this analysis.

A rigorous evaluation of all available covariates influencing the various PD endpoints was beyond the scope of the current work. However, to evaluate the

possibility that the F-ara-A exposure effects were confounded by some other covariate, multivariate GLMs were fit to PD data while including effects for other potentially important covariates. The covariates used were: anti-thymocyte globulin (ATG) use in the preparative regimen (ATG not used = 0, ATG used = 1), advanced age (AGE $< 55 = 0$, AGE $> 55 = 1$), umbilical cord stem cell source (non-cord source = 0, cord source = 1), and comorbidity score (COMOR $\leq 1 = 0$, COMOR $> 1 = 1$). Analogous to equation 4.3, the multivariate model was:

$$\begin{aligned} g(\pi_i) &= \beta_0 + \beta_{AUC}(AUC_i - 6) + \beta_{ATG}ATG_i + \beta_{AGE}AGE_i \\ &\quad + \beta_{CORD}CORD_i + \beta_{COMOR}COMOR_i \end{aligned} \tag{4.8}$$

Only the logit link function was evaluated in the multivariate models.

Information about model parameters was not available prior to starting the analysis. Therefore, noninformative priors were used on β_0 and β_1 (equation 4.3) for all PD GLMs and all link functions. These priors were implemented in OpenBUGS using the `dflat()` function in OpenBUGS.

4.2.4 Model Evaluation and Diagnostics

The convergence of the Gibbs sampler for generalized linear modeling of binary PD endpoints was assessed with multiple approaches. Convergence was assessed

graphically by running multiple chains with over-dispersed starting values and checking to ensure that the chains had mixed adequately by the time production samples were drawn. The correlation in parameter samples was assessed by plotting autocorrelation versus sample lag as output by BUGS software.

The suitability of the different link functions in the generalized linear model was assessed using the deviance information criterion (DIC) [69]. Like other criteria including AIC, DIC is a model selection criteria that balances model fit and complexity. DIC is defined as:

$$DIC = \bar{D} + p_D \tag{4.9}$$

where $\bar{D}$ is the deviance averaged over the posterior samples of θ (a measure of model fit) and p_D is the effective number of parameters in the model (a measure of model complexity). Different models may be compared based on DIC, with preference for models with lower DIC (best fit with least complexity).

Generalized linear model fits were evaluated through posterior-predictive model checking (PPC) [66, 56]. The primary PPC used samples from the posterior distribution of β_0 and β_1 for each PD endpoint and derived predicted success probabilities on new AUC data across the observed AUC range (3.5 to 9.5 mg·hr·L^{-1}). If β_0^* and β_1^* are samples from the posterior predictive distributions of β_0 and β_1, respectively, then the posterior predictive success probability is:

$$\pi_i^* = \frac{exp(\beta_0^* + \beta_1^* \cdot (\tilde{AUC}_i - 6))}{1 + exp(\beta_0^* + \beta_1^* \cdot (\tilde{AUC}_i - 6))} \tag{4.10}$$

where $\tilde{AUC}_i$ is new exposure data (i.e. not previously observed in the analysis data set). For each $\tilde{AUC}_i$, the median π_i and a 95% posterior predictive interval for π_i were graphically compared to model-independent estimates for π. Model-independent estimates for π_i were obtained after binning event data within low (mean [range] = 4.8 [3.5 - 5.5] mg·hr·L^{-1}), medium (6.2 [5.51 - 7] mg·hr·L^{-1}), and high (8 [7.01 - 9.6] mg·hr·L^{-1}) observed AUC groups (table 4.1).

A second form of PPC used the number of events for each endpoint, both overall and conditional on F-ara-A AUC, as the test statistic in the PPC. For each PD endpoint and each β_0^* and β_1^* sample from its respective posterior distribution, event data at observed AUC levels were simulated. For a given PD endpoint, at each iteration of the Gibbs sampler, the overall total number of simulated events (averaged across all observed AUCs) as well as simulated event totals within the binned AUC groups were calculated. The distributions of simulated event totals were compared to observed event totals. A Bayesian p-value (p^{PPC}) was calculated as the fractional number of times the simulated event total exceeded the observed event total (see equation 3.13).

In addition to individual event totals, a composite event was evaluated through PPC. The composite event was defined as:

$$Y_{COMP} = Y_{ENGRAFT} \cdot Y^{C}_{GVHD} \cdot Y^{C}_{TRM} \tag{4.11}$$

In other words, the patient was considered a "success" with respect to the composite event if he or she engrafted, but did't experience grade III or IV GVHD or treatment-related mortality. Using the PPC procedure described above, the distribution of simulated composite events was compared to the number of observed composite events.

4.2.5 Time-to-Event Analysis

In the primary exposure-response analysis, generalized linear models were estimated treating PD response data as binary random variables. As noted above, censoring in the data was ignored in this analysis and subjects were considered "failure" (PD outcome = 0) if the data were censored. A secondary, time-to-event analysis for the TRM and ENGRAFT endpoints was completed to assess the consequences of the assumption regarding censoring in the data. For both endpoints, semi-parametric proportional hazards models were fit to survival times using the `coxph` function within the `survival` package in R. The form of the regression model [70] was the same for TRM and ENGRAFT survival times. Models were

explored where F-ara-A AUC is treated both as a continuous and categorical covariate, where AUC categories were as defined in section 4.2.4 and table 4.1. If AUC was a continuous covariate, the hazard function was:

$$h(t|AUC_i) = h_0(t)\ exp\left(\beta_1\ \left[AUC_i - 6\right]\right) \tag{4.12}$$

where t is the time to PD outcome occurrence, $h(t|AUC)$ is the survival function conditional on F-ara-A AUC, $h_0(t)$ is the estimated baseline hazard function (i.e. the hazard for a subject with AUC of 6 mg·hr· L^{-1}), and β_1 is the estimated parameter describing the effect of changes in AUC on hazzard. The hazard function when AUC was a categorical covariate was:

$$h(t|AUC_i) = h_0(t)\ exp\left(\beta_1\ I_1^{AUC-MID} + \beta_2\ I_2^{AUC-HIGH}\right) \tag{4.13}$$

where $h_0(t)$ is the baseline hazard function for the lowest AUC bin, I are functions indicating membership in the mid AUC bin (I_1) and high AUC bin (I_2), and β_1 and β_2 are the respective regression coefficients. The exponential of β_1 (e^{β_1}) describes the change in the hazard ratio with a unit change in F-ara-A AUC.

The Cox-proportional hazards models were fit for the TRM and ENGRAFT endpoints and the statistical significance of F-ara-A exposure effect parameter estimates were evaluated through a likelihood-ratio test. Plots of predicted six-month survival probabilities for TRM and ENGRAFT conditional on F-ara-A

exposure were generated using the `survfit` function in R and compared to posterior predictive exposure-response relationships from the Bayesian generalized linear modeling of TRM and ENGRAFT as binary responses with censored values set to zeros (section 4.2.3). Median event probabilities from each method were compared at F-ara-A exposures of 5, 6.5, and 8 mg·hr·L^{-1}.

4.3 Results

Generalized Linear PD Modeling

From the pharmacokinetic analysis, the median F-ara-A AUC was 5.8 mg·hr·$L^{-1}$with range of 3.5 to 9.6 mg·hr·L^{-1}. Figure 4.2 shows the distribution of the 87 AUCs used for PD modeling. Table 4.1 summarizes event numbers and probabilities for the different PD endpoints, both overall and conditional on F-ara-A exposure. Of the 87 patients studied, 72 (83%) engrafted, 10 (11%) experienced grade III or IV acute GVHD, and 19 (22%) experienced treatment-related mortality. The model-independent success probabilities trended to increase with F-ara-A AUC for the TRM endpoint, but a clear pattern was lacking for the GVHD or ENGRAFT endpoints. There were no GVHD events in the subgroup with highest of the observed F-ara-A AUC. TRM event probabilities on the logit scale increased approximately linearly with F-ara-A AUC (table 4.1 and figure 4.3).

Generalized linear modeling focused on models using the logit link function (i.e. logistic regression models). Parameter estimates from logistic regression models for each PD endpoint are presented in table 4.3. The estimates show a significant increase in the log odds of TRM with increasing F-ara-A AUC (β_1 estimate [95% CI] = 0.55 [0.15 , 0.7]). Interpreted as an odds ratio, there is a 73% (95% CI: [18% , 163%]) increase in the odds of TRM with each mg·hr·L^{-1}increase in F-ara-A AUC. The model was not able to detect a significant effect of F-ara-A AUC on either the ENGRAFT endpoint (odds ratio: 1.06 [0.69 , 1.64]) or the GVHD endpoint (0.77 [0.41 , 1.28]). Credible intervals for β_1 on F-ara-A AUC excluded zero (no F-ara-A exposure effect on success probability) for the TRM but not the GVHD or ENGRAFT endpoints. Figure 4.4 shows density plots of some samples drawn from the posterior distributions of parameters in the logistic regression models.

Table 4.4 shows DIC values for full (including a parameter for F-ara-A AUC effect - β_1) and reduced (with intercept parameter only - β_1 fixed to 0) models for each PD endpoint using logit, probit, and complimentary log-log link functions. With respect to the TRM endpoint, DIC values were smaller for the full model compared to the reduced model by approximately 5 points and this trend was consistent for all link functions tested. For GVHD and ENGRAFT endpoints, on the other hand, DIC values were smaller for the reduced model (excluding the

F-ara-A AUC effect) for all link functions. For any given model for any given endpoint, DIC values for different link functions were very similar and didn't indicate a strong preference for a particular link function in the GLM. Based on this link function result, modeling estimates for probit and complementary log-log link functions are omitted from further discussion and all GLM development and use will focus on models using the logit link function.

Posterior predictive checks for GLMs using the logit link function and including the F-ara-A AUC effect (equations 4.3 and 4.4) are shown in figure 4.5. The median model-predicted probability of engraftment is slightly greater that 0.8 and changes very little across the observed F-ara-A AUC range. The median model-predicted success probability for the GVHD endpoint slowly increases with decreasing F-ara-A exposure, but remains less than 0.2 at the lowest observed AUC and the relationship is, overall, relatively flat. With respect to the TRM endpoint, a clear increase in probability of TRM is seen with increasing F-ara-A AUC. At lowest exposures, the median model-predicted probability of TRM is less than 0.05, but increases to greater than 0.6 at the highest exposure (a 12-fold increase in probability of TRM). It should be noted that all exposure-response trends are valid only across the observed AUC range of $\sim$ 3.5 to 9.5 mg·hr·L^{-1}. For all endpoints, success probability prediction intervals become wider at the lowest and highest F-ara-A AUC due to fewer observations, but the median predictions

track well with model-independent probabilities (figure 4.5).

A second posterior predictive check based on observed and model-predicted event counts is presented in figure 4.6. At F-ara-A AUCs in lower, middle, and upper parts of the observed range, the model was able to predict observed numbers of events, with p^{PPC} ranging from 0.13 to 0.7. No significant deficiency was noted when posterior predictive checking on the total number of events observed across all AUCs. Figure 4.7 shows results from posterior predictive model checking on the composite event endpoint (equation 4.11). There were 52 composite success events observed in the study population. The median model-predicted number of composite events was 50, with a 95% prediction interval of 37 to 61. The Bayesian p-value for this composite statistic was 0.374, indicating the discrepancy between the observed and model-predicted values could have happened by chance even if the model were correct.

Figure 4.8 shows sample history trace plots for intercept (β_0) and drug effect (β_1) model parameters for each of the three PD endpoints. Two chains are shown, with over-dispersed starting values (β_0 starting at 10 and -10 ; β_1 [F-ara-A effect] started at 2 and -2). The chains show good mixing suggesting that production samples were truly drawn from the parameters' posterior distributions. Figure 4.10 shows autocorrelation plots for one of the chains shown in figure 4.8. The low autocorrelation suggests that production samples were independent draws from

the posterior distribution. Figure 4.9 shows the Brooks, Gelman, Rubin (BGR) statistic [71], the ratio of inter-to intra-chain variability for posterior samples for model parameters for each PD endpoint. All plots show the ratio approaching 1 (inter-chain variability being about equal to inter-chain variability) by the 5000th iteration, again suggesting that the Gibbs sampler had converged to the parameters' true posterior distribution by the time production samples were drawn.

Modeling results from the multivariate generalized linear PD models are shown in table 4.5. Credible intervals covered zero for all covariate effect parameters except for β_{ATG} and β_{AUC} on the TRM endpoint. On average, the log odds of TRM increased by 1.82 units in subjects receiving ATG in the preparative regimen compared with patients not receiving ATG. This translates to a odds ratio of 6.2 (95% CI: 1.8 - 25). Estimates for F-ara-A AUC effects in the saturated model were not different from estimates obtained in the univariate models (compare with 4.3). For ENGRAFT and GVHD endpoints, DIC values were lower for a reduced model without β_{AUC} compared with a full model that included the exposure effect. DIC increased by 6.6 when β_{AUC} was removed from the saturated model.

Time-to-Event PD Modeling

Results from exposure-response modeling of survival times for ENGRAFT and TRM endpoints are shown in tables 4.6 and 4.7 and figures 4.11 and 4.12. Across

all exposures, median time to ENGRAFT and TRM outcomes were 9 and 44 days, respectively [2]. Exploratory Kaplan-Meier survival curves (figure 4.11) show strong signal indicating a decreasing time to TRM with increasing F-ara-A exposure. Time to ENGRAFT event appears to be independent of F-ara-A exposure, with possible increase in time to ENGRAFT in the highest F-ara-A exposure bin. Parameter estimates from proportional hazards modeling shown in tables 4.6 and 4.7 confirm the trend.

Figure 4.12 shows the exposure-response relationships for TRM and ENGRAFT considering the outcome data as both binary outcomes (GLM) and as times to events (PH). Data are shown for AUC used as a continuous covariate and survival probabilities for the PH method are evaluated at 42 and 180 days for ENGRAFT and TRM outcomes, respectively. Median predictions from both the PH and GLM approaches lie within 95% prediction intervals for the other approach. Overall, there is good agreement between the two approaches.

4.4 Discussion

The work in this chapter investigates relationships between F-ara-A exposure and clinical outcomes after HCT. A linear pharmacodynamic model was chosen for modeling the exposure-response relationships for all three endpoints. Exploratory

plots of model-independent event probabilities (figure 4.3) suggest insufficient information in the data to support more complicated PD models (e.g. EMAX model) and that the linear model is not inappropriate. EMAX models are attractive because they can assess the lower and upper limits of drug effects. Unfortunately, the linear PD model predicts direct proportionality between the logit of the event probability and F-ara-A exposure, which is not likely to be the case across all possible exposures. That is, it is highly likely that there is an exposure at which the maximum effect is reached and this feature is not captured by the linear PD model. Therefore, it is extremely important to avoid the temptation to extrapolate the exposure-response modeling beyond the observed range of exposures.

One attractive feature of the GLM PD models presented here is that F-ara-A exposure is treated as a continuous exposure metric. The modeled PD probabilities change continuously with increasing F-ara-A exposure across the observed AUC range and modeled event probabilities were consistent with model-independent probability estimates (figure 4.5). Establishing a continuous, instead of piecewise constant relationship (based on OC) between PD probabilities and F-ara-A AUC increases the informativeness of the PD model and allows quantitative decision-making when targeting F-ara-A exposures. For example, the model does not assume that TRM risk stays constant above or below a single "cutpoint"

but allows evaluation of risk across a range of exposures. Similar information is obtained from the ENGRAFT and GVHD outcomes.

Logistic regression modeling shows a clear increase in the probability of TRM with increases in F-ara-A exposure (table 4.3 and figure 4.5). Under the model, the probability of treatment-related mortality increases from less than 10% at the lowest observed exposures (approximately 4 mg·hr·L^{-1}) to more than 50% at the higher observed exposures (approximately 8.5 mg·hr·L^{-1}). The FLU exposure effect on TRM was not changed in a generalized linear PD model with other relevant covariate effects (equation 4.9 and table 4.5). These results are consistent with previous work [2] that showed increased risk of TRM at exposures exceeding an F-ara-A AUC cutpoint.

Interestingly, the modeled probability of ENGRAFT remained relatively constant over the observed exposure range, with the median of the predictive distribution always greater than 80%. Although the HCT preparative regimen is reduced-intensity (or nonmyeloablative), the concern remains that under-dosing FLU may result in F-ara-A exposures so low that likelihood of engraftment is compromised. While conclusions on ENGRAFT probability cannot be drawn on AUCs less than 3.5 mg·hr·L^{-1}, the modeling work here suggests the maximum FLU drug effect with respect to ENGRAFT has been reached. Increasing FLU doses to more aggressive exposure levels (say, greater than 6 mg·hr·L^{-1}) is not

likely to increase chances of engraftment, but would likely increase the risk of TRM. See chapter 5 for more consideration of optimizing doses based on different PD endpoints.

One concern related to modeling methodology was the handling of the data as discrete binary responses, ignoring censoring in the dataset. Initially, parametric time-to-event models that considered ENGRAFT and TRM event times as Weibull-distributed random variables were explored. This approach was abandoned because of model misspecification at survival times of interest (6 months for TRM). The logistic regression approach that considers these data as binary outcomes seems to sufficiently describe the data. A secondary analysis of survival times using a semi-parametric time-to-event analysis carried out as a check on the appropriateness of the GLM approach. Graphical comparison of typical model-predicted probabilities (figure 4.12) showed similar results for the two approaches.

Another limitation of the exposure-response modeling was the inability to directly address correlation between PD endpoints. The TRM, GVHD, and ENGRAFT outcomes were modeled independently, with parameters for each response estimated separately in independent generalized linear models. This independence approach ignores our intuition that, for example, as the probability of GVHD decreases, the probability of engraftment should decrease as well. While there are well-known and easily-implemented methods for modeling correlation repeated

measures of the same response within subjects in a population, it is not immediately clear how to address correlation of single measures of several different responses. There are examples in the statistics literature and the pharmacometrics literature where latent random effect variables have been used to account for the correlation between repeated measures of several response variables. However, it is not clear how these variables would be used in models for single measures of multiple responses.

While the suspected correlation in the different PD responses couldn't be directly addressed in the model, posterior predictive checking was used to try to evaluate the independence assumption [72]. A single composite success event (Y_{COMP}) was derived from the individual endpoints to indicate success if there was engraftment without TRM or grades 3 or 4 acute GVHD. Because observed Y_{COMP} was derived from observed individual endpoint data, the composite measure implicitly incorporates correlation between the individual endpoints. In the PPC, individual endpoint predictions were obtained from logistic regression models that ignored correlation in responses (a model deficiency if the responses are presumably correlated). If the deficiency due to the independence assumption had a significant negative impact on modeling conclusions from a predictive standpoint, the distribution of the number of model-predicted composite outcomes may potentially look very different than the observed number of composite outcomes.

Figure 4.7 shows that the number of composite successes predicted under the model (median: 50 events) matched up well with the observed total (52 events), with $p^{PPC} = 0.374$ indicating that this difference between model-predicted and observed event totals could happen by chance even if the independence model were correct. This PPC is just one check of the consequences of the independence assumption and the results suggest that the assumption does not seriously affect model predictions. The PPC does not (and can not) show that the model is the "correct" model and it is not able to address other potential model deficiencies [72].

In summary, three generalized linear PD models adequately describe F-ara-A exposure-response relationships for TRM, ENGRAFT, and GVHD outcomes. Consistent with previous work, the models predict increased risk of TRM at higher F-ara-A AUCs with little benefit with respect to ENGRAFT. These models are suitable for simulation-based FLU dose optimization.

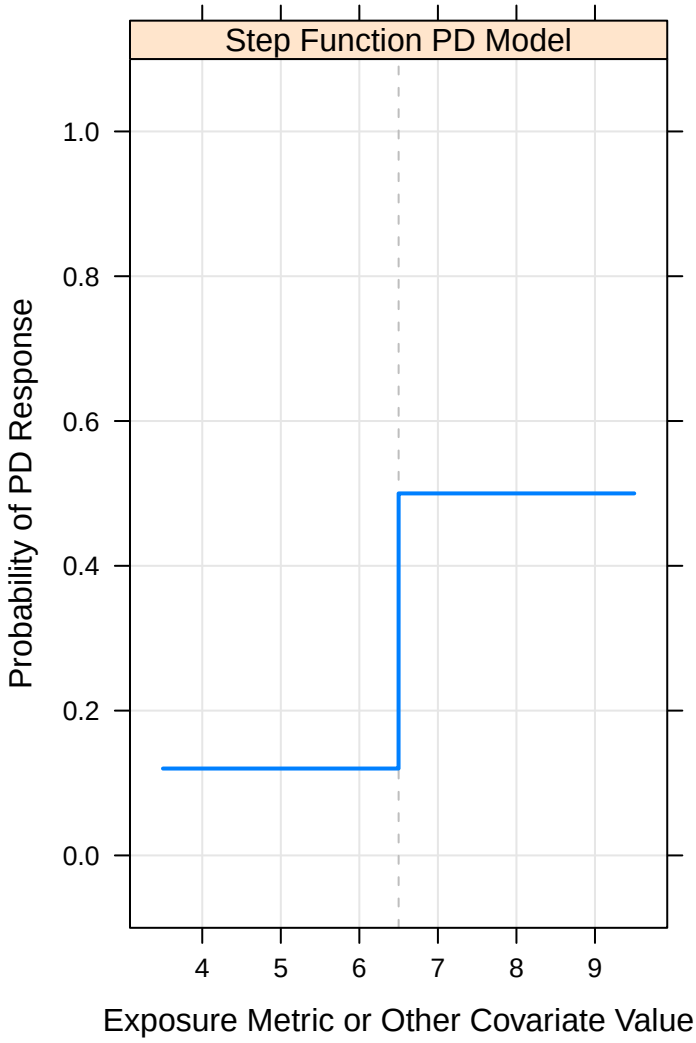

Figure 4.1: A simulated step-function exposure-response relationship (solid line) that might result from recursive partitioning analysis. There is constant probability of experiencing the PD endpoint on either side of the exposure cutpoint (dashed line).

