Modelling Immune Reconstitution following Paediatric Haematopoietic Stem Cell Transplantation and in HIV-Infected Children

Rollo L Hoare

A thesis submitted to University College London for the degree of
DOCTOR OF PHILOSOPHY

Centre for Mathematics and Physics in the
Life Sciences and Experimental Biology
&
Infection, Inflammation and Rheumatology Section, Institute of Child Health

UNIVERSITY COLLEGE LONDON

October 2015

I, Rollo Lupton Hoare confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Mechanistic mathematical modelling can be used to understand the fundamental drivers of the immune system and how the system is affected by medical interventions. Key to this understanding in children is the interplay between age and treatment-related effects. This thesis focusses on immune reconstitution following paediatric haematopoietic stem cell transplantation (HSCT) and following the start of antiretroviral therapy (ART) in children infected with human immunodeficiency virus (HIV). Since quantitative reconstitution is only one aspect of immune function, in the final chapter I develop a model to explore the dynamics of T cell receptor diversity.

Following HSCT, reconstitution of neutrophils and platelets was modelled using a previous mechanistic model. For CD4 T cell reconstitution, a novel mechanistic model was constructed that included age-related changes in T cell dynamics, the delay to thymic output after HSCT and competition for resources. In HIV-infected children starting ART, a simplified previous model for CD4 T cell and HIV dynamics was adapted to include mechanistic elements for multiphasic viral load decline, age-related changes in T cell dynamics and competition for resources. Using nonlinear mixed-effects modelling with these deterministic models allowed parameters to be estimated with the uneven and often sparse data available. The models were then used to find factors that affect reconstitution. The model for CD4 reconstitution following HSCT was then used to make verifiable predictions of reconstitution in a new cohort of paediatric patients.

T cell receptor diversity dynamics were investigated with a stochastic model in which all T cells compete equally for a global resource. The model was simple enough that numerical simulations could be performed with large numbers of cells and clonotypes, and the model could be characterised analytically. Equations were obtained for long-term mean T cell numbers, clonotype numbers, clonotype size distributions and the Gini coefficient as a measure of dispersion. The model was then extended to model host-donor CD8 memory T cell dynamics in bone marrow transplanted mice, showing that biologically simple assumptions could explain the observed dynamics.

Acknowledgements

Over the years, many people have contributed to making this work possible. I would particularly like to thank:

My supervisors, Robin Callard and Joseph Standing. They have been fantastic over the past years – patient, kind and generous with their time. They have introduced me to a wealth of people, enabled me to enjoy my research thoroughly and taught me about the world of research, and I cannot thank them enough.

Paul Veys for sharing all the HSCT data from Great Ormond Street Hospital, for taking such an interest in this project and for his advice on HSCT. Nigel Klein for his advice with T cell dynamics and for allowing me to use the PENTA HIV data. Grant Lythe with whom I have collaborated on the stochastic model for clonotype dynamics, allowing me into a whole world of mathematics about which I had little previous knowledge. He has been incredibly patient and helpful with my lack of understanding and taught me much. Andrew Yates for receiving me at the Einstein College of Medicine in New York and looking after me there. Ben Seddon and Thea Hogan for allowing me the use of the data from Thea's fantastic mouse experiment and for taking the time to make it available to me. Joanna Lewis who originally set me off in the world of reconstitution and T cell dynamics. Sarah Walker for allowing me the use of the ARROW HIV data. Julie Bertrand, Eva Germovsek and Charlotte Barker for all their advice during the London Pharmacometrics Inerest Group meetings and for making the PAGE conferences such fun. Vania de Toledo for enlivening the office every day and generally organising our lives. Ronan Doyle, Lisa Carter, Liam Shaw, Deji Majekodunmi, Felicity Fitzgerald, Julia Kenny and Hannah Jones for all the fun in the office and at ICH events.

Caroline, Isabella, Brian and François for their support over the years and Johnny and Bella for their patience. Finally I must thank Hannah, without whose help I simply could never have managed this.

Contents

List of Figures

List of Tables

List of Abbreviations

AIDS	Aquired immune deficiency syndrome
APC	Antigen presenting cell
ARROW	Antiretroviral Research for Watoto
ART	Antiretroviral therapy
ATG	Antithymocyte globulin
BIC	Bayesian information criterion
BLQ	Below the limit of quantification
BMT	Bone marrow transplant
cART	Combination ART
CBT	Cord blood transplant
CDM	Clinically driven monitoring
CMV	Cytomegalovirus
CT	Continuous treatment
CWRES	Conditionally weighted residuals
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
EM	Expectation-maximisation
FOCE	First order conditional estimation
GCSF	Granulocyte colony stimulating factor
GvHD	Graft versus host disease
HAART	Highly active ART
HIV	Human immunodeficiency virus
HSC	Haematopoietic stem cell
HSCT	Haematopoietic stem cell transplantation
IMP	Importance sampling
IPRED	Individual prediction
ITS	Iterative two stage
K-PD	Kinetic-pharmacodynamic

LCM	Laboratory and clinical monitoring
LOQ	Limit of quantification
MHC	Major histocompatability complex
NK	Natural killer
NLME	Nonlinear mixed-effects
NNRTI	Non-nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside-analog reverse transcriptase inhibitor
ODE	Ordinary differential equation
OFV	Objective function value
PBSCT	Peripheral blood stem cell transplant
pcVPC	Prediction corrected VPC
PD	Pharmacodynamic
PENTA	Pan-European network for the treatment of AIDS
PTI	Planned treatment interuption
RNA	Ribonucleic acid
SAEM	Stochastic approximation expectation maximisation
SCID	Severe combined innumodeficinecy
SDE	Stochastic differential equation
sp-MHC	Self-peptide MHC
TCR	T cell receptor
TREC	T cell receptor excision circle
VPC	Visual predictive check

Chapter 1

Introduction

1.1 The haematopoietic system

The haematopoietic system comprises all cells produced by haematopoietic stem cells. Haematopoietic stem cells (HSCs) are the progenitors for all blood cells in the body. They largely reside in the bone marrow, particularly in the femur, pelvis and sternum, but they are also found in the peripheral blood and in umbilical cord blood. They are self renewing and multipotent and have the common myeloid progenitor and common lymphoid progenitor as offspring (see Figure 1.1).

The common myeloid progenitor produces red blood cells (erythrocytes) and megakaryocytes, which in turn produce platelets (thrombocytes). Red blood cells are involved in oxygen transport around the body and platelets contribute to blood clotting. The rest of the offspring of the common myeloid progenotor are white blood cells (leukocytes), all of which are involved in the innate immune system. Macrophages and neutrophils are phagocytes that engulf particles or pathogens; basophils and eosinophils are closely related to the neutrophil; and mast cells are associated with wound healing and pathogen defence in mucous membranes and connective tissue.

The common lymphoid progenitor produces the rest of the white blood cells:

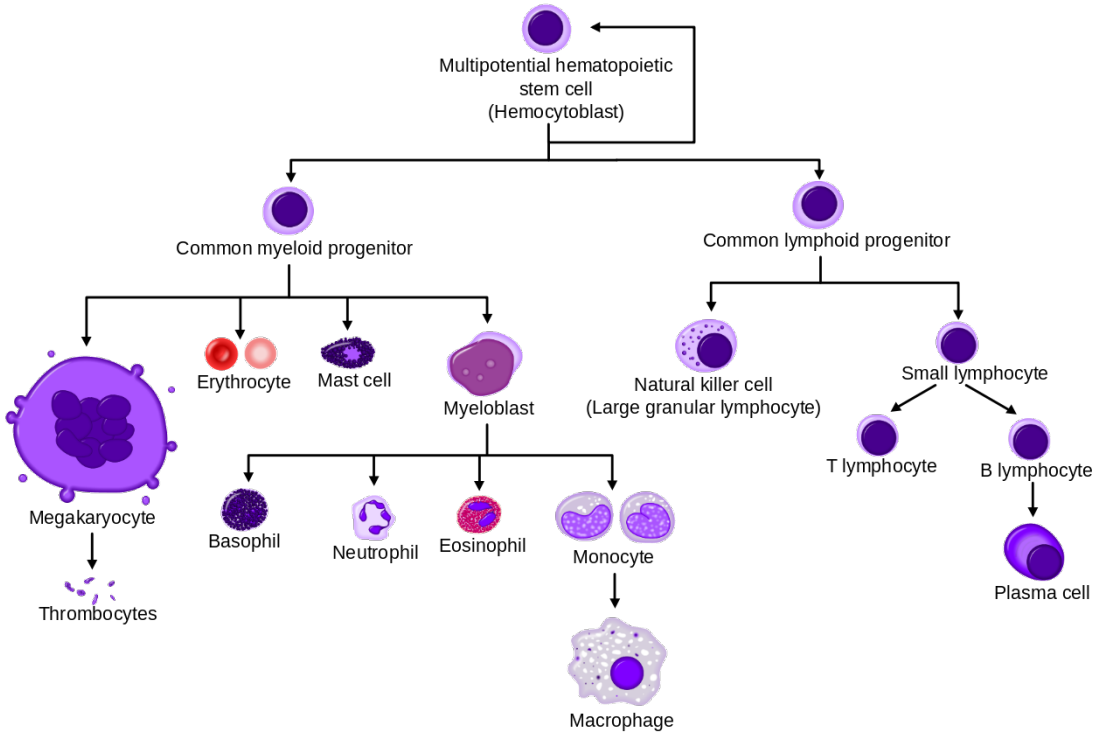

Figure 1.1: The maturation pathway of blood cells in the body from haematopoietic stem cells. Macrophages, eosinophils, neutrophils, basophils, mast cells and NK cells are parts of the innate immune system. T and B cells are part of the adaptive immune system. Erythocytes are red blood cells and megakaryocytes produce platelets for clotting. *Source: Wikimedia Commons.*

natural killer (NK) cells and the lymphocytes, T and B cells. NK cells are a key component of the innate immune system, killing compromised host cells such as virus infected or tumour cells; B and T cells form the foundation of the adaptive immune system.

Dendritic cells are highly heterogeneous antigen presenting cells. They are produced from both the common myeloid and the common lymphoid progenitor [1].

1.2 The immune system

The innate immune system has evolved in animals and plants for immediate protection against infection from other organisms. This first line of defence is a set of non-specific cells and mechanisms, which recognise and respond to many

pathogens in the host in a fast and generic manner. It is largely made up of the complement system and phagocytic white blood cells, such as neutrophils and macrophages. These phagocytes take up a variety of microorganisms into intracellular vesicles where they destroy them with degradative enzymes and other antimicrobial substances.

The adaptive immune system has evolved in addition to the innate immune system in vertebrates, including humans. It is able to respond in a pathogen-specific way. Once the adaptive immune system has encountered and responded to a pathogen in the host, it is able to acquire immunological memory of this pathogen, enabling it to enhance its response to the same pathogen in the future.

The adaptive immune system works by recognising specific antigen proteins — parts of bacteria, viruses and microorganisms — with receptors on the cell surface, the T cell receptor (TCR) and B cell receptor (BCR). T cells require the antigen to be processed inside another cell and presented on the major histocompatability complex (MHC), a cell surface protein. TCRs then recognise this peptide–MHC combination. MHC comes in two types, class I and class II. Class I are found on the surface of most nucleated cells and when infected present peptides from pathogens such as viruses to cytotoxic CD8 T cells, which then kill the infected cell. Class II are mostly found only on antigen presenting cells (APCs) and present peptides to helper CD4 T cells. On recognition of the peptide–MHC combination, the CD4 T cell will provide helper signals to activate the presenting cell [2]. They will also proliferate, differentiate into effector cells and release cytokines to activate other immune cells. Hence CD4 T cells are vital for the adaptive immune system to function fully. The most common APC is the dendritic cell, which are specialist APCs, although macrophages, eosinophils, mast cells, CD8 T cells and B cells can also act as APCs.

In contrast, B cells can recognise antigen in its cognate form with their BCR. On further activation by the CD4 T cells, they will then differentiate into either

memory B cells or into plasma B cells. Memory B cells are long lived cells that have immune memory for the presented antigen peptide. Plasma B cells are large B cells that produce and secrete antibodies in large quantities, which then bind to microbes including bacteria, fungi and viruses, assisting phagocytosis and the activation of the complement system.

T cells, B cells and APCs circulate through the blood and the lymphatic system, made up of the lymph nodes and spleen. Dendritic cells pick up antigen at the site of infection and then migrate to the lymph nodes where T cells are found in very high concentrations. T cells thus sample thousands of peptide-MHC complexes on the APCs every day, on top of which B cells sample cognate antigen directly, ensuring that the immune system has a high probability of encountering pathogen-derived antigen wherever the infection might be in the body. On recognising antigen, T and B cells receive signals to proliferate and differentiate into effector cells and then leave the lymph nodes and spleen in large numbers in order to attack the recognised pathogen. It is the vast diversity of the BCR and TCR repertoires that allows the immune system to respond in a pathogen-specific manner.

In this thesis, I look at three cell types from the haematopoietic system: platelets, which are necessary for haemostasis, neutrophils, vital to the innate immune system and T cells, crucial to the adaptive immune system.

1.3 Platelets

Platelets, or thrombocytes, are fragments of thrombokaryocyte cytoplasm with no nucleus. They are unique to mammals, and their function is to assist coagulation factors in haemostasis, the stopping of a flow of blood at interrupted endothelia.

Platelets gather at the site of interruption and perform primary haemostasis to create a white clot through three processes: adhesion, making bonds outside the

interrupted endothelium; activation, secreting cytokines and activating receptors; and aggregation, attaching to each other through receptor bridges [3]. Primary haemostasis results in the activation of secondary haemostasis and the coagulation cascade. In this process, platelets express thrombin receptors that bind thrombin molecules, which in turn produce polymerised fibrin from soluble fibrinogen in the serum. The fibrin then form a red clot as long strands of insoluble protein, and bind to platelets forming a mixed clot.

Platelets are derived from megakaryocytes in the bone marrow, from where they enter the peripheral blood. A single megakaryocyte can produce thousands of platelets, and in a healthy adult 2×10^{11} platelets are produced per day [4]. Normal platelet concentrations in healthy adults are roughly $200 \times 10^3 / \mu$L, with an average lifespan of 8 to 9 days [4].

1.4 Neutrophils

Neutrophils are a key part of the innate immune response. They are the most abundant of the white blood cells in the body, with 3000 to 5500 cells/μL in the peripheral blood [2]. They are highly motile, and are one of the first-responders in the acute phase of inflammation, particularly in response to bacterial infections, environmental exposure and some cancers. They are attracted to the site of infection through chemotaxis, following cytokines expressed by other activated white blood cells. At the site of the infected tissue, neutrophils are recruited through the induction of adhesion molecules on the endothelial cells of blood vessels and changes to the adhesion molecules expressed on neutrophils. The neutrophils then migrate from the blood vessels into the infected tissue through extravasation, where they survive for 1-2 days [5,6]. Dead neutrophils are the predominant cells in pus, causing its yellow-white appearance.

At the site of infection, neutrophils not only release cytokines, attracting other

white blood cell types and thus amplifying the inflammatory response, but also directly attack pathogens either through phagocytosis, or the release of soluble antimicrobials, or through the generation of neutrophil extracellular traps [7].

Neutrophils are produced in the bone marrow from common myeloid progenitor stem cells through granulopoiesis. This process takes roughly 6.5 days [5], and involves a series of steps, including the formation of granules, eventually leading to the exit of mature neutrophils from the bone marrow. Neutrophils are produced in large numbers, with roughly 10^{11} cells/day leaving the bone marrow [8], and they do not proliferate in the peripheral blood. When circulating in the blood stream, they are short lived, with an average lifespan of 10 – 17 hours [9,10], although this lifespan is increased by steroids, such as glucocorticoid [11].

Neutropoiena, the severe reduction in the concentration of neutrophils, leaves patients highly susceptible to infection with a large range of pathogens, demonstrating the importance of neutrophils in immune defence [2]. Neutrophil precursor production in the bone marrow is dependant on the hormone granulocyte-colony stimulating factor (GCSF), which regulates the production of neutrophil precursors depending on the circulating concentration of neutrophils in the bloodstream. Recombinant human GCSF (rhGCSF) is used as a treatment for neutropoienia.

1.5 T cells

T cells are so called because they develop in the thymus, a small organ near the heart and a part of the lymphatic system.

Pre-T cells leave the bone marrow and travel through the blood stream arriving at the thymus expressing neither CD4 nor CD8. In the thymus they first undergo gene-rearrangement to produce the TCRs. TCRs are formed from one α– and one β–chain or from one γ– and one δ–chain. The β (or δ) chain is formed

first and paired with a pseudo-α (or γ) chain to establish functional rearrangement before pairing with a full α (or γ) chain. The amino acids that make up these chains are formed from variable (V), joining (J) and constant (C) regions, with β– and δ–chains having an additional diversity (D) region. The C region is the same or very similar for all TCRs, and codes for the trans-membrane polypeptides. The DNA encoding the V, J and D regions are composed of many gene segments and the DNA encoding the TCR is then a small number of these segments chosen at random through DNA recombination. This rearrangement results in $\sim 5 \times 10^6$ pairs of combinations of V(D)J regions for α– and β–chains [2]. Further diversity results from junctional diversity, whereby a number of nucleotides can be added or deleted between the gene segments of the V and J regions for α–chains and between the V, D and J regions of the β–chains. This leads to a further diversity of $\sim 2 \times 10^{11}$, resulting in a total of $\sim 10^{18}$ possible TCRs [2]. The DNA in the unused gene segments is excised from the genome and remains in the nucleus as the T cell receptor excision circle (TREC). This piece of DNA is neither replicated in division nor degraded.

Expression of the TCR proteins on the cell surface triggers the expression of both CD4 and CD8, changing them from double negative to double positive thymocytes. These thymocytes then undergo the processes of positive and negative selection. First, cells are positively selected that have a strong enough affinity for self-peptide MHC (sp-MHC) of either class I or II. If they have an affinity for class I, they will drop their CD4 marker and become CD8 cells, and if they have an affinity for class II, they will drop their CD8 marker and become CD4 cells. Those that do not have a strong enough affinity will die. This ensures the released cells will be effective. Then the successful cells are negatively selected for those with too strong an affinity to sp-MHC, with the strongly responding cells forced towards apoptosis. This is to prevent auto-immunity. Thus the cells that survive are in the 'Goldilocks region', with strong enough but not too strong

affinity for sp-MHC [12].

These T cells then leave the thymus and enter the peripheral blood stream and lymphatic systems. At any one time, around 98% of T cells are in the lymphatic system, and only 2% are in the blood stream [13]. The T cells that leave the thymus are naïve T cells, characterised by expression of the cell-surface protein CD45RA. On encountering a foreign antigen that elicits an immune response, the T cells proliferate and differentiate to become effector T cells. After the immune response, some of the T cells are maintained, and become long-lived memory T cells, typically expressing the CD45RO cell-surface protein.

T cell numbers in the body are maintained through homeostatic mechanisms that determine proliferation and death. For T cells to survive and proliferate they require interactions with resources such as cytokines [14,15] and, for naïve cells, sp-MHC [16]. It is thought that there are different thresholds for survival and proliferation whereby a certain threshold number of interactions with resources are required for survival, and a second higher threshold number for proliferation. Insufficient interactions will result in cell apoptosis [17,18]. Hence homeostasis is maintained through competition for these resources [19]; when there are many T cells, there will be few resources per cell, lowering proliferation and raising loss rates, and when there are very few cells, there will be many resources per cell, resulting in low apoptosis and high proliferation. This results in lymphopoenia-induced proliferation [20]. There is evidence to suggest that there is a spread of thresholds within the T cell population, leading to kinetic heterogeneity in proliferation and death rates [21,22].

Both CD4 and CD8 T cells compete for IL-7 cytokines [23–25], while CD8 also compete for IL-15 cytokines [26]. For these cytokines, T cells compete globally. Different clonotypes of TCR will respond to different sp-MHC, although there will be some crossover [27,28]. Naïve T cells thus have intra-clonotype competition for sp-MHC, as well as inter-clonotype competition for sp-MHC with other clonotypes

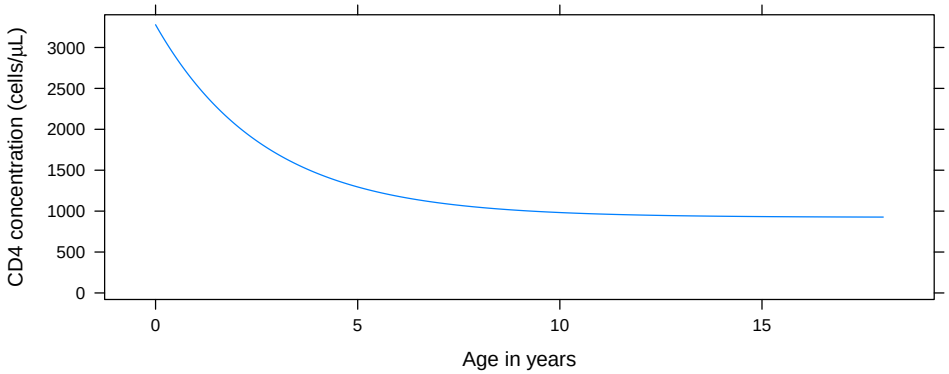

Figure 1.2: The mean expected CD4 concentration with age for a healthy child [31].

that might recognise the same sp-MHC and both naïve and memory T cells have global competition for the cytokines [16,29,30].

Through childhood, the immune system develops rapidly. The thymus reaches full size at 1 year, but the thymic epithelial space involutes by 70% over the first 20 years of life [32]. Because T cells mature in the thymic epithelial space, the number of T cells output by the thymus decreases as the child grows up [33,34]. The concentration of T cells in the blood decreases by a factor of three between 0 and 10 years of age, as can be seen in Figure 1.2 [31]. Also, the proportion of T cells proliferating and dying decreases from early childhood [35,36].

1.6 Immune reconstitution

Medical interventions that cause temporary immunodeficiency or remove a pre-existing condition that causes immunodeficiency are followed by a period in which the immune system recovers (immune reconstitution). The primary theme of this thesis is immune reconstitution following paediatric haematopoietic stem cell transplantation (HSCT). One of the models has been further extended to study immune reconstitution following initiation of antiretroviral therapy in children infected with human immunodeficiency virus (HIV).

1.6.1 Haematopoietic stem cell transplantion

HSCT is used for a range of conditions, which can broadly be split into two categories: red blood cell issues, and immune system disorders. Red blood cell issues include anaemia, Glanzmann's thrombasthenia, and other problems with platelets. Immune system disorders include immunodeficiencies (such as Wiskott–Aldrich syndrome, severe combined innuondeficiency (SCID) or chronic granulomatous disease), leukaemias (myeloid and lymphoblastic), hemophagocytic lymphohistiocytosis, autoimmunities and lymphomas. HSCT is also used as treatment for other diverse conditions, such as inflammatory bowel disease and metabolic defects like Hurler's Syndrome.

Pre-transplant conditioning

Before HSCT, the patient is given a conditioning regimen to eradicate disease and to reduce or ablate the host immune system. This conditioning can be radiotherapy, chemotherapy, anti-lymphocyte antibodies or a combination of the three. This is for three major reasons: (1) to reduce the chance of graft rejection, whereby the remnants of the host immune system attack the donor cells; (2) to lower the rates of graft-versus-host disease (GvHD), whereby the donor immune system attacks the host immune system and the host body; (3) in the case of cancers, to remove cancerous cells and to lower the chances of relapse.

The drugs used in pre-HSCT conditioning and their modes of action are given in Table 1.1. While the anti-lymphocyte antibodies bind to and attack the lymphocytes directly, the mode of action of the chemotherapy drugs and total body irradiation is to prevent the production of new cells. The anti-lymphocyte antibodies will therefore ablate the T cells and B cells, the long lived cells of the haematopoietic system, and the chemotherapy and radiotherapy will ablate the short-lived cells such as neutrophils and platelets as well as the haematopoietic stem cells (HSCs).

Table 1.1: Conditioning regimens and prophylactic drugs, the drug types and their modes of action, adapted from the version written for publication in Barker *et al* [37].

	Drug	Drug type	Mode of action
Conditioning regimen	Alemtuzumab	Monoclonal antibody	Binds to CD54, expressed on the surface of mature lymphocytes but not on haematopoietic stem cells.
	Anti-CD45	Monoclonal antibody	Binds to CD45, expressed on the surface of mature T cells.
	Antithymocyte globulin	Polyclonal antibody	Antibodies harvested from rabbits injected with human lymphatic cells that attack human T cells.
	Busulphan	Alkylating anti-neoplastic agent	Attacks dividing and resting cells. Cell apoptosis by alkylation creating adenine-guanine cross-links.
	Cyclo-phosphamide	Nitrogen mustard alkylating agent	Attacks dividing and resting cells. Cell apoptosis by attaching alkyl group to guanine bases in DNA.
	Fludarabine	Purine analog	Prevents DNA synthesis by interfering with ribonucleotide reductase and DNA polymerase.
	Melphalan	Nitrogen mustard alkylating agent	Attacks dividing and resting cells. Cell apoptosis by attaching alkyl group to guanine bases in DNA.
	Treosulphan	Alkylating anti-neoplastic agent	Attacks dividing and resting cells. Cell apoptosis by alkylation creating adenine-guanine cross-links. Lower toxicity version of busulphan.
Prophylaxis	Cyclosporine	Immuno-suppressant	Lowers T cell immune activity. Prevents IL-2 transcription by binding to lymphocyte cyclophilin, this complex then inhibits calcineurium.
	Methotrexate	Antimetabolite	Mainly suppresses fast proliferating cells. Purine base synthesis inhibited through reduced metabolism of folic acid.
	Mycophenolate	Immuno-suppressant	Inhibits monophosphate dehydrogenase which controls guanine monophosphate synthesis rate in purine base synthesis for B and T cell proliferation.

Mechanism of transplantation

HSCT is the transfer of haematopoietic stem cells from a donor to a host. The HSCs can be extracted from the donor using three methods: (1) bone marrow transplants (BMTs) use stem cells extracted directly from the donor bone marrow, usually from the hip bone with a needle; (2) peripheral blood stem cell transplants (PBSCTs) use circulating stem cells from the peripheral blood of the donor following the administration of granulocyte-colony stimulating factor (GCSF) to stimulate stem cells from the bone marrow into the peripheral blood; (3) cord blood transplants (CBTs) use stem cells in blood taken from the umbilical cord of newborn babies. Sometimes more than one umbilical cord is used to increase the number of stem cells in the graft.

Donor haematopoietic stem cells are transferred into the blood stream of the host. From there, they make their way to the bone marrow where they start to proliferate and re-populate the haematopoietic system. The reconstitution of cells such as neutrophils and platelets is fast, taking a matter of weeks, but is slow for T and B cells, taking months to years for reconstitution.

Allogeneic donors are donors with different genes to the patient and are classed as related, unrelated, family, sibling or haploidentical (sharing half the genes with the host). Donors can also be syngenic, as in identical twins. HSCT can also be done autologously, whereby HSCs are taken from the patient, stored while the patient undergoes treatment, and re-administered after treatment.