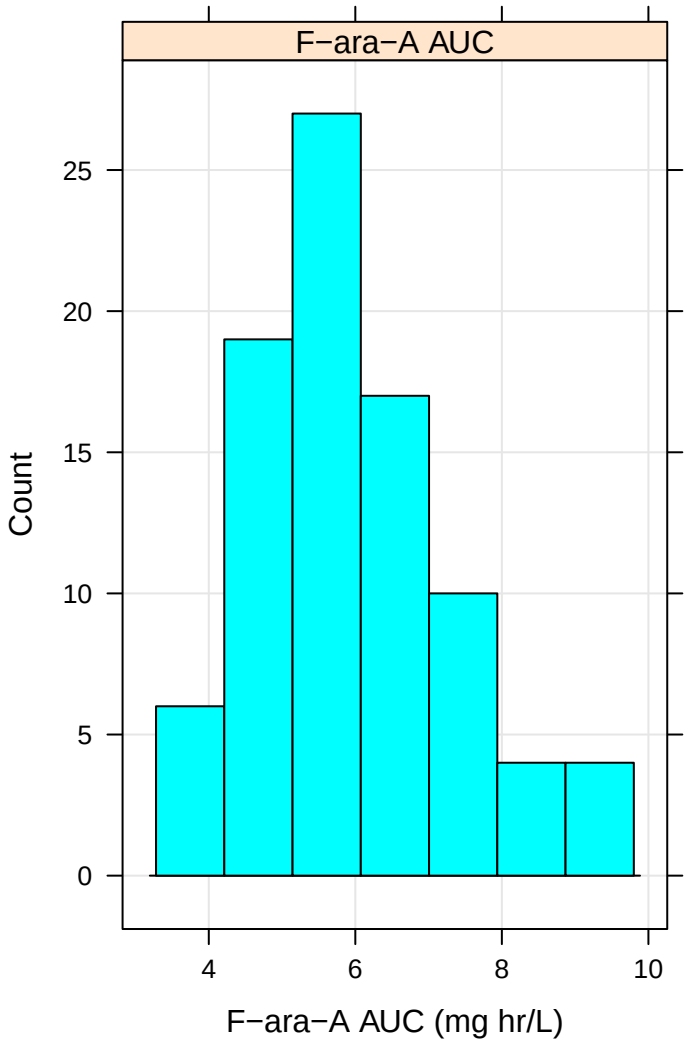

Figure 4.2: Histogram showing distribution of the F-ara-A AUCs used in exposure response modeling. In each of the 87 patients in the study, AUC was calculated as the fludarabine dose divided by the individual's estimate of clearance. Individual post-hoc clearance values from the final NONMEM PK run were used to calculate AUC. As noted in the text, AUC is $\int_0^\infty C(t)dt$

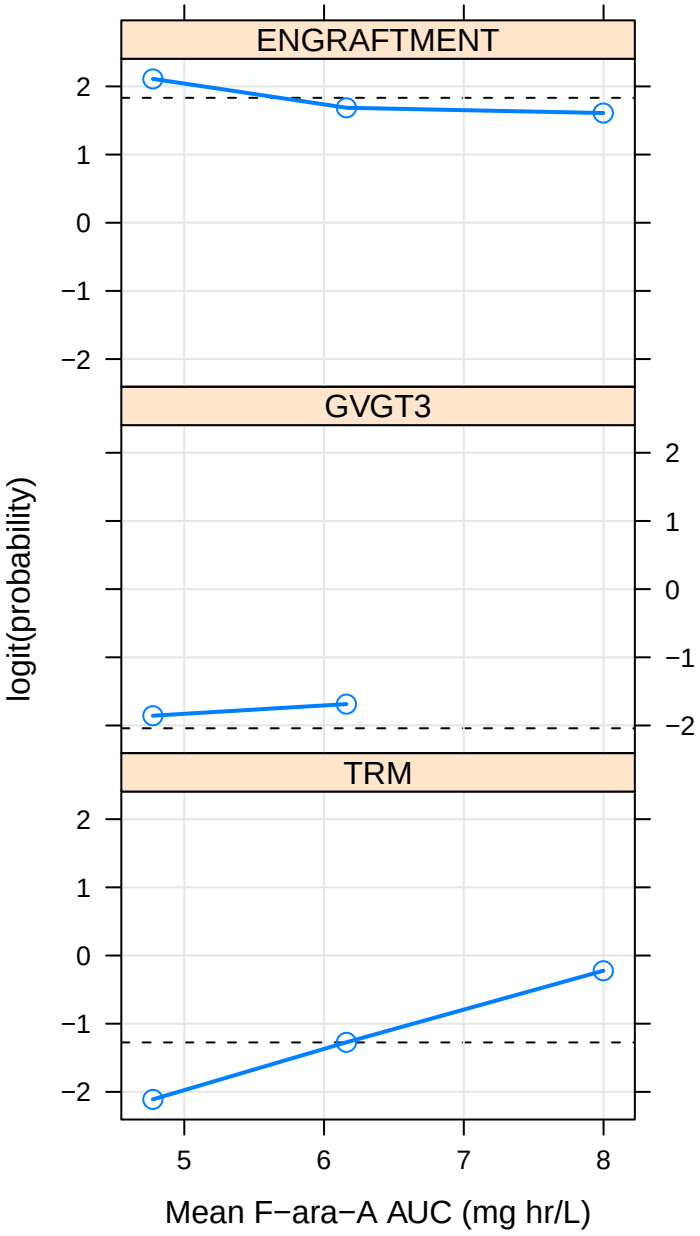

Figure 4.3: Logit-scale success probabilities versus F-ara-A exposure. PD event data were binned as described in table 4.1 and 4.2 and plotted versus F-ara-A AUC. Success probabilities are expressed on the logit scale. Shown are probabilities for engraftment (top panel), grade III or IV acute GVHD (middle panel), and treatment-related mortality (bottom panel). In each panel, a dotted horizontal line indicates the model-independent logit-scale success probability averaged over all F-ara-A AUCs for that endpoint.

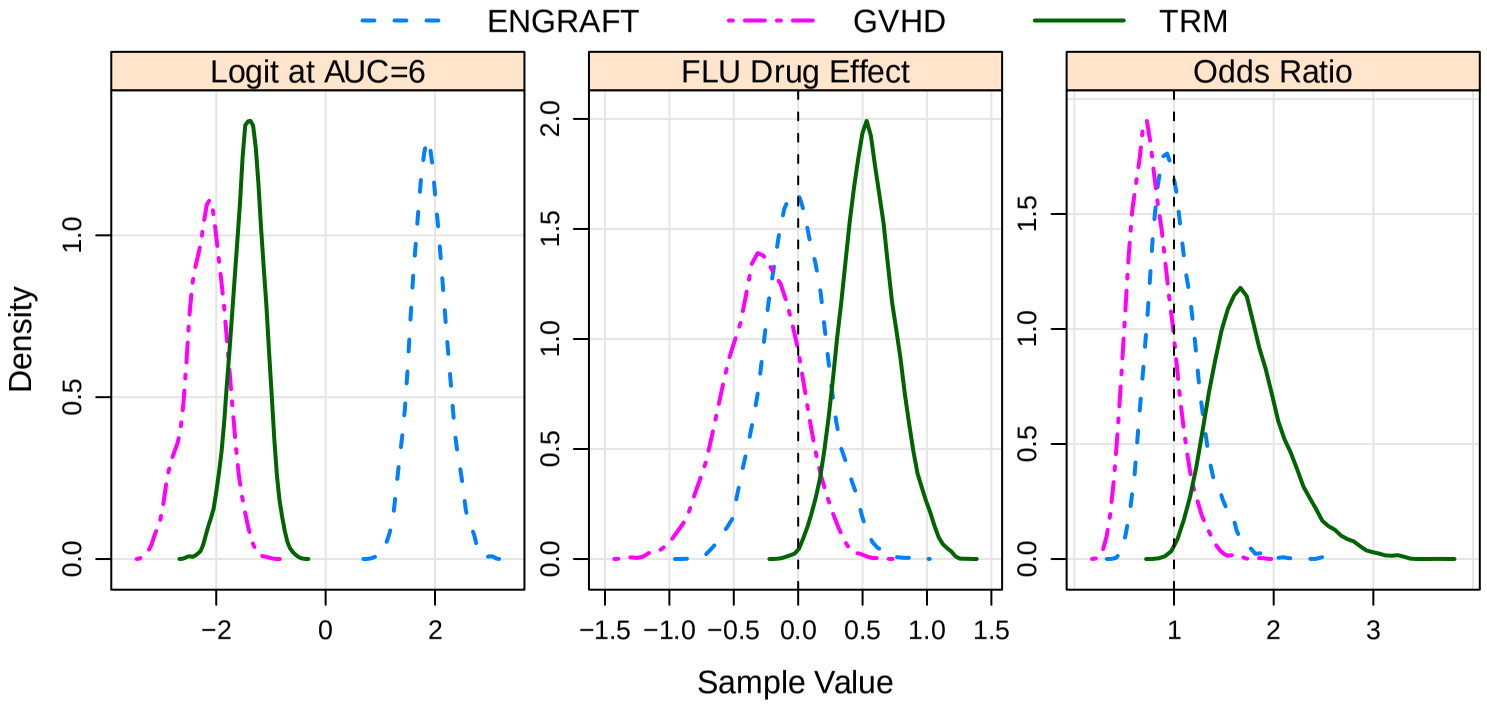

Figure 4.4: Density plots showing samples from posterior distributions of pharmacodynmic model parameters. Samples for intercept (β_0, left panel), F-ara-A slope (β_1, middle), and odds ($exp(\beta_1)$, right) are shown for engraftment, acute GVHD grade III or IV and treatment-related mortality. Dashed vertical lines are placed at 0 and 1 in the middle and right panels, respectively, to indicate the no-effect values for the F-ara-A AUC effect.

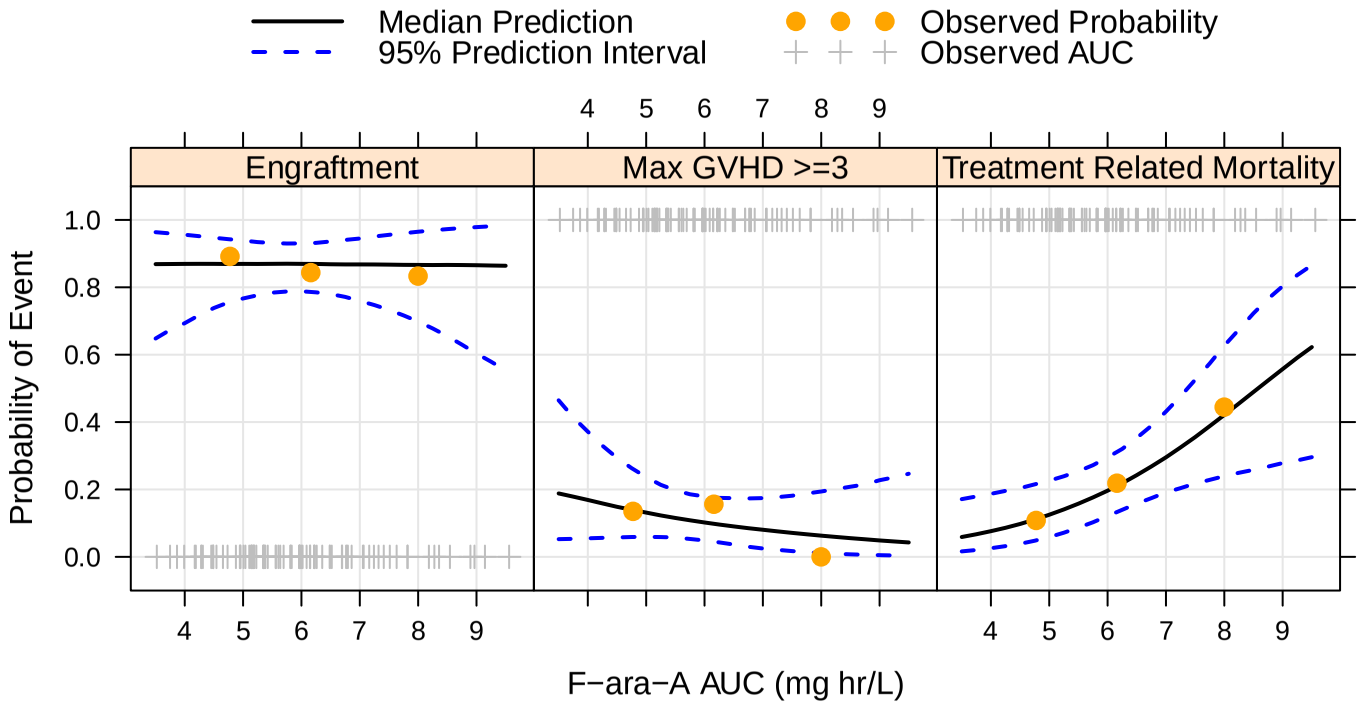

Figure 4.5: Posterior predictive check for logistic regression pharamcodynamic modeling. Engraftment endpoint is shown in the left panel, acute GVHD (grade III or IV) is shown in the middle panel, and treatment-related mortality in the right panel. In each panel, median (solid line) and 95% prediction interval (dashed lines) of posterior predictive success probabilities are plotted versus F-ara-A AUCs across the observed range. Dots indicate model-independent probabilities (table 4.1) and crosses show observed F-ara-A AUCs.

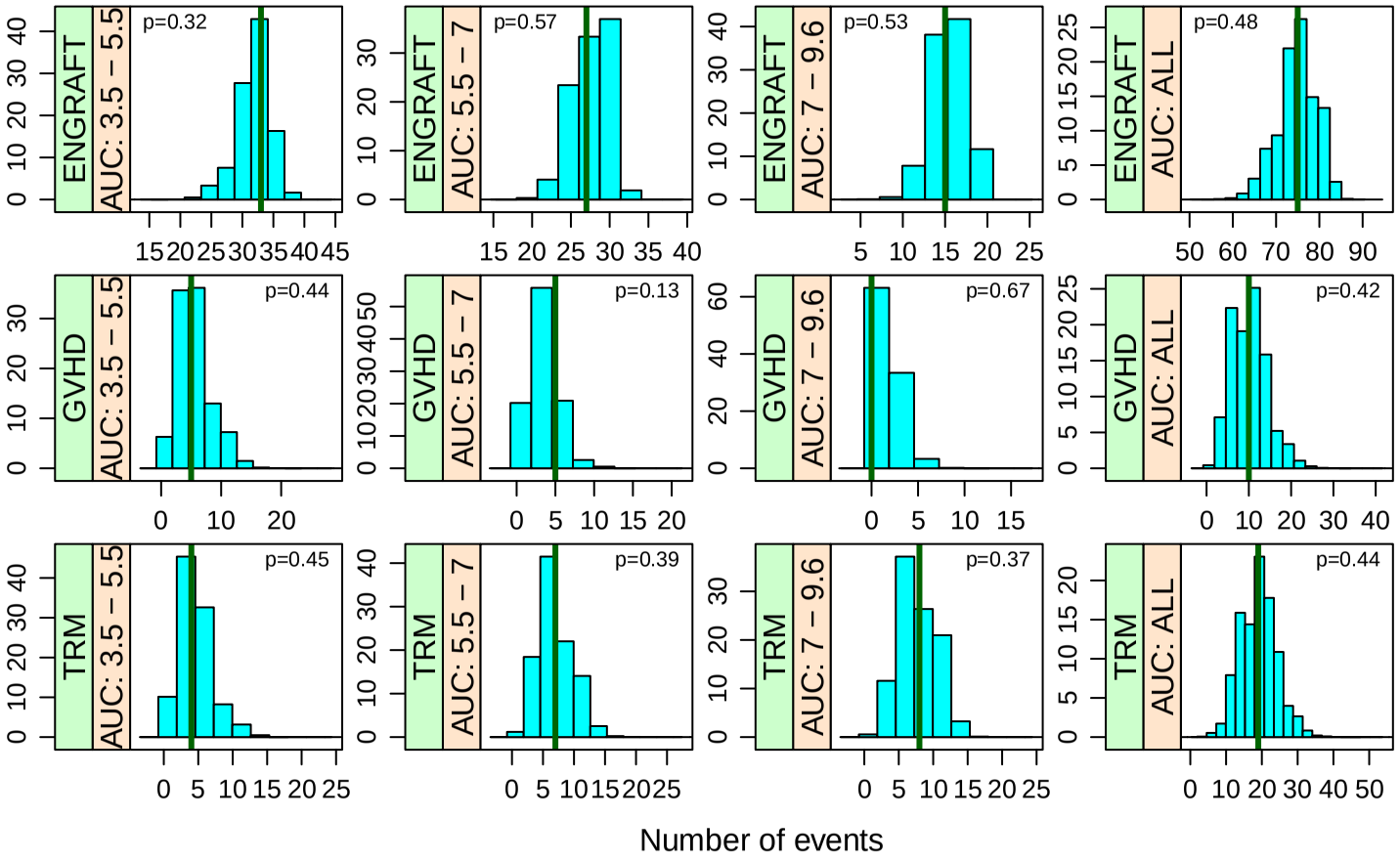

Figure 4.6: Histograms show distributions of simulated event numbers and vertical spikes show number of events observed in the dataset. The left three columns show event subtotals conditional on F-ara-A AUC. The right column shows totals averaged across all AUCs. p^{ppc} is shown in the upper corner.

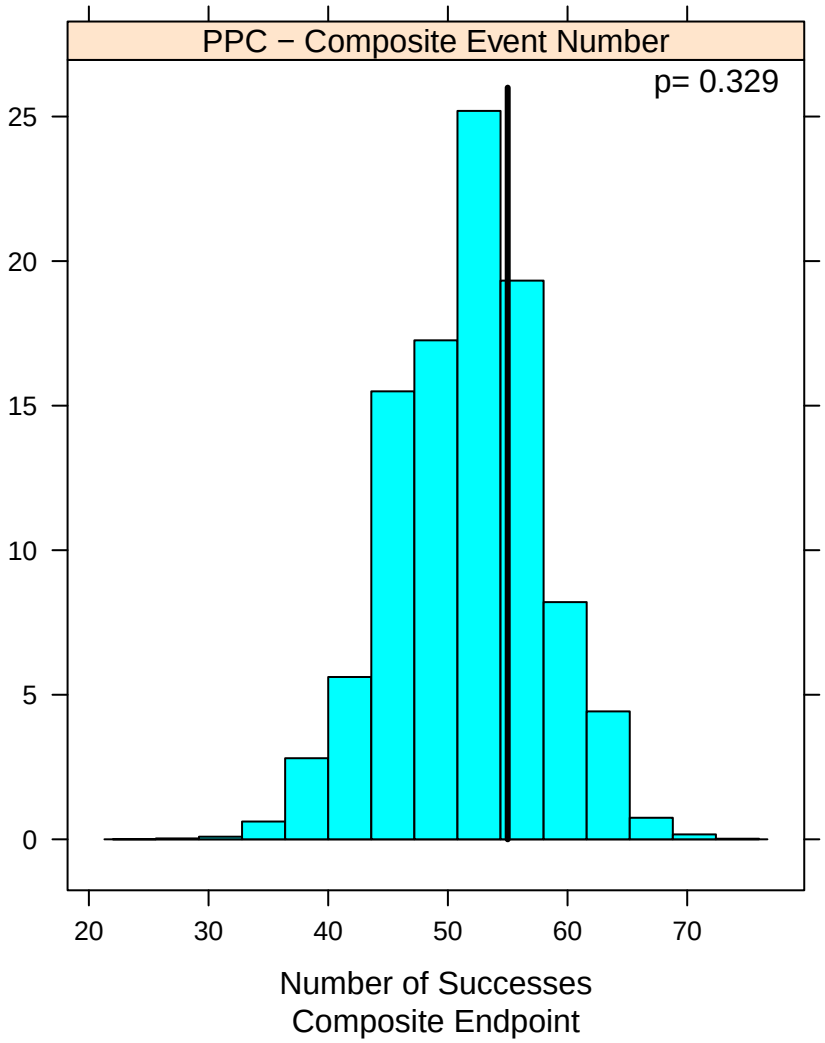

Figure 4.7: Histogram showing distribution of simulated numbers of a composite PD endpoint that considers ENGRAFT, TRM, and GVHD (equation 4.11). The vertical spike indicates the number of observed composite events. p^{ppc} is sown in the upper corner.

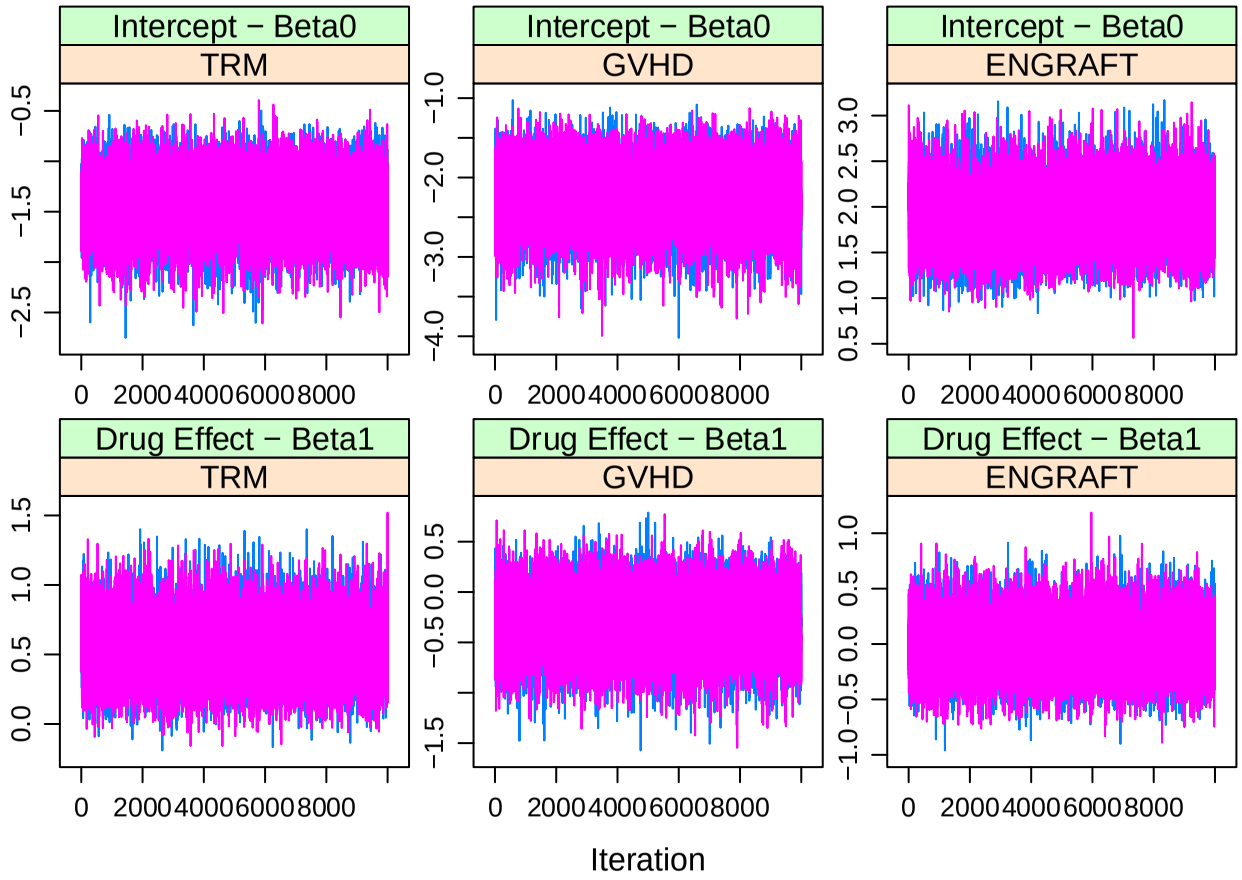

Figure 4.8: Sample trace plots for intercept (β_0, top) and fludarabine drug effect (β_1, bottom). Parameter samples for TRM, GVHD, and ENGRAFT PD models are shown in the left, middle, and right columns, respectively.

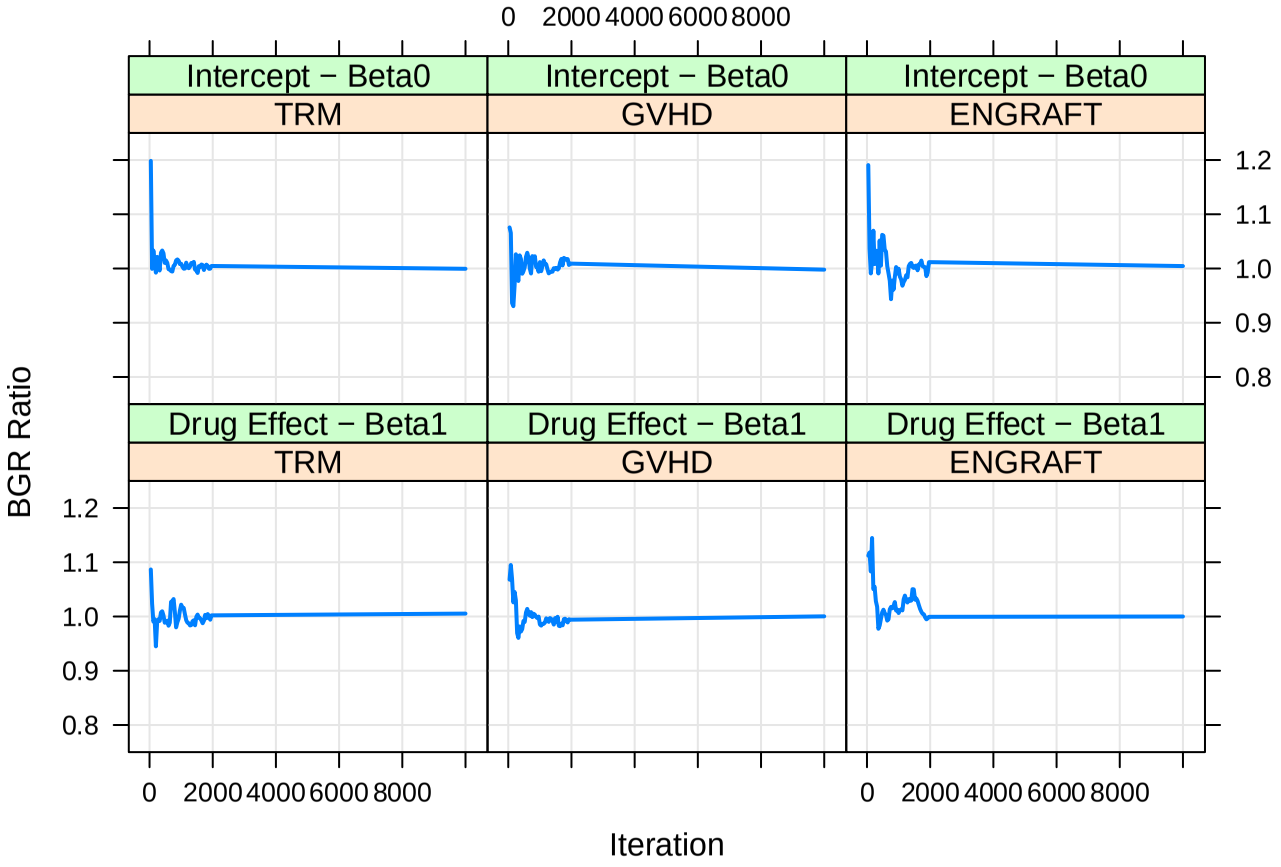

Figure 4.9: BGR statistics for intercept (β_0, top) and fludarabine drug effect (β_1, bottom) are plotted versus iteration number. BGR numbers for TRM, GVHD, and ENGRAFT PD models are shown in the left, middle, and right columns, respectively.

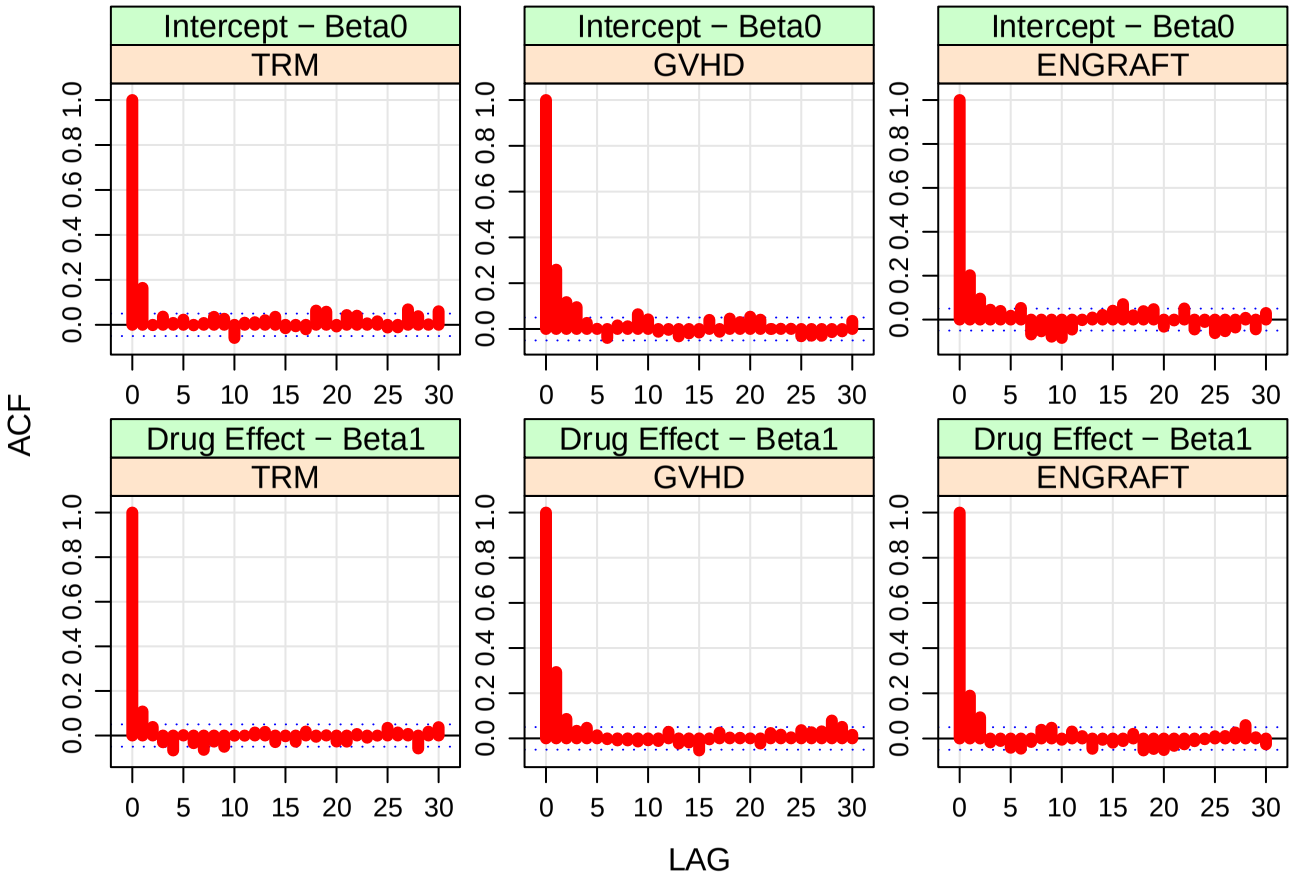

Figure 4.10: Sample autocorrelation for intercept (β_0, top row) and fludarabine drug effect (β_1, bottom row) parameters in logistic regression modeling of TRM (left column), GVHD (middle column), and ENGRAFT (right column) endpoints.

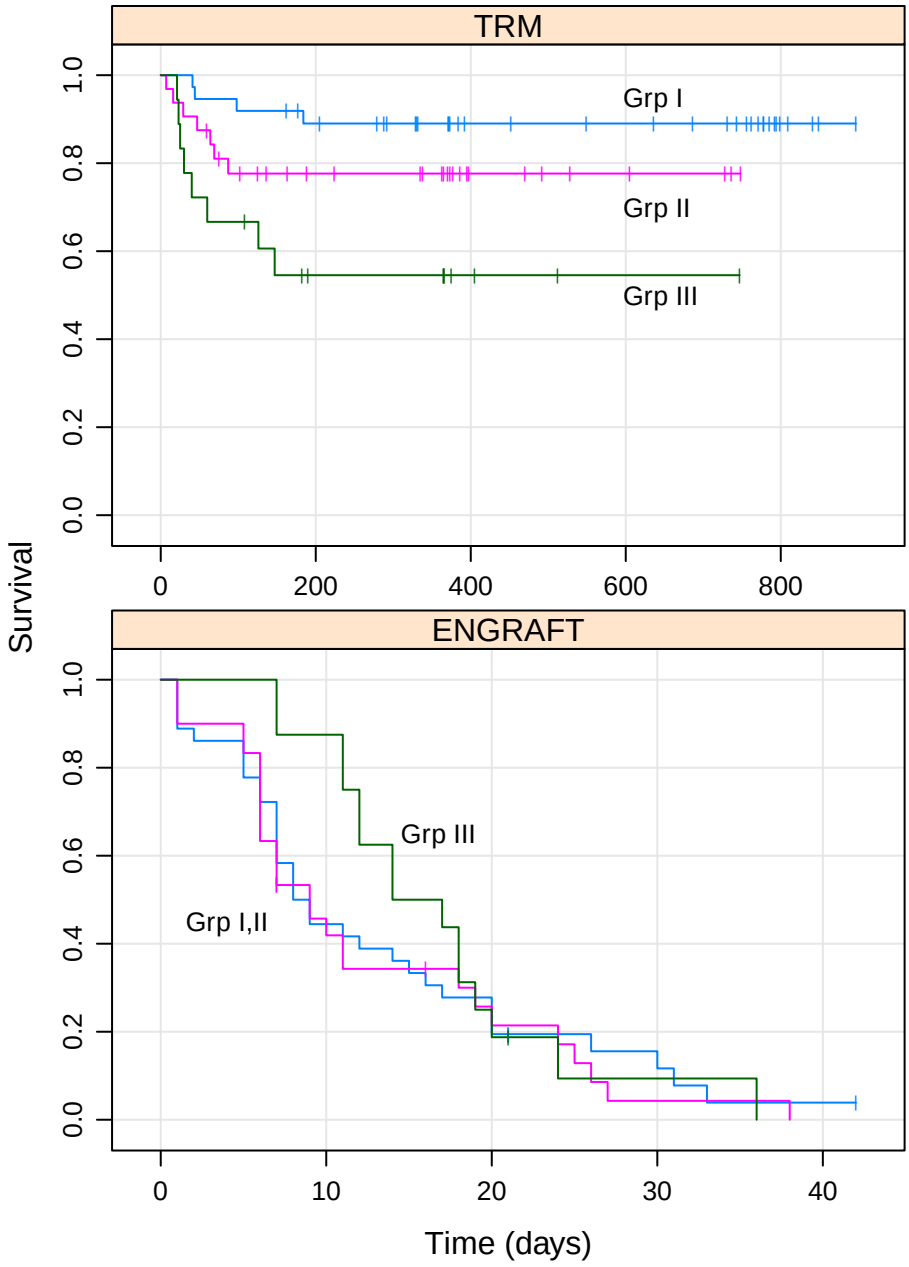

Figure 4.11: Kaplan-Meier plots for treatment-related mortality (top) and engraftment (bottom) PD endpoints. Subjects were split into three F-ara-A exposure groups according to table 4.1 and survival curves were generated using `survfit` function in R. Grp I-III indicate the lowest to highest exposure groups, respectively.

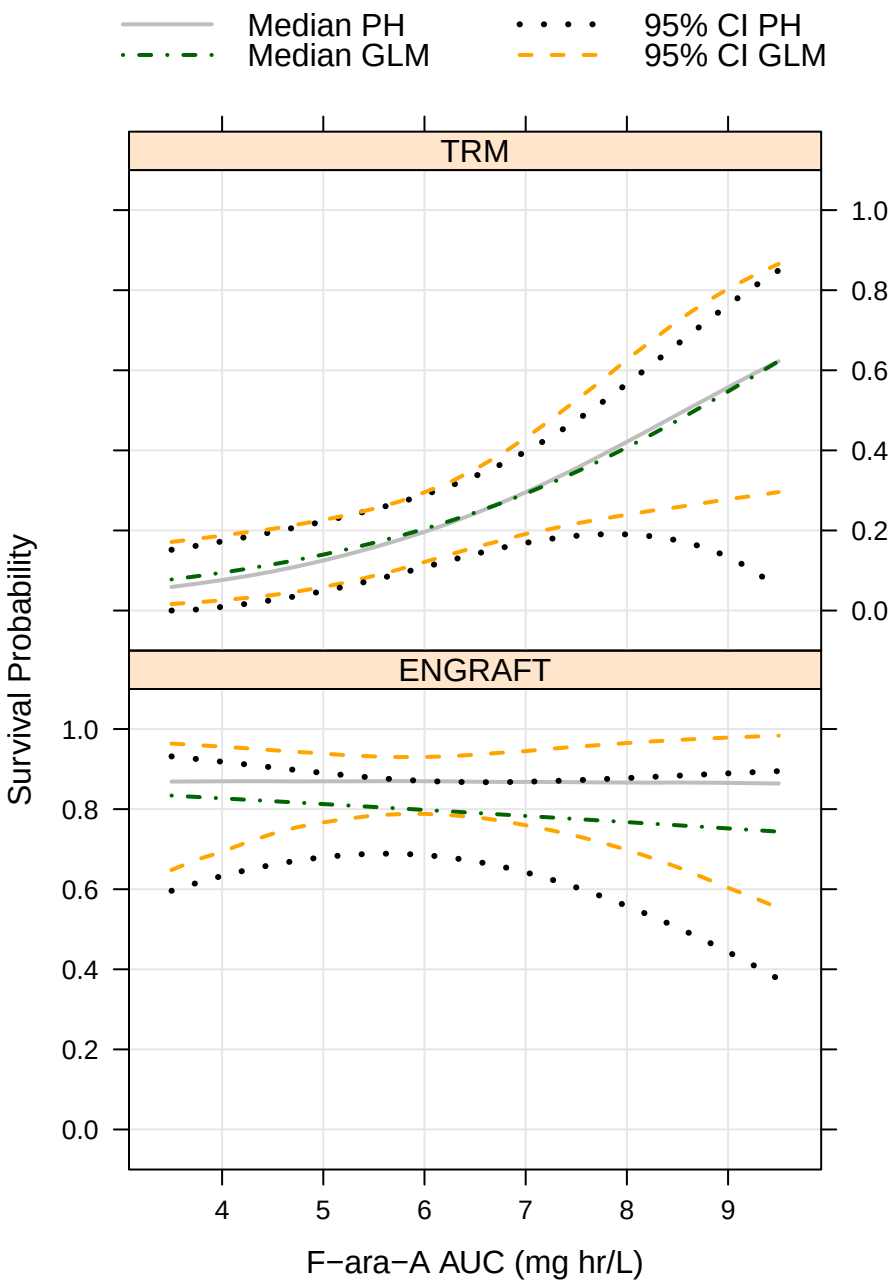

Figure 4.12: Exposure response model for treatment-related mortality (top panel) and engraftment (bottom panel) endpoints estimated using Cox proportional hazards (PH) regression and generalized linear models (GLM). Median and 95% prediction intervals for both methods are shown. PH regression survival probabilities are evaluated at 42 and 180 days for ENGRAFT and TRM, respectively.

Table 4.1: Pharmacodynamic Event Summary

Endpoint	Group	AUC Mean	AUC Range	N	Y	Prob	Logit
TRM	All	5.95	3.5 - 9.6	87	19	0.22	-1.28
	I	4.77	3.5 - 5.4	37	4	0.11	-2.11
	II	6.2	5.6 - 6.9	32	7	0.22	-1.27
	III	8	7.0 - 9.6	18	8	0.44	-0.223
GVHD	All	5.95	3.5 - 9.6	87	10	0.11	-2.04
	I	4.77	3.5 - 5.4	37	5	0.14	-1.86
	II	6.2	5.6 - 6.9	32	5	0.16	-1.69
	III	8	7.0 - 9.6	18	0	0	NA
ENGRAFT	All	5.95	3.5 - 9.6	87	72	0.83	1.57
	I	4.77	3.5 - 5.4	37	32	0.86	1.86
	II	6.2	5.6 - 6.9	32	25	0.78	1.27
	III	8	7.0 - 9.6	18	15	0.83	1.61

Table 4.2: Pharmacodynamic Event Summary

Group	AUC Mean	AUC Range	Endpoint	N	Y	Prob	Logit
All	5.95	3.5 - 9.6	TRM	87	72	0.22	-1.28
			GVHD	87	10	0.11	-2.04
			ENGRAFT	87	19	0.83	1.57
Group I	4.77	3.5 - 5.4	TRM	37	4	0.11	-2.11
			GVHD	37	5	0.14	-1.86
			ENGRAFT	37	32	0.86	1.86
Group II	6.2	5.6 - 6.9	TRM	32	7	0.22	-1.27
			GVHD	32	5	0.16	-1.69
			ENGRAFT	32	25	0.78	1.27
Group III	8	7.0 - 9.6	TRM	18	8	0.44	-0.223
			GVHD	18	0	0	NA
			ENGRAFT	18	15	0.83	-1.61

Table 4.3: Exposure-Response Model Parameter Estimates

Outcome	Intercept (β_0)		Slope (β_1)		Odds	
ENGRAFT	1.89	[1.3 , 2.54]	-0.019	[-0.48 , 0.46]	0.98	[0.62 , 1.6]
GVHD	-2.18	[-3 , -1.5]	-0.27	[-0.89 , 0.25]	0.77	[0.41 , 1.28]
TRM	-1.41	[-2 , -0.87]	0.55	[0.15 , 0.97]	1.73	[1.18 , 2.63]

Table 4.4: Exposure Response Model Selection

Endpoint	Link	pD Full	pD Reduced	DIC Full	DIC Reduced
TRM	Logit	2.010	1.016	88.02	93.36
	Probit	1.954	1.029	87.69	93.39
	Cloglog	1.911	1.040	88.27	93.42
GVHD	Logit	1.997	0.9841	65.34	64.05
	Probit	1.981	1.0280	65.17	64.14
	Cloglong	1.893	0.9709	65.19	64.06
ENGRAFT	Logit	2.018	1.005	84.01	82.00
	Probit	1.929	0.9876	83.80	81.97
	Cloglog	1.913	0.9848	83.77	81.96

Table 4.5: GLM Saturated Model Estimates

Endpoint	Parameter	MEAN	95% CI
ENGRAFT	β_0	3.56	[0.772 , 7]
	β_{AUC}	-0.01	[-0.452 , 0.51]
	β_{ATG}	-1.02	[-2.61 , 0.556]
	β_{AGE}	0.25	[-1.36 , 1.91]
	β_{CORD}	-2.00	[-4.73 , 0.193]
	β_{COMOR}	0.72	[-1.48 , 2.64]
GVHD	β_0	-0.54	[-3.52 , 1.83]
	β_{AUC}	-0.23	[-0.903 , 0.32]
	β_{ATG}	-0.83	[-2.6 , 0.739]
	β_{AGE}	-0.51	[-2.23 , 1.13]
	β_{CORD}	-0.62	[-2.14 , 0.882]
	β_{COMOR}	-1.03	[-3.03 , 1.19]
TRM	β_0	-1.09	[-3.36 , 1.11]
	β_{AUC}	0.67	[0.22 , 1.15]
	β_{ATG}	1.91	[0.584 , 3.4]
	β_{AGE}	-0.48	[-1.81 , 0.857]
	β_{CORD}	-0.43	[-1.86 , 1.05]
	β_{COMOR}	-1.08	[-2.99 , 0.918]

Table 4.6: Survival Analysis - AUC Continuous

Endpoint	$\exp(\beta_1)$	95% CI	Wald Stat	p-value
TRM	1.51	[1.13 , 2.03]	7.71	0.00775
ENGRAFT	0.996	[0.84 , 1.18]	0	0.9623

Table 4.7: Survival Analysis - AUC Categorical

Endpoint	AUC GRP	$\exp(\beta_{(.)})$	95% CI	z	p-value
TRM	II	2.34	[0.685 , 8.01]	1.36	0.17480
	III	5.35	[1.61 , 17.8]	2.733	0.00628
ENGRAFT	II	1.11	[0.651 , 1.88]	0.370	0.711
	III	0.887	[0.476 , 1.65]	-0.379	0.705

Chapter 5

Optimization of fludarabine doses in HCT

5.1 Introduction

Chapters 3 and 4 reported the results of models describing the pharmacokinetics and pharmacodynamics, respectively, of FLU in HCT. In the population PK model, F-ara-A concentrations after 40 mg/m^2 FLU were well-described by clearance and volume of distribution parameters in a 3-compartment model. The typical value of CL in the population varied strongly with the patient's CLCR. In the PD model, the probability of certain HCT outcomes were found to systematically vary with F-ara-A exposures (AUC). Most notably, the probability

of treatment-related mortality greatly increased and probability of severe graft-versus-host disease decreased at higher F-ara-A exposures. The probability of neutrophil engraftment seemed constant over the observed AUC range of 3.5 to 9.5 $mg{\cdot}hr{\cdot}L^{-1}$.