Complications following HSCT

Apart from disease relapse and graft rejection, after HSCT another other major complication is graft-versus-host disease (GvHD). GvHD occurs when the donor cells recognise the body of the host as 'non-self' through the sp-MHC interactions and mount an immune response to the host cells. GvHD is classed as acute if it manifests in the first 100 days after HSCT, and chronic after 100 days. In order to prevent or decrease the likelihood of GvHD, donors are sought whose cells are more likely to recognise the body as self. This is achieved by selecting donors that have similar MHC proteins through the process of human leukocyte antigen (HLA) matching. The HLA system is the loci of the genes which encode the MHC proteins. HLA class I matching is considered more important, as this encodes the sp-MHC combination recognised by cytotoxic CD8 cells, but donors are preferable that also match with HLA class II. Selected donors almost always match on HLA class I, and are then classed as 'matched' or 'mis-matched' depending on the quality of the HLA class II matching.

In order to further moderate the effects of GvHD, after the transplant and during the reconstitution the patient can be further treated with prophylactic drugs. The major drugs used in post-HSCT prophylaxis and their modes of action are also given in Table 1.1. The drugs either lower the activity of the present immune cells or reduce the rate of production of new immune cells. These however leave the patient immunocompromised for longer.

While a patient is immunocompromised, they are susceptible to opportunistic

infections such as fungal infections, acute viral infections (e.g. influenza) and the re-emergence of latent infections such as adenovirus, cytomegalovirus (CMV) and Epstein Barr virus (EBV), which can also be introduced by the donor. These infections are a major cause of death in patients undergoing HSCT. If the patient's immune system does not stage a full recovery, they may need further interventions, such as treatment with GCSF or repeat HSCT. Understanding the rate and extent of immune reconstitution is thus of vital importance.

Prognosis following HSCT

According to the Worldwide Network for Blood and Marrow Transplantation, 2006, there were 50,417 patients undergoing their first HSCT worldwide in 1327 centres in 71 participating countries [38], of which 21,516 were allogeneic transplants (43%) and 28,901 were autologous (57%). In the UK in 2013, there were 3840 HSCTs according to the British Society for Blood and Marrow Transplantation [39]. Of these, 370 were in children, 287 with allogeneic grafts and 83 with autologous grafts.

A survey by the paediatric diseases working party of the European Group for Blood and Marrow Transplantation of 31,713 children between 1970 and 2002 found a cumulative incidence of transplant related mortality at day 100 and at 2 years for children given allo-HSCT was 13% and 21%, respectively [40]. This does not include disease related mortality.

1.6.2 HIV-infected children undergoing ART

Human immunodeficiency virus (HIV) attacks cells that express the CD4 protein on their surface, mainly infecting CD4 T cells. Over a long period of time, HIV causes a decline in CD4 T cell concentration, leaving patients immunocompromised and hence vulnerable to opportunistic infections. If left untreated, this leads to acquired immunodeficiency syndrome (AIDS) and eventually death, usu-

ally from infection. Antiretroviral therapy (ART) is the standard treatment for adults and children infected with HIV. ART uses many mechanisms to suppresses HIV replication, reducing viral load often to undetectable levels. This allows CD4 T cells to reconstitute, but the reconstitution is slow, taking between one and two years.

HIV is however never fully eradicated, with HIV surviving in niches of the body, and so patients have to remain on ART for the rest of their lives. This is problematic, particularly in HIV-infected children that may be receiving ART for many years. Not only can long-term toxicities hinder the child's development, but also ART and the resultant monitoring is expensive. It is therefore of interest to understand the effects of planned treatment interruptions and of less intensive monitoring of the children.

1.7 Why use mathematical modelling?

Modelling is particularly useful in longitudinal datasets, such as the ones largely used in this work. By fitting curves to longitudinal data, modelling makes it possible to find rates and long-term averages of immune reconstitution for patients that have variable data. Furthermore, models can pick up general trends in the data that may not be obvious otherwise, and thus allow a more robust analysis of the factors that affect these trends. The majority of the data used in this work are routine clinical patient data from hospitals. These data are highly variable, and are often sparse and uneven, making analysis difficult. In children, because of the developing immune system, there are rarely sufficient children of any one age to do like for like comparisons. On top of that, treatment regimens are rarely identical, making analysis of the factors that affect recovery yet more complicated.

Empirical modelling, where a curve is selected that matches the trajectory seen in the data and then fitted to the data, is the most frequently used form of

modelling for longitudinal data. This method has limitations, with parameters that are often difficult to interpret, and does not make use of all the available information. In contrast, the work presented here centres around mechanistic modelling, whereby models are intended as a direct mathematical abstraction of the system's biology. The intention is to cut through to the relevant underlying biology of the system, then to abstract this biology into a mathematical framework to construct a model. Using this approach allows the direct interpretation of parameters that reflect components of the biological system and hence, when these models are fitted to data, it allows sensible inferences to be taken from parameter values. Furthermore, mechanistic modelling allows the use of known information about the fundamentals of the system that is being modelled. The immune system is a complex interplay between many components, often relying on cascades for cell production, with competition for resources amongst cells, and a large inter-cell regulation network. The situation in children is yet more complex, with concentrations of different cells changing dramatically and non-linearly with age as the immune system develops.

A further advantage of models is that they allow for extrapolation and prediction. Early data can be used to form a predicted curve, which can then be used to make a prediction for the long-term future of that patient. Mechanistic models allow more confidence in these extrapolations because the biological basis of the model helps to keep the predicted curves within a biological range. Hence, immune reconstitution is particularly well suited to mechanistic modelling.

In this work, two different types of mechanistic modelling are used which have different bases. A stochastic agent-based model is used where each cell is modelled individually, and an event is the division into two cells or death of that cell. These events happen randomly, with certain probabilities per unit time, in a manner very much like they would in the actual body. Hence it is a direct, albeit simplified, representation of the system. Overall dynamics of the

system are then modelled through the outcome of the many events for many cells. However, these models are computationally expensive and so are not well suited to large numbers of cells. As such a useful approximation is to use deterministic models. For these, it is assumed that because there are many cells, so many events are happening at any one time that they can be treated as continuous and deterministic. Accordingly these systems can be modelled using ordinary differential equations.

1.8 Aims of the project

The aim of this project was to investigate with mechanistic modelling immune reconstitution following paediatric HSCT and in HIV-infected children starting ART.

The bulk of this work centres around deterministic modelling of cell concentration data. For paediatric HSCT, the reconstitution of three cell subsets were modelled, platelets, neutrophils and CD4 T cells. For HIV-infected children, CD4 T cells were modelled in conjunction with HIV viral load. The general workflow for each cell subset was to:

1. Compile the relevant data
2. Construct and develop the mechanistic model
3. Apply the model to the data using non-linear mixed-effects modelling
4. Perform covariate analysis to find the factors that affect reconstitution.

Further, predictions of reconstitution were tested for CD4 T cell reconstitution following HSCT.

The rest of this work is set out as follows: Chapter 2 describes the pharmacodynamic modelling of short-term reconstitution of neutrophils and platelets. In this chapter, the main statistical methods that are also used in Chapter 3 and

Chapter 4 are presented. Chapter 3 looks at long-term reconstitution of CD4 T cells, and presents a novel mechanistic model and its applications. Chapter 4 analyses HIV-infected children commencing ART by combining a model for viral load dynamics with mechanistic elements from the model presented in Chapter 3. Chapter 5 presents a stochastic global-competition model for T cell homeostasis, and the effects of inter-clonotype competition for resources.

Chapter 2

Short-term neutrophil and platelet reconstitution following HSCT

2.1 Introduction

Neutrophils are vital for immune defence and neutropaenia (the lack of neutrophils) leaves a patient highly susceptible to infection from a large range of pathogens. A study found that before neutrophil engraftment following HSCT (defined as a neutrophil concentration of greater than 500 cells/μL for three consecutive days, usually in the first 30 days post HSCT), the incidence rate of bloodstream infection was 22% with a 12.5% mortality in infected patients, while in the months after engraftment, the infection rate was 19.5% with a mortality of just 1.7% [41]. Another study found that over 64% presented fever within 30 days of HSCT, of which 26% had a clinically proven infection and 12% resulted in death [42]. As a result, while a patient is neutropaenic, they have to remain in isolation rooms in hospitals in order to reduce the chances of an infection that they cannot fight. Furthermore, they are given antibiotics to help fight infections. Understanding

the rate of reconstitution of neutrophils and what affects it is therefore of major importance.

Platelets are a key component of haemostasis and patients with delayed platelet recovery following HSCT can require transfusions of blood products to alleviate the loss of platelets [43]. This requires significant medical resources and carries risk for the patient. Recently, recombinant human thrombopoietin has been found to be an effective therapy for delayed platelet engraftment [44]. A study from 1996 found that delayed platelet recovery is associated with decreased overall survival [43], with mortality from transplant-related complications at 30% for patients who had platelet reconstitution of greater than 20,000 /μL by day 60 and 56% for patients who did not following allogeneic transplants. Another more recent study also found that delayed platelet recovery was associated with one year survival rates, with survival found to be 77% for patients who had platelet reconstitution above 50,000 /μL and 59% for those who had platelet reconstitution below 50,000 /μL [45]. This study also found that transplant related mortality was higher in patients who had poor reconstitution of platelets at 30% versus 11% [45]. Hence understanding the factors that are associated with slow platelet reconstitution following HSCT is of importance.

Neutrophils and platelets have short mean lifetimes in the peripheral blood (10 – 17 hours [9,10] and 8 to 9 days [4] respectively). This means to maintain equilibrium, neutrophil and platelet production rates are also high, and so reconstitution is fast, taking a few weeks for full reconstitution. Because this time-period is short, intra-individual differences due to age for each measurement can be ignored. This allows mathematical models that were developed in adults to be applied, with any effects from differences in age between children accounted for by inter-individual differences in their random effects.

Neutrophil and platelet concentrations decline on the use of pre-HSCT conditioning as the chemotherapy and radiotherapy prevent the synthesis of new cells,

causing neutropaenia and thrombocytopaenia (the lack of platelets). There is a time delay from drug administration to the measured decline in neutrophils and platelets because measurements are taken of circulating cells and there will be cells that have already been produced in the bone marrow that have not yet reached the circulation (see Figure 2.3). This same delay from production to appearance in the circulation means that concentrations continue to decline for a few days after HSCT. The concentrations then reach a nadir, after which the they increase with reconstitution. Pharmacodynamic models then try to find the relationship between the drugs used and these concentration profiles.

2.2 Modelling short-term reconstitution

In the section below, I discuss the mechanistic mathematical models for neutropaenia and thrombocytopaenia that have previously been proposed.

2.2.1 Neutrophils

Neutropaenia and the subsequent reconstitution has been widely studied as the most common adverse dose-limiting toxicity of chemotherapy drugs [46]. Early quantitive analysis identified a minimum desirable neutrophil concentration of 1×10^9 cells/μL, below which the number of days of infection increases sharply [47]. Because neutrophil concentration measurements in oncology clinical trials are often not frequent enough to pick up the actual nadir, other analyses used summary variables to describe the extent of neutropaenia that also include information on its duration. These include time to nadir [48], the number of days spent below a certain neutrophil concentration, or the area between the curve of neutrophil concentration and a certain fixed concentration [49].

Summary variables, however, inevitably waste information contained in the data, whereas whole time-course modelling makes use of more of the available

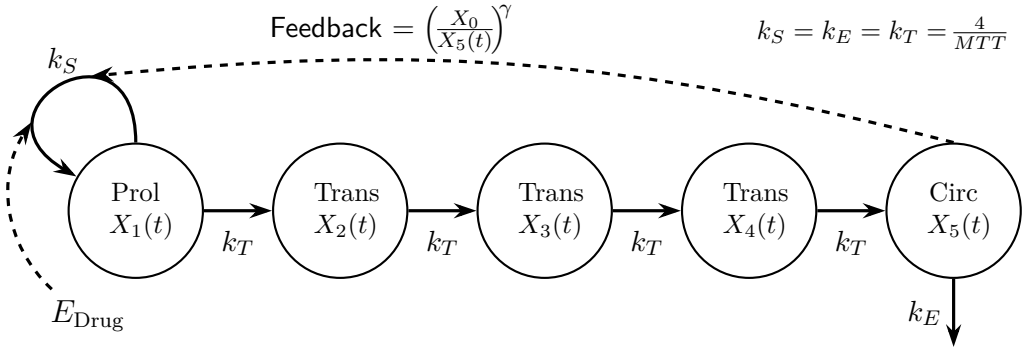

Figure 2.1: The basis of the model used for neutrophil reconstitution, from Friberg *et al* [52]. Cells are produced in the proliferationg compartment X_1 and then transfer through the transit compartments $X_{2\to4}$ to the circulating compartment X_5 where they are measured. Because of the feedback loop, the proliferation rate in the proliferating compartment is affected by the circulating concentration of neutrophils, with feedback strength given by γ. In this model, the elimination rate k_E, synthesis rate k_S and transfer rate k_T are all equal to $\frac{4}{MTT}$ where MTT is the mean transfer time of the neutrophils. The drug acts with strength E_{Drug} to reduce the production rate in the proliferating compartment.

data. Empirical models break up the time-course; one used three sections, a horizontal line for baseline concentration, a line with negative gradient for the decrease in concentration up until the nadir and a logistic curve for reconstitution [50]; another used a cubic spline function with three break points [51].

Mechanistic models are advantageous because of their more interpretable parameter values and their greater predictive usefulness. To make it possible to fit these models to data however they need to be simple enough with few enough parameters such that the parameters can be estimated. Mechanistic models of neutropaenia all have common features, with cells produced in one compartment, representing the bone marrow, and then some sort of maturation of these cells as they move to the circulating compartment, where cell concentrations are measured. This maturation causes a time delay from drug administration to the changes in observed concentration, which was explained in the mechanistic models using either a time-lag [53] or transit compartments [54–56]. These models culminated in the model of Friberg *et al* [52], a schematic of which is shown in Figure 2.1.

The Friberg model

The model of Friberg *et al* [52] differed from the previous models by having both self-renewal of cells in the bone marrow compartment (rather than a zero-order input of cells), and a feedback loop whereby the concentration of circulating cells affects the rate of cell self-renewal in the bone marrow in order to model the effects of granulocyte-colony stimulating factor (GCSF). This model has five compartments: a proliferating compartment $X_1(t)$, representing stem cells in the bone marrow, three transit compartments $X_{2\to4}(t)$, representing stages of neutrophil development before they appear in the peripheral blood, and a circulating compartment $X_5(t)$, representing the concentration of circulating neutrophils in the peripheral blood. The model has four parameters to be fitted: X_0, the initial and long-term steady-state concentration of neutrophils in the absence of drugs; $MTT = \frac{4}{k_S} = \frac{4}{k_T} = \frac{4}{k_E}$, the mean transfer time for neutrophils through the transit compartments; γ, the strength of the feedback effects from GCSF; and E_{Drug}, the effect of the myeloablative drugs. The equations for the dynamics with time t are given by,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}X_1(t) &= k_S X_1(t)\left(\frac{X_0}{X_5(t)}\right)^{\gamma}(1-E_{\text{Drug}}) - k_T X_1(t) && X_1(0) = X_0 \\
\frac{\mathrm{d}}{\mathrm{d}t}X_2(t) &= k_T X_1(t) - k_T X_2(t) && X_2(0) = X_0 \\
\frac{\mathrm{d}}{\mathrm{d}t}X_3(t) &= k_T X_2(t) - k_T X_3(t) && X_3(0) = X_0 \\
\frac{\mathrm{d}}{\mathrm{d}t}X_4(t) &= k_T X_3(t) - k_T X_4(t) && X_4(0) = X_0 \\
\frac{\mathrm{d}}{\mathrm{d}t}X_5(t) &= k_T X_4(t) - k_E X_5(t) && X_5(0) = X_0. \qquad (2.1)
\end{aligned}$$

With this model, the system starts at steady state concentration, X_0, and then for the days where E_{Drug} is non-zero, the concentration falls. With $k_S = k_T = k_E$, the concentrations are the same in all five compartments at steady state, but on perturbation by the drug effect, there is a delay to the dynamics of the measured

circulating compartment.

After the publication of the model of Friberg *et al*, two further models for neutropaenia were proposed, which were largely based on the previous models. The model of Panetta *et al* [57] is very similar to that of Friberg but differed by having two transit compartments and an altered form for the feedback loop. While more recently, the model of Bulitta *et al* [58] uses time-delay differential equations. These two models were compared by Soto *et al* [59] to the models of Friberg *et al* [52], Minami *et al* [53] and Zamboni *et al* [56]. Soto *et al* found that none of the models showed superior performance in comparison to the original Friberg model.

This chapter describes the use of the Friberg model for neutrophil dynamics. This model has three major advantages over other models. Firstly, other models do not contain the feedback loop, modelling the rebound in neutrophil concentration. Secondly, delay-differential equation models are difficult to code into software and are unstable when fitting to data. Thirdly, more complicated models involve more parameters to estimate but have not shown significantly improved performance. I allow the long-term neutrophil concentration X_∞ to differ from the initial neutrophil concentration X_0 in order to allow for the differences resulting from the donor haematopoietic system following HSCT. Furthermore, I test the effects of allowing the elimination rate k_E to differ from the synthesis and transfer rates K_S and k_T. Differences between these rates were not visible in the sparser data on which this model was originally developed, but might be with the richer data used for this analysis.

Modelling the effects of steroids

The administration of corticosteroids causes a transient increase in the concentration of circulating neutrophils, lasting for a few hours after the dose and returning to a normal level after about one day [60–62]. This effect has also been observed

in children, and levels remained high for as long as corticosteroids were being administered [63]. The reason for this increase is not fully understood, but the cells are known to be almost entirely mature neutrophils [61], which implies that neutrophils are moving out of the vasculature, mediated by endothelial cell adhesion molecule interactions [64].

This effect has been modelled with an extension to the Friberg model. The method used an additional input compartment that starts with a certain number of cells and releases cells into the circulating compartment at a first-order rate [64]. This method has also been applied successfully to other datasets for neutrophil reconstitution [65,66].

2.2.2 Platelets

Thrombocytopaenia and subsequent platelet reconstitution has not been widely studied through mathematical modelling. The majority of studies into platelet reconstitution have tended to use the Friberg model [67–71]. When fitting this model, these studies made no adaptions to the model to account for the differences between platelet and neutrophil dynamics, but they manage a good fit to the data. When using the Friberg model, the transit compartments become the production of megakaryocytes and the subsequent production of platelets.

More recently, Hayes *et al* [72] developed a simpler model for platelet dynamics shown in Figure 2.2. This has a zero-order input of cells into the bone marrow, and has removed the feedback loop. This model has been successfully applied to other datasets for platelets [73–76].

The Hayes model

The model for platelets as developed by Hayes *et al* [72] has four compartments: a production compartment $X_1(t)$, representing stem cells in the bone marrow, two transit compartments $X_{2\to3}(t)$, representing stages of development between

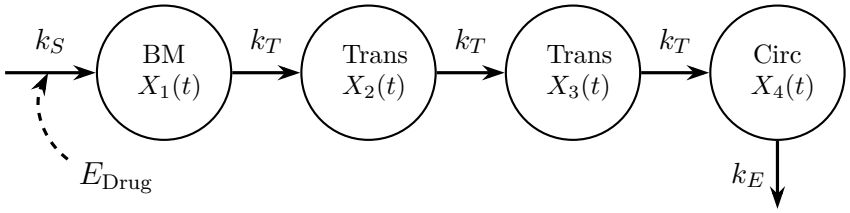

Figure 2.2: The basis of the model used for platelet reconstitution, from Hayes *et al* [72]. Platelets are produced into the bone marrow (BM) compartment X_1 at rate k_S. They then transfer through the transit compartments $X_{2\to3}$ at rate k_T to the circulating compartment X_4 where they are measured. They are then eliminated at rate k_E. The drugs act with strength E_{Drug} to reduce the production of platelets into the bone marrow compartment.

bone marrow and platelet including megakaryocyte production, and a circulating compartment $X_4(t)$, representing the concentration of circulating platelets in the peripheral blood. The model has five parameters: X_0, the initial and long-term steady-state concentration of platelets; k_S, the zero-order input of megakaryocyte precursors into the bone marrow compartment; k_T the transfer rate of cells between the compartments; k_E the death rate of circulating platelets; and E_{Drug}, the effect of the myeloablative drugs preventing cell synthesis. The equations for the dynamics are given by,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}X_1(t) &= k_S(1-E_{\text{Drug}}) - k_T X_1(t) & X_1(0) &= \frac{k_E}{k_T}X_0 \\
\frac{\mathrm{d}}{\mathrm{d}t}X_2(t) &= k_T X_1(t) - k_T X_2(t) & X_2(0) &= \frac{k_E}{k_T}X_0 \\
\frac{\mathrm{d}}{\mathrm{d}t}X_3(t) &= k_T X_2(t) - k_T X_3(t) & X_3(0) &= \frac{k_E}{k_T}X_0 \\
\frac{\mathrm{d}}{\mathrm{d}t}X_4(t) &= k_T X_3(t) - k_E X_4(t) & X_4(0) &= X_0.
\end{aligned} \tag{2.2}$$

Similarly to the model above, the system starts at steady state concentration with $X_4 = X_0$, and then for the days where E_{Drug} is non-zero, the concentrations in the compartments fall, with a delay to the dynamics of the measured circulating compartment. Similarly to neutrophils, I allow the long-term platelet concentration X_∞ to differ from the initial platelet concentration X_0 in order to allow for

the differences resulting from the donor haematopoietic system following HSCT. Because of the inter-dependence of the parameters, one parameter can be defined in terms of the others, with $k_S = k_E X_0$, leaving five parameters to be estimated.

In this chapter, I will compare the results of fitting both the Friberg and the Hayes models to the platelet data.

2.2.3 Conditioning drug protocols & pharmacodynamics

Patients received a large range of protocols for their pre-transplant conditioning. Not only did they receive a range of drugs as well as radiation, they were also given these drugs on different days and in different combinations. Because exact information is not available for every patient in the dataset, a summary table was created from the available protocols for the general method of use of each drug, shown in Table 2.1.

For each patient the information as to which combination of drugs they received was available. This information combined with the protocol summary table was then used to define for each patient the days on which they were expected to receive neutropaenia- and thrombocytopaenia-inducing conditioning treatment.

Table 2.1: Conditioning drugs for HSCT protocols summary. The conditioning regimen is formed from a combination of the therapies listed below. All the therapies except for alemtuzumab and antithymocyte globulin (ATG) are neutropaenia- and thrombocytopaenia-inducing. This table is a summary of the available protocols that were used, with each dot giving the days on which these drugs were normally given.

		Days from HSCT								
Drug	N / T	-9	-8	-7	-6	-5	-4	-3	-2	-1
Alemtuzumab	No		•	•	•	•	•			
Anti-thymocyte globulin	No					•	•	•		
Busulfan	Yes		•	•	•	•				
Cyclophosphamide	Yes					•	•	•	•	
Fludarabine	Yes				•	•	•	•	•	
Melphalan	Yes							•	•	•
Treosulphan	Yes				•	•	•			
Total body irradiation	Yes	•	•	•	•					

N/T: Neutropoenia/thrombocytopoenia inducing conditioning.

Furthermore, methylprednisolone, a steroid, is always given in conjunction with alemtuzumab and ATG. So the patients that were given these drugs were assumed to receive methylpredisone on the days as described in Table 2.1.

Pharmacodynamic models represent how the body is affected by the drug concentration. They are based on receptor theory [77], whereby a drug (D) binds with a receptor (R) to form a drug-receptor complex (DR) that initiates a sequence of events leading to a drug effect, E_{Drug}. The drug, receptor, and drug-receptor complex form a dynamic equilibrium with the concentrations related by,

$$[DR]k_{off} = [D][R]k_{on} \tag{2.3}$$

where $[D]$, $[R]$ and $[DR]$ are the concentration of the drug, receptor and drug-receptor complex respectively, k_{on} is the binding rate of the drug and receptor and k_{off} is the unbinding rate of the receptor-drug complex. If we assume that there are a finite number of receptors, then the total concentration of receptors, $[R_{tot}] = [R] + [DR]$. As a result, the dependence of $[DR]$ on $[R]$ can be removed leaving,

$$[DR] = \frac{[D][R_{tot}]}{[D] + \frac{k_{off}}{k_{on}}}. \tag{2.4}$$

With E_{Drug} proportional to drug-receptor concentration $[DR]$, and $E_{\max}$ the drug effect when all receptors are bound such that $[DR] = [R_{tot}]$, then with drug concentration redefined as C, (2.4) can be re-written as,

$$E_{\text{Drug}} = E_{\max}\frac{C}{C + EC_{50}}, \tag{2.5}$$

where $EC_{50} = \frac{k_{off}}{k_{on}}$ is the concentration at which the effect size is half. This gives a sigmoidal drug effect with concentration whereby either the receptors are in such excess that changes in concentration produce no measurable effect or the drug concentration is in excess and all receptors are used, resulting in no increase

in drug effect.

In the case where the concentration $C \ll EC_{50}$, (2.5) can be approximated as a linear model,

$$E_{\text{Drug}} \approx \frac{C}{EC_{50}} E_{\max} = \text{Slope } C. \tag{2.6}$$

The drug effect E_{Drug} can vary between 0 (fully ineffective) and 1 (fully effective) for both the linear and sigmoidal models.

In this chapter, because I have no information about drug concentrations in the blood, I assume a pharmacokinetic model for drug concentration that follows an exponential decay with time after drug dose, referred to as a K-PD (kinetic-pharmacodynamic) approach [77,78]. I then test the effects of both the sigmoidal and linear K-PD models. The half lives of the neutropaenia- and thrombocytopaenia-inducing conditioning drugs are short: busulfan 2.3 hrs [79]; cyclophosphamide 5 – 9 hrs [80]; fludarabine 8 – 10 hrs [81]; melphalan 1.3 hrs [82]; and treosulphan 1.7 – 2.2 hrs [83]. With the sigmoidal K-PD model I therefore assume that the drug falls well below the EC_{50} within one day and, according to (2.5), the corresponding drug effect therefore falls to zero one day after a dose. For the linear model, I assume that the drug effect decreases in some proportion to the drug concentration and so follows an exponential decay in the days following the last drug administration so that,

$$E_{\text{Drug}} = E_{\max} \mathrm{e}^{\frac{-t-t_D}{T_{\text{Drug}}}}, \tag{2.7}$$

where t_D is the time at which the patient received their last dose of conditioning and T_{Drug} gives the time for which the drug effect persists following this last dose.

2.3 Methods

2.3.1 The data

The data was collected as part of routine clinical practice between 2005 and 2011 from children undergoing HSCT at the Blood and Marrow Transplant Unit at Great Ormond Street Hospital for Children NHS Trust. The parents of the patients whose data is in the database have provided written informed consent for the use of the data. It comprises blood concentrations of many cell types taken at regular intervals for up to seven years after the transplant, including neutrophil and platelet concentrations, analysed in this chapter. Conditioning regimens usually start nine days before HSCT and these cells types reconstitute quickly, so the data was cut to 14 days before HSCT up until 100 days after HSCT for the work in this chapter. The dataset has 299 patients, who have had 337 transplants between them. The demographics of the data are summarised in Table 2.2. As can be seen, there is huge heterogeneity in the diagnoses, patient characteristics and forms of treatment.