Thus, the PK model describes relationships between patient-specific covariates (including FLU dose) and F-ara-A exposures. The PD model describes the relationship between F-ara-A exposures and therapeutic outcomes. Clinical decisions about selecting FLU doses in HCT could be made with a quantitative basis if the PK and PD models were linked to provide information about the relationship between FLU dose and PD outcomes.

The objective of work presented in this chapter is to leverage PK and PD modeling results to optimize FLU doses in HCT.

5.2 Methods

The general method for simulation-based FLU dose optimization is shown in figure 5.1. F-ara-A PK and PD models were used to simulate exposures (PK) and resulting predicted outcome probabilities (PD) across different FLU doses and in different patient populations. The probability of achieving a certain PD criteria given the FLU dose and patient-specific characteristics was calculated through

integration of the posterior predictive distribution for each PD outcome success probability. Optimization of FLU doses simultaneously across both positive (ENGRAFT) and negative (TRM and GVHD) outcomes was done by calculating an overall utility index, again conditional on dose and covariate. Description of the optimization details are as follows.

5.2.1 PK Simulation

F-ara-A exposures were simulated from the population PK model presented in chapter 3. Based on PD modeling results, simulations focused on AUC, calculated by:

$$AUC_i^* = \frac{DOSE}{CL_i^*} = \frac{DOSE}{TVCL \cdot e^{\eta^*_{CL,i}}} \tag{5.1}$$

where $\eta_i^* \sim N(0, \omega^2_{CL})$ and the AUC_i^* and CL_i^* indicate exposure and clearances, respectively, simulated in the ith *in silico* individual. Population clearances were calculated according to equation 3.9:

$$TVCL = \theta_{COMOR} \cdot \left[\theta_{CLnr} + \theta_{CLr} \cdot \frac{RF}{RF_{std}}\right] \cdot \left[\frac{WGT}{WGT_{std}}\right]^{0.75} \tag{5.2}$$

Several FLU doses of interest were explored for simulated patients in a variety

of covariate combinations. Weight was fixed to 80 kg and height fixed to 170 cm so that BSA in the simulation (1.91 m^2) was close to the typical BSA in the study population (1.93 m^2). To explore the effect of changes in renal function on dose optimization, subjects were simulated with CLCR ranging from 50 to 150 ml/min. To evaluate the effect of comorbidity score (COMOR) on dosing decisions, subjects with COMOR = 0/1 (reference clearance) or COMOR $\geq$ 2 (10% reduction in clearance) were simulated at several CLCR. In all other simulations, patients were simulated with COMOR $\geq$ 2, the majority ($\sim$75%) of the observed study population. Because doses were always normalized to BSA, the influence of weight in dose optimization was not addressed and was kept fixed at 80 kg for all simulations. Final model parameter values are shown in table 3.2. For the purposes of the dose-optimization simulations, PK model parameters were assumed to be fixed and known without uncertainty.

5.2.2 PD Simulation

Simulated F-ara-A AUC were used for posterior prediction of each PD outcome probability (π_{TRM}, π_{GVHD}, and $\pi_{ENGRAFT}$). Consistent with the Bayesian estimation approach, PD model parameters were not fixed, known quantities, but rather random draws from their respective posterior predictive distributions. For a given PD outcome type, the posterior predictive success probability (π_i^*) as a

function of model parameter samples (β_0^*, β_1^*) and a simulated exposure (AUC_i^*) was:

$$\pi_i^* = \frac{exp[\beta_0^* + \beta_1^*(AUC_i^* - 6)]}{1 + exp[\beta_0^* + \beta_1^*(AUC_i^* - 6)]} \tag{5.3}$$

Thus, PD success probabilities conditional on exposures are averaged over the posterior uncertainty of model parameters. Section 4.3 and figure 4.5 addressed the posterior predictive checking of the generalized linear PD models.

5.2.3 Optimization on Individual PD Outcomes

In order to optimize doses with respect to individual outcomes, PD criteria were set to define a "successful" dose. In general, a FLU dose was considered to be successful if it was associated with a high probability of neutrophil engraftment, but low probability of serious graft-versus-host disease and treatment-related mortality. Specifically, the FLU dose was successful with respect to ENGRAFT endpoint if $\pi_{ENGRAFT}^* > 80\%$. The dose was successful with respect to the TRM endpoint if $\pi_{TRM}^* < 20\%$. Finally, a successful dose with respect to the GVHD endpoint had $\pi_{GVHD}^* < 30\%$. According to these criteria, TRM and ENGRAFT endpoints have equal "weight" in the optimization, with less weight on the GVHD endpoint. Optimizations were also carried out with equal weight across all endpoints. That

is weights for π^*_{TRM} and $\pi^*_{ENGRAFT}$ were as above, but π^*_{GVHD} had to be $< 20\%$ for the dose to be successful.

Because Bayesian estimation and prediction allows direct probability statements about the uncertainty of model parameters and predictions, the probabilities that the success criteria are met may be obtained through integration of the posterior predictive distributions of π^*_i for each outcome, covariate, and dose combination. Figure 5.2 shows an example posterior predictive distribution for π^*_i and the tail area indicating the probability that π^*_i is greater than a critical value that defines a successful dose [1] . The integration was performed by dividing the number of posterior-predictive π^*_i samples meeting the criteria by the total number of samples drawn. A BUGS code section that does the posterior predictive simulation for dose optimization is shown in the Appendix (A).

5.2.4 Utility Index

In addition to optimizing on individual PD criteria, it was desired to find a single dose that optimized across multiple PD endpoints. Thus, there is potentially one

[1] The probability of a probability is admittedly confusing notation. A note of clarification: As noted in section 4.2.3 and 4.2.4, π^*_i is a posterior predictive success probability for a binary PD event (e.g. TRM or no TRM). The probability that π^*_i exceeds some value of interest is obtainable as a consequence of using a Bayesian approach. That is, π^*_i isn't known as a fixed estimate (as in a frequentist approach), but rather as probability distribution describing the likelihood that π^*_i takes certain values. Given the data and the dose, some values for π^*_i are more likely than others and the posterior predictive distribution can be integrated to find the probability that π^*_i lies in a certain range of values of interest (see figure 5.2). This tail-area probability is referred to in this chapter as the *dose success probability*.

dose that simultaneously balances the benefit of ENGRAFT versus the risk of TRM and GVHD. A utility index (UI) was defined as a single measure of the net benefit of a dose with respect to multiple PD endpoints. The primary utility index used for dose optimization was the product of the tail area dose success probabilities for the individual PD outcomes (see above and figure 5.2) :

$$UI_{TGE} = P_{\pi_{TRM<0.2}} \cdot P_{\pi_{GVHD<0.3/0.2}} \cdot P_{\pi_{ENGRAFT>0.8}} \tag{5.4}$$

where the TGE notation indicates that TRM, GVHD, and ENGRAFT are all considered in the utility index. Because the index is a product of probabilities, UI is constrained to the interval [0,1]. However, since the individual probabilities are potentially correlated, UI should not be considered itself to be a joint probability of individual events. For a given covariate combination, the utility index was plotted versus FLU dose. The dose that maximized the utility index was the optimized dose for that patient covariate combination. In addition to this full utility index, doses were optimized for utility indices that took into account only two of the three PD endpoints (e.g. TRM and ENGRAFT only, ignoring GVHD). Utility indices used in FLU dose optimizations are shown in table 5.1.

5.2.5 Simulation Scope

As noted above, all *in silico* patients had a weight of 80 kg, height of 170 cm, and BSA of about 1.9 m^2. Creatinine clearances were explored from 50 to 150 ml/min. FLU doses were tested from 20 to 60 mg/m^2 in 2.5 mg/m^2 increments. The effect of COMOR status in the optimizations was reflected as a 10% decrease in F-ara-A clearance for subjects with COMOR ≥ 2.

Dose success probabilities and utility indices were evaluated across all combinations of FLU doses and patient CLCRs. However, some dose-covariate combinations were clinically unrealistic and resulted in F-ara-A exposures outside of the observed range. Therefore, final conclusions about dose optimization were only made for dose-covariate combinations where at least 70% of the realized exposures were within the observed F-ara-A AUC range.

5.3 Results

Figures 5.3 and 5.4 show the distributions of F-ara-A AUC simulated from the PK covariate model for clearance. There was only one simulated AUC in the observed range for the 20 mg/m^2 dose (at CLCR= 50 ml/min) and three in range for the 25 mg/m^2 dose. Exposures from 50 and 60 mg/m^2 FLU doses were within the observed range at several levels of renal function (figure 5.4). As expected,

exposures from 40 mg/m^2 FLU covered the observed range at typical observed CLCR.

Figure 5.5 shows distributions of posterior predictive outcome success probability (π) for TRM, GVHD, and ENGRAFT endpoints. Simulated doses of 30 and 40 mg/m^2 and CLCR of 75 and 125 ml/min are shown. These plots are similar in concept to figure 5.1. Dashed lines indicate the criteria probabilities defining a "successful" FLU dose. Tail areas to the left (TRM and GVHD) or right (ENGRAFT) of the criteria points are the dose success probabilities (see below). For ENGRAFT, neither the most likely (mode) success probability nor the tail area probabilty are sensitive to changes in dose or CLCR, consistent with the relatively flat exposure-response curve shown in chapter 4. For TRM, higher doses and lower CLCR (1) increase the π_{TRM} mode and (2) decrease the critical tail-area probability, both indicating lower dose success probability. It is also shown in the GVHD panel for CLCR of 125 ml/min that the posterior mode for π_{GVHD} can change only slightly while tail-areas are changing more dramatically when doses or CLCR is changed. This change in tail-area is due to increased uncertainty in the predictions, demonstrating that both changes in point estimates as well as changes in posterior predictive uncertainty can influence the dose success probability.

Figure 5.6 shows results of dose optimization simulations for individual PD

endpoints at different levels of renal function and with ≥ 2 comorbidity score. Results are shown for simulated F-ara-A AUCs restricted to the observed exposure range as well as for all simulated exposures. The discussion here focuses on outcomes for exposures within the simulated range, except where noted. The probability of dose success with respect to GVHD increases slightly at higher FLU doses, but is generally high ($> 90\%$) independent of dose. Dose success probability with respect to ENGRAFT increases slightly over the lower dose ranges, and then begins to decline at higher doses due to increasing uncertainty in prediction at higher exposure levels (see figure 4.5). The probability of successful dose for the TRM endpoint decreases clearly over the 20 to 60 mg/m^2 range of doses used for simulation. Figure 5.7 shows the same optimization data as shown in figure 5.6, but this time grouped by PD endpoint instead of CLCR. The plot emphasizes the sensitivity of the dose optimization to TRM, but not ENGRAFT or GVHD. Also, figure 5.7 shows the increased influence of GVHD on dose success probability when weighting on GVHD is increased (pGVHD < 0.2 panel) compared to the lower weight scenario (pGVHD < 0.3 panel).

Figure 5.8 shows the results of FLU dose optimization with respect to the utility index considering dose success probabilities for all PD endpoints simultaneously (UI_{TGE}). Consistent with the prominence of TRM in individual endpoint optimizations, there is a marked decline in the utility index with increasing FLU

doses at all levels of renal function in the simulation after a plateau close to the lowest realistic dose. When all F-ara-A AUCs are considered, there is a decline in dose success probability at doses lower than the plateau due to increasing influence by the GVHD endpoint.

Figure 5.9 and 5.10 show the dose-utility profile for all utility indices used (see table 5.1). When TRM is omitted from the utility index (UI_{GE}), there is little preference for a specific dose. While the utility index is maximized at decreasing doses for decreasing CLCR, there is only a small penalty for dosing higher or lower than the maximum utility value. The pattern for utility indices that account for TRM (UI_{TGE} , UI_{TG} , UI_{TE}) are all qualitatively similar, showing rapid decline in UI with increasing dose after a plateau just above the lowest doses.

Figure 5.11 shows the effect of considering the effect of comorbidity score on F-ara-A exposure when optimizing FLU doses. At all dose-CLCR combinations, simulated clearance was on average 10% lower in *in silico* subjects with comorbidity score ≥ 2. Optimized doses were slightly lower for subjects with 0/1 comorbidity score.

Figure 5.12 shows the effect of differences in GVHD endpoint weighting on FLU dose optimization. Optimizations are shown at multiple CLCR levels. Optimized doses were slightly higher when weight on the GVHD endpoint was increased to equally weight all endpoints.

Table 5.2 shows optimized FLU doses for the different covariate combinations tested. For UI_{TGE} and COMOR ≥ 2, optimized doses decreased from 35 mg/m^2 for patients with CLCR of 150 ml/min to 20 mg/m^2 for patients with CLCR of 50 ml/min. Doses optimized while ignoring TRM (UI_{GE}) were consistently higher by 5 to 12.5 mg/m^2 and doses were slightly lower (0 to 2.5 mg/m^2) when ignoring GVHD (UI_{TE}) or ENGRAFT (UI_{TG}). Considering comorbidity score in the dose optimization changed dosing recommendations under UI_{TGE} moderately, with $\sim$2.5 mg/m^2 decrease for subjects with score ≥ 2 compared to patients with 0/1 comorbidity score. When all PD endpoints were equally weighted (by increasing dose success criteria from 0.3 to 0.2), the optimized doses under UI_{TGE} increased consistently by 2.5 mg/m^2 . Doses optimized under other indices and with equal weighting increased by 2.5 to 5 mg/m^2, except for UI_{TE} which didn't change as expected.

5.4 Discussion

One of the most important questions that clinicians face in making pharmacotherapeutic decisions is: "What is the right dose?". Many different types of patient characteristics may affect the outcome of therapy (such as age, comorbid conditions, disease history), but getting the dose right for a given patient when drug

therapy is the primary mode of treatment is of particular importance [73, 61, 74]. Doses first of all must be sufficiently high to achieve a therapeutic concentration at the site of action. But doses that lead to supratherapeutic exposures could possibly harm the patient by inducing some toxic effect. Thus a successful dose is one that balances the positive, therapeutic effects of the drug against the negative, toxic effects. The question implies the need to establish the relationship between *dose* and *outcome* in order to individualize therapy for that patient.

In chapter 3, a PK model was presented that described the relationship between patient-specific covariates (such as weight, CLCR, and FLU dose) and F-ara-A exposures. This established the link between *dose* and *exposure*. The pharmacodynamic model presented in chapter 4 provided a link between F-ara-A *exposures* and clinical *outcomes* such as treatment-related mortality and neutrophil engraftment. But, alone, these two models do little to help the clinician who is sitting at the bedside in need of guidance in selecting a *dose* that is most likely to lead to the best *outcome* for the patient.

The work in this chapter is an attempt to provide such guidance for dosing fludarabine in HCT. All available information obtained from PK and PD models were leveraged to evaluate the likelihood of successful outcomes after different FLU doses in HCT using a utility index (equation 5.4) that allowed optimization across both positive and negative drug-therapy outcomes. Other studies have

reported the use of a clinical utility index to optimize drug doses against multiple competing endpoints [75, 76, 77, 78, 79]. Ouellet et. al. [77] used a "clinical utility index" with clinician-assigned weights describing the relative desirability of positive drug therapy outcomes and undesirability of negative outcomes. The weighted negative outcomes are subtracted from the positive ones, providing the overall net benefit at a given dose. In the current work, weights are assigned to different outcomes by specifying a desired success probability (criteria) with respect to each outcome in the optimization, and the product of the probabilities that these criteria were met formed the overall utility index.

Simulation scenarios for the optimizations included many dose-covariate combinations that realized F-ara-A exposures outside of the range observed in the data set. A lack of evidence for nonlinearity in the PK model suggest that these exposures could realistically occur in patients, but caution must be used when using these extreme exposures to make outcome predictions with the PD model. The generalized linear PD model predicted a linear relationship between F-ara-A exposure and success probability (on logit scale) across the observed range (equation 4.3). Extrapolation should not be attempted because the linear PD relationship may not remain linear outside of the observed AUC range. Thus, while optimization results for these extreme exposures are shown to let the reader see more of the shape of the optimization curve, it is important to limit dose optimization

conclusions to those based on the restricted set of observed F-ara-A exposures.

The optimization results presented here provide a framework suggesting clinically relevant guidance for dosing FLU in HCT. Because a significant fraction of the FLU dose is cleared through the kidney (as F-ara-A), dose optimizations were done across a range of normal and impaired renal function (CLCR 50 to 150 ml/min). It is critical to note that all CLCR calculations were performed according to the equation published by Cockcroft and Gault [44] without modification [59]. As expected, patients with greater degrees of renal impairment required greater reductions in FLU dose (figure 5.8, table 5.2) to achieve optimized outcomes. Also, optimized doses were slightly lower for subjects with comorbidity score ≥ 2. These results could be used to provide data-driven guidance for adjusting FLU doses according to renal function status and comorbid conditions in an HCT population.

The risk of treatment-related mortality is clearly the driver for the dose optimization process presented here. This is evident by the clear increase in optimized doses under UI_{GE} compared to UI_{TGE} (table 5.2, figure 5.10) as well as the as the negligible decrease in probability of meeting dose success criteria for GVHD and ENGRAFT (figure 5.7) and UI_{GE} (figure 5.9) when doses were evaluated away from the optimum. Consistent with PD modeling, consideration of the ENGRAFT endpoint in the utility index did not significantly change dosing recommendations

across the observed range of F-ara-A AUC. This suggests that the 40 mg/m^2 FLU dose is achieving exposures that are at the top of the dose-response curve and that the maximal drug effect has been achieved. Presumably, no further dose increase will improve probability of ENGRAFT. Furthermore, doses leading to exposures in the lower end of the observed range are equally likely to result in positive ENGRAFT outcomes as are doses leading to higher exposures. Considering GVHD in the utility index was able to influence dosing recommendations, providing limited pushback against the influence of TRM. This may be the critical adjustable factor for finally deciding on a dose (e.g. allowing equal or less weight in the decision process).

The target FLU dose used in the study (40 mg/m^2) optimized UI_{TGE} for patients with a 0/1 comorbidity score and the highest levels of renal function (table 5.2). Simulations suggested lower doses, on average, across all levels of renal function and comorbidity status. About 40% of the patients in the study had an average CLCR in the 50 to 100 ml/min range and would qualify for 25 to 50% reduction in FLU dose on a mg/m^2 basis according to these optimization results. Even when GVHD was weighted equally with the other endpoints, the optimization results still suggest, on average, a dose reduction to 25 to 32.5 mg/m^2 for patients with CLCR between 75 and 125 ml/min. The overwhelming conclusions of this work is that FLU doses should be de-escalated in order to optimize

outcomes by reducing the risk of TRM.

Fludarabine doses have been optimized here under exposure-response models for three endpoints: treatment-related mortality, grade III or IV acute GVHD, and neutrophil engraftment. No attempt was made here to evaluate other outcomes of potential interest. Consistent with the current work, Long-Boyle et. al. [2] found no significant relationship between F-ara-A exposure and grade III or IV acute GVHD. But when grade II-IV acute GVHD was considered, there was a significant relationship with F-ara-A exposure. If grade II-IV acute GVHD were considered in the dose optimization as presented here, there would likely be more influence by the GVHD endpoint to keep doses higher. Furthermore, only acute GVHD was considered in this work. As mentioned in chapter 2, therapeutics that work to prevent acute GVHD also help to prevent chronic GVHD [21]. Thus, it is not clear what would happen to the chronic GVHD rate and severity if doses were severely reduced to potentially allow some GVHD. It is unknown how other endpoints not considered here like overall survival or relapse rates would affect dosing recommendations.

Finally, the optimization results presented here were derived from models estimated on a limited data set (N=87). Before implementation in routine or widespread use, the dosing guidance presented here should be evaluated and confirmed in a controlled clinical setting.

The question of correlation between PD endpoints initially raised in chapter 4 must be revisited here. PD endpoints are most-likely correlated to some degee. For example, the reliance on a graft-versus-tumor effect for successful engraftment in a nonmyeloablative preparative regimen suggests that $\pi_{ENGRAFT}$ might increase as π_{GVHD} increases. Despite this potential correlation, the utility indices used to optimize FLU doses were the products of posterior predictive tail-area probabilities arising from independent generalized linear PD models for the different PD endpoints (table 5.1). Clearly, the utility index cannot be quantitatively understood as a joint probability and the operating characteristics of this utility index are not totally known. However, two lines of reasoning suggest that the UI is not inappropriate for use in the dose-optimization.

First, posterior-predictive check presented in chapter 4 (section 4.3 and figure 4.7) evaluated the ability of the three independent GLMs (ignoring correlation in endpoints) to predict one composite endpoint reflecting all three endpoints. The independence models were able to predict the number of composite clinical success events adequately ($p^{PPC} = 0.374$). The results failed to reject the model with respect to the composite endpoint and suggest that it is not inappropriate to use these models in a joint utility index to optimize doses across all endpoints.

Second, the dose optimization results are consistent across all utility indices used. The most basic utility index presented was for the individual endpoints

presented in figure 5.7. These tail areas are true probabilities of success for the dose with respect to the individual endpoints and are the basic units for constructing the composite utility indices. Thus, the UI_{TG} dose-index profile should look like a compromise between the dose-success probability profiles for TRM and GVHD. The UI_{TGE} dose-index profile should capture the compromise between the univariate dose-success probability profiles for TRM, GVHD, and ENGRAFT all together. The patterns and optimized doses for the composite indices (figure 5.9 and table 5.2) do seem like reasonable compromises between the individual dose-success probabilities at a given dose (figure 5.7). The utility indices maintain the prominence of TRM in the optimizations keeping doses low as well as the small effect of GVHD and ENGRAFT pushing doses slightly up. Therefore, even though the exact numerical meaning of the UI is still not clear, the suggested doses seem to appropriately balance the outcomes and truly represent optimized doses.

In summary, the optimization process utilized information in population PK and PD models to make clinically-relevant suggestions for optimal FLU doses with respect to outcomes after HCT. FLU doses should be reduced to optimize outcomes and prevent TRM.

Table 5.1: FLU Dose-Optimization Utility Indices

Utility Index	Endpoints Considered	Expression
UI_{TGE}	TRM, GVHD, ENGRAFT	$P_{\pi_{TRM}<0.2} \cdot P_{\pi_{GVHD}<0.3/0.2} \cdot P_{\pi_{ENGRAFT}>0.8}$
UI_{TE}	TRM, ENGRAFT	$P_{\pi_{TRM}<0.2} \cdot P_{\pi_{ENGRAFT}>0.8}$
UI_{TG}	TRM, GVHD	$P_{\pi_{TRM}<0.2} \cdot P_{\pi_{GVHD}<0.3/0.2}$
UI_{GE}	GVHD, ENGRAFT	$P_{\pi_{GVHD}<0.3/0.2} \cdot P_{\pi_{ENGRAFT}>0.8}$

Table 5.2: Optimized Fludarabine Doses

π_{GVHD}	θ_{COMOR}	CLCR	Optimized FLU Dose (mg/m^2)			
Criterion	(*% decr*)	(*ml/min*)	UI_{TGE}	UI_{TE}	UI_{TG}	UI_{GE}
0.3	10	50	20.0	20.0	20.0	25.0
0.3	10	75	22.5	22.5	22.5	32.5
0.3	10	100	27.5	25.0	25.0	35.0
0.3	10	125	30.0	30.0	30.0	42.5
0.3	10	150	35.0	32.5	32.5	45.0
0.2	10	50	20.0	20.0	20.0	27.5
0.2	10	75	25.0	22.5	25.0	32.5
0.2	10	100	30.0	25.0	27.5	40.0
0.2	10	125	32.5	30.0	32.5	42.5
0.2	10	150	37.5	32.5	35.0	50.0
0.3	0	50	22.5	20.0	20.0	30.0
0.3	0	75	25.0	25.0	25.0	35.0
0.3	0	100	30.0	27.5	27.5	40.0
0.3	0	125	35.0	32.5	32.5	45.0
0.3	0	150	37.5	35.0	37.5	50.0

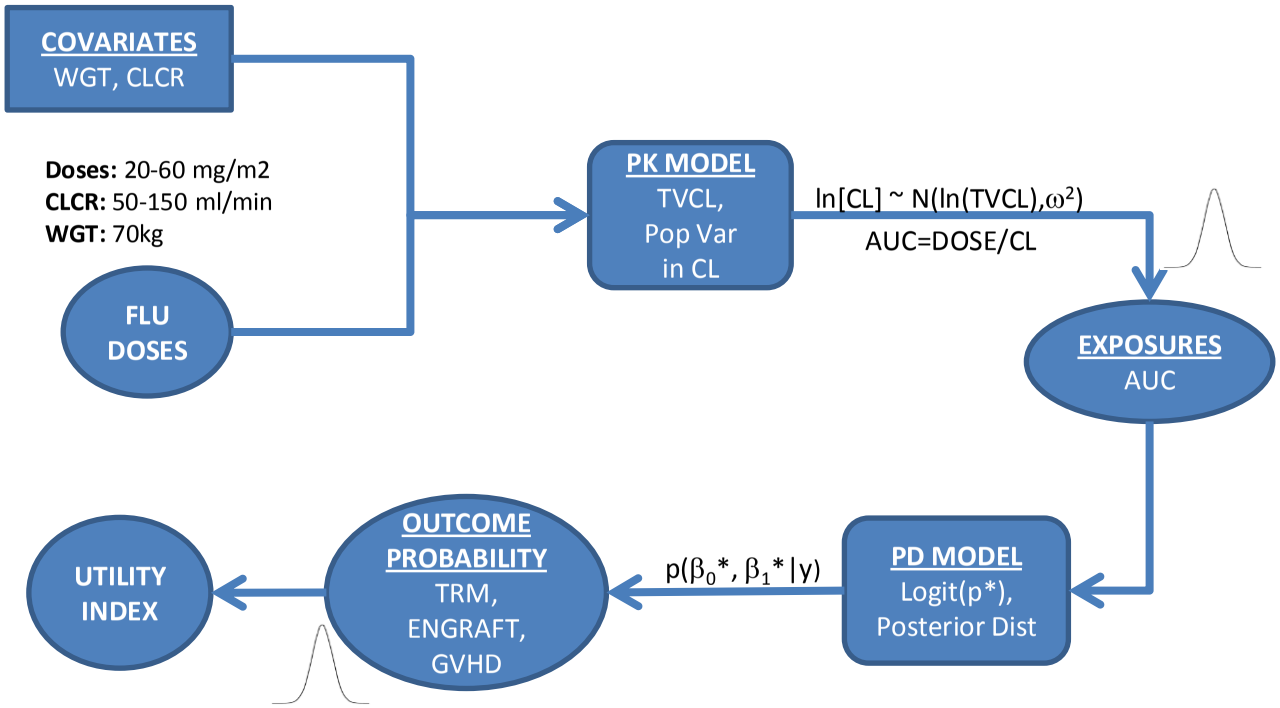

Figure 5.1: An overview of the dose-optimization procedure. Exposures are simulated via the PK model at different dose-covariate combinations. Simulated exposures are used to obtain simulated PD outcome success probabilities through the PD model. Success probabilities are converted into a utility index as a summary of the net magnitude of benefit or risk associated with the dose.

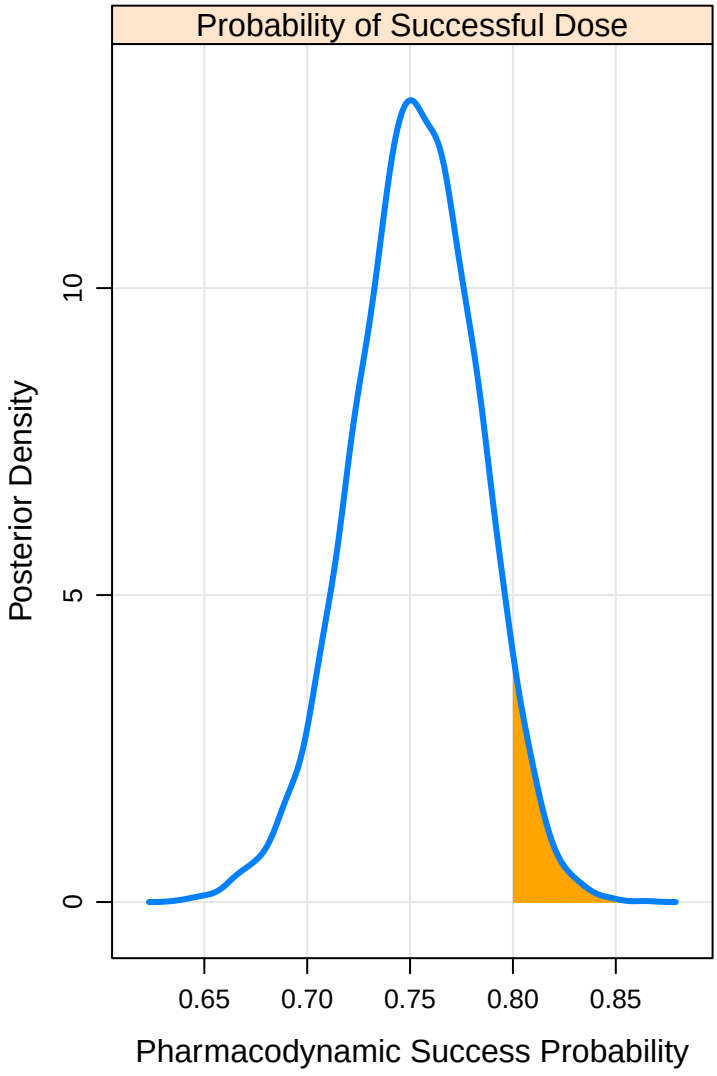

Figure 5.2: A simulated posterior-predictive distribution for a hypothetical PD outcome success probability, π_i^*, at some combination of dose and patient-specific covariates. The density is shown by the solid line. The integrated tail-area is the "dose success probabiliy" and is shown by the shaded area under the density at PD success probability greater than 0.8. It is unlikely that this dose will be successful with respect to the individual PD endpoint in these patients.

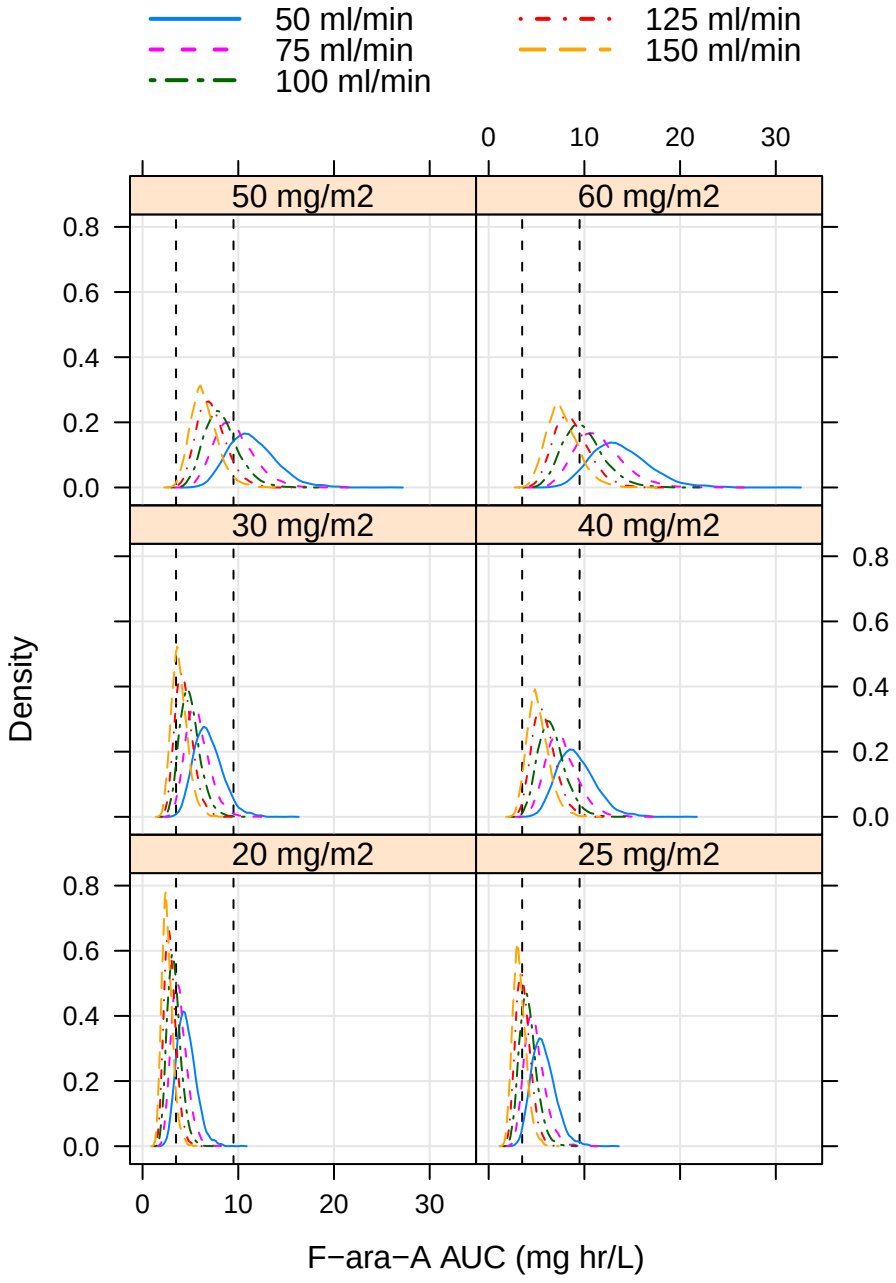

Figure 5.3: Distributions of simulated F-ara-A AUC at all combinations of FLU doses and patient CLCR. Dotted vertical lines indicate the range of observed F-ara-A AUC.

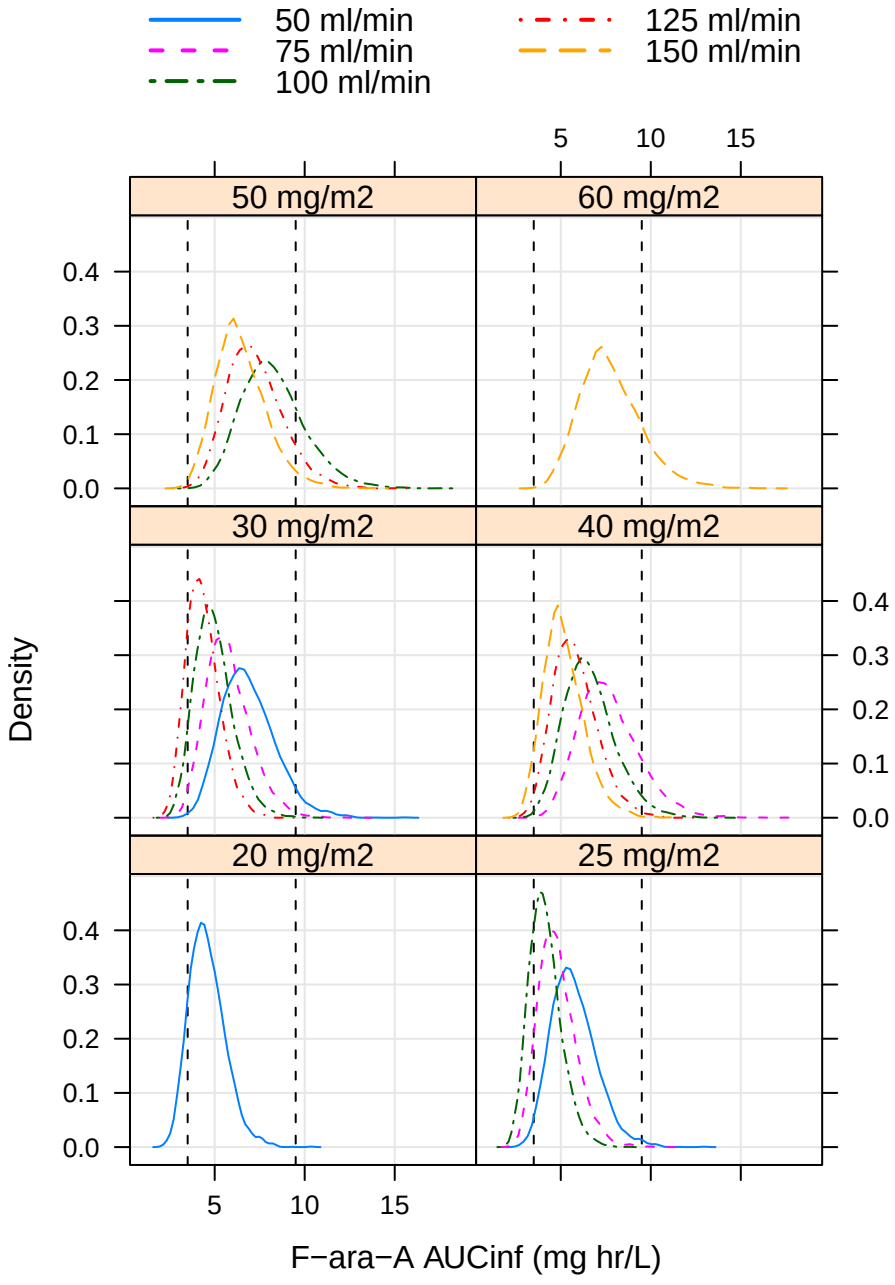

Figure 5.4: Distributions of simulated F-ara-A AUC showing only dose-CLCR combinations that were consistent with the exposure range observed in study patients, indicated by the dotted vertical lines.

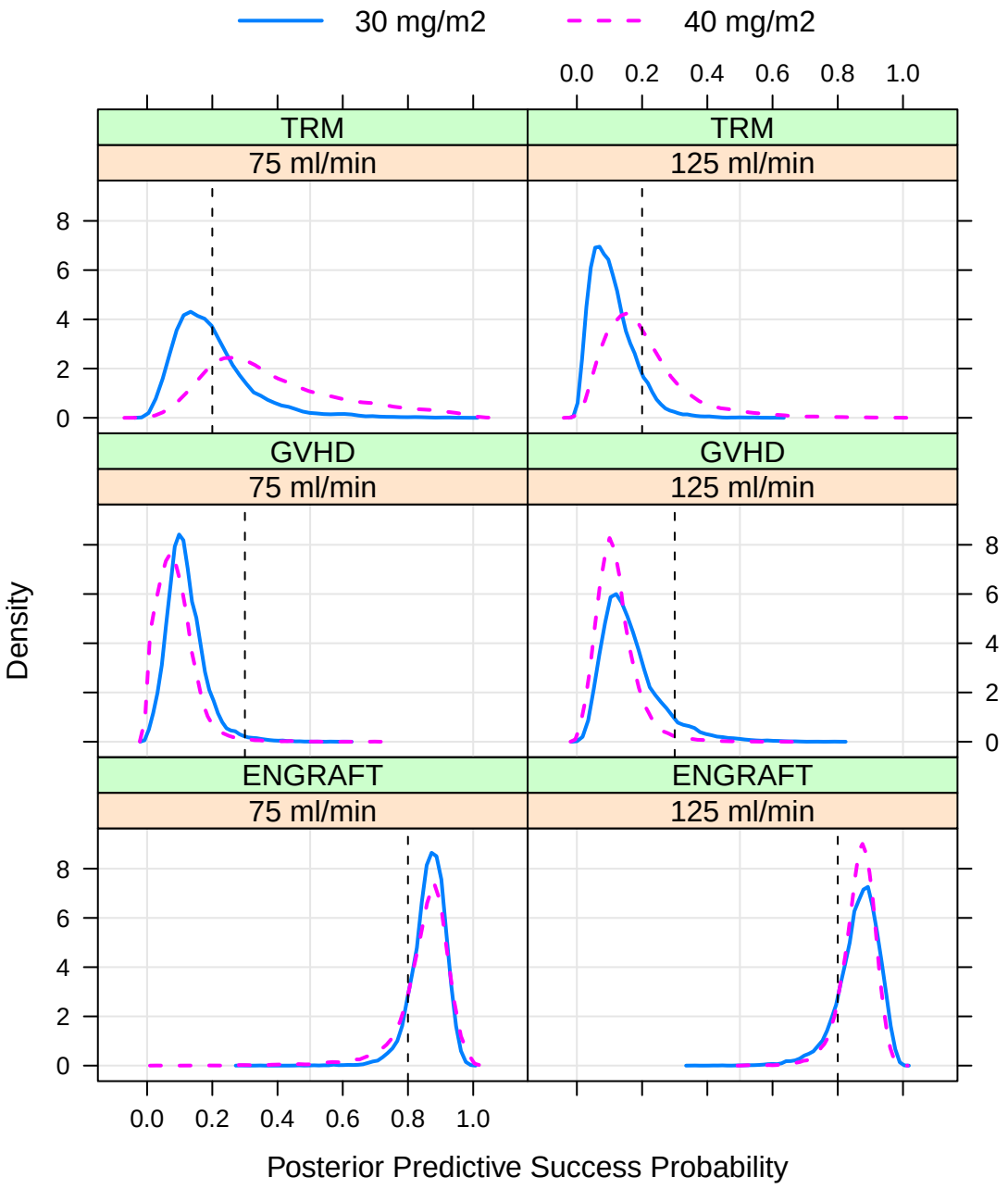

Figure 5.5: Density plots of simulated outcome success probabilities (π_{TRM}, π_{GVHD}, and $\pi_{ENGRAFT}$) at 30 and 40 mg/m^2 and CLCR 75 and 125 ml/min. Dashed vertical lines mark the PD outcome criteria defining a successful dose. These plots are similar to figure 5.1, but use data simulated from the PK/PD model.