For studies of this kind, the data are unusually rich. For the neutrophil dataset, there are 19,118 measurements in the period from 14 days before to 100 days after the transplant, while in the platelet dataset, there are 22,149 measurements in the same time-frame. Plots of the concentration profiles with time from HSCT are shown in Figure 2.3. As can be seen, there is a brief increase in neutrophil concentrations just before HSCT caused by the steroids given to some patients with their conditioning. Concentrations then decrease in response to conditioning and subsequently increase rapidly, returning to a similar concentration to previously. In contrast, for platelets, while the decrease is rapid, the recovery is slower and levels do not on average return to those before HSCT. The local regression can however be misleading as the reconstitution may be fast but with varying delays from HSCT, leading to a gradual increases in the average

Table 2.2: Breakdown of the demographics and the drugs used for the patients in the short-term reconstitution dataset.

	No	%		No	%
Age at HSCT (years)			**Donor type**		
$0 \rightarrow 1$	75	22	Matched	215	63
$1 \rightarrow 2$	62	18	Sibling	93	27
$2 \rightarrow 5$	72	21	Family	18	5.3
$5 \rightarrow 10$	79	23	Unrelated	104	30
$10 \rightarrow$	52	15	Mis-matched	107	31
Sex			Sibling	2	0.6
Male	216	63	Family	6	1.8
Female	126	37	Unrelated	99	29
HSCT			Haploidentical	15	4.4
1^{st}	291	85	Autologous	5	1.5
2^{nd}	47	14	**Viruses**		
3^{rd}	4	1.2	Cytomegalovirus		
Diagnosis			Positive	111	32
Immunodeficiencies	145	42	Negative	227	66
SCID	89	26	Unknown	4	1.2
Wiskott-Aldrich	12	3.5	Epstein Barr virus		
CGD	12	3.5	Positive	88	26
Leukaemia	105	31	Negative	128	37
ALL	50	15	Unknown	126	37
AML	40	12	Adenovirus		
HLH	37	11	Positive	106	31
Anaemia	22	6.4	Negative	236	69
Autoimmune	9	2.6	**Conditioning**		
Lymphomas	5	1.5	Alemtuzumab	165	48
GvHD			Anti-CD45	13	3.8
Reported	102	30	Anti-thymocyte globulin	13	3.8
I	39	11	Busulphan	80	23
II	38	11	Cyclophosphamide	153	45
III	18	5.2	Fludarabine	166	49
IV	7	2.0	Melphalan	101	30
Stem cells			Treosulphan	70	20
Bone marrow	158	46	Total body irradiation	51	15
Peripheral blood	129	38	None	44	13
Cord blood	53	15	**Prophylaxis**		
Combinations	2	0.6	Ciclosporin	298	87
			Methotrexate	72	21
			Mycophenolate	164	48

Abbreviations: SCID: severe combined immunodeficiency sydrome; CGD: chronic granulomatous disease; ALL: acute lymphoblastic leukaemia; AML: acute myeloid leukaemia; HLH: hemophagocytic lymphohistiocytosis; GvHD: graft versus host disease.

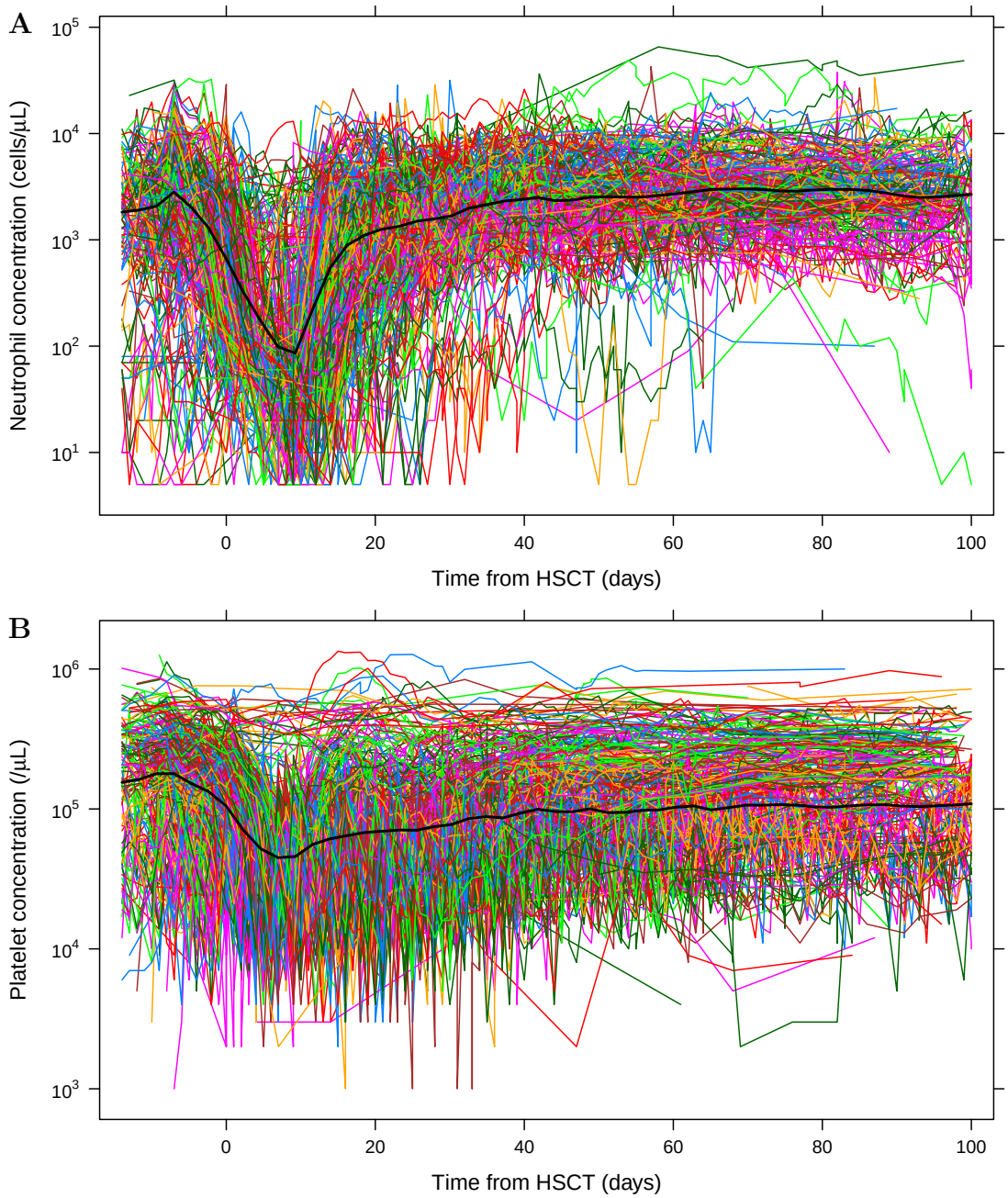

Figure 2.3: Time courses of reconstitution following HSCT for (**A**) neutrophil concentration and (**B**) platelet concentration plotted with a log concentration scale. Each coloured line represents an individual patient's data. The black lines give local regression curves for the data.

of neutrophil and platelet concentrations across the population (see Figure 2.4). An advantage of nonlinear mixed-effects modelling (NLME) is its ability to find parameter estimates that can account for this.

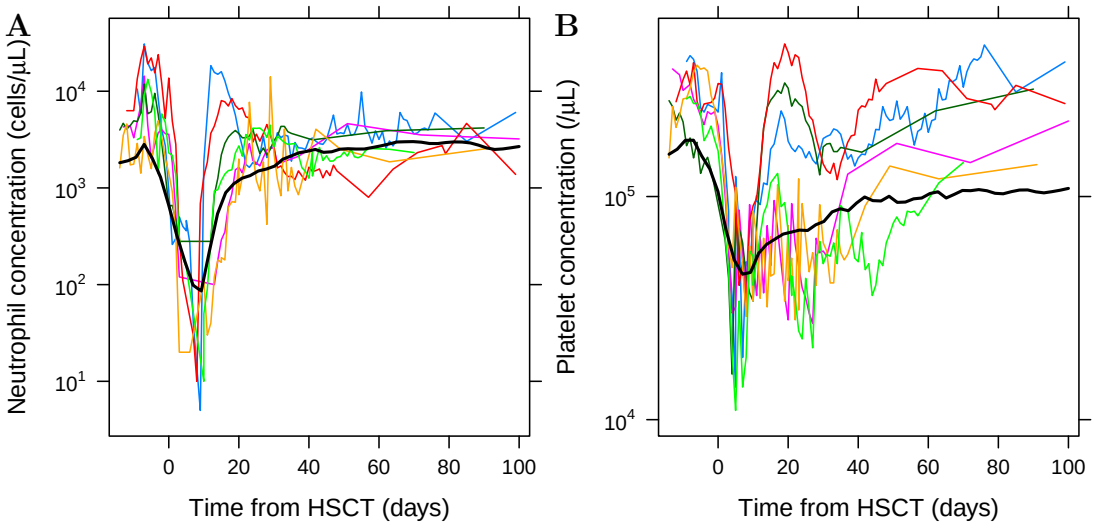

Figure 2.4: Differences between individual dynamics and the local regression for (**A**) neutrophil and (**B**) platelet concentrations. For each dataset, the first six patients in the dataset have been plotted to demonstrate that while the local regression might reconstitute slowly, some of the individual patients reconstitute quickly but with different time delays.

2.3.2 Model fitting

Nonlinear mixed effects modelling

Nonlinear mixed-effects modelling was used to fit the models to the data. This approach involves estimating population typical values, along with multiple levels of variability. In the simplest case, this variability comprises inter-individual variability at the level of the model parameters and residual variability at the level of observations. As a result, this approach is useful because it takes into account the correlation between the measurements of an individual, which is necessary for the uneven data available for this type of analysis.

NLME can be thought of as a hierarchical model [84]. At the highest level, taking a general differentiable function $f(\boldsymbol{\phi}, t)$ dependent on the parameter vector $\boldsymbol{\phi}$ and time t, the fixed effects, which give the population average, are modelled as,

$$y_j = f(\boldsymbol{\phi}, t_j) + \epsilon_j \qquad \epsilon_j \sim \mathcal{N}(0, \sigma^2) \tag{2.8}$$

where y_j is the jth observation of the dependant variable y at time t_j, and ϵ_j is

residual error for each observation, which is assumed to be normally distributed and independent.

At the next level, inter-individual variability can be included in the form of random effects. Random effects allow the parameter values to vary for each individual. They are assumed to be normally distributed around the mean 0 with variance Ω. Thus the vector of parameters for the ith individual $\boldsymbol{\phi}_i$ can be decomposed as

$$\boldsymbol{\phi}_i = \boldsymbol{\beta} + \boldsymbol{b}_i \qquad \boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{\Psi}) \tag{2.9}$$

where $\boldsymbol{\beta}$ is the vector of fixed effects, $\boldsymbol{b}_i$ is the random effect vector for individual i, and $\boldsymbol{b}_i$ is normally distributed according to the variance–covariace matrix $\boldsymbol{\Psi}$. If the parameters in the model are bounded, then the random effects can be log or logit transformed.

Hence the full mixed-effects model for observation y_{ij} at time t_{ij} is given by

$$y_{ij} = f(\boldsymbol{\phi}_i, t_{ij}) + \epsilon_{ij}, \qquad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \tag{2.10}$$

where the residual error, ϵ_{ij} now accounts for measurement error, model misspecification and noise. The residual error ϵ_{ij} is assumed to be normally distributed with variance σ and to be independent. The elements of the parameter vector are referred to as θs, the variances of the random effects as Ωs and the variance on the residual error as σ.

Fitting NLME models

Parameter estimation was carried out using NONMEM 7.3 [85] This has the capability to use various algorithms to maximise the likelihood by minimising the negative log of the total marginal density. The marginal density of $\boldsymbol{l}$ is given by

$$p(\,\boldsymbol{l}\,|\,\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Psi}) = \int p(\,\boldsymbol{l}\,|\,\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2)\;\; p(\,\boldsymbol{b}\,|\,\boldsymbol{\Psi})\; d\boldsymbol{b}, \tag{2.11}$$

where $p(\boldsymbol{l} \mid \boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{\sigma}^2)$ is the conditional density of $\boldsymbol{l}$ given the random effects $\boldsymbol{b}$, and $p(\boldsymbol{b} \mid \boldsymbol{\Psi})$ is the marginal distribution of $\boldsymbol{b}$. The most simple and widely used algorithm, first order conditional estimation (FOCE), uses a gradient-based method based on a first order Taylor expansion of the model function. This method can however struggle with more complex models, being computationally expensive and getting stuck in local minima.

Other algorithms use expectation-maximisation (EM) methods, whereby a two-step process is used to find the optimum parameters. These methods are more stable for complex models such as those with many parameters and random effects, unstable model output or local minima because they are sampling-based rather than gradient-based methods. The algorithm alternates between an expectation (E) step, whereby a function is created for the expectation of the log-likelihood using the current parameter estimates, and a maximisation (M) step, whereby parameter values are calculated that maximise the expected log-likelihood found in the E step. In the case of NONMEM, in the E step, the parameters (θs), random effect variances (Ωs) and residual error variance (σ) are fixed, while for each individual expected values and variances of the random effects (ηs) are evaluated. Then in the M step, the θs, the Ωs and σ are updated using these expected values of the ηs. EM methods have the advantage that fitting for all off-diagonal elements in the variance-covariance matrix of random effects takes no longer than just fitting for the diagonal elements.

The simplest of these methods is iterative two stage (ITS), where the the E step is the same as FOCE, and the conditional mode and the first order approximation of the ηs are found by maximising the posterior density.

The E step in importance sampling (IMP) evaluates the conditional mean and variance of ηs through Monte Carlo sampling. It uses the posterior density, which incorporates the likelihood of parameters relative to θs and ηs with the individual's observed data. For each iteration, the normal density near the

mean of the posterior (from the previous iteration) is used as a proposal density. From this, Monte Carlo samples are generated, and conditional means and variances evaluated. In the M step, population parameters are then updated from subjects' conditional mean parameters and variances by single iteration maximisation steps.

Finally, in stochastic approximation expectation maximisation (SAEM), similar to IMP, for the E step, random samples of the ηs are generated from normal distributions, but instead of being centred at the mean of the posterior density (like IMP), the sample is centred on the previous sample position. New samples are accepted with a certain probability. This method uses two phases: a burn-in phase, where SAEM evaluates a highly stochastic approximation of individual parameters, and in the M step population parameters are updated from subjects' conditional mean parameters and variances by single iteration maximisation steps. Then an accumulation phase, where individual parameter samples from previous iterations are averaged together, converging towards the true conditional parameter means and variances.

FOCE and Importance sampling can also be carried out in an 'Expectation Only' form, whereby multiple iterations of the expectation step are carried out. This is useful for finding an objective function following SAEM, or for fitting to a new individual, whilst keeping population parameters fixed.

The objective function value (OFV) produced by all methods in NONMEM apart from SAEM is $-2\ln(\text{likelihood})$ and is thus log multivariate normal. Thus the difference between the OFVs of two separate models is the division of two multivariate normals. If the difference between two models involved the addition of k parameters, and these parameters are nested, the difference is approximately χ^2_k distributed. This means for the addition of one parameter, an improvement in OFV of 3.84 points is equivalent to $p < 0.05$. Following model fitting with SAEM, an expectation only process needs to be carried out, whereby population-

level parameter means and their variance-covariance matrix are held fixed while individual-level parameters are calculated and an OFV is produced. This is usually done with FOCE or importance sampling.

2.3.3 Covariate model-building

In order to find what affects the reconstitution, factors contained in the dataset can be tested as covariates. These factors can be continuous, such as age, or categorical, such as diagnosis, drugs, donor type or stem cell source. Categorical covariates are assumed to be independent of time for each individual. These covariates are then tested to ascertain whether they significantly affect reconstitution.

Nesting

Covariates are added into the model by including an extra parameter, the covariate coefficient. These covariate coefficients alter the fixed effects parameters for the different individuals in the population according to the values of the covariates. For continuous covariates, there is only ever one coefficient for each covariate. For categorical covariates, if there are more than two possible states for the covariate, one extra coefficient is added for each of the extra states beyond the most common state. The coefficients are included in nested models whereby if the coefficient equals zero, it makes no change to the parameter to which it is applied. The null hypothesis is therefore that the coefficient equals zero. With the coefficients in nested models, the likelihood ratio of the model with and without the covariate will be asymptotically χ^2_n distributed with n degrees of freedom, where n is the number of new coefficients included in the model. Hence, the significance of the covariate can be calculated.

Stepwise covariate model building

Covariates are included into the models using stepwise covariate model-building (SCM) [86]. In SCM, firstly a forward selection is carried out, whereby covariates are added into the model one by one for each parameter and the covariate that causes the largest improvement in the fit of the model is kept in, as long as it meets the minimum criterion of a p-value $p < p_1$ where p_1 is often in the region of 0.05. Then all the remaining covariates are tested again for each parameter, and similarly that which causes the largest improvement is kept in so long as it meets the criterion and so on until all covariates have been tested and no more offer a large enough improvement. Then SCM carries out a backwards selection whereby covariates that have been included in the forward selection are removed one by one, and those that do not cause a loss in model fit that meets the stricter criterion of $p < p_2$, where $p_2 < p_1$, are then removed. These criteria can be adjusted to make selection more or less likely.

2.3.4 Data below the limit of quantification

Tests for the quantities of biomarkers such as neutrophil, platelet and CD4 concentrations and viral loads have a limit of quantification (LOQ) below which the measurements become unreliable. These data are referred to as below the limit of quantification (BLQ) data, and are usually reported as the LOQ for that assay. There are three major methods for handling data of this form. BLQ data can be omitted from the dataset, the BLQ data can be substituted with LOQ/2 for the purposes of modelling, or a method assessing the likelihood of data classed at BLQ as being BLQ can be used. Beal [87] introduced a series of methods to handle BLQ data, which were compared to each other by Ahn *et al* [88] and to data omission and substitution by Bergstrand and Karlsson [89]. They found using the so-called M3 method produced the least biased fit. In this method, data are assessed for their likelihood to be below the LOQ given the parameter values.

Run-times for this method, however, are slow and one improvement is to treat data reported as above the LOQ as observations, and only to assess data reported as BLQ for its likelihood to be BLQ [90]. Furthermore, it is only necessary to use this method when a significant proportion of the data is BLQ, otherwise, LOQ/2 can be used as a useful approximation.

In this chapter, because only a small proportion of these data are BLQ, I substitute BLQ data with LOQ/2.

2.3.5 Diagnostic plots

Alongside changes in OFV, quality of fit for the models to the observed data is assessed with diagnostic plots [91]. Diagnostic plots can give information related to model biases and model mis-specification.

Residuals plots

The residual is the difference between the observed data and the individual model trajectory for each data point. These differences are usually normalised by dividing by the standard deviation of that individual's data to produce weighted residuals. During fitting, models are linearised and this linearisation is conditioned around the post-hoc individual empirical Bayes estimates of the inter-individual random effects. As a result, the standard deviation used in the normalisation needs to use the FOCE approximation, producing conditional weighted residuals (CWRES) [92]. These are given by,

$$\text{CWRES} = \frac{\mathbf{y}_i - E_{\text{FOCE}}[\mathbf{y}_i]}{\sqrt{\mathbf{Cov}_{\text{FOCE}}(\mathbf{y}_i)}} \tag{2.12}$$

where $\mathbf{y}_i$ is the observed data for individual i, $E_{\text{FOCE}}(\mathbf{y}_i)$ is the individual model prediction for patient i and $\mathbf{Cov}_{\text{FOCE}}(\mathbf{y}_i)$ is the variance of the inter-individual random effects.

Residuals plots can be used to assess model mis-specification. Firstly, the CWRES should be normally distributed with mean 0, variance 1 and 95% of the CWRES between ±2, demonstrating that normality assumptions are valid. Secondly, CWRES are assumed to be independent and so by plotting them against time, population model predictions or other factors such as age it is possible to assess inherent biases in the model with any of these variables.

Individual and population predictions

Plots of the model predictions at the time points at which there are observed data against the observed data themselves are also used to assess model mis-specification. These can be done with both population-level and individual-level predictions. Individual predictions (IPRED) demonstrate whether there are mis-specifications in the model at either high or low observed data and non-normality.

Visual predictive checks

Whilst the previous diagnostics asses whether the model fits the observed data, visual predictive checks (VPCs) are used to assess whether data simulated from the model matches the observed data, both in the overall trends and in their variability [93,94]. Using the population level parameter means and the variance covariance matrix of the random effects, a large number (usually between 300 and 1000) of data points are simulated from the model for each data point in the observed data, with the simulated data having the same characteristics (such as covariates and ages and times at data observations) as the observed data. The data are binned by, and plotted against, an independent variable such as time and the median and percentiles of the observed data in these bins are then compared to the corresponding prediction interval of the median and percentiles of the simulated data. Significant differences between medians of the bins can demonstrate issues with the fixed effects, while the differences in the extent of the

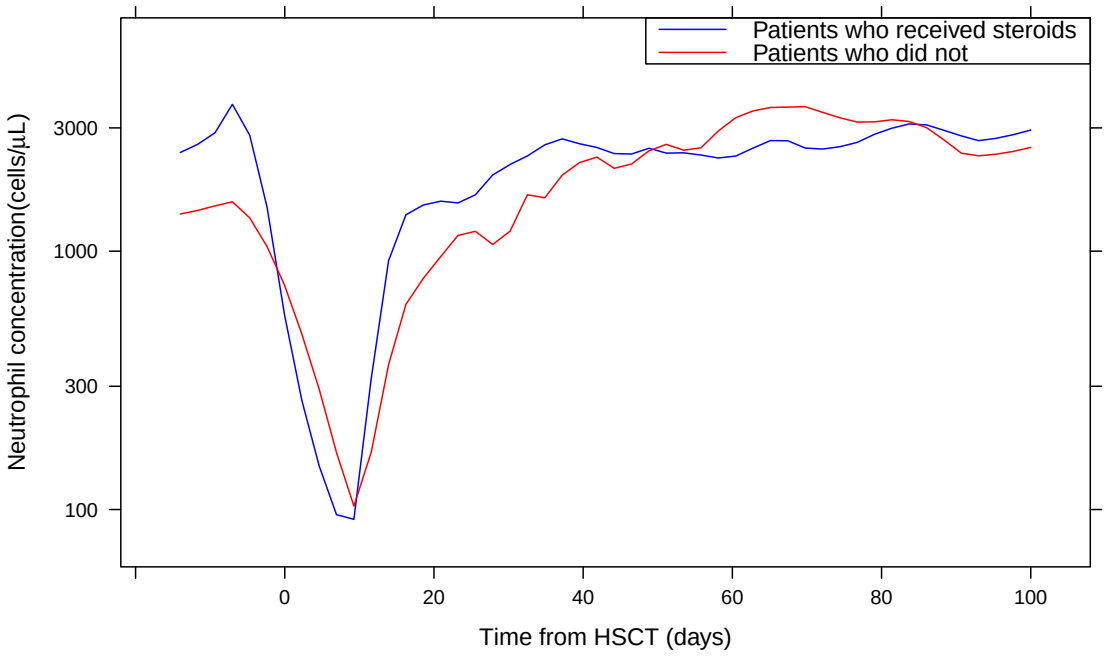

Figure 2.5: Comparison of the local regression curves of neutrophil concentrations for patients who receive steroids and patients who do not. As can be seen, patients that received steroids during conditioning had a spike in neutrophil count in the days before HSCT.

inter-percentile range can diagnose issues with the sizes of the random effects.

Predictions within a single bin, however, can differ from each other because of differences in other independent variables such as age or other covariates, making interpretation of a VPC difficult and lowering the ability to diagnose model mis-specification and poor random effects [94]. Bergstrand *et al* [95] have proposed a prediction corrected VPC (pcVPC) as a solution to this problem. The dependant variable in a bin is prediction corrected, which corrects for differences in a bin from independent variables.

2.4 Results

2.4.1 Neutrophils

Excluding patients with steroids

As mentioned in Section 2.2.1, steroids cause a transient increase in neutrophil concentration, which can last for up to a day after the dose [60–63]. Patients who are conditioned with alemtuzumab or ATG are given steroids before each dose of these conditioning drugs in order to mitigate some of their toxicity. Local regression curves of the raw data for those who received steroids and those who did not are compared in Figure 2.5, making the effect of the steroids clear. Although there is still a small rise in neutrophil concentrations in the days before transplant in patients who do not receive steroids, the patients who do receive steroids have a much more substantial increase. In order to make fitting models to these data simpler, at first in this section patients who received either of these conditioning drugs, and hence received steroids, were removed from the dataset. There were 161 patients left in this subset of the main dataset with 9324 measurements of neutrophil concentration in the time between 14 days before and 100 days after HSCT.

I compared four variants of the model: linear and sigmoidal K-PD with the original Friberg model, and then linear and sigmoidal K-PD with a variant of the Friberg model allowing elimination rate k_E to differ from synthesis and transfer rates k_S and k_T. The results of this comparison are summarised in Table 2.3.

Base model with sigmoidal K-PD

With the sigmoidal K-PD model, it is assumed that the drug is effective on days that it is given (according to the protocols in Table 2.1), and has no effect otherwise. This assumes that the EC_{50} is sufficiently high that the drug concentration has fallen to below its effective concentration within a day.

The parameter estimates were mostly sensible (see Table 2.3), although E_{Drug} was at its upper bound, with a large random effect, demonstrating the difficulty that this variant of the model had with sufficient cell loss to create the size of the decline in neutrophil concentration in the observed data.

Base model with linear K-PD

The K-PD model was tested as linear, assuming that the drug concentration follows an exponential decay and that the drug effect follows in proportion such that,

$$E_{\text{Drug}} = E_{\max} \mathrm{e}^{-\frac{(t-t_D)}{T_{\text{Drug}}}} \tag{2.13}$$

where t_D is the protocol-dependent time of last drug dose and T_{Drug} is a new parameter to be estimated, related to the lifetime of the drugs' effects on the

Table 2.3: Comparison of parameter values, random effect variances and objective function values for the different versions of the model for neutrophil reconstitution excluding patients who had steroids

		Base model		Differing k_E	
	K-PD model	Sigmoidal	Linear	Sigmoidal	Linear
Parameters	X_0 (cells/μL)	1100	1150	1180	1180
	X_∞ (cells/μL)	2800	2550	2550	2460
	MTT (days)	6.37	6.39	4.72	5.24
	γ	0.107	0.163	0.920	0.163
	$E_{\max}$	1.00	0.746	0.975	0.813
	T_{Drug} (days)	-	9.06	-	9.12
	k	-	-	2.32	0.569
Random effects	X_0	1.91	1.93	2.01	1.97
	X_∞	1.33	0.859	1.03	0.889
	MTT	0.486	0.666	0.683	0.898
	γ	0.967	0.845	0.982	0.773
	$E_{\max}$	66.5	2.22	4.04	1.90
	T_{Drug}	-	0.786	-	0.706
	k	-	-	0.593	7.62
	σ	0.591	0.495	0.547	0.481
	OFV	6442	5086	5888	4913

σ is the variance of the residual error. The random effects were fitted as log-normal distributions for all parameters because they are all bounded at 0, except $E_{\max}$, which was modelled as a logistic transformation of a normal distribution because $E_{\max}$ is bounded at both 0 and 1.

body.

This alteration produced a much improved quality of fit with OFV falling by 1355 points, a decrease in Bayesian Information Criterion (BIC) of 1346 points where $\text{BIC} = -2\ln(\text{likelihood}) + k\ln(n)$, where k is the number of new parameters and n is the number of datapoints. It is notable that the random effect variance for $E_{\max}$ is much more sensible value with the model in this form and that $E_{\max}$ is no longer equal to 1.00. The drug concentration half life is $T_{\text{Drug}}\ \ln(2) = 6.28$ days. The other parameters are broadly similar.

Differing elimination rate from synthesis and transfer rates, sigmoidal K-PD
The base model assumes that production, transfer and elimination rates are the same. This means that there are also the same number of cells in each compartment at steady state. Keeping the synthesis rate equal to the transfer rate $k_S = k_T$ while allowing the elimination rate $k_E = \frac{k_S}{k}$ to differ keeps the concentration in the proliferation and transfer compartments the same, while allowing the concentration to be different in the circulating compartment. This affects the shapes of the downward and upward trajectories of neutrophils.

Allowing for this in the model improved the OFV by 554 points, an improvement in BIC of 546 points, not as much as the linear K-PD model, but still a significant improvement.

Differing elimination rate from synthesis and transfer rates, linear K-PD
The combination of $k_E \neq k_{S,T}$ and linear K-PD produced the best fit in terms of OFV and BIC and in terms of parameter estimates. The OFV is 1529 points lower than that for the original base model, with a drop in BIC of 1510 points. The parameter estimate for $E_{\max}$ is also sensible at 81.3%, as is the estimate for the drug lifetime with a half-life of 6.32 days. Following HSCT, the final neutrophil concentration is expected to be more than double that before HSCT at 2460 cells/μL as opposed to 1180 cells/μL. The parameter giving the differences

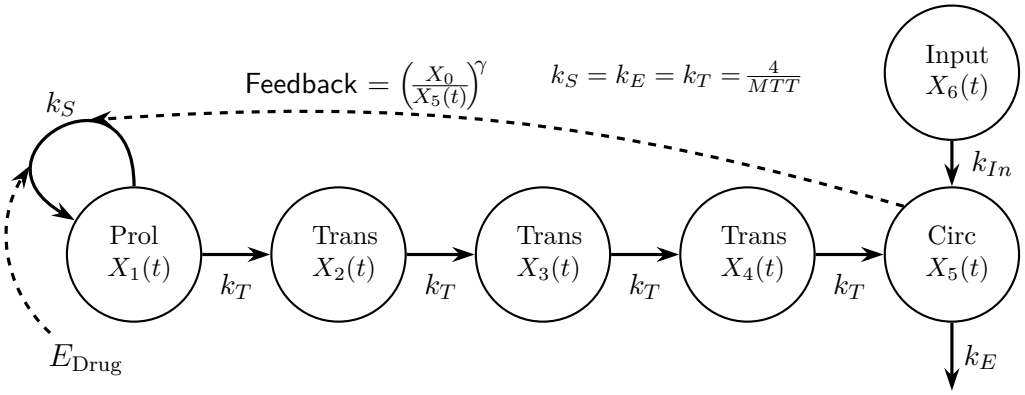

Figure 2.6: The basis of the model used for neutrophil reconstitution including the effects of steroids on neutrophil concentration, from Ozawa *et al* [64]. The dynamics are the same as in the model of Friberg *et al* [52], with the addition of another compartment, the Input compartment from which cells enter the circulating compartment at rate k_{In} for the days on which the patients were given steroids.

between the elimination and synthesis rates, $k = \frac{k_S}{k_E}$, differs between the two K-PD models, with a much smaller random effect with the linear K-PD model. In the model with linear K-PD, $k = 0.569$, finding that the rate of elimination is higher than that of synthesis or transfer.

In this variant of the model with linear K-PD and differing elimination rate from synthesis and transfer rates, the random effects are much more sensible sizes for most parameters. Furthermore the value of the residual error σ falls between this variant of the model and all other variants, implying that this variant explains more of the variability in the data than the others. I therefore use this variant of the model for the analysis of neutrophil dynamics.

Re-introducing patients with steroids

In order to model the full dataset for neutrophil reconstitution, it is necessary to account for the transient effects of steroids on the neutrophil concentration. One method is to model the patients that receive steroids during their conditioning regimen using an input compartment, as demonstrated in Figure 2.6, where k_{In} is zero, except for the days following the administration of steroids. The input compartment $X_6(t)$ starts with a fixed concentration of neutrophils, and on the

administration of steroids, a proportion per day of the neutrophils will leave into the circulating compartment, boosting the measured neutrophil concentration. Because this effect is transient with drug application, it does not affect long term concentrations. As such, the equations for the dynamics of the circulating and input compartments are,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}X_5(t) &= k_T X_4(t) - k_E X_5(t) + K_{In} X_6(t) & X_5(0) &= X_0 \\ \frac{\mathrm{d}}{\mathrm{d}t}X_6(t) &= -k_{In} X_6(t) & X_6(0) &= IP_0, \end{aligned} \tag{2.14}$$

where IP_0 is the concentration of neutrophils in the input compartment. The dynamics for the rest of the compartments are unchanged.

A comparison of the results of fitting this six-compartment model and the original five-compartment Friberg model to the full dataset for neutrophil concentration reconstitution is given in Table 2.4. As can be seen, the addition of the sixth input compartment reduces the OFV by 350 points, a highly significant decrease, implying the six-compartment version produces a much better fit to these data.

Final structural model

The final model for reconstitution of neutrophil concentration is then a combination of the six-compartment model including the effects of steroids with the adaption of the Friberg model to allow a different elimination rate from the transfer and synthesis rates and using a linear K-PD model. The inclusion of these adaptions further decreased the OFV by 3244 points (see Table 2.4). The dy-

namics of the final model are therefore given by the following equations,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}X_1(t) &= k_S X_1(t)\left(\frac{X_0}{X_5(t)}\right)^{\gamma}(1-E_{\text{Drug}}) - k_T X_1(t) & X_1(0) &= \frac{X_0}{k} \\
\frac{\mathrm{d}}{\mathrm{d}t}X_2(t) &= k_T X_1(t) - k_T X_2(t) & X_2(0) &= \frac{X_0}{k} \\
\frac{\mathrm{d}}{\mathrm{d}t}X_3(t) &= k_T X_2(t) - k_T X_3(t) & X_3(0) &= \frac{X_0}{k} \\
\frac{\mathrm{d}}{\mathrm{d}t}X_4(t) &= k_T X_3(t) - k_T X_4(t) & X_4(0) &= \frac{X_0}{k} \\
\frac{\mathrm{d}}{\mathrm{d}t}X_5(t) &= k_T X_4(t) - k_E X_5(t) + K_{In} X_6(t) & X_5(0) &= X_0 \\
\frac{\mathrm{d}}{\mathrm{d}t}X_6(t) &= -k_{In} X_6(t) & X_6(0) &= IP_0, \qquad (2.15)
\end{aligned}$$

Table 2.4: Comparison of parameter estimates for different models for neutrophil reconstitution, including patients with steroids.

		Base model	Input compartment	Input compartment & differing k_E
	K-PD model	Sigmoidal	Sigmoidal	Linear
Parameters	X_0 (cells/μL)	1590	1240	1560
	X_∞ (cells/μL)	2830	2710	2740
	MTT (days)	5.41	5.69	4.86
	γ	0.101	0.108	0.159
	E_{Drug}	1.00	0.990	0.814
	K_{In} (/day)	-	0.527	0.673
	IP_0 (cells/μL)	-	2900	866
	T_{Drug} (days)	-	-	6.94
	k	-	-	0.681
Random effects	X_0	1.91	2.16	2.00
	X_∞	0.862	0.891	0.561
	MTT	0.312	0.321	0.859
	γ	0.762	0.839	0.716
	E_{Drug}	40.6	12.8	2.88
	K_{In}	-	5.79	0.167
	IP_0	-	5.02	2.89
	T_{Drug}	-	-	0.801
	k	-	-	3.16
	σ	0.556	0.541	0.447
	OFV	12144	11794	8550

σ is the variance of the residual error. The random effects were fitted as log-normal distributions for all parameters because they are all bounded at 0, except $E_{\max}$, which was modelled as a logistic transformation of a normal distribution because $E_{\max}$ is bounded at both 0 and 1.

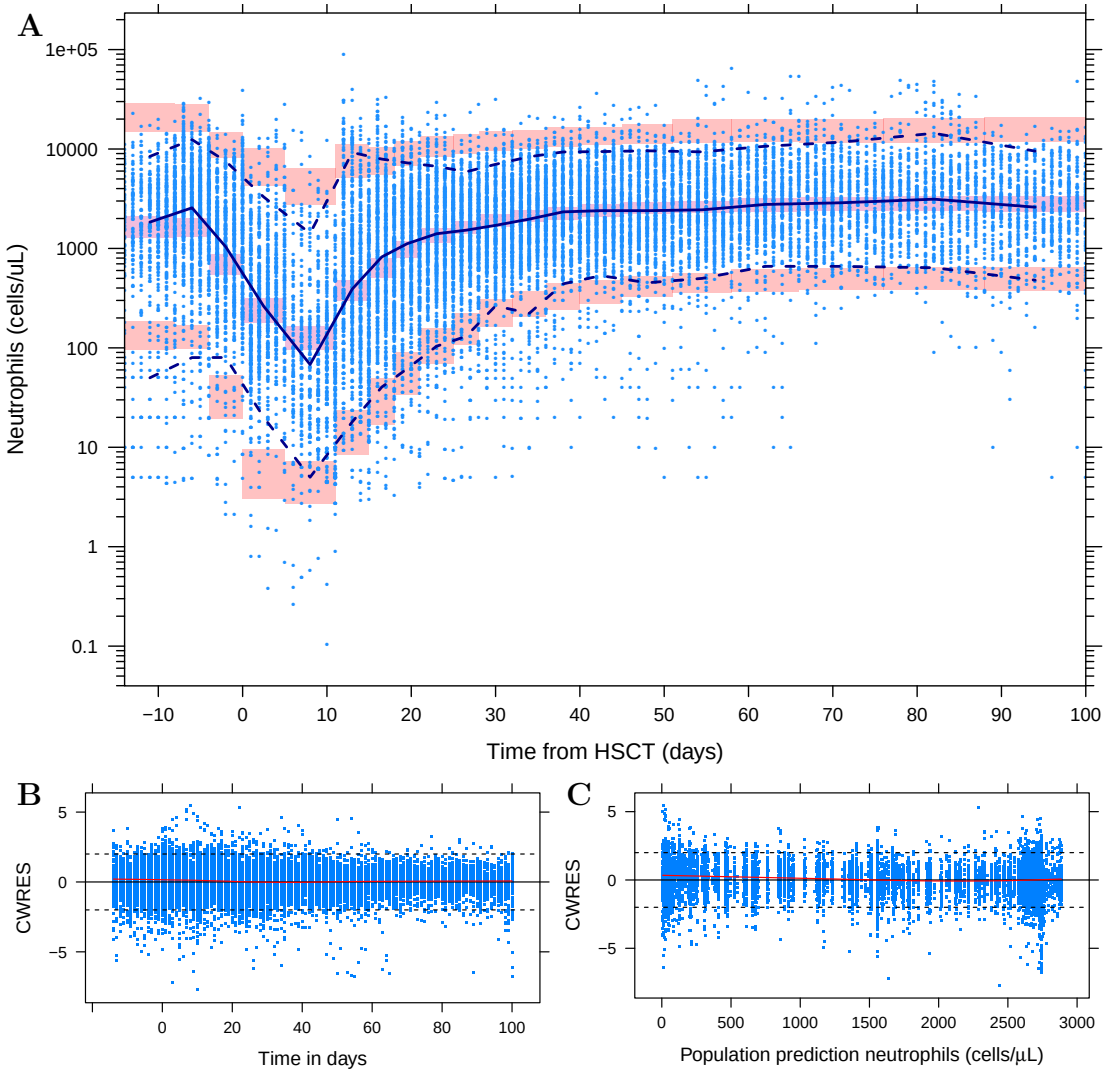

Figure 2.7: Diagnostic plots for the final model of neutrophil reconstitution including patients that received steroids, allowing different elimination and transfer rates and using a linear K-PD model. **A** gives a visual predictive check. The dots give the observed data and the filled and dashed lines give the median and 2.5th and 97.5th percentile of the observed data. The pink shaded regions give the 95% confidence intervals of the simulated median and 2.5th and 97.5th percentiles. **B** and **C** give the CWRES against time and population prediction respectively, with the blue dots giving the residuals, the red lines local regression curves and the black dashed lines giving ±2.

where

$$E_{\text{Drug}} = E_{\max} \mathrm{e}^{-\frac{(t-t_D)}{T_{\text{Drug}}}} \tag{2.16}$$

and

$$k_S = k_T = \frac{4}{MTT} = k \times k_E. \tag{2.17}$$

Parameter estimates for the final model are given in Table 2.4. The model finds

that long-term neutrophil concentration is almost double the concentration before HSCT (2740 cells/μL after compared to 1560 cells/μL before). With the linear K-PD model the drug effect decreases because the drug action is spread over a longer time period. The drug median half life was found to be 4.81 days. The input compartment was estimated to have a concentration of 866 cells/μL, with 68% of the concentration leaving the compartment each day that the patient was given steroids. The rate of elimination of circulating neutrophils was found to be higher than the rates of transfer and synthesis of neutrophils.

Diagnostic plots are shown in Figure 2.7. The visual predictive check from the model in Figure 2.7A demonstrates that the model simulated output reflects the observed data well, capturing the median and the variability of the data for the majority of the neutrophil concentration trajectory. The plots of conditionally weighted residuals (CWRES) demonstrate that the residuals are approximately normally distributed as they should be. With the local regression line falling along the x-axis, they also demonstrate that there is no bias either with time or with population prediction and that the residuals are independent.

Covariate analysis

Once the structural model was fixed, a covariates analysis was performed in order to ascertain which factors significantly affected the reconstitution trajectories. The objective function value from the model was unstable while carrying out stepwise covariate model building, and so a univariate analysis was performed. The covariates that were tested are listed in Table 2.2. All covariates were tested on X_0, X_∞, MTT and k, just the drug-related covariates were tested on E_{Drug} and T_{Drug}, while the steroid-related conditioning drugs, alemtuzumab and ATG, were tested as covariates on K_{In} and IP_0. The results of the covariate analysis are given in Table 2.5 and shown in Figure 2.8. The patients that received cord blood stem cells on average had a decreased initial concentration of neutrophils

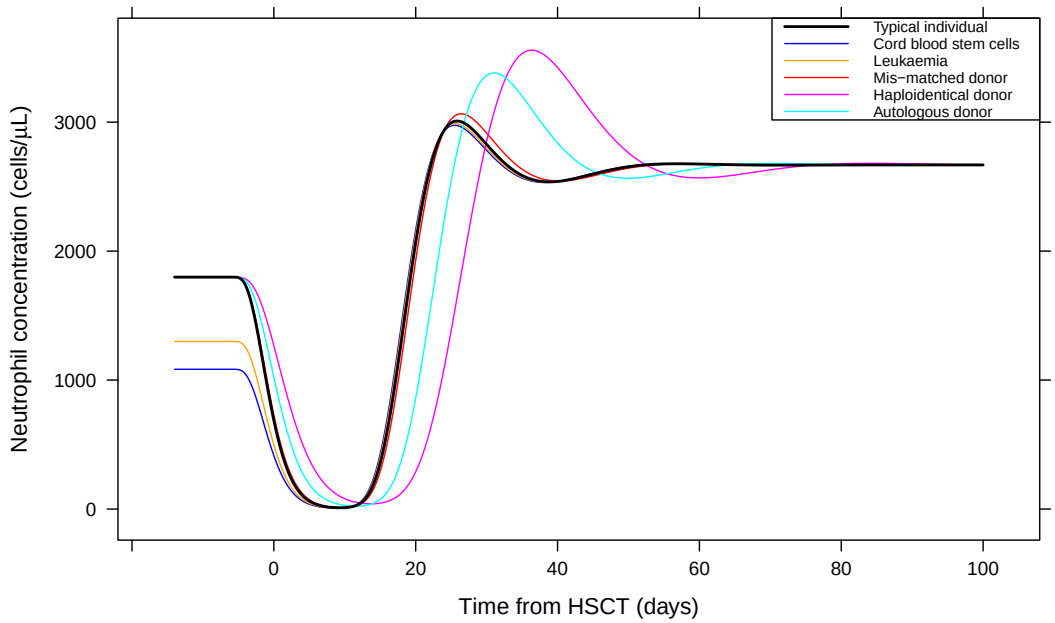

Figure 2.8: The effects of covariates on neutrophil dynamics following HSCT. The typical individual is the model output for a patient that does not have each or any of the covariates listed here, so a transplant of either peripheral blood or bone marrow stem cells from a matched donor and another diagnosis apart from leukaemia.

X_0, as did patients that had leukaemia as opposed to another diagnosis. The type of donor that was used for the HSCT was also found by the model to affect the recovery of neutrophils in patients. Reconstitution following matched and mismatched donors were similar to each other but reconstitution following haploidentical and autologous transplants had a longer mean transfer time MTT resulting in delayed reconstitution.

2.4.2 Platelets

The two models that have been used previously for platelet dynamics, the models of Hayes *et al* and Friberg *et al*, were fitted to the data of platelet concentrations around HSCT. The dynamics are simpler than for neutrophils in that there are no effects from steroids. As such, a sixth input compartment is not necessary and the models can be fitted straight to the full data.

The Hayes model and the base Friberg model were both fitted to the data

Table 2.5: Covariates included in the model for neutrophil dynamics

Parameter	Covariate	Effect size	p-value
X_0	Cord blood stem cells	-0.397	$\ll 0.001$
X_0	Leukaemia	-0.277	$\ll 0.001$
MTT	Donor type		< 0.001
	Matched	$0^\dagger$	
	Mis-Matched	0.0274	
	Haploidentical	0.481	
	Autologous	0.240	

The parameter for patients who had the respective covariate is multiplied by (1 + Effect size). So an effect below 0 decreases the parameter for that covariate and an effect greater than 0 increases that parameter for that covariate. The null hypothesis is then that the effect size is zero. † Typical individual.

using a sigmoidal and a linear K-PD model, and the variant of the Friberg model with elimination different from synthesis and transfer was also fitted using a linear K-PD model.

The parameter estimates from fitting these models are given in Table 2.6. The Hayes model, particularly with the sigmoidal K-PD model did not achieve a good fit to the data, with a high OFV and a high residual error in comparison to the Friberg model. With the linear K-PD model it achieved a much higher quality of fit, but still did not manage to attain an OFV of a similar standard to that of the Friberg model. Of the variants of the Friberg model, those with the linear K-PD model have a better fit. There is not a large difference however between the two, and the addition of a differing rate of elimination does not improve fit enough that it is worth the extra parameter (ΔBIC = 11) in the model. Furthermore, the variant of the model where $k_E = k_S = k_T$ has the lowest residual error of all the variants.

An explanation for the large difference in OFV between the Hayes and Friberg models is demonstrated in Figure 2.9 which compares individualised output for a subset of patients for the Hayes and Friberg models (with $k_E = k_S = k_T$), both with linear K-PD models. These patients have been chosen to demonstrate that the platelet reconstitution appears to form oscillations in some patients during

reconstitution. The Hayes model cannot recreate these whereas the Friberg model can because of the feedback term. This model therefore seems to be the better representation of platelet reconstitution. And so for the rest of this section I use the Friberg model with linear K-PD and with elimination, transfer and synthesis rates the same ($k_E = k_S = k_T$).

The parameters for long-term and initial platelet concentration are very similar, with the long-term platelet concentration slightly higher at $X_\infty = 166{,}000\ /\mu\text{L}$ than the initial platelet concentration of $X_0 = 163{,}000\ /\mu\text{L}$. The mean transfer time MTT is 5.73 days, while the drug effect $E_{\max}$ is low at just 28.6%. The

Table 2.6: Comparison of parameter estimates for different models for platelet reconstitution.

	Model	Hayes		Friberg		
						Differing k_E
	K-PD model	Sigmoidal	Linear	Sigmoidal	Linear	Linear
Parameters	$X_0\,(/\mu\text{L})$	124,000	163,000	161,000	163,000	164,000
	$X_\infty\,(/\mu\text{L})$	125,000	206,000	179,000	166,000	165,000
	k_E (/day)	0.252	0.359	–	–	–
	k_T & k_S (/day)	0.238	0.445	–	–	–
	MTT (days)	–	–	6.69	5.73	6.33
	γ	–	–	0.0835	0.109	0.127
	$E_{\max}$	0.988	0.988	0.490	0.286	0.330
	T_{Drug} (days)	–	15.5	–	5.24	5.20
	k	–	–	–	–	0.475
Random effects	X_0	0.922	0.840	0.814	0.832	0.825
	X_∞	0.949	1.34	0.899	0.833	0.840
	k_E	0.761	0.579	–	–	–
	k_T & k_S	0.951	0.718	–	–	–
	MTT	–	–	0.413	0.419	0.540
	γ	–	–	1.51	1.10	1.20
	$E_{\max}$	0.439	4.07	2.70	1.05	1.26
	T_{Drug}	–	5.04	–	3.78	3.23
	k	–	–	–	–	0.241
	σ	0.395	0.305	0.285	0.278	0.280
	OFV	4400	-577	-1897	-2248	-2269

σ is the variance of the residual error. The random effects were fitted as log-normal distributions for all parameters because they are all bounded at 0, except $E_{\max}$, which was modelled as a logistic transformation of a normal distribution because $E_{\max}$ is bounded at both 0 and 1.

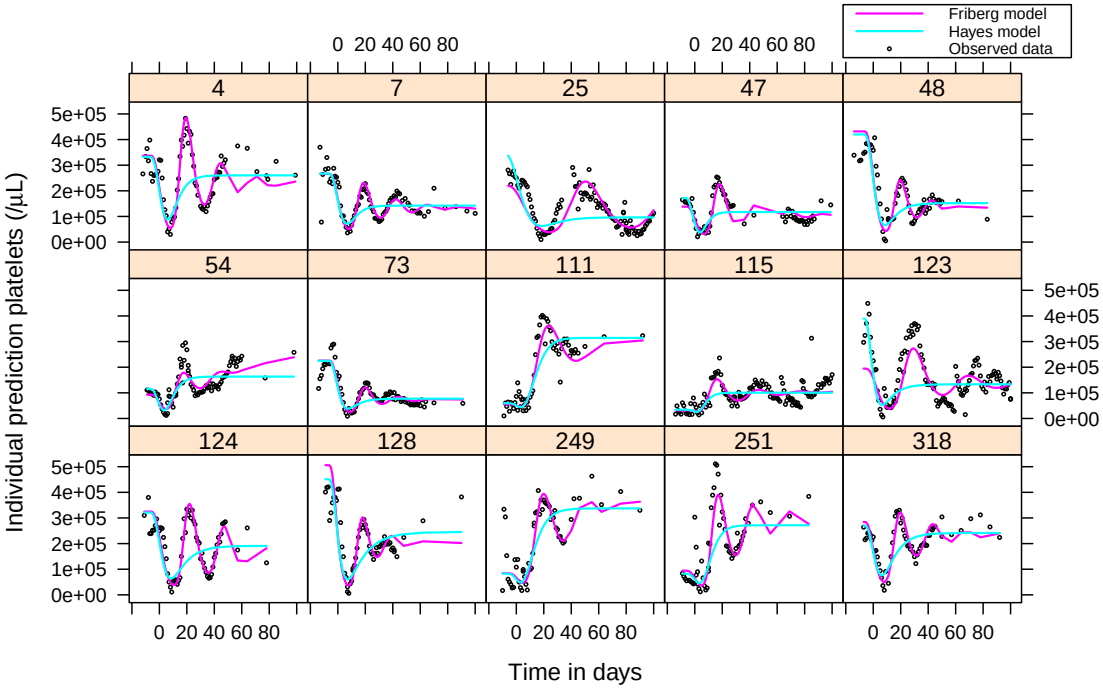

Figure 2.9: Comparison of the two models for platelet reconstitution for a subset of patients, demonstrating the differences between individual predictions. The black dots give the observed data, and it can be seen that while the Friberg model can recreate the observed oscillations because of the feedback loop, the Hayes model cannot and draws a straight line through the data.

drug half-life was estimated at 3.63 days.

Diagnostic plots for the Friberg model are shown in Figure 2.10. The visual predictive check in Figure 2.10A demonstrates that the model simulated output does capture the median of the observed data well and most of the variability, although it over-predicts the reconstitution at later time points and for the 95th percentile. The plots of the conditionally weighted residuals in Figures 2.10B and C demonstrate that the residuals are roughly normally distributed with mean zero and standard deviation one, and that there are no obvious biases with time or population prediction and so the residuals are independent.