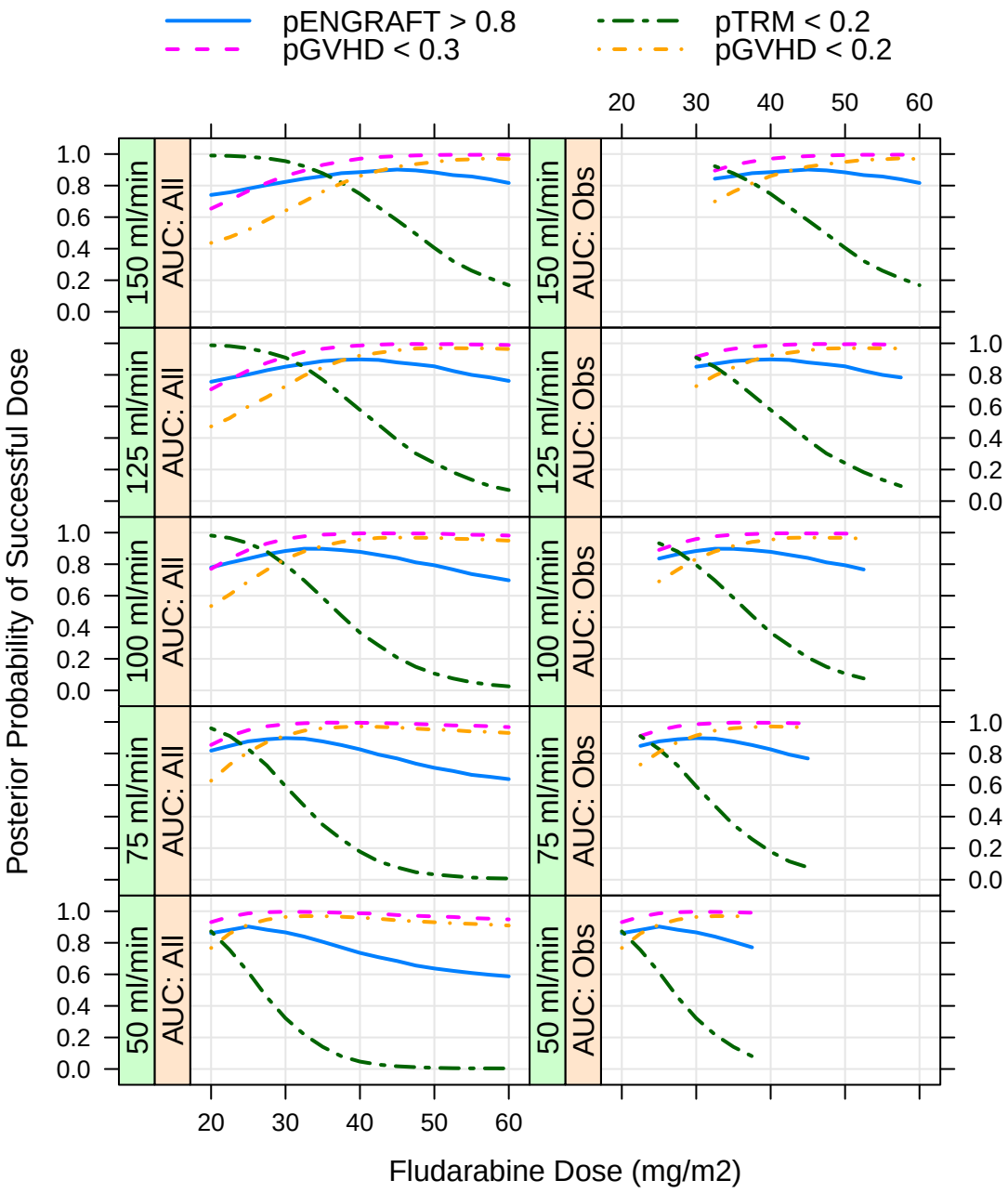

Figure 5.6: Posterior-predictive dose success probability versus FLU dose, by CLCR (50 to 125 ml/min [bottom to top]). The left column shows results for F-ara-A AUC restricted to the observed range and the right column shows results at all simulated AUCs.

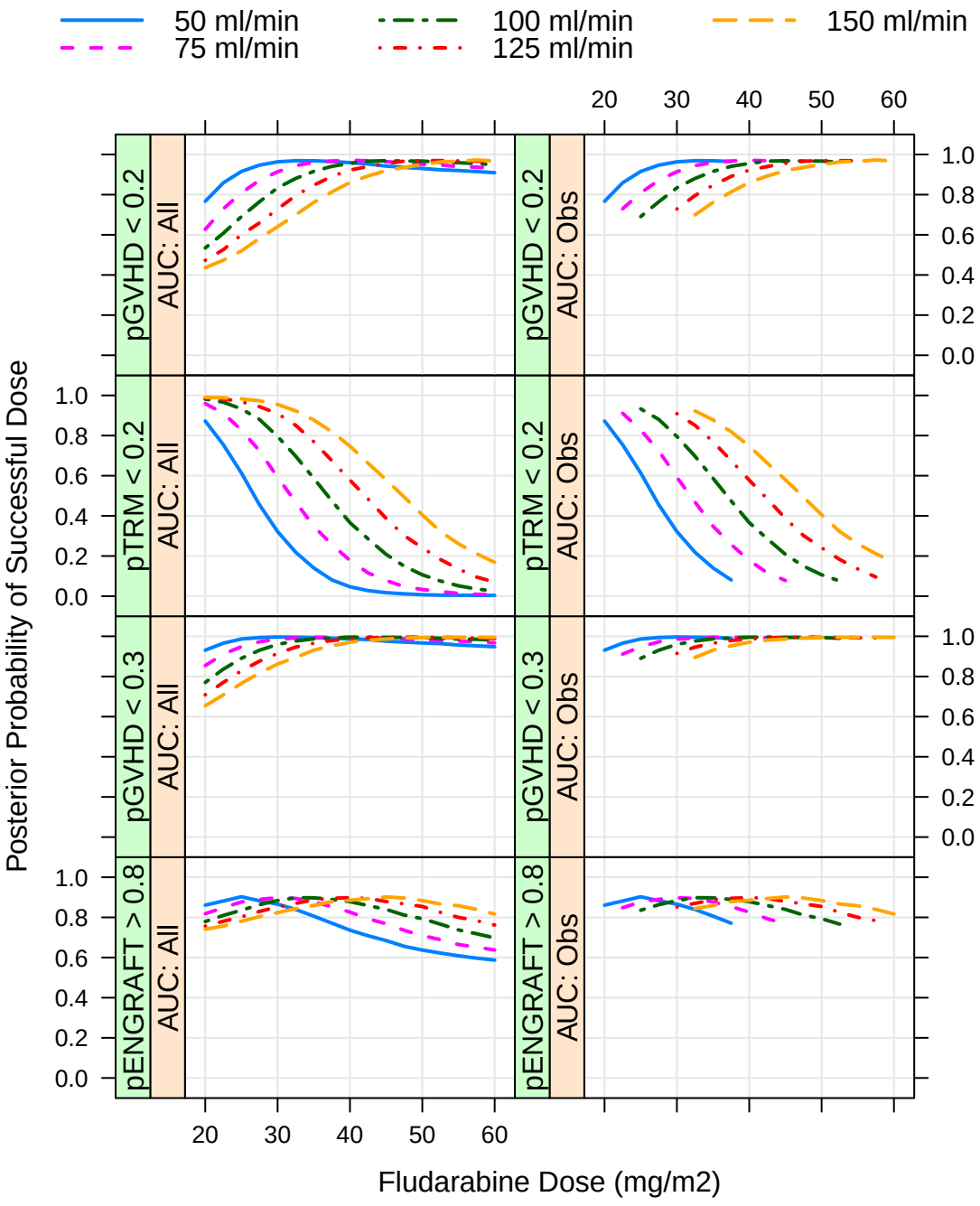

Figure 5.7: Posterior-predictive dose success probability versus FLU dose, grouped by PD endpoint (ENGRAFT, GVHD, TRM). Left column: results for F-ara-A AUC restricted to observed range. Right column: results for all simulated AUCs.

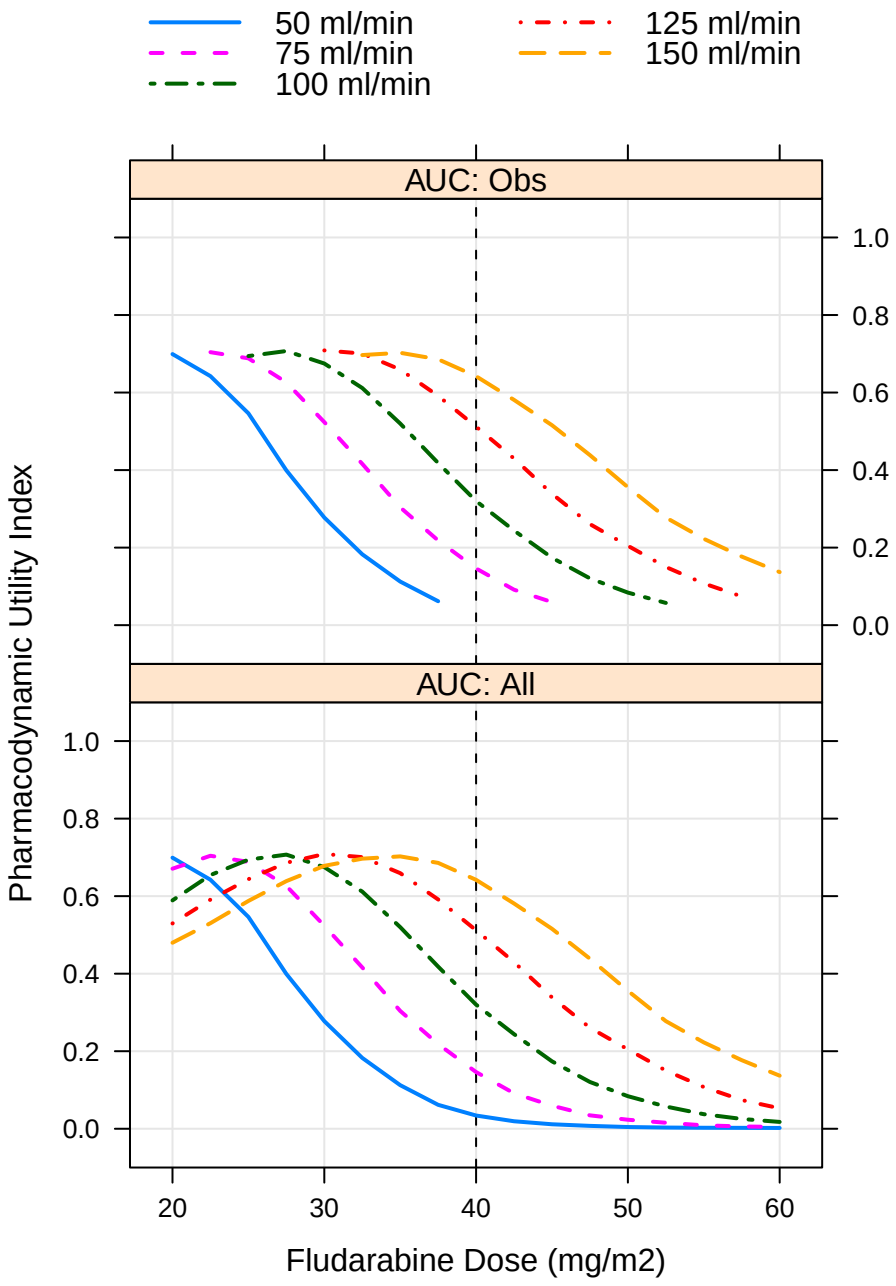

Figure 5.8: Pharmacodynamic utility index (UI_{TGE}) versus FLU dose for CLCR of 50 to 150 ml/min. The bottom panel shows results for all simulated F-ara-A exposures and the top panel shows results for exposures restricted to the observed range. Optimized doses are listed in table 5.2. The vertical dashed line indicates 40 mg/m^2 dose.

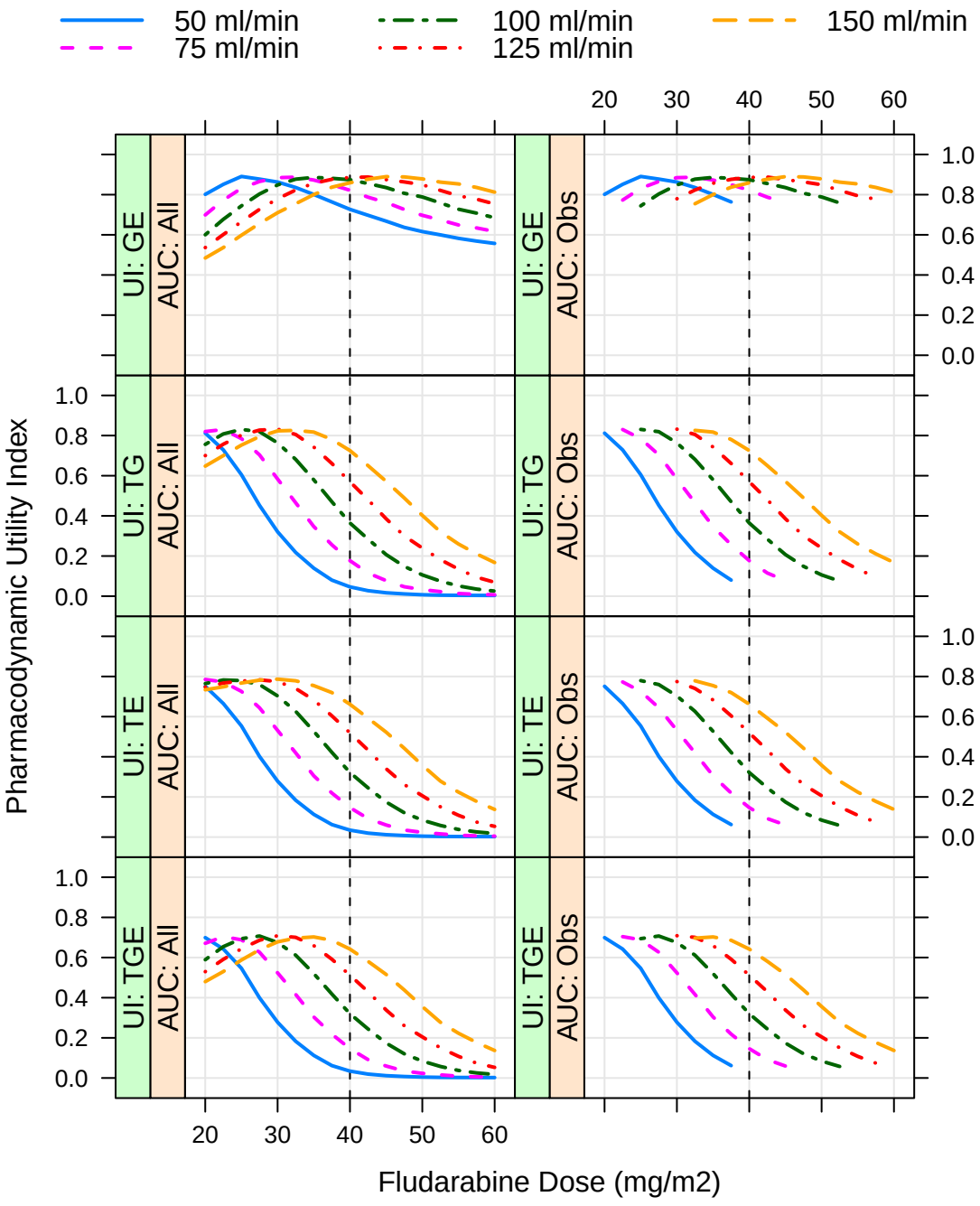

Figure 5.9: Utility index versus FLU dose for all utility indices (table 5.1) at CLCR from 50 to 150 ml/min. The left column shows results for all simulated F-ara-A exposures. The right column shows results for only simulated exposures within the observed range. The vertical dashed line indicates 40 mg/m^2 dose.

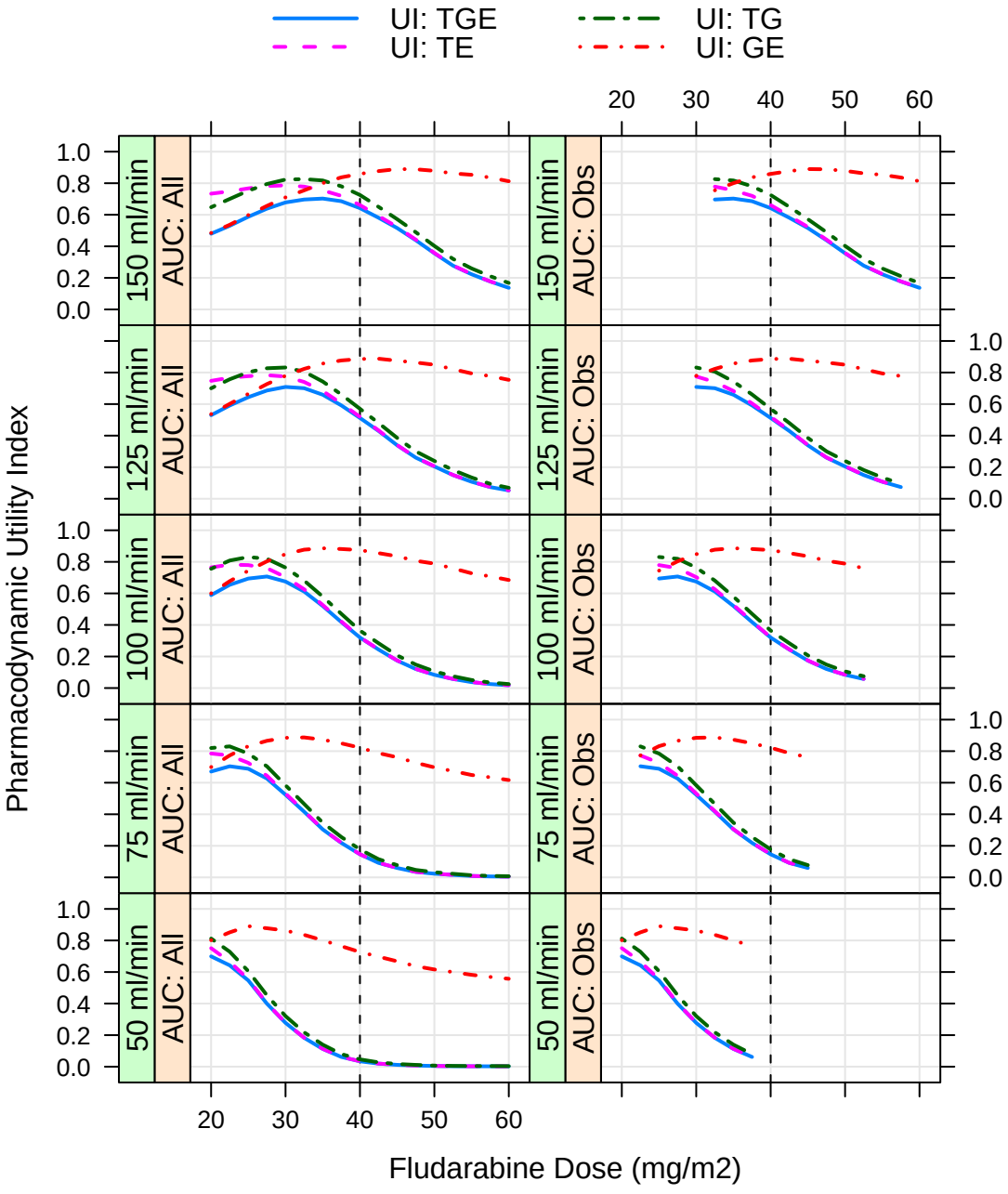

Figure 5.10: Utility index versus FLU dose for all utility indices (table 5.1) at CLCR from 50 to 150 ml/min. This is the same data presented in figure 5.9, but grouped by CLCR, not utility index.