Covariate analysis

Once the structural model was fixed at the Friberg model with linear K-PD, covariates analysis was performed using the SCM procedure to find the factors

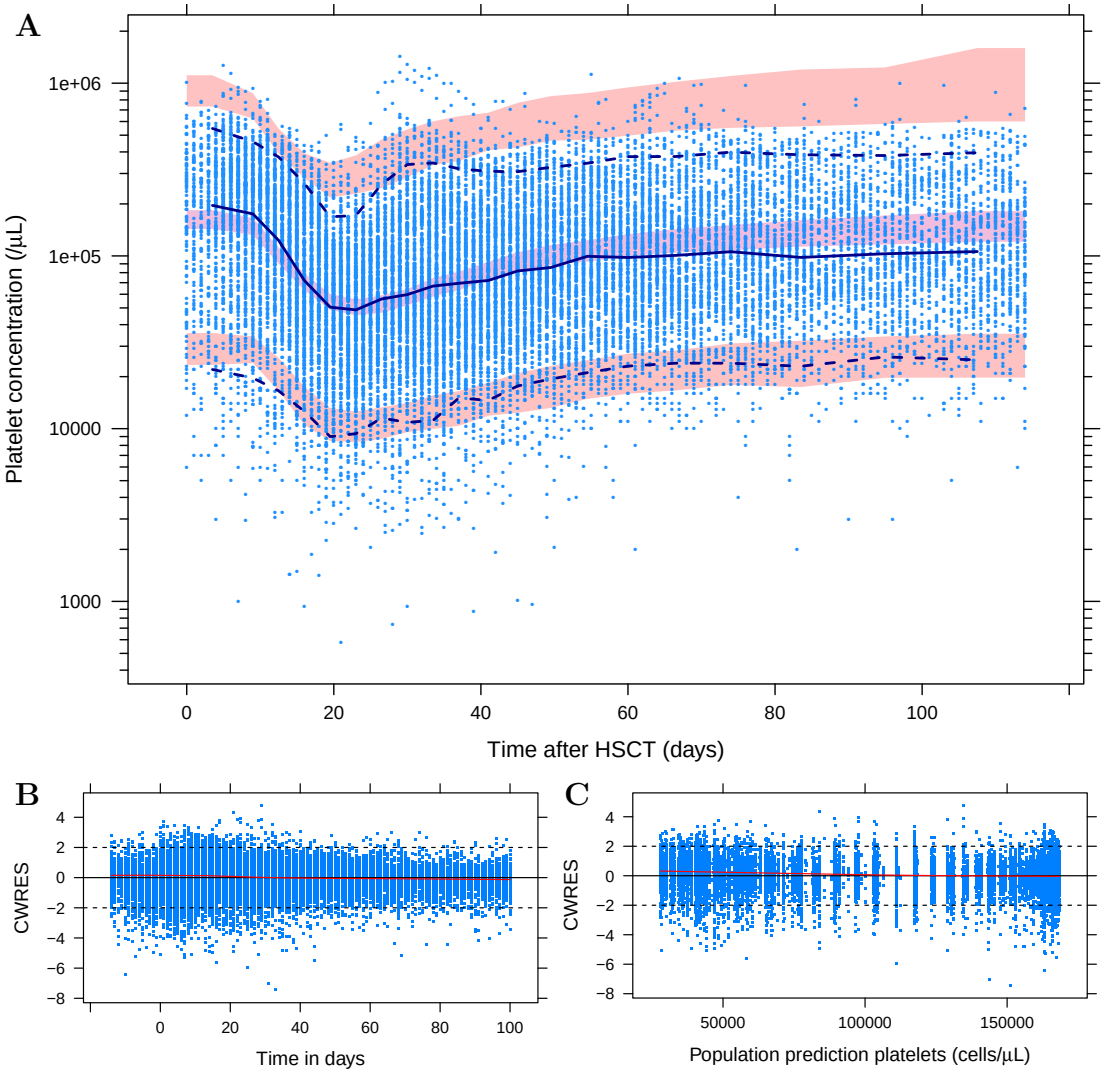

Figure 2.10: Diagnostic plots for the final model of platelet reconstitution, using a linear K-PD model. **A** gives a visual predictive check. The dots give the observed data and the filled and dashed lines give the median and 2.5th and 97.5th percentile of the observed data. The pink shaded regions give the 95% confidence intervals of the simulated median and 2.5th and 97.5th percentiles. **B** and **C** give the CWRES against time and population prediction respectively, with the blue dots giving the residuals, the red lines local regression curves and the black dashed lines giving ± 2.

that significantly affect the reconstitution of platelets. The covariates that were tested are listed in Table 2.2. All covariates were tested on X_0, X_∞ and MTT and just the drug-related covariates were tested on E_{Drug} and T_{Drug}. The results of the covariate analysis are given in Table 2.7 and shown in Figure 2.11. As can be seen, the significant covariates all affected either the initial platelet concentration X_0 or the long-term platelet concentration X_∞. Patients that received busulfan

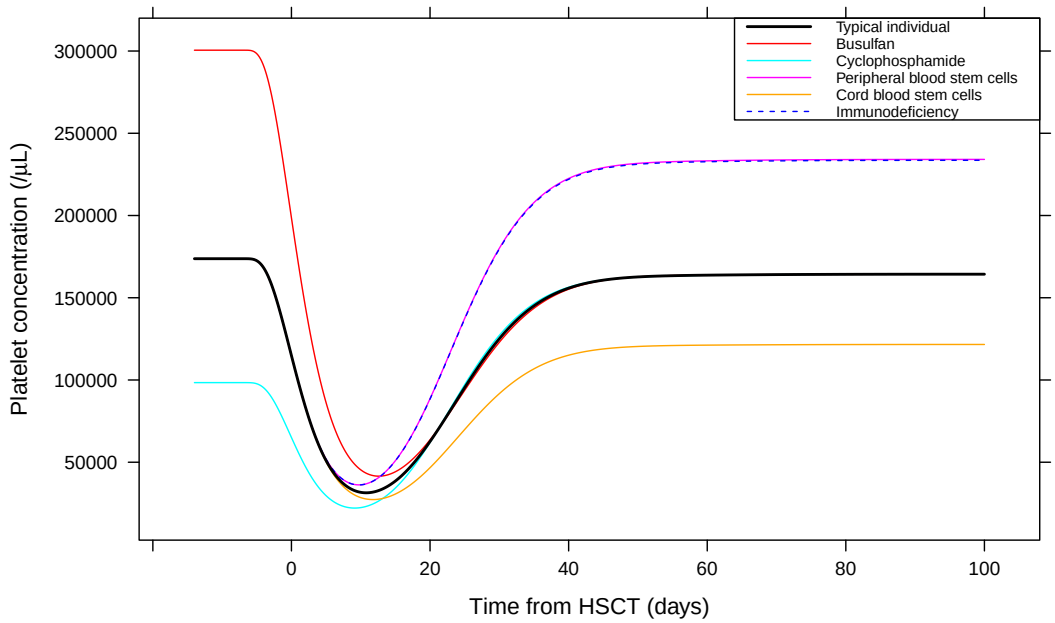

Figure 2.11: The effects of covariates on platelet dynamics following HSCT. The typical individual is the model output for a patient that does not have each or any of the covariates listed here, so received bone marrow stem cells, did not receive either busulfan or total body irradiation as conditioning, and had another diagnosis apart from an immunodeficiency.

were found through the model to have on average a higher platelet concentration pre-HSCT, while those who received cyclophosphamide were found to have a lower pre-HSCT platelet concentration. Patients that received peripheral blood stem cells or were diagnosed with immunodeficiency were found by the model to have on average a raised long-term platelet concentration while patients that received cord blood stem cells as opposed to bone marrow or peripheral blood stem cells were found by the model to have on average a decreased long-term platelet concentration.

2.5 Discussion

In this chapter, I have successfully applied previous models for neutrophil and platelet reconstitution to the context of paediatric HSCT. The mechanistic models were similar in that they had a proliferation compartment, on which the drugs

Table 2.7: Covariates included in the model for platelet dynamics

Parameter	Covariate	Effect size	p-value
X_0	Busulfan	0.730	$\ll 0.001$
X_0	Cyclophosphamide	-0.434	$\ll 0.001$
X_∞	Stem cell source		< 0.001
	Bone marrow	$0^\dagger$	
	Peripheral blood	0.425	
	Cord blood	-0.260	
X_∞	Immunodeficiency	0.421	< 0.001

The parameter for patients who had the respective covariate is multiplied by (1 + Effect size). So an effect below 0 decreases the parameter for that covariate and an effect greater than 0 increases that parameter for that covariate. The null hypothesis is then that the effect size is zero. † Typical individual.

acted, multiple transit compartments and finally a circulating compartment where the cell concentrations were measured. A comparison of models found that the linear kinetic-pharmacodynamic (K-PD) model was the better K-PD model for all model variants. For neutrophils, the model variant of the Friberg model with elimination rate differing from synthesis and transfer rates provided the best fit to the data. For platelets, the Friberg model provided a much better fit than the Hayes model because of the presence of a feedback loop, and the model variant in which the elimination rate was the same as the synthesis and transfer rates was selected as that which was the best balance between objective function value and number of parameters.

After discovering that the Friberg model gave such a large improvement in fit over the Hayes model for the platelet data, closer analysis of individual trajectories demonstrated that platelet concentrations in some individuals oscillated during reconstitution. The feedback term was therefore necessary to explain the data from these patients. These oscillations had been unexpected; none of the other studies that have used this model have reported oscillations, although the data that they were using may not have had the resolution to see these effects [67–71]. Oscillatory dynamics for platelets have been reported in patients with cyclic thrombocytopenia, and healthy volunteers were found to have fluctuations with

a time period in the order of 25 – 30 days [96], similar to this analysis. They have not however been reported as part of reconstitution. One option for the cause of this feedback is transforming growth factor beta (TGF-beta) levels producing a negative feedback loop on megakaryocyte production [97].

The model parameters for initial and final concentration were different for neutrophils, with the final concentration almost double the initial concentration ($X_\infty = 2740\,\text{cells}/\mu\text{L}$ compared with $X_0 = 1560\,\text{cells}/\mu\text{L}$). This implies that on average the population has a much more healthy innate immune system following HSCT than before the transplant, which is not surprising given that some patients were given HSCT for neutropaenic immunodeficiencies. For platelets, graphically the local regression curves of Figure 2.3 appears to show that platelets do not recover to the same level at which they started. The model however implied that on average platelets can be expected to recover to the level at which they started within three months (X_∞ =166,000 /μL cf. X_0 =163,000 /μL). This difference is because the local regression curves cannot account for the fact that patients recover quickly, but at different delays from the HSCT (see Figure 2.4). For both neutrophils and platelets these values of X_0 and X_∞ are well within the ranges for normal children [98,99].

The value of the mean transfer time for neutrophils $MTT = 4.86$ days agrees well with other analyses that have used this model on adults, with values in the range of 3.8 to 5.6 days [52,59]. The mean transfer time for platelets was longer with $MTT = 5.73$ days. Platelets have a longer lifetime in the periphery, and so this was to be expected [4,9,10]. Previous uses of this model have found an MTT in the range of 4.3 to 9.2 days, which agrees with the results presented here [67–69].

With the linear K-PD model, the maximum drug effect was larger for neutrophils with $E_{\max} = 81.4\%$ than for platelets with $E_{\max} = 28.6\%$. This means the drugs were roughly 2.8 times as effective in neutrophils, which is in line with the

observed data where the neutrophil concentrations undergo a 10-fold decline and platelet concentrations undergo just a 3-fold decline. This parameter is usually reported as the '*Slope*' parameter in other uses of this model, where $Slope \simeq \frac{E_{\max}}{EC_{50}}$ (see Section 2.2.3). However, given that we have no drug concentration data, we use a slightly different form of the parameterisation. One study that uses this model to estimate parameters for both neutrophil and platelet concentration profiles with the same drug finds that the drug is 2.8 times as effective for neutrophils as platelets [67], which agrees with the results of this analysis.

Half lives of the drugs used in conditioning are short: busulfan 2.3 hrs [79]; cyclophosphamide 5–9 hrs [80]; fludarabine 8–10 hrs [81]; melphalan 1.3 hrs [82]; and treosulphan 1.7-2.2 hrs [83]. As such, it is interesting that the drug effect was found to have a half life of 4.8 days for neutrophils and 3.6 days for platelets after the end of conditioning treatment. This implies that the model may not be fully capturing the effects of these drugs. One explanation could be the structure of the proliferating compartment. At the moment it is assumed that these proliferating stem cells form a resident population that produce neutrophil precursors. These proliferating stem cells however are in dynamic equilibrium with their own precursors and these precursors will also have been prevented from proliferating by the conditioning drugs. This would result in an extra delay following the drugs before neutrophil precursor production increases and hence an apparently longer drug effect. Future work with higher resolution pharmacokinetic drug data could tease out this relationship.

These models are however very much semi-mechanistic. They take a complex system — the production of new platelets and neutrophils from stem cells — and simplify them to a number of linear compartments. While the key features are mechanistic, with new cells only produced in the bone marrow compartment, cell concentrations measured in the circulating compartment and the existence of a feedback mechanism, the exact meaning of the compartments themselves is

difficult to interpret. Furthermore, it is difficult to extract information (such as neutrophil or platelet peripheral lifetimes) from this model.

Once the structural models were fixed, it was possible to analyse the factors that affected the reconstitution of both neutrophils and platelets. For neutrophils, it was not possible to perform multivariate analysis because the objective function value proved too unstable to perform SCM reliably. SCM is usually performed on datasets much smaller than these here, and with so many data, small fluctuations in model parameters can lead to very large fluctuations in objective function value which may explain the difficulty found here. It was however possible to perform a univariate analysis. The patients that received cord blood stem cells were found by the model to have a lower pre-HSCT neutrophil concentration than those that received either peripheral blood or bone marrow stem cells from the donor. In the observed data, patients that had a cord blood transplant had a median neutrophil concentration before the start of conditioning 22% lower than the other patients, which can explain a part of the difference. Patients that had leukaemia were also found by the model to have a reduced pre-HSCT neutrophil concentration. Again, this is backed by the observed data, with leukaemia patients having median pre-conditioning neutrophil concentrations reduced by 59% from the rest of the patients. Covariate analysis also identified donor type as a factor that affects reconstitution, with patients that had either haploidentical donor or autologous transplants having a later and slower recovery than those that had either mis-matched or matched donors. The patients that had haploidentical donors were more likely to have had no conditioning (50% — 6 out of 12 — in comparison to 12% in the rest of the patients), and as a result the profile will be quite different, which could explain some of the difference.

It was possible to perform multivariate analysis to find the factors that affect platelet reconstitution, finding that the conditioning drugs busulfan and cyclophosphamide affected the initial platelet concentration X_0 and that donor stem

cell source and immunodeficiency affected long-term platelet concentration X_∞. The differences in the initial concentration are apparent in the observed data, with the median initial concentration in patients that had busulfan 19% higher than those that did not and the median initial concentration in patients that had cyclophosphamide 33% lower than those that did not. Whilst this explains some of the difference found through the model, it does not explain all of it, and it is likely that some of the differences found by the model are due to mis-specification in the protocol for the days on which conditioning drugs were given (Table 2.1).

The high long-term platelet concentration for patients diagnosed with immunodeficiency as opposed to other diagnoses is an interesting effect that is visible in the observed data, with median platelet concentration between days 60 and 100 post HSCT almost double in patients that had immunodeficiencies. Finally, the differences between the stem cell sources are also visible in the data, an explanation for these differences could be the make-up of the donor cell grafts, where cord blood grafts are likely to be smaller with lower concentrations of platelet-producing megakaryoctes, whereas peripheral blood is likely to be more diverse, potentially providing more platelets early on. The significance of this covariate on long-term concentration therefore potentially implies that the increase in platelets following HSCT might be more to do with platelets produced by megakaryoctes found in the donor cell graft, rather than stem-cell derived megakaryoctes.

In this chapter, known mechanistic models for the short term reconstitution of neutrophils and platelets have been fitted to data from paediatric HSCT using nonlinear mixed effects modelling. To our knowledge, this is the first time this analysis has been done in this way. The models were complex enough that they were able to recreate these systems, but simple enough that it was possible to estimate parameters. The models have been used to perform a covariate analysis to find the factors that affect the reconstitution of these vital cell subsets from the haematopoietic system.

Chapter 3

Long-term CD4 T cell reconstitution following HSCT

3.1 Introduction

Full CD4 T cell reconstitution takes between one and three years following HSCT. Over this time period a child's immune system will have matured appreciably and the expected CD4 concentration will have changed, with CD4 concentration expected to fall by $\frac{2}{3}$ between 0 and 10 years old (see Figure 1.2). Because these changes occur within an individual, any modelling attempts need to account for age-related effects before the underlying trends can be ascertained.

In this chapter, I develop a model for long-term CD4 T cell reconstitution. In the model development, different methods were used to attempt to account for age, which finally led to the use of mechanistic modelling. The mechanistic models had the advantage of being able to take into account other areas of relevant biology as well, such as competition for resources. The final model was then used to make predictions of long-term reconstitution on an individual basis, using the patient's covariates and early data. Predictions of long-term reconstitution could greatly assist clinicians by giving early warnings of potential long-term problems

following the HSCT and indicating the need for further interventions, such as prophylactic antimicrobials, donor lymphocyte infusion or repeat HSCT.

3.1.1 Quantifying immune reconstitution

Two methods have been used in previous studies to assess reconstitution: lymphocyte subset concentrations at pre-determined time points after HSCT, and the time after HSCT that lymphocyte subsets reach pre-determined concentrations.

Variations in the extent of reconstitution at particular times are assessed through the pre-determined time-points method. Differences in reconstitution following transplantation with the three different stem cell sources have been widely reported. Increased B cell reconstitution has been found to occur earlier after cord blood transplantation (CBT) compared with bone marrow transplantation (BMT) [100] and greater B cell and NK cell reconstitution but reduced T cell reconstitution at 3 months following unrelated CBT compared with matched sibling BMT and unrelated BMT [101]. Studies looking at the effects of the donor types on reconstitution, have found increased reconstitution with matched sibling donors rather than mismatched family or unrelated donors [102] and reduced T cell reconstitution at 6 months in patients having autologous peripheral blood stem cell transplantation (PBSCT) compared with allogeneic BMT or PBSCT [103]. Age effects have been identified, with younger patients reconstituting fewer CD8 T cells at one month following allogeneic HSCT [104]. Further, Epstein-Barr virus DNAemia was found to have a negative impact on the reconstitution of T cells at one year [104]. A study into the effects of reduced intensity conditioning showed that the lower concentrations of the conditioning drugs compared to full myeloablative conditioning resulted in increased T and NK cell reconstitution at four months, concluding that immune reconstitution is accelerated by reduced conditioning [105].

The pre-determined time-points method is also used to find the leukocyte

concentration thresholds at certain times that indicate improved survival probability. Studies into CD4 concentrations have used 115 cells/μL at 20 days after HSCT [106] and 86 cells/μL at 35 days [102], while a study into absolute lymphocyte count used 150 cells/μL at 30 days [107]. One study used a more sophisticated method, classifying people into high- and low-mortality risk groups by forming an ellipsoid reference domain for the normal concentrations of three lymphocyte subsets [108]. Most simply, patients with higher CD4 T cell concentrations at predetermined time points had improved survival probability [101].

Changes in the rate of reconstitution are assessed through the time to reach pre-determined concentrations method. Studies into stem cell sources have found faster B and slower CD8 T cell reconstitution following unrelated CBT as opposed to unrelated BMT [109] and faster reconstitution following PBSCT compared with BMT [110], while another found faster reconstitution following CBT with high concentrations of CD45 cells [111]. One study that looked at the effects of conditioning drugs found that anti-thymocyte globulin (ATG) almost doubles the time to reach normal CD4 T cell concentrations, but that CD8 T cells were unaffected [112]. High doses of ATG were also associated with increased incidence of life-threatening infections [112]. Finally T cell reconstitution was found to slow with age across childhood [109].

The pre-determined concentrations method has also been used to find cut-off times to reach concentrations that result in improved survival probability. Significant improvements in survival probability were found for patients who reconstituted above the 5th percentile of normal CD4 T cell concentration within one year [110]. Survival probability was also conditioned by the CD8 T cell reconstitution rate and the time to reach the 10th percentile of normal CD4 T cell concentration [113].

For assessments of concentrations at pre-determined time points, the data must be rigorously collected on the day specified. Meanwhile for time to event

analysis of pre-determined concentrations, the event must be monitored frequently or the time is missed. In contrast, mathematical modelling is advantageous because it allows data collection at any time, offering greater flexibility. Furthermore, mathematical modelling can analyse both the rate and extent of the reconstitution, whereas both the approaches described above only evaluate specific factors that affect either the rate or the extent of the reconstitution, not both.

In this chapter, I construct a mechanistic mathematical model for CD4 T cell concentration in order to model the long-term reconstitution following HSCT and determine the factors that affect this reconstitution.

3.2 Methods

3.2.1 The data

The CD4 T cell concentration data used in this chapter were collected as routine clinical practice by the Blood and Marrow Transplant Unit at Great Ormond Street Hospital for Children NHS Trust. Two separate datasets were used; a model-building dataset and a validation dataset. The model-building dataset uses CD4 T cell concentrations from the patients studied in Chapter 2. This dataset, shown in Figure 3.1, was used for developing the model and covariate analysis. The validation dataset was a separate cohort of patients who had transplants between 2010 and 2014. This dataset was used to validate the predictive ability of the model.

The modelling dataset has 288 patients who had 319 transplants between them. The median age at transplant was 37 months, with a range of 16 days to 16 years. In this dataset, 24% of the patients died within the 1–6 year follow-up period; of these patients, 36% died from infection, 35% from disease relapse, and 15% from acute GvHD.

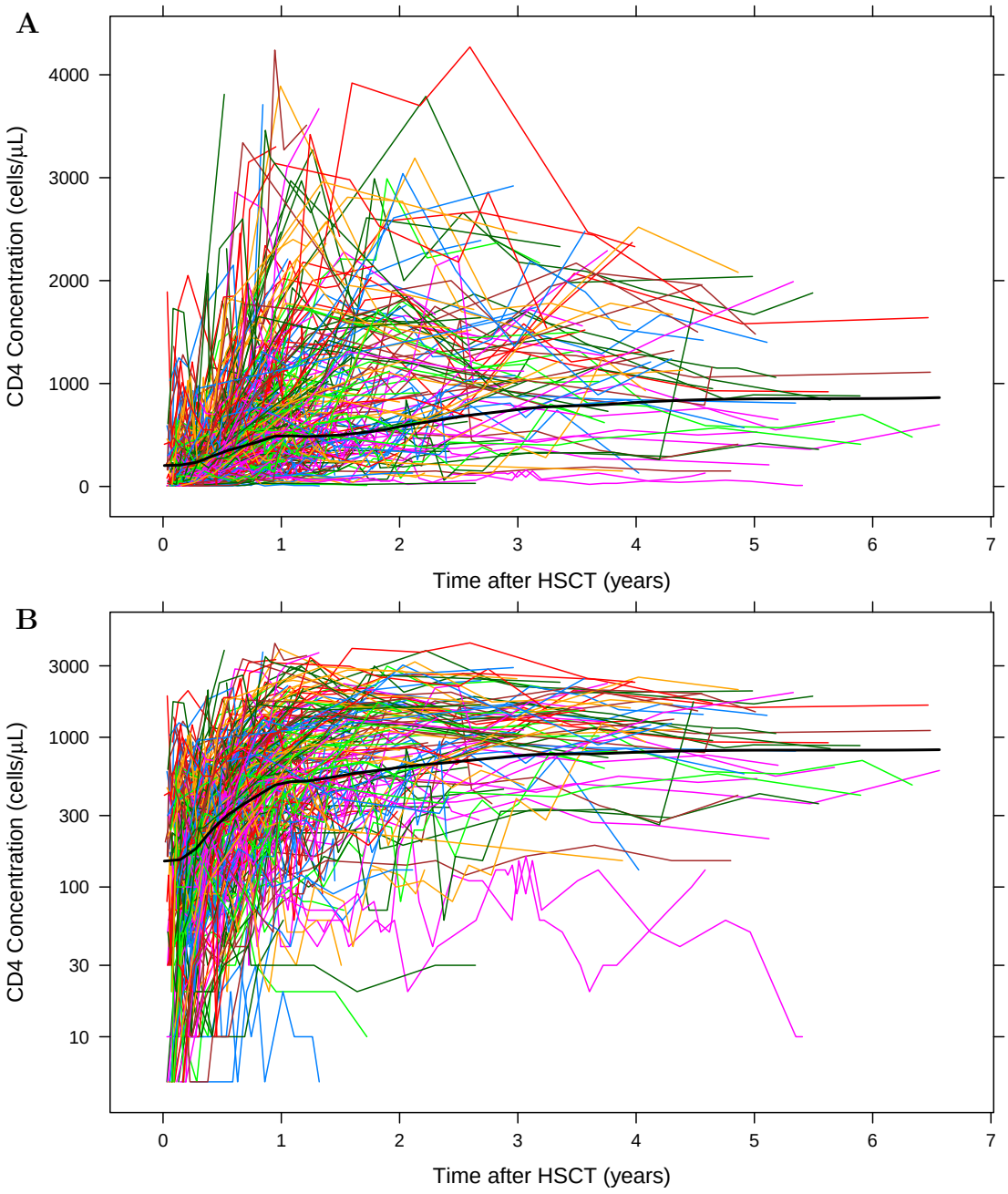

Figure 3.1: Data for CD4 T cell reconstitution following HSCT used for model-building. Each coloured line gives the data for a single individual, while the black line gives the local regression curve for the data. **A** is plotted with a linear y-scale, while **B** is the same data, plotted on a log y-scale.

The validation dataset is a subset of 75 patients from 132 patients who had 144 transplants between them. The patients chosen for the validation dataset were those for which there were at least three data points before six months post HSCT and at least three data points after six months post HSCT. A breakdown

Table 3.1: Breakdown of the demographics and the drugs used for the patients in the long-term, CD4 reconstitution datasets.

	M	V		M	V
Age at HSCT (years)			**Donor type**		
$0 \rightarrow 1$	51	14	Matched	201	39
$1 \rightarrow 2$	67	12	Sibling	86	14
$2 \rightarrow 5$	74	16	Family	16	5
$5 \rightarrow 10$	76	23	Unrelated	99	20
$10 \rightarrow$	51	10	Mis-matched	102	28
Sex			Sibling	3	0
Male	202	51	Family	6	1
Female	117	24	Unrelated	93	27
HSCT			Haploidentical	13	2
1^{st}	272	66	Autologous	3	6
2^{nd}	40	8	**Viruses**		
3^{rd}	3	1	Cytomegalovirus		
Diagnosis			Positive	102	12
Immunodeficiencies	138	30	Negative	213	61
SCID	82	18	Unknown	4	2
Wiskott-Aldrich	12	5	Epstein Barr virus		
CGD	12	6	Positive	82	12
Leukaemia	95	17	Negative	120	48
ALL	45	8	Unknown	117	2
AML	35	8	Adenovirus		
HLH	35	5	Positive	106	—
Anaemia	21	0	Negative	213	—
Autoimmune	8	0	**Conditioning**		
Lymphomas	5	0	Alemtuzumab	158	30
GvHD			Anti-CD45	12	2
Reported	102	45	Anti-thymocyte globulin	10	12
I	39	25	Busulphan	75	31
II	38	15	Cyclophosphamide	140	12
III	18	4	Fludarabine	67	55
IV	7	1	Melphalan	97	17
Stem cells			Treosulphan	67	18
Bone marrow	149	27	Total body irradiation	44	6
Peripheral blood	120	28	None	41	4
Cord blood	48	20	**Prophylaxis**		
Combinations	2	0	Ciclosporin	280	66
			Methotrexate	66	12
			Mycophenolate	158	51

M: Model-building dataset, used for model building and covariate analysis; V: Validation dataset, used for assessing the predictive ability of the model. *Abbreviations*: SCID: severe combined immunodeficiency sydrome; CGD: chronic granulomatous disease; ALL: acute lymphoblastic leukaemia; AML: acute myeloid leukaemia; HLH: hemophagocytic lymphohistiocytosis; GvHD: graft versus host disease.

of the demographics of both datasets, and the drugs used for conditioning and prophylaxis, is given in Table 3.1.

3.2.2 Model building

In contrast to the short-term reconstitution seen for platelets and neutrophils, the time scale for CD4 T cell reconstitution is similar to the expected changes in CD4 T cell concentration as children get older. As such, individuals can be expected to have intra-individual age-related changes in their data and age cannot just be accounted for in the inter-individual random effects. Furthermore, CD4 T cells proliferate in the periphery which neither neutrophils nor platelets do. As such the models used in Chapter 2 cannot apply here and so a new model for long-term reconstitution following HSCT is constructed.

Empirical modelling

Early modelling attempts used an empirical model for CD4 T cell reconstitution, with three parameters: long-term CD4 T cell concentration asy, initial CD4 T cell concentration int, and the rate of reconstitution c,

$$y = asy - (asy - int)\, e^{-ct} \,. \tag{3.1}$$

In order to fit this model to the data, two different approaches were tried to account for the effects of development, pre-adjusting the data and adjusting the model. Pre-adjusting the data involved attempting to remove the effects of age from the data before fitting the model by standardising to some sort of expected concentration for age. The first option for this was to use z-scores, where the observed concentration is compared to a distribution of CD4 concentration for the same age [114]. A score is then assigned, with the median concentration giving a score of 0, and a score of ± 1 corresponding to ± 1 standard deviation

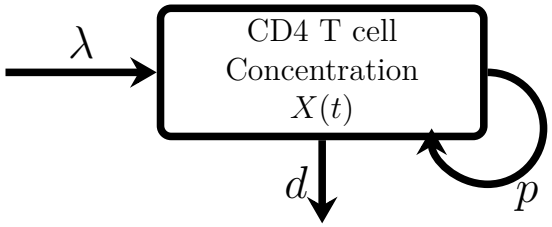

Figure 3.2: Schematic of the long-term CD4 T cell reconstitution model. CD4 T cells in the compartment have concentration $X(t)$. They enter the compartment from the thymus at zero order rate λ. Cells then die at rate d and proliferate at rate p.

from the median. Outside of a relatively confined range around the median, such as with the concentrations immediately after HSCT, these scores are however of limited validity. Instead, another option is to use the ratio of the observed CD4 concentration to the expected CD4 concentration for that age [31], with one of three transformations, log, 2nd root or 4th root of this ratio.

Adjusting the model involved accounting for the changing CD4 concentration within the model itself, while fitting to the raw data. This was achieved by having age-dependance on asy in one of four forms, linear, bi-linear, exponential decay and a ratio of the expected concentration for age.

On comparison, the results from these forms of the data and model, however, were very inconsistent and it was not possible to distinguish which version was the least biased analysis of the data. It was thus decided that the use of a mechanistic modelling approach could provide a better explanation for the effects of normal age-related changes in a child's immune system.

Mechanistic modelling

The structural model as applied in this chapter is a one compartment turnover model as given in Figure 3.2. In this model the central compartment represents the CD4 T cell concentration $X(t)$ with time t. New cells from the thymus enter this compartment at zero-order rate λ. Cells can proliferate into two cells or die at first order rates p or d respectively. This gives the following ordinary differential

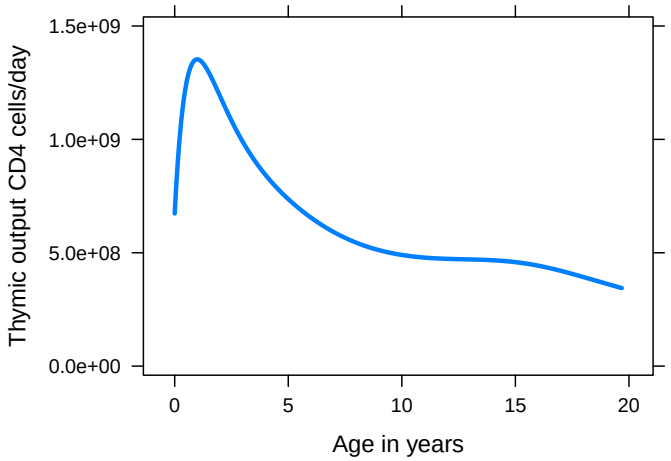

Figure 3.3: Expected thymic output of CD4 T cells with age using the formula from Bains *et al* [33] given in Eq (3.3). Thymic output peaks at around 1 year of age and then deceases as the thymus involutes.

equation for the dynamics,

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \lambda - dX + pX. \tag{3.2}$$

The model is then made more mechanistic by using mathematical functions in place of some of the parameters to better represent the underlying biology.

Accounting for the effects of age

Because the concentration of CD4 T cells decreases across childhood by ⅔ between 0 and 10 years old (see Figure 1.2), it is important the effects of immune development are taken into account.

Thymus development

Thymic output first increases with age as the thymus grows, peaking at about one year of age, and then decreasing with age as the thymus involutes [33]. These changes with age were recently characterised mathematically by Bains *et al* [33]. Their work makes use of the fact that TRECs are neither degraded nor replicated in division to calculate the number of cells produced by the thymus with age. They then remove the dilution in TREC numbers caused by cell division, using

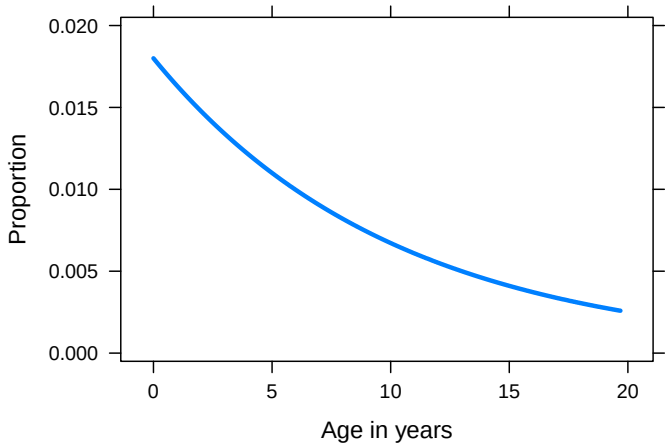

Figure 3.4: Proportion of CD4 T cells expressing Ki67 with age using the formula in (3.4) [33]. Ki67 is used as a marker for proliferation and so this decline can be used for the rate of decline in turnover rates of CD4 T cells with age through childhood.

levels of Ki67 with age as a marker for cell division. This leaves the following functional form for thymic output with time t and age τ (see Figure 3.3),

$$\lambda(t,\tau) = \lambda_0 \times \frac{y(\tau)\, V(\tau)\, \gamma}{\eta(c-\gamma)}, \tag{3.3}$$

where $V(\tau)$ is the expected CD4 T cell concentration with age; $\eta = 0.52$ is the duration of Ki67 expression; $c = 0.25$ and $\gamma = 0.08$ are constants related to the TREC content of CD4 cells as they leave the thymus; and $y(\tau)$ is the proportion of CD4 cells expressing Ki67 with age,

$$y(\tau) = 0.02\, e^{(-0.00027\,\tau)} \tag{3.4}$$

shown in Figure 3.4. The parameter λ_0 is included such that the proportion of the expected thymic output for age is estimated but the shape of the changes with age are maintained.

Changes in loss and proliferation with age

It is thought that the cell proliferation and loss rates decrease with increasing age through childhood [33,35,36]. Ki67 expression can be used as a marker for CD4 T cell proliferation, and so the change in Ki67 expression with age, which

has recently been characterised, can be used to inform on the time-scales of the decrease in proliferation and loss throughout childhood [36]. Ki67 expression with age is given by $y(\tau)$ shown Figure 3.4 and equation (3.4).

Homeostatic mechanisms and competition for resources

T cells are thought to compete for resources such as cytokines and sp-MHC (see Section 1.5). This means when there are very few T cells, there is much more resource for each T cell so that proliferation increases and loss decreases. This has been observed in a cohort of thymectomised children that had reduced T cell concentrations in whom Ki67 expression levels, a marker for proliferation, was found to be raised [115].

A mechanistic mathematical model for the homeostatic mechanisms and competition for resources has previously been proposed [116]. In this two-compartment homeostatic model, T cells produced by the thymus enter the resting compartment X. From this, cells are activated at rate a into the dividing compartment Y from which two cells return at rate r to the resting compartment. Cells have different loss rates in the resting and dividing compartments, d_X and d_Y respectively. Homeostatic mechanisms from competition for resources result in density dependant activation and death rates. The full model is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \lambda - X\Big(d_X - a\Big) + 2rY \tag{3.5}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = aX - d_Y Y - rY, \tag{3.6}$$

where

$$a = a_0 \mathrm{e}^{-c_a(X+Y)} \tag{3.7}$$

$$d_X = d_0 \mathrm{e}^{c_d(X+Y)} \tag{3.8}$$

$$d_Y = \mu Y, \tag{3.9}$$

λ is thymic output, and t is time after HSCT. a_0 and d_0 could be considered as the activation and loss rates at zero CD4 T cell concentration; c_a and c_d as the strength of the competition effects.

Simplifying the homeostatic model

This model was simplified in order to reduce it from two compartments to one (see Figure 3.2), thus making the model parameters more identifiable, and simpler to fit to the data. To achieve this, the following two assumptions were used.

Quasi steady state approximation: With $r \gg a$, it can be assumed that the dynamics of the dividing compartment, Y are fast relative to the dynamics of the resting compartment, X. Thus the resting compartment is in quasi-steady state and, as a result, the dynamics are driven by the concentration in the resting compartment, X, such that from the ordinary differential equation, (3.6),

$$\left.\frac{\mathrm{d}Y}{\mathrm{d}t}\right|_x = aX - rY - d_Y Y = 0 \tag{3.10}$$

which gives,

$$\mu Y^2 + rY - aX = 0, \tag{3.11}$$

which can be rearranged to get Y in terms of X,

$$Y = \frac{1}{2\mu}\left(-r \pm r\sqrt{1 + \frac{4\mu a}{r^2}X}\right). \tag{3.12}$$

Low death rate in dividing compartment: With the parameters values obtained on fitting the full model to data for CD4 concentrations [116], the death rate in the dividing compartment, μ, was very small, and in particular, the proportion of cells dying was much lower than those returning to the resting compartment, namely $\mu \ll r$. Thus the term $\frac{4\mu a}{r^2}X \ll 1$, and the Binomial theorem can be used to say,

$$\sqrt{1 + \frac{4\mu a}{r^2}X} \approx \left(1 + \frac{2\mu a}{r^2}X\right), \tag{3.13}$$

allowing the approximation of (3.12) to become:

$$Y = \frac{1}{2\mu}\left(-r \pm r\left(1 + \frac{2\mu a}{r^2}X\right)\right) \tag{3.14}$$

giving

$$Y = \frac{a}{r}X, \tag{3.15}$$

where the negative root has been discarded as both X and Y are positive. Substituting (3.15) into (3.5) thus gives,

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \lambda - X\left(d_X - a\right), \tag{3.16}$$

removing the dependance on the 2nd compartment completely. Activation from the original model is now better described as proliferation, so the nomenclature is changed such that,

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \lambda - X\left(d - p\right), \tag{3.17}$$

where,

$$p = p_0 \mathrm{e}^{-c_p X} \tag{3.18}$$

$$d = d_0 \mathrm{e}^{c_d X}. \tag{3.19}$$

Effects of changing cell concentrations on competition for resources

Because cell concentrations are changing through childhood and in order to make the parameter values more interpretable, these functions were altered such that they scale with CD4 T cell concentration,

$$p = p_0 \mathrm{e}^{c_p\left(1 - \frac{X(t)}{V(\tau)}\right)} \tag{3.20}$$

$$d = d_0 \mathrm{e}^{c_d\left(\frac{X(t)}{V(\tau)} - 1\right)}, \tag{3.21}$$

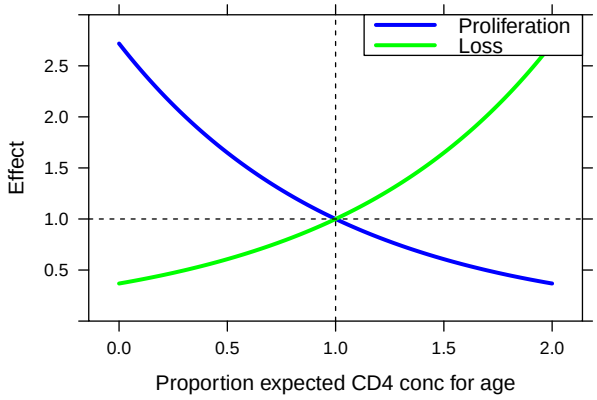

Figure 3.5: Competition effects with proportion of expected CD4 T cell concentration using formulas (3.18) and (3.19). The effect is to multiply the proliferation and loss rates for age by the blue and green lines respectively. The effect is dependent on the ratio of the CD4 concentration to that expected of a child of the same age. When the concentration equals that expected of a healthy child, the effect is 1.

where $V(\tau)$ is the expected CD4 concentration for age. This means the parameters, p_0 and d_0 can be interpreted as the proliferation and loss rates respectively when the CD4 concentration is equal to that expected of a healthy child, while c_p and c_d are the strength of the effects of competition for resources on loss and proliferation respectively. These functions are shown in Figure 3.5.

Effects of the HSCT on thymic output

Following HSCT thymic output of T cells does not recover for between 6 and 12 months, as demonstrated by evidence from both TREC analysis [117–119] and studies of recent thymic emigrants using CD31 expression [100]. This lack of production leads to a delay in the reconstitution of CD4 T cells. A sigmoidal function with time after transplant was chosen to model this effect, such that thymic output would increase from 0 cells/day on the day of transplant, to that expected of a healthy child after some time delay. Several functions were tested to achieve this, including a simple logistic, generalised logistic, and Hill function as well as adaptions thereof. They were compared by BIC and their properties

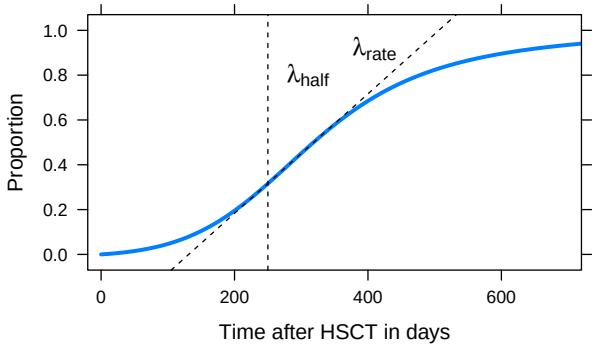

Figure 3.6: The proportion of the expected thymic output with time following HSCT, modelling the effects of impaired thymic output using formula (3.22). Immediately after the transplant thymic output is impaired. With time after HSCT, thymic output recovers, with λ_{half} giving the time at which this increase is at its maximum rate and λ_{rate} the rate of this increase.

and the the final selected function is given by (Figure 3.6),

$$\Delta_{\text{thymus}} = \frac{1 - \exp\left[-2t/\lambda_{half}\right]}{1 + \exp\left[-\lambda_{rate}\left(1 - t/\lambda_{half}\right)\right]} \tag{3.22}$$

where λ_{half} and λ_{rate} are two new parameters to be estimated. λ_{half} gives the time after HSCT that the thymic output increases at its maximum rate, and λ_{rate} gives the rate of this increase.

The complete structural model

The structural model is then given by:

$$\frac{\mathrm{d}}{\mathrm{d}t}X(t,\tau) = \lambda(t,\tau) - \mathrm{d}(t,\tau)X(t,\tau) + p(t,\tau)X(t,\tau)\ ,$$

where

$$\lambda(t,\tau) = \lambda_0 \frac{y(\tau)\,V(\tau)\,\gamma}{\eta(c-\gamma)}\Delta_{\text{thymus}}(t) \tag{3.23}$$

$$p(t,\tau) = y(\tau)p_0 \mathrm{e}^{c_p\left(1-\frac{X(t)}{V(\tau)}\right)} \tag{3.24}$$

$$d(t,\tau) = y(\tau)d_0 \mathrm{e}^{c_d\left(\frac{X(t)}{V(\tau)}-1\right)}\ , \tag{3.25}$$

leaving eight parameters to be estimated: X_0 the initial concentration of T cells; λ_0 proportion of the expected thymic output with age; p_0 and d_0 the proliferation and loss rates respectively when the concentration is that of a healthy child; c_p and c_d the strength of the effects of competition for resources on proliferation and loss respectively; λ_{rate} the rate of change of thymus recovery; and λ_{half} the time at which thymic recovery is half.

3.2.3 Sensitivity and identifiability analysis

Before fitting the model to data, the model was assessed for its parameter sensitivity and its theoretical identifiability using the FME package [120] in R 3.1.3 [121]. With this package, it is possible to get a plot of the model's sensitivity to perturbations in the model parameters. In Figure 3.7A, the effects of increasing each of the parameters by 10% have been plotted against time, and it can be seen that while the CD4 T cell concentration is sensitive to all parameters, some parameters cause changes which are either very similar or exact opposites of each other.

The package was also used to carry out collinearity analysis. In this process, a collinearity index is calculated based on the extent to which similar model output can be produced from several different parameter combinations, based on Omlin *et al* [122]. The collinearity index can be interpreted as the amount of the change in model output through altering one of the parameters that can be compensated through changes in other parameters. The index is given by collinearity $= 1 - 1/k$ where k is the fraction of the changes that can be compensated. This means that with a collinearity index of 20, changes in one parameter can be compensated by 95% by changes in other parameters, while a model with a collinearity of 1 would have perfectly orthogonal parameters. In Figure 3.7B the collinearity index is given for the eight parameter combinations for which all three of the key parameters (λ_0, d_0, p_0) are represented and the col-

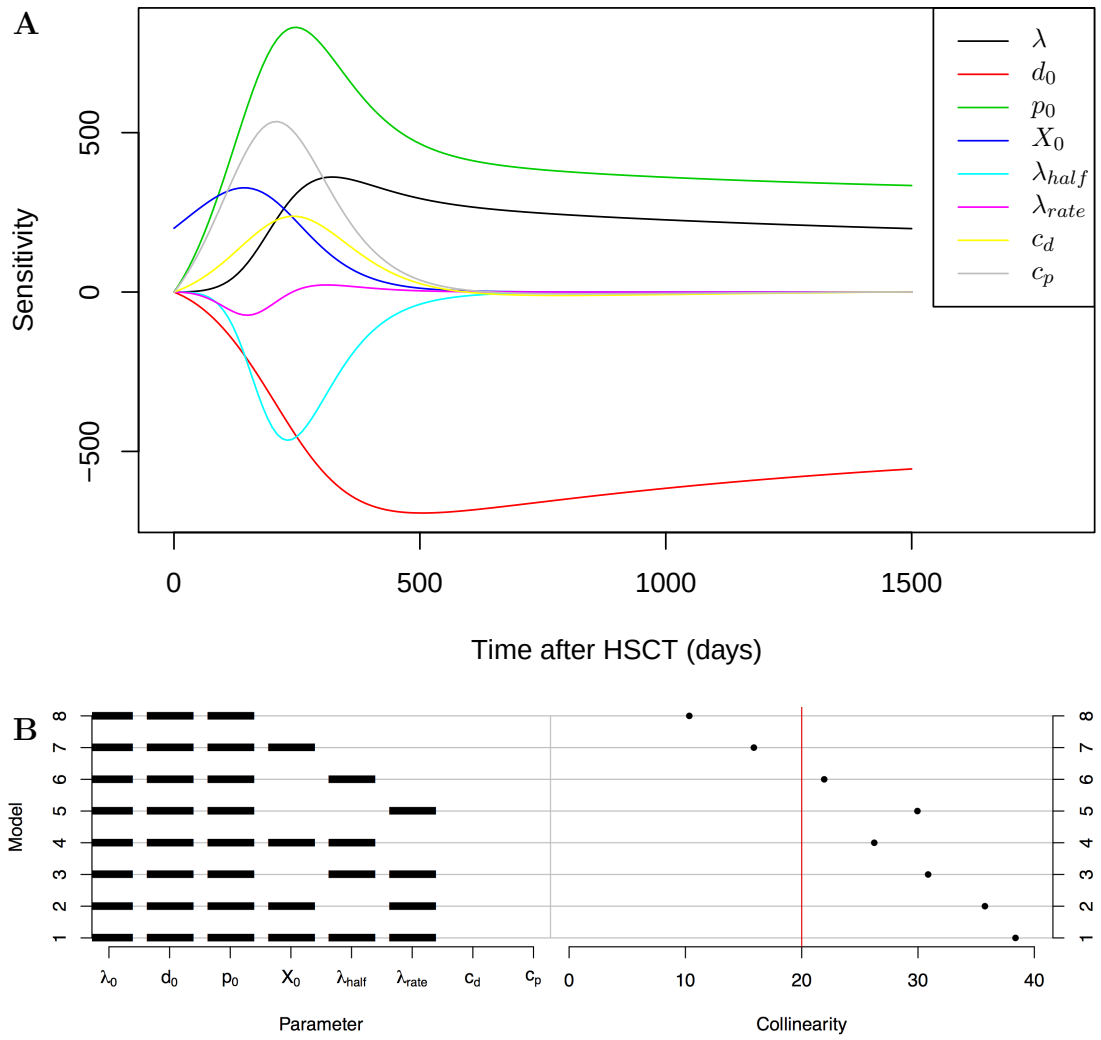

Figure 3.7: Assessing the structural properties with the model with the FME package [120]. **A** gives the sensitivity of the model with time after HSCT to perturbations in each of the parameters. Each line represents the effect of increasing that parameter by 10%, demonstrating the co-dependence of parameters in the model. **B** gives the collinearity of different combinations of the model parameters. Each model tests a different combination of parameters given by the black bars and finds a collinearity index for the co-dependance of the parameters tested. In this plot, just the 8 models that included all of λ_0, d_0 and p_0 and had a collinearity less than 40 are shown.

linearity is under 40. As can be seen, not a single model in which the parameters for the strength of competition for resources c_p and c_d are estimated are represented. As such, these parameters are highly collinear, with any combination of parameters that includes one of c_p or c_p having a collinearity score greater than 40. They are therefore difficult to estimate alongside the other parameters, and their effects can be compensated for by the other parameters. In the rest of the

analysis in this chapter they are therefore fixed to 1.

This package has tested theoretical identifiability. Practical identifiability, the limitations of what can be fitted to a given dataset, also needs to be taken into account for model fitting, which may further reduce the parameters that it is possible to fit. In the case of this analysis however, with the inter-individual changes in parameters with age and the use of mixed effects modelling (neither of which are taken into account by the package), it is likely that it will be possible to fit further parameters than might be expected.

3.2.4 Model fitting and covariate model building

Nonlinear mixed-effects (NLME) modelling, as outlined in Section 2.3.2 on Page 48, was used to fit the model to the data. The model-fitting was carried out using NONMEM 7.3 [85], using the importance sampling expectation-maximisation algorithm and the ADVAN13 (general nonlinear kinetics) subroutine [123]. Furthermore, covariate analysis was carried out using stepwise covariate model building, outlined in Section 2.3.3 on page 52 using PsN 3.5.3 [124]. Similarly to Chapter 2, only a small proportion (1.2%) of the data are below the limit of quantification and so, as set out in Section 2.3.4, these data were substituted with half the limit of quantification, LOQ/2.

3.2.5 Making predictions

The intention was to make predictions for the long-term CD4 T cell reconstitution of children on an individual patient basis, using just their early data, their age and their other relevant covariates.

Once the covariate model had been finalised through SCM, population mean parameter estimates and the variance-covariance (var-covar) matrix of the random effects were estimated by fitting the model to the model-building dataset. Using this as the basis, the model could then be fitted to an individual's data, us-

ing an "expectation-only" process, whereby the population-level parameter means and var-covar matrix were held fixed and the individual-level parameters and the individual var-covar matrix were estimated. In this process, Monte Carlo sampling is used to evaluate the individual-level conditional (posterior) mean and the var-covar matrix. The likelihood of these are then maximised given the individual's observed data and the population means and var-covar matrix [123].

Trajectories were then produced by simulating a large number ($\sim$500) of parameter sets from these individual-level estimates, from which were found the median and confidence intervals for that individual's CD4 reconstitution trajectory, forming a graphical output.

In order to validate the predictive ability of the model, a validation dataset was used, described in Section 3.2.1. For each of the 75 patients in this dataset, only data from the first six months post HSCT as well as their relevant covariates were used to make the predicted trajectory. This trajectory was then compared to the rest of their observed data for up to three years after HSCT.

3.3 Results

3.3.1 Model fit

Following fitting the model to the model-building dataset, the typical trajectories of patients of different ages at time of HSCT are given in Figure 3.8. As can be seen, there is an initial delay to reconstitution as thymic output is impaired [100,117–119] and then the typical patient reconstitutes to 90% of the expected CD4 T cell concentration of a healthy child of that age, and then tracks that concentration throughout childhood. It can also be seen that the rate of recovery is age-dependent, with younger children recovering more quickly in comparison to older children. The median-aged child (37 months old) took 22 months to reconstitute to 90% CD4 for age, while a 1 year old was predicted to take 17

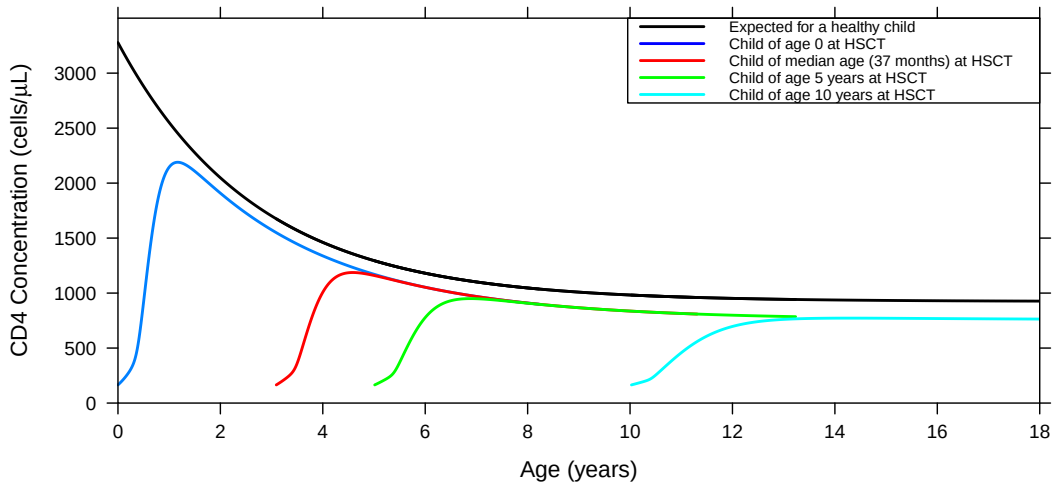

Figure 3.8: Model predicted CD4 reconstitution following HSCT for children of different ages (coloured lines) compared to the expected CD4 concentration of a healthy child (black line).

months and a 10 year old 33 months.

The parameter estimates in Table 3.2 are from fitting the model to the model-building dataset. It was found to take 5 months before thymic output recovered to 50% of the long-term thymic output for age, with recovery in thymic output happening quickly, recovering from 10% to 90% output between 3 and 7 months post HSCT. Following this recovery in thymic output, the mean thymic output for age was estimated as 23% of the previously predicted output [36].

Table 3.2: Estimated parameter means with standard deviations, and random effect variances with standard deviations.

Parameter		Estimate	s.d.	Ω	s.d.
λ_0	Proportion theoretical thymic output [33] (cells/day)	0.227	0.0714	1.59	0.59
d_0	Proportion expected loss (/day)	0.454	0.0912	1.67	0.382
p_0	Proportion expected proliferation (/day)	0.204	0.0234	0.294	0.113
X_0	Initial concentration of T cells (cells/μL)	165	21.9	1.28	0.202
λ_{half}	Time to recovery for thymic output (days)	136	21.9	1.24	0.270
λ_{rate}	Rate of recovery for thymic output	9.00	1.28	1.27	0.427
σ	Variance of the residual error	0.219	0.0164	—	—

Parameter estimates and the random effect variances (Ωs) are found through fitting the model to the model-building dataset with NLME modelling. The standard deviations (s.d.) has been found through bootstrap using PsN 3.5.3 [124].

Both diagnostic plots and changes in the Bayesian information criterion (BIC) were used to assess the impact of removing the mechanistic elements from the model. A substantial increase in BIC of 255 points resulted from the removal of the post-HSCT delay to thymic output; an increase of 115 points was found from the removal of the effects of competition for resources on proliferation and loss; and an increase in BIC of 12.0 points resulted from the effects of age on proliferation and loss. These increases, as well as changes in diagnostic plots, demonstrate the necessity of these mechanistic components, and that they significantly improve the ability of the model to describe the data.

3.3.2 Covariate analysis

In order to find what factors affected reconstitution, covariate analysis was carried out using the stepwise covariate model-building (SCM) procedure. In the forward search, for testing which covariates to include in the model a p-value of $p_1 < 0.01$ was used, while in the backwards search for testing which of these included covariates to retain, a p-value of $p_2 < 0.005$ was used (see Section 2.3.3 in Chapter 2). The covariates tested are listed in Table 3.1 and these covariates were tested on each of the parameters that were estimated in the model. The included parameters and the size of their effects are given in Table 3.3, while the the effect that these covariates have on the mean trajectory of a child of median

Table 3.3: Covariates included in the model for long-term CD4 reconstitution

Parameter	Covariate	Effect size	s.d.	p-value
X_0	Alemtuzumab	-0.840	0.025	$\ll 0.001$
X_0	Antithymocyte globulin	-0.933	0.107	$\ll 0.001$
X_0	Acute GvHD	0.328	0.201	< 0.001
λ_0	Leukaemia	1.26	0.442	< 0.001
p_0	No conditioning	-0.907	0.022	$\ll 0.001$

The parameter for patients who had the respective covariate is multiplied by (1 + Effect size). So an effect below 0 decreases the parameter for that covariate and an effect greater than 0 increases that parameter for that covariate. The null hypothesis is then that the effect size is zero.