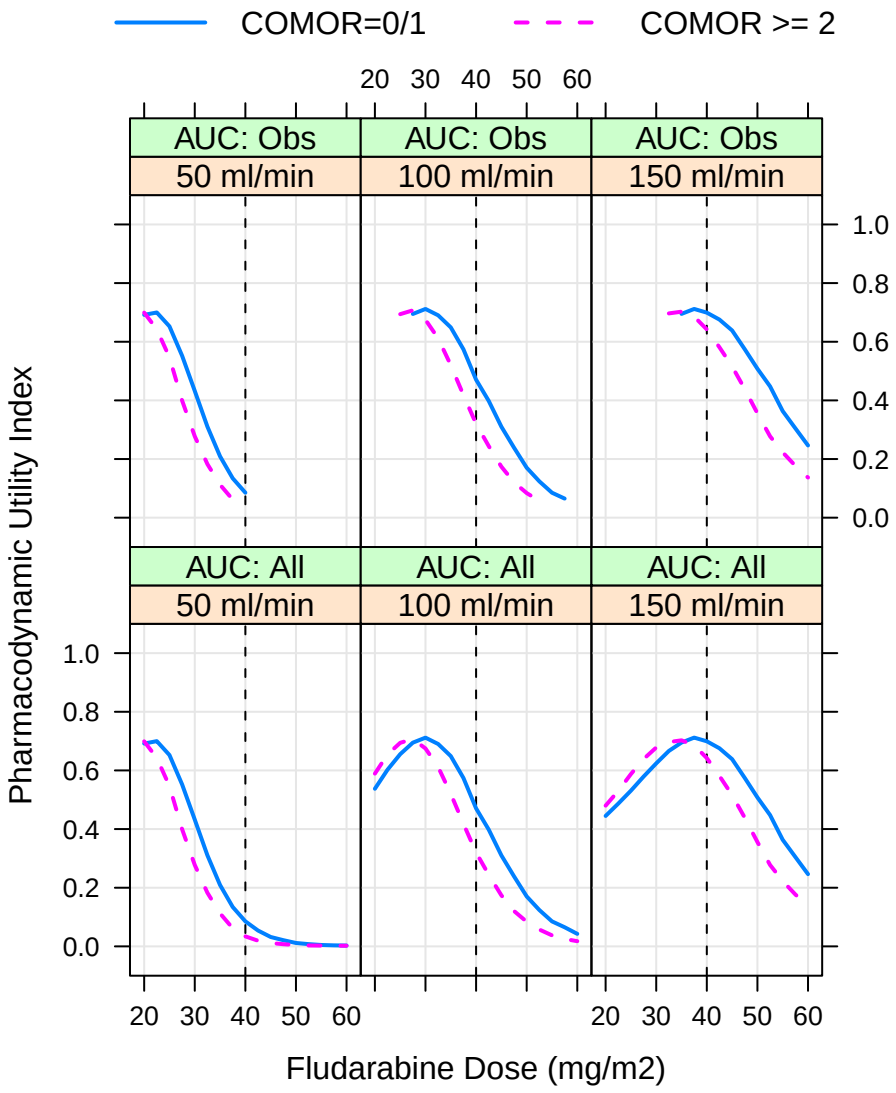

Figure 5.11: Utility index (UI_{TGE}) versus FLU dose for zero- and non-zero comorbidity score at CLCR of 50, 100, and 150 ml/min. The dashed vertical line marks the 40 mg/m^2 dose. AUC: Obs - simulated AUCs restricted to observed range. AUC: All - AUCs from all dose-covariate combinations.

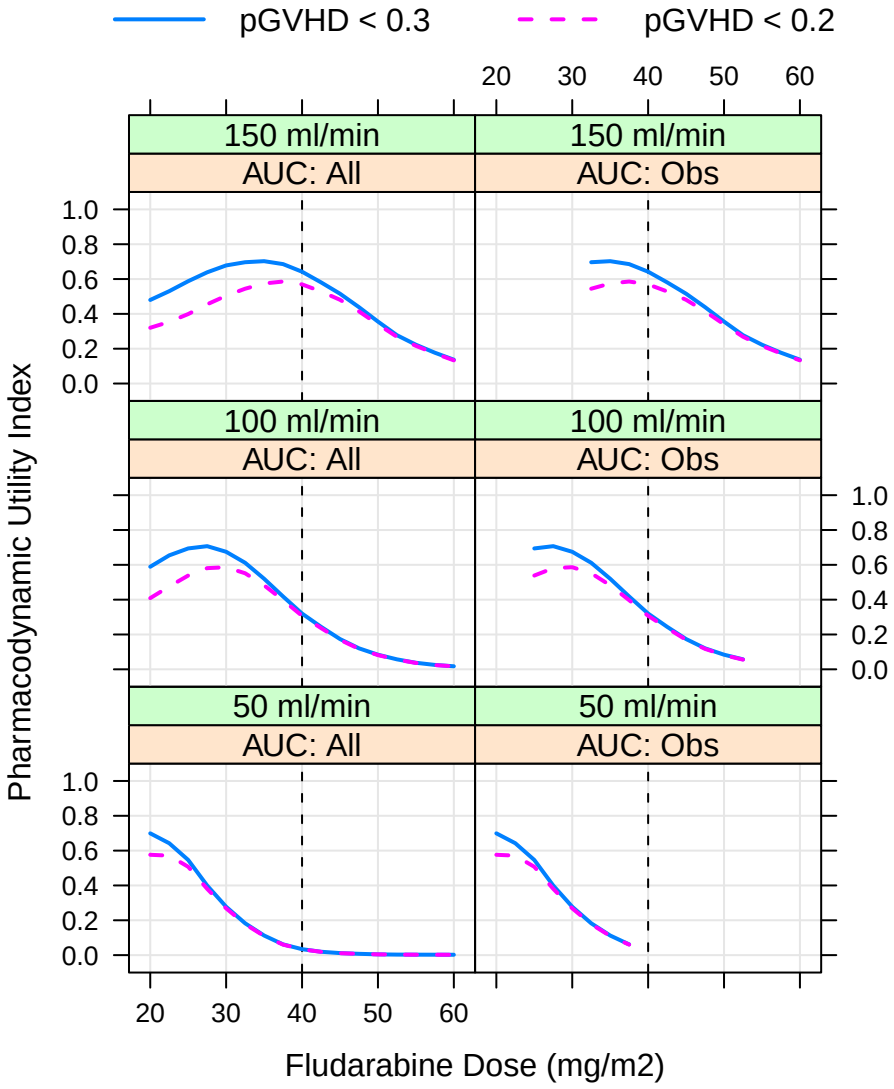

Figure 5.12: Utility index (UI_{TGE}) versus FLU dose for lower (pGVHD < 0.3) and higher (pGVHD < 0.2) weights on GVHD outcome in the optimization. Results are shown at CLCR of 50, 100, and 150 ml/min. The dashed vertical line marks the 40 mg/m^2 dose. AUC: Obs - simulated AUCs restricted to observed range. AUC: All - AUCs from all dose-covariate combinations.

Chapter 6

Summary and Future Direction

6.1 Summary

The modeling and dose-optimization work presented in this thesis builds upon and extends the previous findings of Long-Boyle et. al. [2] that established a link between F-ara-A exposure and outcomes after nonmyeloablative HCT. Particularly important was their finding of increased treatment-related mortality in patients with high F-ara-A exposure after 40 mg/m^2 fludarabine. This was a critical *confirming* step that established the need to evaluate F-ara-A exposures when fludarabine is used for conditioning in nonmyeloablative HCT.

Upon starting the current work, the overall aim was to answer the question, "How can we best individualize fludarabine doses in HCT to optimize outcomes,

including minimizing treatment-related mortality?". Answering the question requires *learning* more about fludarabine clinical pharmacology in HCT. This work used population pharmacokinetic modeling to identify sources of variability in F-ara-A PK and Bayesian pharmacodynamic modeling and to develop graded, quantitative exposure-response relationships to accomplish this learning.

In the PK model, creatinine clearance was found to be an extremely important covariate for predicting F-ara-A renal clearance. This was not a surprising result, given that the fraction excreted unchanged as F-ara-A is almost 60%. However a previous population analysis of F-ara-A failed to identify creatinine clearance as an important factor for predicting F-ara-A clearance [39]. Also, comorbidity score was found to have a small effect on the average F-ara-A clearance in the population. Weight was included in the model for clearance according to an allometric relationship, thus facilitating use of this model to predict F-ara-A PK in a pediatric population.

Consistent with the optimal cutpoint methodology used before [2], the pharmacodynamic modeling showed a clear increase in treatment related mortality with increasing F-ara-A exposure. At the lowest exposures, TRM probability was also very low ($\sim$10%). But toward the highest observed exposures, typical model-predicted TRM probability was over 50%. Clearly, there was not a wide enough exposure range to establish a complete exposure-response profile for the

neutrophil engraftment endpoint. But modeled engraftment rates were high across all observed exposures, suggesting that there was little or no engraftment benefit from pushing doses higher. Of course, the severe acute GVHD rate was lower as exposures increases, suggesting some benefit with respect to the GVHD endpoint.

Leveraging the information that was extracted from the data in the form of PK and PD models, a utility index was used to optimize fludarabine doses in HCT. As expected from the PK modeling work, dose optimization work suggested significant dose changes according to the level of renal function impairment. This work provides detailed and data-driven guidance for reducing fludarabine doses in case of renal dysfunction. Furthermore, only very small dose reductions were recommended for patients with significant comorbidity scores. This also was expected after PK modeling showed only a 10% reduction in clearance for patients with higher comorbidity scores. Overall, using a utility index than considered treatment-related mortality, acute grade III-IV GVHD, and neutrophil engraftment, the dose optimization process clearly indicated doses should be reduced from the 40 mg/m^2 that was given in the study protocol.

6.2 Future Directions

6.2.1 Fludarabine Pharmacokinetics

One potential future direction for fludarabine PK investigation is the cellular distribution of fludarabine. Although intravenous doses of fludarabine are delivered directly to the blood and fludarabine is quickly and completely activated, the circulating F-ara-A is hardly located at the ultimate site of action upon administration. Distribution into tissues, including circulating tumor cells and bone marrow, and metabolism to the active F-ara-ATP must take place before immunosuppressive activity is realized. Thus, in addition to organ clearances and passive tissue distribution, there could be molecular sources of variability in these transport and intracellular metabolic processes that are important for fully understanding F-ara-A PK.

6.2.2 Fludarabine Pharmacodynamics

There are at least two clear areas that warrant future study with respect to fludarabine pharmacodynamics. Going hand-in-hand with the pharmacokinetic aim discussed above, exposure-response relationships for intracellular F-ara-ATP should be investigated. In this work, the exposure-response models were proposed that linked F-ara-A exposure in the plasma to outcomes after HCT. Naturally,

it is easiest to measure F-ara-A in the plasma. However, given uncertainties in cellular transport and metabolism, it is not clear how the exposure-response relationship would change if it were modeled using intracellular F-ara-ATP exposure to establish the relationship.

Second, as noted in the chapter 5 discussion, there are many clinical endpoints to consider when evaluating outcomes after HCT that were not evaluated here. Chronic GVHD, relapse, and overall survival may be important factors to incorporate in to future fludarabine pharmacodynamic modeling work.

6.2.3 Fludarabine Dose Optimization

As mentioned above, future dose optimization work could potentially include more endpoints than the three used here. Furthermore, as with all simulation-based recommendations, the dosing suggestions presented here should be evaluated and their accuracy confirmed in controlled clinical use before being routinely used in clinical care.

Beyond these items, there is one more potentially important avenue that should be addressed in future work: dose individualization based on individualized PK. In the modeling and simulation work presented here, doses have been optimized on a population level. That is, the optimal dose is the dose that, on average, will lead to the best outcome. For any one particular patient, however, the "on-average"

optimal dose may not be the dose that is optimal for him or her. One reason for this could be imprecise exposure targeting when dosing based on the typical value of clearance in the population. Despite a strong PK covariate model, the low unexplained variability in clearance may not be low enough to precisely target exposures in individual patients. However, the population pharmacokinetic model would provide strong prior information on F-ara-A PK, allowing good empirical Bayes estimates of individual PK with only a few sparse plasma concentrations. These PK samples could be obtained after a small, subtherapeutic "test-dose" given prior to starting therapy and used to determine individual PK. Once individualized PK or PD parameters become known, the individual success rate will likely improve greatly due to better targeting to a therapeutic and non-harmful exposure.

References

[1] C. G. Brunstein, J. N. Barker, D. J. Weisdorf, T. E. DeFor, J. S. Miller, B. R. Blazar, P. B. McGlave, J. E. Wagner, Umbilical cord blood transplantation after nonmyeloablative conditioning: impact on transplantation outcomes in 110 adults with hematologic disease., Blood 110 (8) (2007) 3064–70.

[2] J. Long-Boyle, K. Green, C. Brunstein, Q. Cao, J. Rogosheske, D. Weisdorf, J. Miller, J. Wagner, P. McGlave, P. Jacobson, High fludarabine exposure and relationship with treatment-related mortality after nonmyeloablative hematopoietic cell transplantation, Bone Marrow Transplantat.

[3] E. A. Copelan, Hematopoietic stem-cell transplantation., N Engl J Med 354 (17) (2006) 1813–26.

[4] D. Weisdorf, GVHD the nuts and bolts., Hematology Am Soc Hematol Educ Program (2007) 62–7.

[5] E. Petersdorf, Stem Cell Transplantation for Hematologic Malignancies, in: R. J. Soiffer (Ed.), Hematopoietic stem cell transplantation, 2nd Edition, Springer, Totowa, NJ, 2008, pp. 19–38.

[6] M. J. Laughlin, M. Eapen, P. Rubinstein, J. E. Wagner, M.-J. Zhang, R. E. Champlin, C. Stevens, J. N. Barker, R. P. Gale, H. M. Lazarus, D. I. Marks, J. J. van Rood, A. Scaradavou, M. M. Horowitz, Outcomes after transplantation of cord blood or bone marrow from unrelated donors in adults with leukemia., N Engl J Med 351 (22) (2004) 2265–75.

[7] C. Cutler, Stem Cell Sources for Allogeneic Transplantation, in: R. J. Soiffer (Ed.), Hematopoietic stem cell transplantation, 2nd Edition, Springer, Totowa, NJ, 2008, pp. 225–244.

[8] N. Schmitz, D. C. Linch, P. Dreger, A. H. Goldstone, M. A. Boogaerts, A. Ferrant, H. M. Demuynck, H. Link, A. Zander, A. Barge, Randomised trial of filgrastim-mobilised peripheral blood progenitor cell transplantation versus autologous bone-marrow transplantation in lymphoma patients., Lancet 347 (8998) (1996) 353–7.

[9] J. M. Vose, G. Sharp, W. C. Chan, C. Nichols, K. Loh, D. Inwards, R. Rifkin, P. J. Bierman, J. C. Lynch, D. D. Weisenburger, A. Kessinger, J. O. Armitage, Autologous transplantation for aggressive non-Hodgkin's lymphoma: results

of a randomized trial evaluating graft source and minimal residual disease., J Clin Oncol 20 (9) (2002) 2344–52.

[10] J. E. Wagner, J. N. Barker, T. E. DeFor, K. S. Baker, B. R. Blazar, C. Eide, A. Goldman, J. Kersey, W. Krivit, M. L. MacMillan, P. J. Orchard, C. Peters, D. J. Weisdorf, N. K. C. Ramsay, S. M. Davies, Transplantation of unrelated donor umbilical cord blood in 102 patients with malignant and nonmalignant diseases: influence of CD34 cell dose and HLA disparity on treatment-related mortality and survival., Blood 100 (5) (2002) 1611–8.

[11] F. Baron, R. Storb, Allogeneic hematopoietic cell transplantation following nonmyeloablative conditioning as treatment for hematologic malignancies and inherited blood disorders., Mol Ther 13 (1) (2006) 26–41.

[12] S. Slavin, A. Nagler, E. Naparstek, Y. Kapelushnik, M. Aker, G. Cividalli, G. Varadi, M. Kirschbaum, A. Ackerstein, S. Samuel, A. Amar, C. Brautbar, O. Ben-Tal, A. Eldor, R. Or, Nonmyeloablative stem cell transplantation and cell therapy as an alternative to conventional bone marrow transplantation with lethal cytoreduction for the treatment of malignant and nonmalignant hematologic diseases., Blood 91 (3) (1998) 756–63.

[13] D. G. Maloney, B. M. Sandmaier, S. Mackinnon, J. A. Shizuru, Nonmyeloablative transplantation., Hematology Am Soc Hematol Educ Program

(2002) 392–421.

[14] F. Baron, R. Storb, Allogeneic hematopoietic cell transplantation as treatment for hematological malignancies: a review., Vol. 26, 2004, pp. 71–94.

[15] R. F. Storb, R. Champlin, S. R. Riddell, M. Murata, S. Bryant, E. H. Warren, Non-myeloablative transplants for malignant disease., Hematology Am Soc Hematol Educ Program (2001) 375–91.

[16] M. M. Horowitz, R. P. Gale, P. M. Sondel, J. M. Goldman, J. Kersey, H. J. Kolb, A. A. Rimm, O. Ringden, C. Rozman, B. Speck, Graft-versus-leukemia reactions after bone marrow transplantation., Blood 75 (3) (1990) 555–62.

[17] M. M. Horowitz, R. P. Gale, P. M. Sondel, J. M. Goldman, J. Kersey, H. J. Kolb, A. A. Rimm, O. Ringden, C. Rozman, B. Speck, Graft-versus-leukemia reactions after bone marrow transplantation., Blood 75 (3) (1990) 555–62.

[18] R. H. J. Collins, O. Shpilberg, W. R. Drobyski, D. L. Porter, S. Giralt, R. Champlin, S. A. Goodman, S. N. Wolff, W. Hu, C. Verfaillie, A. List, W. Dalton, N. Ognoskie, A. Chetrit, J. H. Antin, J. Nemunaitis, Donor leukocyte infusions in 140 patients with relapsed malignancy after allogeneic bone marrow transplantation., J Clin Oncol 15 (2) (1997) 433–44.

[19] H. J. Kolb, A. Schattenberg, J. M. Goldman, B. Hertenstein, N. Jacobsen, W. Arcese, P. Ljungman, A. Ferrant, L. Verdonck, D. Niederwieser, F. van Rhee, J. Mittermueller, T. de Witte, E. Holler, H. Ansari, Graft-versus-leukemia effect of donor lymphocyte transfusions in marrow grafted patients., Blood 86 (5) (1995) 2041–50.

[20] J. Antin, Reduced-Intensity Stem Cell Transplantation, Hematology 2007 (2007) 47–54.

[21] J. Ferrara, J. Levine, P. Reddy, E. Holler, Graft-versus-host disease, The Lancet 373 (9674) (2009) 1550–1561.

[22] S. J. Lee, M. E. D. Flowers, Recognizing and managing chronic graft-versus-host disease., Hematology Am Soc Hematol Educ Program (2008) 134–41.

[23] V. Gandhi, W. Plunkett, Cellular and clinical pharmacology of fludarabine, Clinical Pharmacokinetics 41 (2) (2002) 93–103.

[24] S. Johnson, Clinical pharmacokinetics of nucleoside analogues: focus on haematological malignancies, Clinical Pharmacokinetics 39 (1) (2000) 5–26.

[25] W. Plunkett, P. Saunders, Metabolism and action of purine nucleoside analogs, Pharmacology & Therapeutics 49 (3) (1991) 239–268.

[26] W. Yin, E. V. Karyagina, A. S. Lundberg, D. J. Greenblatt, J. Lister-James, Pharmacokinetics, bioavailability and effects on electrocardiographic parameters of oral fludarabine phosphate., Biopharm Drug Dispos 31 (1) (2010) 72–81.

[27] M. R. Hersh, J. G. Kuhn, J. L. Phillips, G. Clark, T. M. Ludden, D. D. V. Hoff, Pharmacokinetic study of fludarabine phosphate (NSC 312887)., Cancer Chemother Pharmacol 17 (3) (1986) 277–80.

[28] L. Danhauser, W. Plunkett, M. Keating, F. Cabanillas, 9-beta-D-arabinofuranosyl-2-fluoroadenine 5'-monophosphate pharmacokinetics in plasma and tumor cells of patients with relapsed leukemia and lymphoma., Cancer Chemother Pharmacol 18 (2) (1986) 145–52.

[29] L. Malspeis, M. R. Grever, A. E. Staubus, D. Young, Pharmacokinetics of 2-F-ara-A (9-beta-D-arabinofuranosyl-2-fluoroadenine) in cancer patients during the phase I clinical investigation of fludarabine phosphate., Semin Oncol 17 (5 Suppl 8) (1990) 18–32.

[30] J. Zhang, F. Visser, K. M. King, S. A. Baldwin, J. D. Young, C. E. Cass, The role of nucleoside transporters in cancer chemotherapy with nucleoside drugs., Cancer Metastasis Rev 26 (1) (2007) 85–110.

[31] V. Gandhi, A. Kemena, M. Keating, W. Plunkett, Cellular pharmacology of fludarabine triphosphate in chronic lymphocytic leukemia cells during fludarabine therapy, Leukemia & Lymphoma 10 (1-2) (1993) 49–56.

[32] Anonymous, Fludara Prescribing Information, Pfizer, 2001.

[33] F. Baron, B. M. Sandmaier, Current status of hematopoietic stem cell transplantation after nonmyeloablative conditioning., Curr Opin Hematol 12 (6) (2005) 435–43.

[34] W. Knebel, J. C. J. Davis, W. D. Sanders, B. Fessler, C. Yarboro, F. Pucino, D. T. Boumpas, The pharmacokinetics and pharmacodynamics of fludarabine in rheumatoid arthritis., Pharmacotherapy 18 (6) (1998) 1224–9.