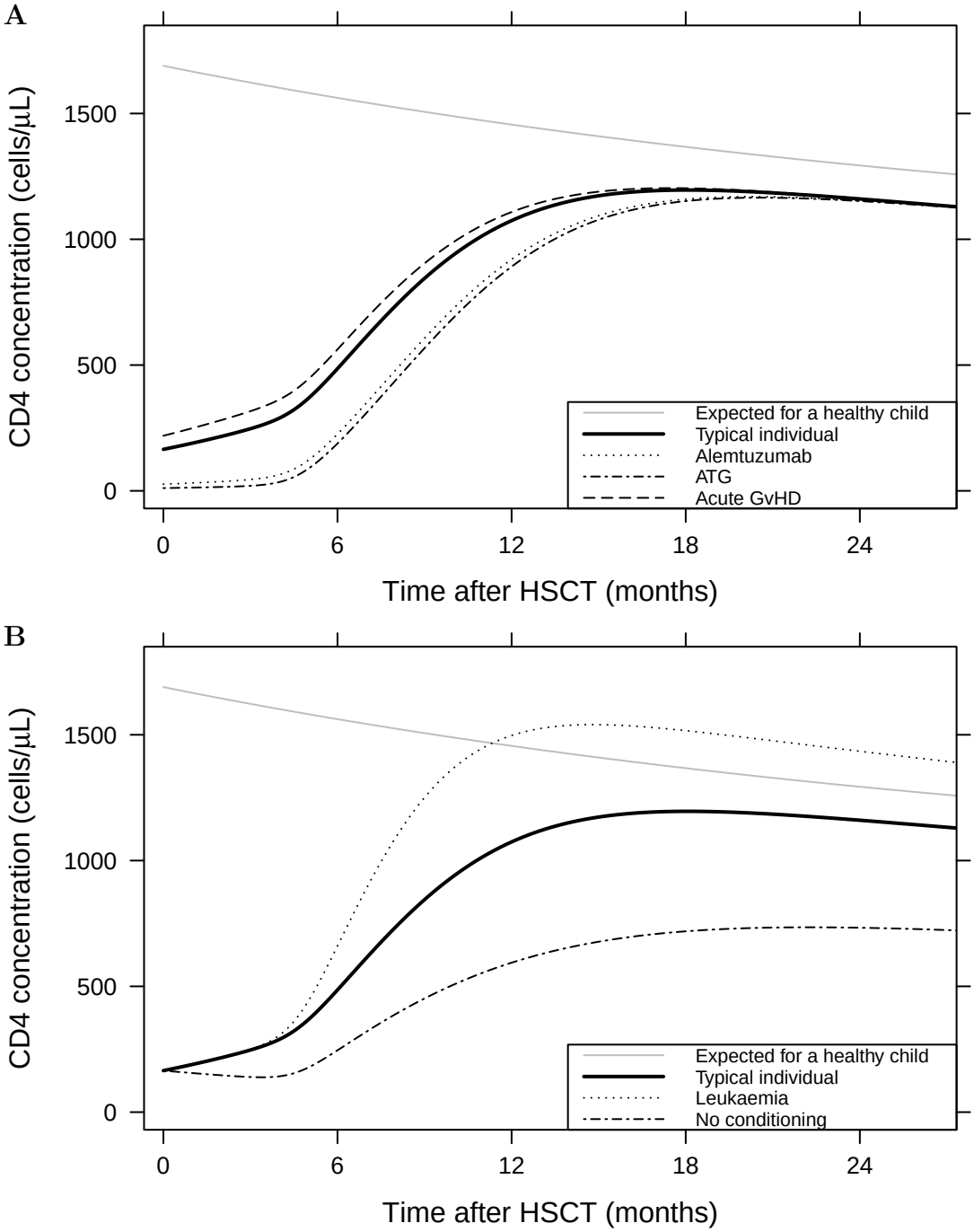

Figure 3.9: The effects of the significant covariates on CD4 reconstitution of a patient of median age at the time of HSCT (37 months). **A** Effects of the conditioning drugs alemtuzumab and ATG and acute GvHD on initial number of cells. **B** Effects of leukaemia and having no conditioning on long-term reconstitution. The covariates have been included through the SCM procedure. A typical individual is one who is not in each of the covariate group listed.

age (37 months) are demonstrated in Figure 3.9.

Effects of conditioning

Two conditioning drugs alemtuzumab and antithymocyte globulin (ATG) were found to significantly decrease the concentration of CD4 cells at the time of transplant. Patients who received neither of these drugs ($n = 151$) had an estimated mean initial CD4 concentration of 178 cells/μL, while patients receiving alemtuzumab ($n = 158$) had an estimated mean of 30.6 cells/μL, a decrease of 83%, and patients receiving ATG had an estimated mean of 8.4 cells/μL, a decreased of 95%. This decrease in initial concentration also resulted in reconstitution to any fixed concentration being delayed by a few months (see Figure 3.9A).

Patients who received no pre-transplant conditioning were affected in a different manner. As demonstrated in Figure 3.9B, while the initial CD4 concentration was unaffected, having no conditioning was found to alter the reconstitution trajectory by decreasing the expected long-term CD4 concentration. The model thus predicts that these patients will have a sub-optimal long-term CD4 concentration.

Effects of leukaemia

The covariate analysis found that patients who had leukaemia ($n = 95$) had a different reconstitution to those that had other diagnoses as shown in Figure 3.9B, with long-term CD4 concentration found to be higher than that expected of a healthy child. This difference was observed in both myeloid leukaemia patients ($n = 50$) and lymphoblastic leukaemia patients ($n = 45$), with no significant difference between these patients ($p = 0.23$).

Effects of acute GvHD

Patients who had acute GvHD ($n = 102$) were estimated to have a higher CD4 concentration at the time of the transplant with a 33% increase compared with those with no reported GvHD. This meant that those patients were predicted by the model to have a marginally earlier reconstitution than other patients.

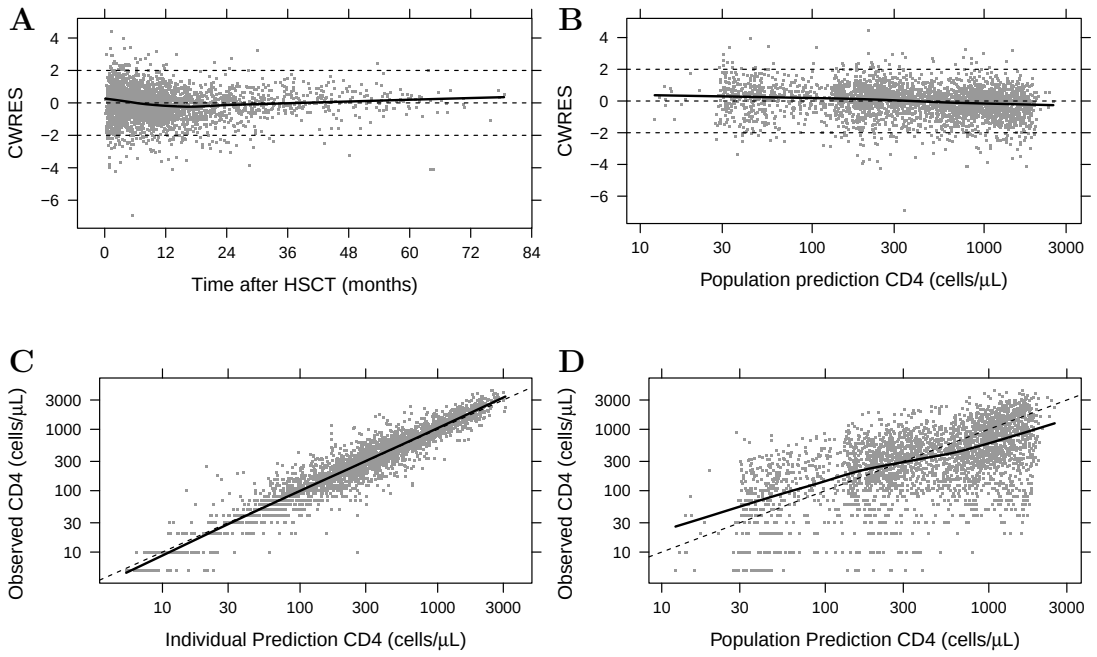

Figure 3.10: Diagnostic plots for the model of long-term CD4 reconstitution. **A** and **B** give the CWRES against time and population prediction respectively, **C** and **D** give the observed data against the individual-level and population-level predictions respectively. Black lines give local regression curves.

3.3.3 Diagnostic plots

Diagnostic plots were used throughout the model building process to assess model misspecification. The diagnostic plots for the structural and covariate model are shown in Figure 3.10. As the plots of the conditionally weighted residuals (CWRES) demonstrate, the residuals are roughly normally distributed with mean 0: they are evenly spread around 0 and more than 95% of them are within ±2. Furthermore, there is no apparent bias either with time after transplant (Figure 3.10A) or with population prediction (Figure 3.10B), as demonstrated by the local regression line, and so the residuals are independent. The plots of the individual predictions and population level predictions against the observed data in Figures 3.10C and 3.10D respectively also demonstrate that there is no inherent bias or model mis-specification, with the individual prediction against the observed data falling along the line of unity.

The prediction corrected visual predictive check (pcVPC) in Figure 3.11 was

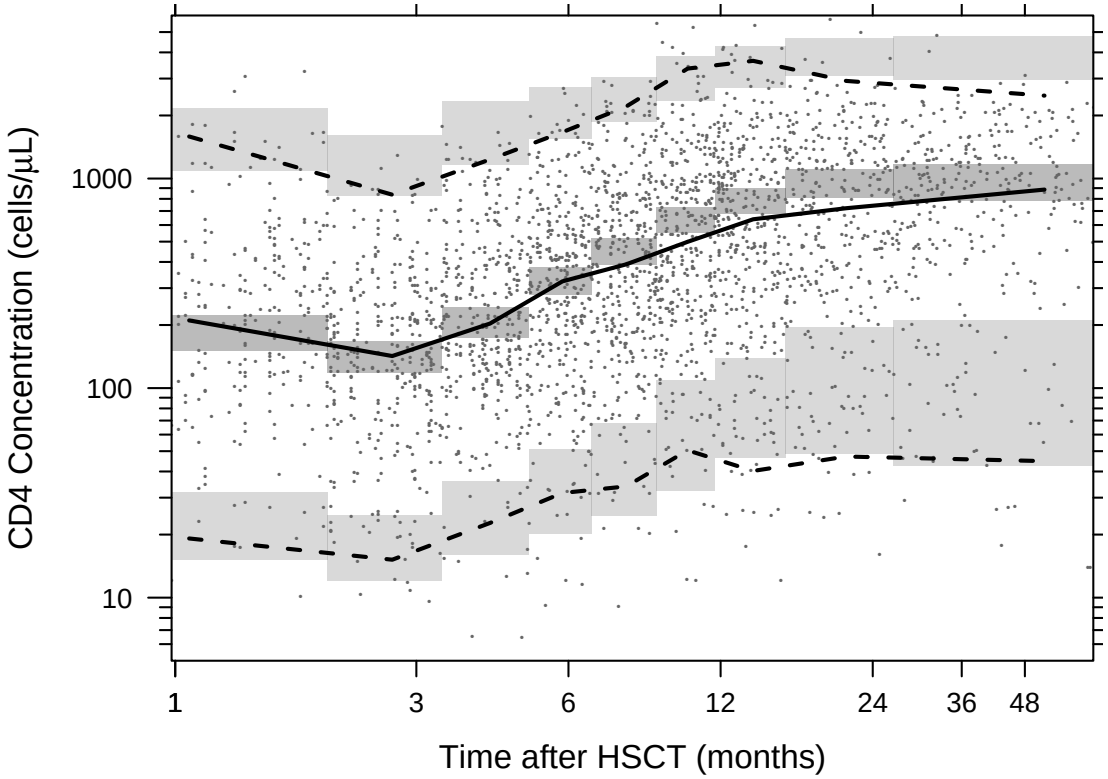

Figure 3.11: Prediction corrected visual predictive check for CD4 concentration reconstitution model. The grey dots give the observed data, with the corresponding median, 2.5th and 97.5th percentiles given by the filled and dashed black lines respectively. The 95% confidence intervals around the model-simulated median and percentiles are given by the grey shaded regions.

performed in PsN version 3.5.3 [124]. 400 new datasets were simulated from the observed data with the parameter estimates obtained from fitting the model to the observed data. As demonstrated, the model captures both the central trend in the observed data and the extent of the variability in the data. The VPC was prediction corrected in order to account for differences caused by the covariates.

3.3.4 Predicting reconstitution in new patients

Having constructed the structural model and the covariate model, the aim was to assess whether the model could be used to make predictions of the reconstitution of new children undergoing HSCT. Using an individual patient's covariates and early observed data following the transplant (the first six months), the method described in Section 3.2.5 was used to form predictions of the individual parameters

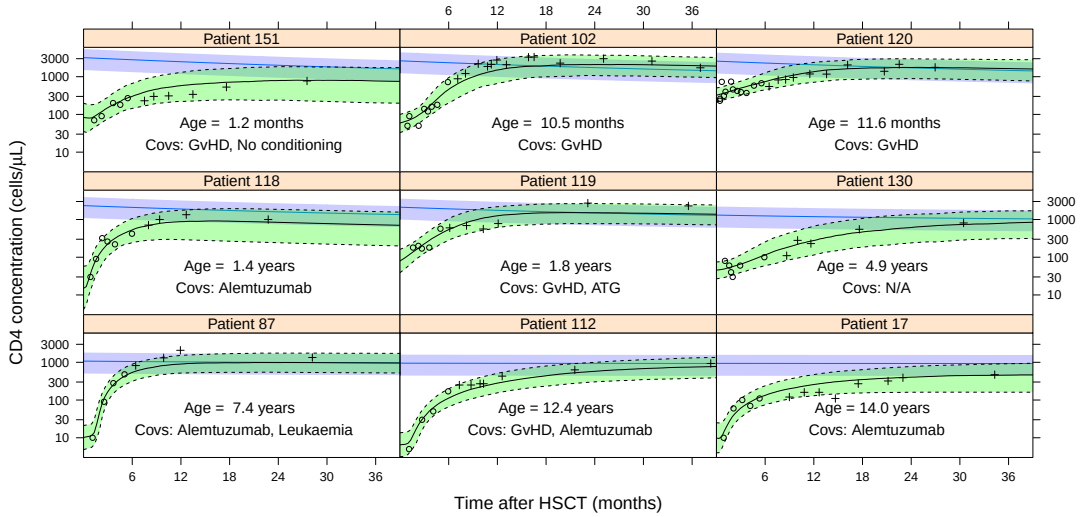

Figure 3.12: Nine examples of predicted reconstitution where the model achieved a good prediction, listed in age order. The circles are the data points that were used to make the predictions, and the crosses are the data not used in forming predictions, for comparison to the predictions. The line is the median prediction, with the green shaded area giving the 90% confidence intervals. The blue line and shaded area are the median and 90% confidence intervals of the expected CD4 concentration of a healthy child of this age.

of each individual in the validation dataset given the population parameter estimates and the variance-covariance (var-covar) matrix for the random effects found from fitting the model to the model-building dataset. From these individual-level parameter estimates and var-covar matrix, 400 sample trajectories were simulated, giving a mean trajectory and confidence intervals for that individual. This trajectory was then compared to the rest of that observed data for that individual in order to assess the predictive ability of the model.

Good predictions were formed by the model in 61 of the patients (81%), with the predicted confidence intervals covering over 75% of the observed data and the trend in the CD4 reconstitution trajectory correctly identified. The model's predictions were largely similar in a further 8 patients (11%), with either the trend in the observed data identified or the majority of the data within the confidence intervals, but not both. In only 6 patients (8%) was the predicted reconstitution trajectory substantially different from the observed data.

The nine examples from the validation dataset of good predictions shown

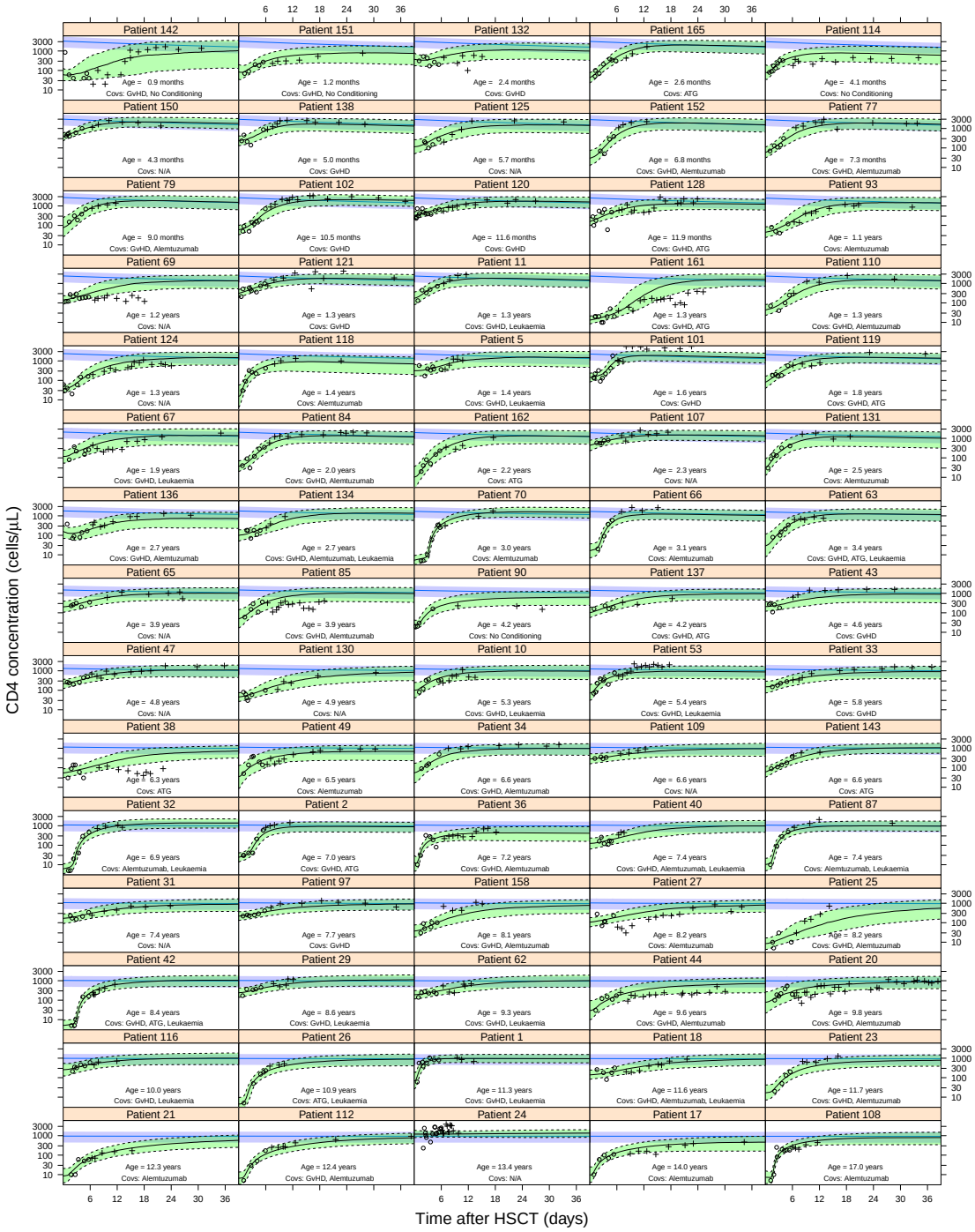

Figure 3.13: Predicted trajectories of all patients in the validation dataset, listed in age order. The circles are the data points that were used to make the predictions, and the crosses are the data not used in forming predictions, for comparison to the predictions. The line is the median prediction, with the green shaded area giving the 90% confidence intervals. The blue line and shaded area are the median and 90% confidence intervals of the expected CD4 concentration of a healthy child of this age.

in Figure 3.12 were chosen as patients with many observed data points and a spread of ages and covariates. All the reconstitution predictions are shown in

Figure 3.13. The predicted reconstitution trajectories vary widely between patients with various initial concentrations, delays to reconstitution and rates of reconstitution. Whilst most patients are predicted to recover to close to the expected CD4 concentration for age, for some, such as patients 151, 118 and 17, their median predicted concentration falls below this level while for others, such as 102, the median rises above the expected concentration. Furthermore, by using a small number of early observed data, patients with the same covariates and similar ages can have substantially different reconstitutions, as demonstrated by patients 102 and 120. Finally, because the model accounts for the age-related changes, it is able to predict accurately the slower reconstitution for older children such as patient 130, who might have been thought to be at risk using a non-modelling technique, but whose slow reconstitution to a normal level could be expected, as is confirmed by the observed data.

3.4 Discussion

In this chapter, I describe the development of a mechanistic mathematical model for CD4 T cell reconstitution following paediatric HSCT. The structural model accounts for the changes in thymic output, loss and proliferation with age, homeostatic mechanisms such as competition for resources and the impairment in thymic output in the months following transplant. Including the age-related changes has made it possible to fit the model to the raw data for CD4 concentration, thus avoiding the need to alter the data to account for age. The model was however simple enough to allow parameters to be estimated from observed CD4 T cell concentration data. The model was then used to perform a covariate analysis. The combination of the structural model and the covariate model has the ability to form individualised predicted trajectories for long-term CD4 reconstitution, providing the basis for a clinical decision support tool.

The results of previous studies that measured CD4 concentrations at fixed time points post HSCT and other studies that measure time taken from HSCT to reach fixed CD4 concentrations both agree well with the output from the model described here for a child of median age (37 months). At three months and one year post HSCT, a reported median CD4 concentration in the range of 100–150 cells/μL [100,101,104,125] and 500–1000 cells/μL [100,104,113] respectively match the output from the model described here of 105 cells/μL and 984 cells/μL at three months and one year post HSCT respectively. Further, a mean time from previous studies of 10.1 months (range 1.1–55.3 months) to reach 500 cells/μL [109] matches the output from the model described here of 7.5 months for a median-aged child, ranging from 5.3 months to 14.3 months for children aged 1 to 10 years.

Estimates for T cell lifespan from the model described here vary across age, increasing from 130 days to 300 days for children aged 1 to 10 years, and to 550 days for a child aged 18 years. These estimates for cell lifespan agree well with the deuterium and BrdU labelling studies that have been conducted more recently and attempt to account for kinetic heterogeneity in the T cell population, finding lifetime estimates between 222 and 611 days (range 167 to 1245) [21,22,126,127].

The absolute values of the function used here for the changes in thymic output with age are uncertain due to a constant related to the length of the time-period of Ki67 expression following cell division [33]. Recause recent studies have implied that the actual thymic output could be as low as 10% of that predicted by Bains *et al.* [22,128], and so in the model a scaling factor λ_0 is used. This allows the absolute values of thymic output to be determined by the model, whilst retaining the theoretical shape of thymic ouput with age. The parameter estimate found with this model gave a thymic output of 23% of the originally predicted output [33], in agreement with the later work [22,128].

Studies of TREC analysis [117,119] and CD31 expression as a marker for re-

cent thymic emigrants [100] have demonstrated that there is a delay of between 5 and 10 months after HSCT before thymic output recovers. This agrees well with the model estimate of 7 months to recover to 90% of the production for age expected by this model. This delay in thymic output has severe immune implications for patients as poor thymic output results both in low T cell concentrations and poor diversity of T cell receptors, inhibiting the ability to fight infection.

The SCM procedure was used to test the available factors for inclusion in the model, finding that the conditioning drugs alemtuzumab and ATG, having leukaemia and acute GvHD status affected the reconstitution. Alemtuzumab and ATG were found to reduce the mean initial CD4 concentration in the model. This initial CD4 T cell concentration mostly reflects donor cells. These drugs are given pre-transplant but have long half lives (15–21 days for alemtuzumab [129] and 30 days for ATG [130]), which means they remain in the body long after transplant. Because they deplete circulating lymphocytes, the reduction in initial concentration found through the model quantifies the extent of the depletion of donor CD4 cells. Though the rate of reconstitution is unaffected, the model predicts that this depletion results in delayed reconstitution to fixed CD4 concentrations, because of the reduced stem-cell independent reconstitution and the lower starting point. This is in agreement with previous studies in both adults and children in which alemtuzumab and ATG were associated with slower and later reconstitution [125,131–133]. This analysis did not find a longer delay to reconstitution from alemtuzumab as opposed to ATG, which was found in another study [134]. This might be due to the quantity of patients that received ATG in the model-building dataset being small.

In contrast to previous work which has found that reduced conditioning results in increased CD4 concentrations [125,131–133], the model found that receiving no pre-transplant conditioning resulted in reduced CD4 reconstitution. An explanation could be that without the space created by pre-transplant ablation with

conditioning, donor T cells and stem cells cannot expand as efficiently. This could also explain the increased long-term CD4 concentrations found with leukaemia patients, who are typically given full myeloablative pre-transplant conditioning to eradicate disease, potentially resulting in more T cell space.

The association found with the model between acute GvHD incidence at raised initial CD4 concentration agrees with previous studies in which T cell depleted grafts were associated with decreased acute GvHD incidence [135–138].

In the observed data, patients who received stem cells derived from cord blood (n=48) as opposed to peripheral blood or bone marrow had faster reconstitution over the months following HSCT. Cord blood transplantation (CBT), however, was not a significant covariate in this analysis, agreeing with Fernandes *et al* [139], because the differences between these patients and the rest could be explained through other covariates. Firstly alemtuzumab and ATG were given to a lower proportion of CBT patients with 40 patients (83%) receiving neither as opposed to 41% of the rest of the patients. Secondly CBT patients were usually younger because of the small grafts available for CBT, with a median age of 1.5 years compared to 3.6 years and 60% under two as opposed to 37% at the time of transplant. Previous work has also demonstrated that the changes in reconstitution following CBT can be explained either by age [109] or the omission of ATG [140].

From the model-building dataset two year mortality could not be predicted, which is unsurprising given that only 20 patients out of the 288 patients in the dataset died from infection. Furthermore, of those patients some of the infections would have been due to lack of innate immunity, leaving only a few for whom low CD4 concentration would have been the cause thus further reducing the possibility of detecting an effect.

A previously unused validation dataset was used to assess the predictive ability of the model. Predictions were formed on an individual basis using only the

patient's significant covariates and data up to six months following HSCT. Accurate predictions were produced in 81% of cases and for up to three years post HSCT. Predictions were based on the parameter means, the var-covar matrix of the random effects and the size of the covariate effects, estimated from fitting the model to the model-building dataset. The quality and quantitiy of the data used in the model development and covariate analysis thus affects the accuracy of the predictions and as more data become available, it is likely that the accuracy could be further improved.