[35] M. Bonin, S. Pursche, T. Bergeman, T. Leopold, T. Illmer, G. Ehninger, E. Schleyer, M. Bornhauser, F-ara-A pharmacokinetics during reduced-intensity conditioning therapy with fludarabine and busulfan., Bone Marrow Transplant 39 (4) (2007) 201–6.

[36] J. M. Foran, D. Oscier, J. Orchard, S. A. Johnson, M. Tighe, M. H. Cullen, P. G. de Takats, C. Kraus, M. Klein, T. A. Lister, Pharmacokinetic study of single doses of oral fludarabine phosphate in patients with "low-grade"

non-Hodgkin's lymphoma and B-cell chronic lymphocytic leukemia., J Clin Oncol 17 (5) (1999) 1574–9.

[37] G. M. Kuo, D. T. Boumpas, G. G. Illei, C. Yarboro, F. Pucino, A. H. Burstein, Fludarabine pharmacokinetics after subcutaneous and intravenous administration in patients with lupus nephritis., Pharmacotherapy 21 (5) (2001) 528–33.

[38] S. M. Lichtman, E. Etcubanas, D. R. Budman, P. Eisenberg, G. Zervos, P. D'Amico, V. O'Mara, K. Musgrave, P. Cascella, A. Melikian, P. H. Hinderling, J. M. Ferrer, G. J. Williams, The pharmacokinetics and pharmacodynamics of fludarabine phosphate in patients with renal impairment: a prospective dose adjustment study., Cancer Invest 20 (7-8) (2002) 904–13.

[39] D. H. Salinger, D. K. Blough, P. Vicini, C. Anasetti, P. V. O'Donnell, B. M. Sandmaier, J. S. McCune, A limited sampling schedule to estimate individual pharmacokinetic parameters of fludarabine in hematopoietic cell transplant patients., Clin Cancer Res 15 (16) (2009) 5280–7.

[40] R. Carroll, D. Ruppert, Power transformations when fitting theoretical models to data, Journal of the American Statistical Association 79 (386) (1984) 321–328.

[41] M. O. Karlsson, L. B. Sheiner, The importance of modeling interoccasion variability in population pharmacokinetic analyses., J Pharmacokinet Biopharm 21 (6) (1993) 735–50.

[42] M. O. Karlsson, R. M. Savic, Diagnosing model diagnostics., Clin Pharmacol Ther 82 (1) (2007) 17–20.

[43] S. Janmahasatian, S. B. Duffull, S. Ash, L. C. Ward, N. M. Byrne, B. Green, Quantification of lean bodyweight., Clin Pharmacokinet 44 (10) (2005) 1051–65.

[44] D. W. Cockcroft, M. H. Gault, Prediction of creatinine clearance from serum creatinine, Nephron 16 (1) (1976) 31–41.

[45] M. L. Sorror, M. B. Maris, R. Storb, F. Baron, B. M. Sandmaier, D. G. Maloney, B. Storer, Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT., Blood 106 (8) (2005) 2912–9.

[46] B. Agoram, A. C. Heatherington, M. R. Gastonguay, Development and evaluation of a population pharmacokinetic-pharmacodynamic model of darbepoetin alfa in patients with nonmyeloid malignancies undergoing multicycle chemotherapy., AAPS J 8 (3) (2006) E552–63.

[47] M. Gastonguay, A full model estimation approach for covariate effects: inference based on clinical importance and estimation precision, AAPS J 6 (S1) (2004) W4354.

[48] P. Bonate, Pharmacokinetic-pharmacodynamic modeling and simulation, Springer Verlag, 2005.

[49] I. Matthews, C. Kirkpatrick, N. Holford, Quantitative justification for target concentration intervention–parameter variability and predictive performance using population pharmacokinetic models for aminoglycosides., Br J Clin Pharmacol 58 (1) (2004) 8–19.

[50] D. R. Mould, N. H. G. Holford, J. H. M. Schellens, J. H. Beijnen, P. R. Hutson, H. Rosing, W. W. ten Bokkel Huinink, E. K. Rowinsky, J. H. Schiller, M. Russo, G. Ross, Population pharmacokinetic and adverse event analysis of topotecan in patients with solid tumors., Clin Pharmacol Ther 71 (5) (2002) 334–48.

[51] H. Boxenbaum, Interspecies scaling, allometry, physiological time, and the ground plan of pharmacokinetics., J Pharmacokinet Biopharm 10 (2) (1982) 201–27.

[52] B. J. Anderson, K. Allegaert, N. H. G. Holford, Population clinical pharmacology of children: modelling covariate effects., Eur J Pediatr 165 (12) (2006) 819–29.

[53] B. J. Anderson, N. H. G. Holford, Mechanism-based concepts of size and maturity in pharmacokinetics., Annu Rev Pharmacol Toxicol 48 (2008) 303–32.

[54] Y. Yano, S. L. Beal, L. B. Sheiner, Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check., J Pharmacokinet Pharmacodyn 28 (2) (2001) 171–92.

[55] N. Holford, The visual predictive checksuperiority to standard diagnostic (Rorschach) plots [abstract 738], in: Annual Meeting of the Population Approach Group in Europe (PAGE), Vol. 14, 2005.

[56] A. Gelman, J. Carlin, H. Stern, D. Rubin, Bayesian Data Analysis, CRC Press, Boca Raton, FL, 2004.

[57] A. Hooker, M. Karlsson, Conditional weighted residuals. A diagnostic to improve population PK/PD model building and evaluation, AAPS J 7 (S2) (2005) W5321.

[58] A. C. Hooker, C. E. Staatz, M. O. Karlsson, Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method., Pharm Res 24 (12) (2007) 2187–97.

[59] K. Baron, A. Birnbaum, J. Cloyd, J. St. Peter, R. Brundage, Evaluation of assumptions in the clinical use of the cockcroft-gault equation [abstract 1636], in: Annual Meeting of the Population Approach Group in Europe (PAGE), Vol. 18, 2009.

[60] G. Casella, An introduction to empirical Bayes data analysis, American Statistician 39 (2) (1985) 83–87.

[61] L. B. Sheiner, Learning versus confirming in clinical drug development., Clin Pharmacol Ther 61 (3) (1997) 275–91.

[62] D. G. Altman, Suboptimal analysis using 'optimal' cutpoints., Br J Cancer 78 (4) (1998) 556–7.

[63] D. G. Altman, P. Royston, The cost of dichotomising continuous variables., BMJ 332 (7549) (2006) 1080.

[64] P. Royston, D. G. Altman, W. Sauerbrei, Dichotomizing continuous predictors in multiple regression: a bad idea., Stat Med 25 (1) (2006) 127–41.

[65] D. Przepiorka, D. Weisdorf, P. Martin, H. G. Klingemann, P. Beatty, J. Hows, E. D. Thomas, 1994 Consensus Conference on Acute GVHD Grading., Bone Marrow Transplant 15 (6) (1995) 825–8.

[66] B. Carlin, T. Louis, Bayesian Methods for Data Analysis, CRC Press, New York, 2008.

[67] A. Agresti, Categorical Data Analysis, 2nd Edition, Wiley-Interscience, Hoboken, NJ, 2002.

[68] J. GABRIELSSON, D. WEINER, Pharmacokinetic & pharmacodynamic Data Analysis: Concepts & Applications with CD-ROM, 2007.

[69] D. Spiegelhalter, N. Best, B. Carlin, A. van der Linde, Bayesian measures of model complexity and fit, Journal of the Royal Statistical Society: Series B 64 (4) (2002) 583–639.

[70] J. Klein, M. Moeschberger, Survival Analysis: Techniques for Censored and Truncated Data, 2nd Edition, Springer Verlag, New York, NY, 2003.

[71] S. Brooks, A. Gelman, Alternative methods for monitoring convergence of iterative simulations, Journal of Computational and Graphical Statistics 7 (4) (1998) 434–455.

[72] A. Gelman, Y. Goegebeur, F. Tuerlinckx, I. Van Mechelen, Diagnostic checks for discrete data regression models using posterior predictive simulations, Applied Statistics 49 (2) (2000) 247–268.

[73] L. J. Lesko, Paving the critical path: how can clinical pharmacology help achieve the vision?, Clin Pharmacol Ther 81 (2) (2007) 170–7.

[74] D. R. Stanski, M. Rowland, L. B. Sheiner, Getting the Dose Right: report from the Tenth European Federation of Pharmaceutical Sciences (EUFEPS) conference on optimizing drug development., J Pharmacokinet Pharmacodyn 32 (2) (2005) 199–211.

[75] A. A. Khan, I. Perlstein, R. Krishna, The use of clinical utility assessments in early clinical development., AAPS J 11 (1) (2009) 33–8.

[76] D. Ouellet, Benefit-risk assessment: the use of clinical utility index., Expert Opin Drug Saf 9 (2) (2010) 289–300.

[77] D. Ouellet, J. Werth, N. Parekh, D. Feltner, B. McCarthy, R. L. Lalonde, The use of a clinical utility index to compare insomnia compounds: a quantitative basis for benefit-risk assessment., Clin Pharmacol Ther 85 (3) (2009) 277–82.

[78] B. Poland, F. L. Hodge, A. Khan, R. T. Clemen, J. A. Wagner, K. Dykstra, R. Krishna, The clinical utility index as a practical multiattribute approach to drug development decisions., Clin Pharmacol Ther 86 (1) (2009) 105–8.

[79] V. Goel, T. Hanson, R. Corrigan, R. Miller, R. Brundage, A bayesian multivariate model for repeated measures of correlated data [abstract 1621], in: Annual Meeting of the Population Approach Group in Europe (PAGE), Vol. 18, 2009.

Appendix A

Code

Care has been taken in this thesis to minimize the use of jargon and acronyms, but this cannot always be achieved. This appendix defines jargon terms in a glossary, and contains a table of acronyms and their meaning.

A.1 NONMEM Control Streams

A.1.1 Final PK Model NONMEM Control Stream

```
$PROBLEM FLU 3CMT FIX EXP LBW-V  COMOR on CL LOG EST  IIV ON Q2
$INPUT C ID NUM SAMP TIME TAD AMT RATE DV LDV=DROP AGE WGT SEX BSA=DROP
       CYC=DROP TBIL=DROP DX=DROP COND=DROP GVHD=DROP CLCR=DROP CGCL
       SCR IBW ATG=DROP WBC=DROP DTD=DROP COMO BMI=DROP LBW CMT MDV EVID

;  NUM = RECORD NUMBER
;  SAMP = PLASMA CONC OBS NUMBER
;  TIME = PK TIME SINCE FIRST DOSE GIVEN
;  TAD = TIME AFTER DOSE
```

```
;  AMT = DOSE
;  RATE = INFUSION RATE
;  DV = PLASMA CONCENTRATION
;  LDV = LN PLASMA CONCENTRATION
;  AGE: IN YEARS
;  WGT = ACTUAL BODY WEIGHT IN KG
;  SEX: MALE=0, FEMALE=1
;  BSA = BODY SURFACE AREA
;  CYC = CYSTATIN C
;  TBIL = TOTAL BILLIRUBIN
;  DX = DIAGNOSIS, CODED BY INTEGER
;  COND: 1 IF SUBJECT GOT ATG, 0 OTHERWISE
;  GVHD: 1 IF SUBJECT DEVELOPED GVHD, 0 OTHERWISE
;  CLCR = COCKCROFT-GAULT CREATININE CLEARANCE USING IBW
;  CGCL = COCKCROFT-GAULT CREATININE CLEARANCE USING ABW
;  SCR = SERUM CREATININE (MG%)
;  IBW = IDEAL BODY WEIGHT (50+2.3*(HGT-60) - 4.5**SEX)
;  ATG: 1 IF SUBJECT RECEIVED ATG, 0 OTHERWISE
;  WBC = WHITE BLOOD CELL COUNT
;  DTD = DAYS TO DEATH
;  COMO = COMORBIDITY SCORE; EITHER 100, 200, OR 300


$DATA ../data/flu.nm.all.csv  IGNORE=C
; Change to the following if data file is in same directory as ctl file:
; $DATA flu.nm.all.csv IGNORE=C

$SUBROUTINES ADVAN11 TRANS4

$PK
   CRCL = 0.85**SEX*(140-AGE)*WGT/72/SCR

   LBWST = 63
   WGTST = 70

   CLNRST = THETA(1)
   CLRST =  THETA(7)

   RF = CRCL*(WGTST/WGT)

   TVCL = (CLNRST + CLRST*(RF/100))*(WGT/WGTST)**0.75

   IF(COMO.NE.100) TVCL = TVCL*THETA(8)

   TVQ2 = THETA(2)*(WGT/WGTST)**0.75
   TVQ3 = THETA(3)*(WGT/WGTST)**0.75
```

```
   TVV1 = THETA(4)*(LBW/LBWST)
   TVV2 = THETA(5)*(LBW/LBWST)
   TVV3 = THETA(6)*(LBW/LBWST)

   CL = TVCL*EXP(ETA(1))
   Q2 = TVQ2*EXP(ETA(5))
   Q3 = TVQ3*EXP(ETA(6))
   V1 = TVV1*EXP(ETA(2))
   V2 = TVV2*EXP(ETA(3))
   V3 = TVV3*EXP(ETA(4))

   IF(CL.LE.0) EXIT 1 100

   SC = V1/1000

$ERROR(OBS)
   IF(ICALL.EQ.4) THEN
      IF(DV.GT.0) Y = LOG(DV)
      RETURN
   ENDIF

   SIGPRO = THETA(9)
   SIGADD = THETA(10)

   IF(TIME.GE.50) THEN
      SIGPRO = THETA(11)
      SIGADD = THETA(12)
   ENDIF

   IPRED = 1E-5
   IF(F.GT.0) IPRED=LOG(F)
   IPRD = F
   W = SQRT(SIGPRO**2 + SIGADD**2/IPRD**2)
   Y=IPRED+W*ERR(1)
   OBS=EXP(DV)
   IWRE = (DV-IPRED)/W


$THETA
(0,6)      ; 1 - CLNR
(0,5.5)    ; 2 - Q2
(0,25)     ; 3 - Q3

(0,60)     ; 4 - V1
(0,80)     ; 5 - V2
(0,35)     ; 6 - V3
```

```
(0,7)      ; 7 - CLR

(0,1)      ; 8 - COMOR

(0,0.5)    ; 9 - Prop Error TIME.LT.50
(0,10)     ; 10 - Add Error TIME.LT.50

(0,0.5)    ; 11 - Prop Error TIME.GE.50
(0,10)     ; 12 - Add Error TIME.GE.50



$OMEGA BLOCK(3)
0.1                    ; 1 - Exponential IIV on CL
0.03 0.1               ; 2 - Exponential IIV on V1
0.03 0.06 0.1          ; 3 - Exponential IIV on V2

$OMEGA
0.2
0.2
0,FIX


$SIGMA
1, FIX

$SIM(12345)
$EST METHOD=COND INTERACTION  PRINT=5 MAXEVAL=9999 NOABORT NOTBT
NOOBT  MSFO=174.msf
$COV PRINT=E
$TABLE ID NUM AMT TIME TAD TVCL TVV1 TVV2 TVV3 TVQ2 TVQ3 CL V1 V2 V3
Q2 Q3 ETA1 ETA2 ETA3 ETA4 ETA5 IPRED IWRE  NOPRINT ONEHEADER FILE=174.TAB
$TABLE ID TVCL TVV1 TVV2 TVV3 TVQ2 TVQ3 CL V1 V2 V3 Q2 Q3 ETA1 ETA2
 ETA3 ETA4 ETA5 FIRSTONLY NOPRINT ONEHEADER FILE=174.FIR
```

A.1.2 WinBUGS code - Generalized Linear Pharmacodynamic Models with PPC Code

```
MODEL:
# nsub = number of subjects in study (87)
#
# TRM[] = treatment-related mortality responses
#     1 = TRM
#     0 = NO TRM
#
# GVHD[] = grade III or IV acute graft-versus-host disease responses
#   1 = grade III or IV GVHD
#   0 = no grade III or IV GVHD
#
# ENGRAFT[] = neutrophil engraftment
#  1 = engraftment criteria met
#  0 = engraftment criteria not met
#
# AUC[] = F-ara-A exposures (AUC{zero-->infinity})
#   AUC[] range: 3.5 to 9.5
#
# p.trm = modeled probability of TRM
# p.gvhd = modeled probability of GVHD
# p.engraft = modeled probability of engraftment
#
# beta.zero[] = intercept parameters for GLMS
#   beta.zero[1] for TRM
#   beta.zero[2] for GVHD
#   beta.zero[3] for ENGRAFT
#
# slope[] = F-ara-A covariate effect on logit(p)
#   slope[1] for TRM
#   slope[2] for GVHD
#   slope[3] for ENGRAFT
#
# n.AUC.star = number of AUCs for posterior predictions
#
# AUC.star[i] = ith AUC for posterior predictions
#    AUC.star[] range: 3.5 to 9.5
#
# TRM.star[i]  = ith posterior TRM predictions at AUC.star[i]
# GVHD.star[i] = ith posterior GVHD predictions at AUC.star[i]
# ENGRAFT.star[i] = ith posterior ENGRAFTpredicitions at AUC.star[i]
```

```
#      i from 1 to n.AUC.star
#
# p.trm.star[i]  = ith posterior predicted success prob TRM
# p.gvhd.star[i] = ith posterior predicted success prob GVHD
# p.engraft.star[i] = ith posterior predicted success prob ENGRAFT
#      i from 1 to n.AUC.star

for(i in 1:nsub) {

   # TREATMENT-RELATED MORTALITY
   TRM[i] ~ dbern(p.trm[i])
   logit(p.trm[i]) <- beta.zero[1] + slope[1]*(AUC[i]-6)

   # MAX GVHD GRADE >= 3
   GVHD[i] ~ dbern(p.gvhd[i])
   logit(p.gvhd[i]) <- beta.zero[2] + slope[2]*(AUC[i]-6)

   # ENGRAFTMENT
   ENGRAFT[i] ~ dbern(p.engraft[i])
   logit(p.engraft[i]) <- beta.zero[3] + slope[3]*(AUC[i]-6)

}

# Prior Distributions
# Flat everyting
for(i in 1:3) {
   beta.zero[i] ~ dflat()
   slope[i] ~ dflat()
}

# PK STUFF
#
# BODYSIZE:
#    Assume WEIGHT = 80 kg
#    Assume HEIGHT = 170 cm
#    Then BSA will be about 1.9 m2
#
# From the Pop-PK model, unexplained variability in CL is 0.0427
#    and non-renal clearance is 6.27  L/hr
#    and renal clearance is 7.25 L/hr when CLCR is 100 ml/min and WGT = 70kg
#
# wgt70 is the simulated weight normalized to 70 kg individual
# clcr70[] is the creatinine clearance per 70 kg of body weight


BSA <- pow(WGT,0.425)* pow(HGT, 0.725)*0.007184
sig2 <- 0.0427
```

```
tau2 <- 1/sig2
clnr <- 6.24
clr <- 7.25

# n.clcr = number of different CLCR we are gertting predictions at
# clnr = non-renal clearance
# clr = renal clearance
# n.doses = number of different doses we are getting predictions at
# comor = effect of comorbidity score on clearance
#     12% decrease in clearance for non-zero comorbidity score

wgt70 <- WGT/70

for(i in 1:n.clcr) {

   clcr70[i]  <- clcr[i]*70/WGT

   for(j in 1:n.comor) {

      TVCL[j,i]<-(clnr+clr *(clcr70[i]/100))*pow(wgt70,0.75)*(1-comor[j])

      for(k in 1:n.doses) {

         this.dose[k,j,i] <- doses[k]*BSA

         # EXPOSURE
         eta.cl[k,j,i] ~ dnorm(0,tau2)
         CL[k,j,i] <- TVCL[j,i]*exp(eta.cl[k,j,i])
         AUC.star[k,j,i]<- this.dose[k,j,i]/CL[k,j,i]

         AUC.above[k,j,i] <- step(AUC.star[k,j,i]-3.5)
         AUC.below[k,j,i] <-  step(9.5 -AUC.star[k,j,i])
         AUC.in[k,j,i] <- AUC.above[k,j,i]*AUC.below[k,j,i]

         # POSTERIOR PREDICEIVE PROBABILITIES

         # Treatment-related mortality
         TRM.star[k,j,i] ~dbern(p.trm.star[k,j,i])
         logit(p.trm.star[k,j,i]) <-
               beta.zero[1] + slope[1]*(AUC.star[k,j,i]-6)

         # Graft-versus-host disease
         GVHD.star[k,j,i] ~ dbern(p.gvhd.star[k,j,i])
         logit(p.gvhd.star[k,j,i]) <-
               beta.zero[2] + slope[2]*(AUC.star[k,j,i]-6)

         # Neutrophil engraftment
```

```
      ENGRAFT.star[k,j,i] ~ dbern(p.engraft.star[k,j,i])
      logit(p.engraft.star[k,j,i]) <-
            beta.zero[3] + slope[3]*(AUC.star[k,j,i]-6)


      # Counters for success with respect to trm and engraftment
      success.trm[k,j,i] <- step(0.2 - p.trm.star[k,j,i])
      success.engraft[k,j,i] <- step(p.engraft.star[k,j,i]-0.8)

      # success.gvhd1 and success.gvhd2 are counters for
      # different dose-success criteria with respect to gvhd
      #    gcrit[1] was 0.3 (relaxed weighting on gvhd outcome)
      #    gcrit[2] was 0.2 (equal weighting on gvhd outcome)
      success.gvhd1[k,j,i] <- step(gcrit[1] - p.gvhd.star[k,j,i])
      success.gvhd2[k,j,i] <- step(gcrit[2] - p.gvhd.star[k,j,i])
    }

  }
}
```