For each individual, parameter means and a var-covar matrix was formed, from which trajectories and confidence intervals were simulated. As the number of observations for that individual increases, the accuracy of the prediction will increase. With few observations, the confidence intervals are wide but as more observations are used, the predictions can be updated and the confidence intervals narrow. With the validating dataset, the confidence intervals narrowed by 11% by using observed data from the first 6 months rather than 3 months post HSCT, and narrowed by a further 12% when using data from the first 12 months.

In conclusion, a mechanistic model has been developed for CD4 T cell reconstitution following paediatric HSCT. Using biological prior knowledge of the effects of age on model components has allowed factors that affect reconstitution to be identified, and knowledge of competition for resources has allowed loss and proliferation to be identified separately. The model's predictive ability was also validated on a separate dataset. To our knowledge, this is the first mechanistic model to be used to form long-term predicted trajectories for the CD4 T cell reconstitution in children following HSCT. These predictions can give clinicians more information about the long-term recovery of the patient's immune system, and hence the potential need for a change in that patient's treatment regimen. Finally, as electronic hospital records become more accessible, it is possible that a clinical tool could be developed that would automatically provide real-time up-

dates of the patient's expected long-term CD4 reconstitution trajectory following HSCT each time a new observation is made.

Chapter 4

Long-term CD4 T cell reconstitution in HIV-infected children starting antiretroviral therapy

4.1 Introduction

Human immunodeficiency virus (HIV) infects cells that express the CD4 protein on their cell surface, which include CD4 T cells, macrophages and dendritic cells. It is a retrovirus, meaning that its genome is encoded in RNA. HIV uses gp120, a protein with a high affinity for CD4, to bind to CD4+ cells, followed by gp41 to fuse the viral envelope into the cell membrane. The viral components then enter the cell where the RNA is reverse-transcribed into double-stranded DNA. This DNA integrates into the host cell's DNA and hijacks the transcriptional machinery of the infected cell to synthesise the necessary protein and RNA components of the HIV virus. These components, having self-assembled into the HIV virus, are released from the infected cell, re-entering the circulation where they

can infect new cells.

In the first few weeks of disease progression, HIV is controlled by the adaptive immune system through antibodies and cytotoxic killing by CD8 T cells, with the initial decline in CD4 T cell concentrations reversed and HIV viral load in the blood contained. Over many years, however, HIV causes a decline in CD4 T cell concentrations. Three major mechanisms for this decline have been proposed. One mechanism is reduced T cell production due to damage either to the bone marrow [141] or to the thymic epithelial space decreasing thymic output [142–145]. Another is increased T cell loss due to increased T cell activation and thus activation induced loss [146–148]. Finally, the third mechanism is an altered T cell distribution between peripheral blood and secondary lymphoid organs, due either to HIV-induced damage to the lymphoid structure [149–151] or the retention of activated CD4 T cells in the lymph nodes for increased time [152]. A combination of these three mechanisms results in a cycle of CD4 T cell decline. Damage to the bone marrow, lymphoid organs and thymus causes increased T cell loss and decreased T cell production. The resulting decreased adaptive immune response leads to increased pathogen burden including higher levels of HIV and persistent infections, which in turn causes increased inflammation and hence further damage to the lymphatic system, bone marrow and thymus [153]. If left untreated the CD4 T cell concentration will eventually decline to the point where the patient has acquired immune deficiency syndrome (AIDS), defined as a CD4 T cell concentration of lower than 200 cells/μL, and so cannot fight infection, resulting in death.

Patients are given antiretroviral therapy (ART) in order to reduce viral loads, thus decreasing the rate of progression towards AIDS by allowing the CD4 T cell concentration to reconstitute. The most common ART drugs used are reverse transcriptase inhibitors, which prevent the process of viral RNA reverse-transcribing into DNA and incorporating into the infected cell's DNA. The two

classes of these are nucleoside-analogue reverse transcriptase inhibitors (NRTIs) such as zidovudine, lamivudine and abacavir and non-nucleoside reverse transcriptase inhibitors (NNRTIs) such as efavirenz and nevirapine. NRTIs create defective deoxynucleotides which lack a 3′-hydroxyl group on the deoxyribose moiety and compete with the natural deoxynucelotides for incorporation into the growing viral DNA and cause chain termination. NNRTIs interfere with the HIV transcriptase enzyme required for viral DNA synthesis. The other common drugs used are protease inhibitors such as ritonavir, lopinavir and nelfinavir, which interfere with the enzymes that assemble new virus particles. There are also drugs which inhibit various other parts of the HIV replication cycle, including cell entry (fusion), integration and maturation.

With the suppression of HIV viral load, often to undetectable levels, fewer CD4 T cells are infected by virus. Homeostatic mechanisms can then allow the CD4 T cell concentration to recover. Furthermore, some of the mechanisms described above that result in CD4 T cell decline might reverse, with evidence suggesting CD4 T cell proliferation rates return to normal [154], thymic output improves [142–145] and CD4 T cells return to circulation from the lymphoid organs [155]. CD4 reconstitution is however often incomplete on ART for reasons that are poorly understood, but are likely to be due to a combination of incomplete recovery from effects of some of the mechanisms described above, and continued reaction to the residual low-level HIV viral load.

The HIV virus is never completely eliminated by ART. The remaining HIV has a very high mutation rate because reverse transcription has a low accuracy, with the HIV mutation rate $\sim 10^{-5}$ mutations/base-pair/cycle, roughly one million times more than that of DNA polymerases $\sim 10^{-9}$–10^{-12} mutations/base-pair/cycle [156]. Resistance to a particular ART drug or class of drugs therefore occurs frequently. To reduce the chances of resistance, drugs from different classes are given in combination because mutations would be required that evade

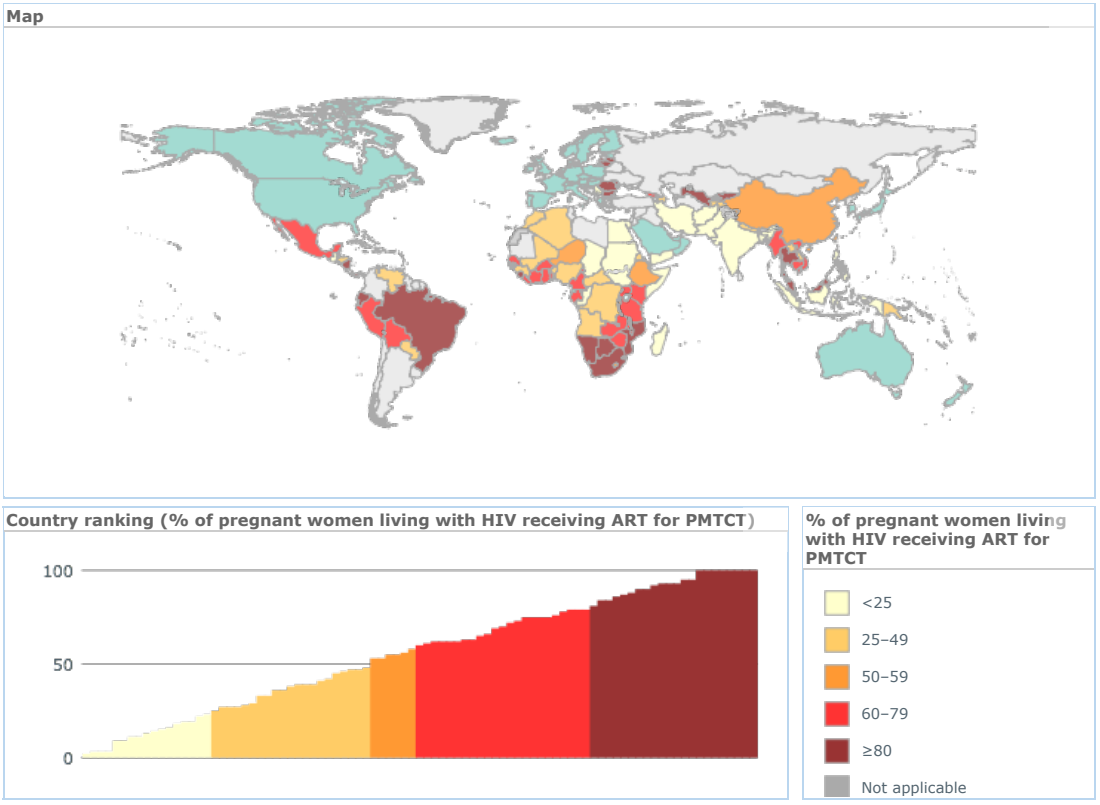

Figure 4.1: Percentage of mothers in middle- and low- income countries around the world receiving ART to prevent mother to child transmission (PMTCT) of HIV. The countries shaded grey (not applicable) are those for which the WHO have no data. The countries shaded green are high-income countries in which usage is very high. From the World Health Organisation [157].

multiple processes simultaneously. This is termed combination-ART (cART) or highly active ART (HAART).

4.1.1 HIV in children

Whilst in adults HIV infection is usually through sexual contact, in children it is usually through mother-to-child transmission *in utero*, at birth or during breastfeeding. Maternal ART can reduce this rate of transmission from the 15% to 45% seen without intervention to 1.2% [158,159]. Of this remaining 1.2% transmission, 65% is thought to occur *in utero* in the last six weeks of pregnancy [160]. Figure 4.1 shows that rates of usage of ART to prevent mother to child HIV transmission is still low in many low- and middle-income countries.

While HIV infection does not seem to affect pre-natal development [161], perinatally HIV-infected children who were untreated had a poor prognosis, with about 1/4 developing AIDS by one year [160,161] and a mean age at death of 9.4 years old [162].

With ART, children now survive well into adulthood. This survival and particularly the resulting longer times on treatment however brings its own problems. On the one hand, multiple studies have found that the younger a patient is at the start of ART, the greater the speed of the recovery and the higher the long-term CD4 T cell concentration [163–170]. On the other hand, the longer they spend on treatment, the more exposed they are to side-effects and long-term toxicities of the drugs, such as lipodystrophy, fatigue, psychiatric symptoms and heart disease. Furthermore, the CD4 T cell concentration of patients receiving ART do not fully reconstitute to the level expected of a healthy child [171,172] and low CD4 T cell concentrations are associated with a higher tumour incidence rate [173] and cardiovascular disease [174,175]. Thus with children potentially living longer with reduced CD4 T cell concentrations, understanding what affects the rate and extent of their immune reconstitution is very important.

Modelling, and in particular mechanistic modelling, has the ability to improve this understanding. Because the CD4 concentration and HIV viral load are coupled by their interactions, it is important to understand what affects the viral load as part of understanding what affects the CD4 concentration; mechanistic models make it possible to include these interactions. Furthermore, mechanistic models offer the possibility of ascertaining which part of the immune system is affected causing the CD4 concentration not to stage a full recovery on ART. While an empirical model has been used to assess CD4 reconstitution in children on ART [171,172] and mechanistic models developed for adults have been applied to children [176] (see below), to our knowledge a mechanistic model that takes into account age-related changes in the immune system has not yet been developed for

HIV-infected children. A model of this form could be used to perform covariate analysis to find the factors that affect HIV viral decline and CD4 reconstitution. It could also be used in combination with pharmacokinetic models to inform the expected effects of drugs on both viral load and CD4 concentration. This could improve both the development of personalised medicine approaches with individual dose-adjustments to minimise toxicity whilst maintaining CD4 concentrations and keeping the incidence of resistance low, and the development of novel therapies such as new ART drugs and other therapies such as gene therapy.

4.1.2 Modelling the dynamics of the immune system and HIV

The simplest modelling used to quantify the dynamics in HIV infected patients starting ART is empirical modelling. For viral load, Ho *et al* [177] and Wei *et al* [178] approximated viral decline on ART with an exponential decay. They had to allow however for a time-lag before the decay started in order to fit this model. For CD4 T cell reconstitution, empirical modelling has successfully been applied in the paediatric context, although it required the adjustment of the data to account for age, either using CD4 T cell z-scores [171] or the ratio of the observed CD4 T cell concentration to that expected for a child of that age at the time of the measurement [31,172]. The models for CD4 reconstitution cannot easily incorporate information on HIV viral load and so lose information that affects outcome such as drug efficacy, which is mainly inferred from viral suppression. Furthermore, patients that did not fit the empirical model had to be excluded [172]. Viral load information, where available, might have provided some explanation as to the differences between these patients and those that did fit the model.

In contrast to empirical models, mechanistic models can incorporate viral load, CD4 concentration and their inter-dependence. Many models have been

proposed, with various mechanistic variations, discussed below. At the most basic level, CD4 T cells are infected by HIV and in turn produce more HIV. Many modifications and complications have been proposed and tested in order to explain both the short- and long-term dynamics of HIV and CD4 reconstitution.

The basic dynamic model

The simplest of the mechanistic models is that of Nowak *et al* [179] and Bonhoeffer *et al* [180]. In this model, there are three compartments: uninfected CD4 cells which can be infected by free HIV to become infected CD4 cells, and these infected CD4 cells then produce more free HIV. They used this model to look at the incidence of resistance to ART.

Splitting infected CD4 T cells

The earlier mechanistic models of McLean *et al* [181] and Perelson *et al* [182] split the infected T cell pool into actively infected and latently infected T cell populations. This was to represent mechanistically the split between T cells that produce virus and those that are infected and do not. Funk *et al* [183] further split the infected cells into persistently infected and defectively infected CD4 cells. These subsets represent very small percentages of the total infected cells, and mechanistically represent the CD4 cells that emit small amounts of HIV and survive for long time-periods, and cells in which the virus is defective respectively.

Splitting free virus

Another model of Perelson *et al* [184–186] split the free virus into infectious and non-infectious virus. This was done to model the effect of ritonavir, a protease inhibitor that renders the free virus produced non-infectious. Hence in the presence of ritonavir infected CD4 cells only produce non-infectious virus, which dies at the same rate as the infectious virus, resulting in the decline in viral load.

Splitting the uninfected CD4 T cells

In the model of de Boer and Perelson [187], the uninfected CD4 T cells were split into active and quiescent, resting, cells. The model aimed to represent the notion that only activated cells could be infected by virus, and so these quiescent cells could not be infected.

Homeostatic modelling

Another branch of mechanistic modeling involved developing models that included homeostatic mechanisms for T cell production. Ribeiro *et al* [188] developed a model for T cell dynamics with two compartments, a resting and an activated compartment. Cells can be activated from the resting compartment into the activated compartment where they can proliferate, die or return to the resting compartment. Yates *et al* [189] adapted this model, with cells in the resting compartment also able to die and cells in the active compartment proliferating as they return to the resting compartment, and the rates density dependent. They then extended the model to include firstly an HIV-infected T cell compartment and finally an immune activated T cell compartment as well. They applied this model to explain the long-term depletion of CD4 T cells in HIV infection.

Hapuarachchi *et al* [116] started with the simplest of the models for CD4 dynamics from Yates et al [189], but used different functions for the density dependence of proliferation and death in the resting compartment to explain competition for resources during homeostasis, as well as a quadratic function for cell death in the activated compartment to model FaS-FaS ligand activation induced fratricide [190].

Fitting mechanistic models to data

Unlike the empirical models, the mechanistic models were often written in order to study the dynamics of the system, rather than to model data. As such, while some of them compared the model output to observed data for HIV viral load as validation, only Perelson *et al* [184] and Funk *et al* [183] attempted to estimate

some parameters from data, using nonlinear regression analysis. This however only worked where it was possible to find closed-form solutions. The rest of the studies above, in particular with models where closed-form solutions were not possible, fixed all parameters to literature values.

Recently studies have used nonlinear mixed effects modelling in order to improve parameter identifiability while fitting these models to data. They have also applied the models to both CD4 concentration and viral load data in parallel, again to improve parameter identifiability. Early attempts to fit full-parameter models of this kind to data used Bayesian approaches [191,192]. These were however highly computationally expensive. More recently, maximum-likelihood approaches have been used. Guedj *et al* [193] performed a full likelihood inference by adapting a Newton-like algorithm. They tested a simplified version of the model of de Boer and Perelson [187] with four compartments: viral load, uninfected active CD4 cells, uninfected quiescent CD4 cells and infected CD4 cells, and were able to obtain parameter estimates. Prague *et al* [194,195] also used the same system to form predictions of CD4 concentrations following changes in ART.

Drylewicz *et al* [196] compared two models, fitting both to data using NLME, the simpler of which has three compartments: viral load, uninfected CD4 cells and infected CD4 cells. For the other, the infected cells were divided into active infected and latent infected CD4 cells, where only the active proportion of the infected CD4 cells can produce more virus. They found a better fit with the model that splits the infected cells into active and latent CD4 cells.

Laveille *et al* [197] compared three of the previously proposed models, applying each to the same dataset for HIV-infected adults and immune reconstitution. They compared the basic dynamic model of Nowak *et al* [179] and Bonhoeffer *et al* [180] to the model that splits uninfected CD4 cells into quiescent uninfected and active uninfected CD4 cells of de Boer and Perelson [187] to the model that

splits infected CD4 cells into active and latent CD4 cells of McLean *et al* [181] and Perelson *et al* [182]. In comparing these variants, for all three the free virus is split into infectious and non-infectious virus in the manner of Perelson *et al* [184–186]. Lavielle *et al* find that the model with the infected cells split into latent and active provides the best fit to data as judged by Bayesian Information Criterion (BIC), similarly to results found by Guedj *et al* [193].

Apart from the model of Haupuarachchi *et al* [116], none of the mechanistic models account for the age-related changes in the immune system during immune system development. Bouzza *et al* [176] applied the basic dynamic model to paediatric data with no changes to the model to account for age. They did not investigate any age-related effects and so it is not possible to tell whether their results were biased with age.

Why alter the previous models?

Infected CD4 cells only represent a very small fraction of the total CD4 cells, with the the mean observed in one study at 0.14% and the maximum observed at 1% [198]. With the parameter estimates found in the modelling of paediatric patients by Bouzza *et al* [176], the mean initial concentration of infected CD4 cells is low at 9.3 cells/μL. However, this still represents 3.8% of the initial CD4 T cells, nearly four times the highest observed infected CD4 T cell proportion [198]. As such, it is questionable whether the infected cells compartments in these models are representing the mechanisms they intend to. In this chapter I therefore simplify the models by removing the infected cells compartment.

The models described above assume that viral load follows almost entirely a mono-phasic exponential decay, whereas current evidence suggests that it follows a multi-phasic decline in the first few months after the start of ART [199]. Free virus enters the blood stream when released from an infected cell, and so the first phase of viral load decline is the result of declining viral production in activated

infected CD4 cells, the second phase in partially activated CD4 T cells, macrophages and dendritic cells, and the third phase in resting memory CD4 T cells. As such a model that allows for this multi-phasic decline would be preferable.

The models above also display oscillatory behaviour in certain circumstances [179,180,182], particularly when viral load increases with the development of resistance. These oscillations cause a rapid decline in CD4 concentration on the development of resistance. In the studies where these models have been applied to data, there is little evidence that either these oscillations or the sharp decline in CD4 concentration exist *in vivo* [183,184,193–195]. The data that I have do not appear to support such oscillatory behaviour in viral load or CD4 concentration. Furthermore, there is not the sharp decline in CD4 concentration that results from the oscillations. The model is therefore altered to remove the oscillations. Finally, since no previous model has included the effects of age-related changes to the immune system into the models, I have constructed a new model by adapting the previous models for HIV virus and CD4 concentration dynamics and including the effects of age.

4.2 The data

The work in this chapter was conducted using two datasets: the Paediatric European Network for the Treatment of AIDS (PENTA) 11 and the Antiretroviral Research for Watoto (ARROW) clinical trials. HIV-infected children were studied in both trials.

4.2.1 PENTA 11

PENTA 11 was a randomised Phase II trial which recruited 109 HIV-infected children from nine countries (France, Germany, Italy, Poland, Spain, Switzerland, Thailand, USA, and the UK) between 2004 and 2006 [200,201]. The aim of the

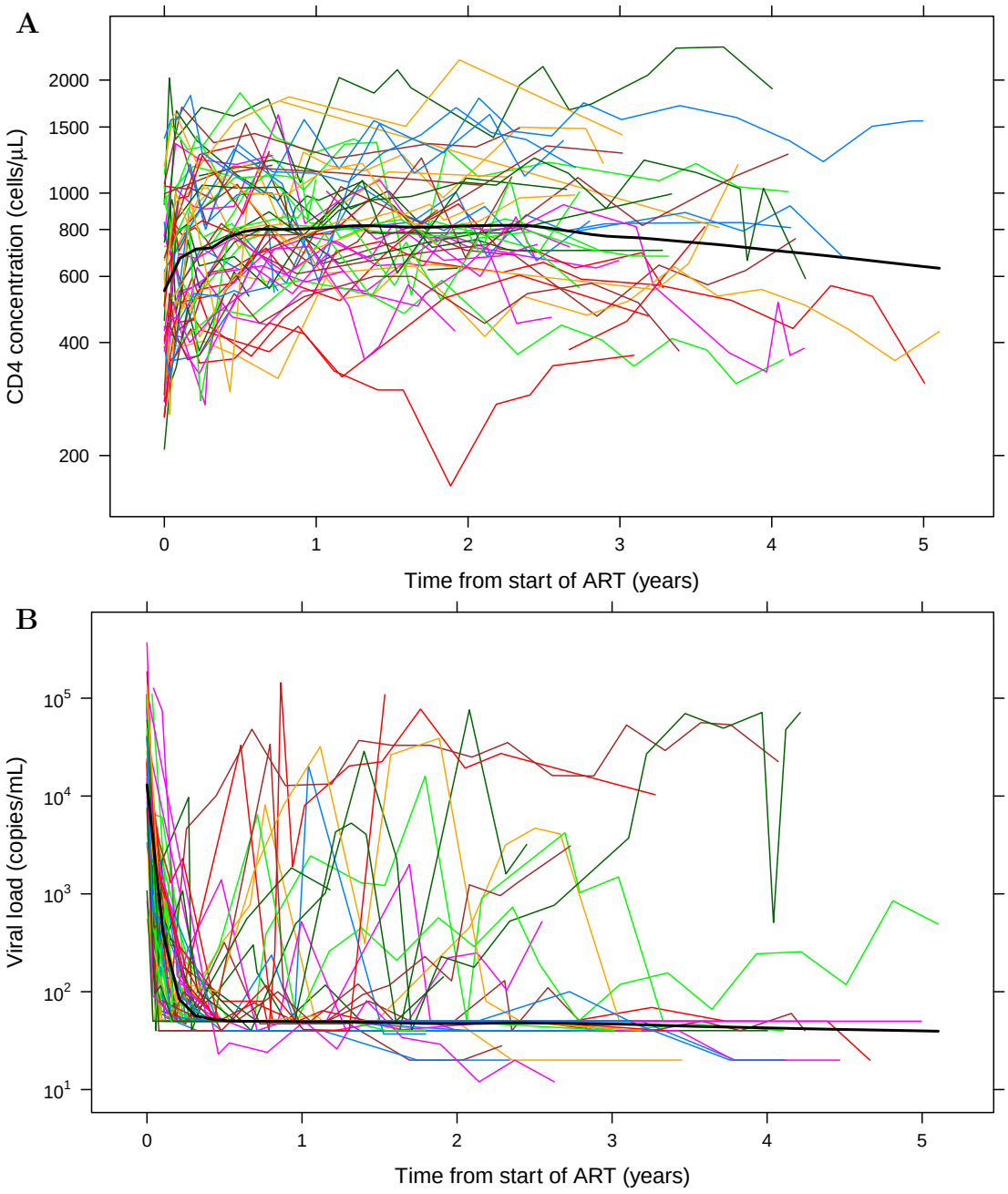

Figure 4.2: PENTA data for children in the PTI arm as they restart ART following the PTI for (**A**) CD4 T cell reconstitution and (**B**) viral load. Each coloured line gives the data for one individual's response to ART. The black line gives a local regression curve.

study was to ascertain whether planned treatment interruption (PTI) in ART disadvantaged the children in terms of their immunology, virology or clinical status. Children in the trail were randomised into continuous treatment (CT) or PTI arms.

Table 4.1: Breakdown of the demographics and the drugs used for the patients in the PENTA dataset for HIV-infected children.

	No	%		No	%
Sex			**Age**		
Male	32	48	4→7	12	18
Female	34	52	7→10	26	39
Ethnicity			10→14	16	24
Asian/Thai	11	17	14→	12	18
Black African/other	19	29	**Number of PTIs**		
Mixed black/white	2	3	1	36	55
Mulato	5	8	2	30	45
White	29	44	**Reason for end of PTI**		
Country			PTI 1		
Germany	1	2	Reached end time	60	91
Spain	15	23	CD4 decline	4	6
France	3	5	Carer request	2	3
Italy	16	24	PTI 2		
Poland	2	3	Reached end time	26	39
Switzerland	2	3	Viral load failure	2	3
Thailand	11	17	Other	2	3
UK/ROI	16	24			

ART Drugs						
	Before PTI		After PTI 1		After PTI 2	
	No	%	No	%	No	%
Abacavir	17	26	24	36	14	21
Didanosine	18	27	13	20	6	9
Efavirenz	17	26	26	39	14	21
Fosamprenavir	0	0	7	11	4	6
Lamivudine	45	68	48	73	24	36
Lopinavir	11	17	16	24	8	12
Nelfinavir	12	18	3	5	0	0
Nevirapine	25	38	21	32	8	12
Ritonavir	3	5	3	5	0	0
Stavudine	30	45	15	23	4	6
Tenofovir	0	0	2	3	0	0
Zalcitabine	1	2	0	0	0	0
Zidovudine	19	29	20	30	8	12

In order to enter the trial, children had to be aged 2–15 years old, be chronically infected with HIV, have been on ART for longer than 24 weeks, and have an undetectable HIV viral load (< 50 copies/ml). Children aged 2–6 years had to have a CD4 T cell concentration above 30% of the expected CD4 T cell concentration of a healthy child of that age, and children aged 7 or over, above

25% of the expected CD4 T cell concentration as well as a CD4 concentration $\geq$ 500 cells/μL [200,201].

During the trial, there were procedures in place to protect the children from excessively low CD4 concentrations. In the children on PTI, ART was restarted in children 2–6 years old if the CD4 concentration dropped below 20% of the expected CD4, and in children 7 or over below 20% of the expected CD4 or less than 350 cells/μL. Furthermore, no PTI exceeded 48 weeks in duration and only children who had had a PTI duration of $>$ 10 weeks during their first PTI and had been back on ART for $\geq$ 24 weeks could undertake a further PTI. Children were followed for up to five years [200,201].

Of the 109 patients, 53 were randomised into the CT arm and 56 into the PTI arm. The analysis of this chapter concerns the children in the PTI arm as patients that restarted ART after a period of no therapy. The median age at restart of ART was 9 years (range 2–16 years). Patients were followed for up to five years after resumption of ART, and in the PTI arm, there were 787 measurements of CD4 T cell concentration and 776 measurements of viral load of which 540 were below the limit of quantification. There were two limits of quantification for viral load (40 copies/mL and 80 copies/mL) depending on the assay used. Patient demographics are shown in Table 4.1, and the data are shown in Figure 4.2.

4.2.2 ARROW

ARROW was an open-label randomised controlled clinical trial which enrolled 1206 HIV-infected children who were receiving ART for the first time at four centres: three in Uganda and one in Zimbabwe [202]. Children aged between 3 months and 17 years were enrolled over an 18 month period between 2006 and 2008 and followed up for 3½–5 years. ARROW had two main objectives: the first was to assess whether ART can be used effectively and safely without monitoring the effects of ART through regular blood tests; the second was to assess the long-

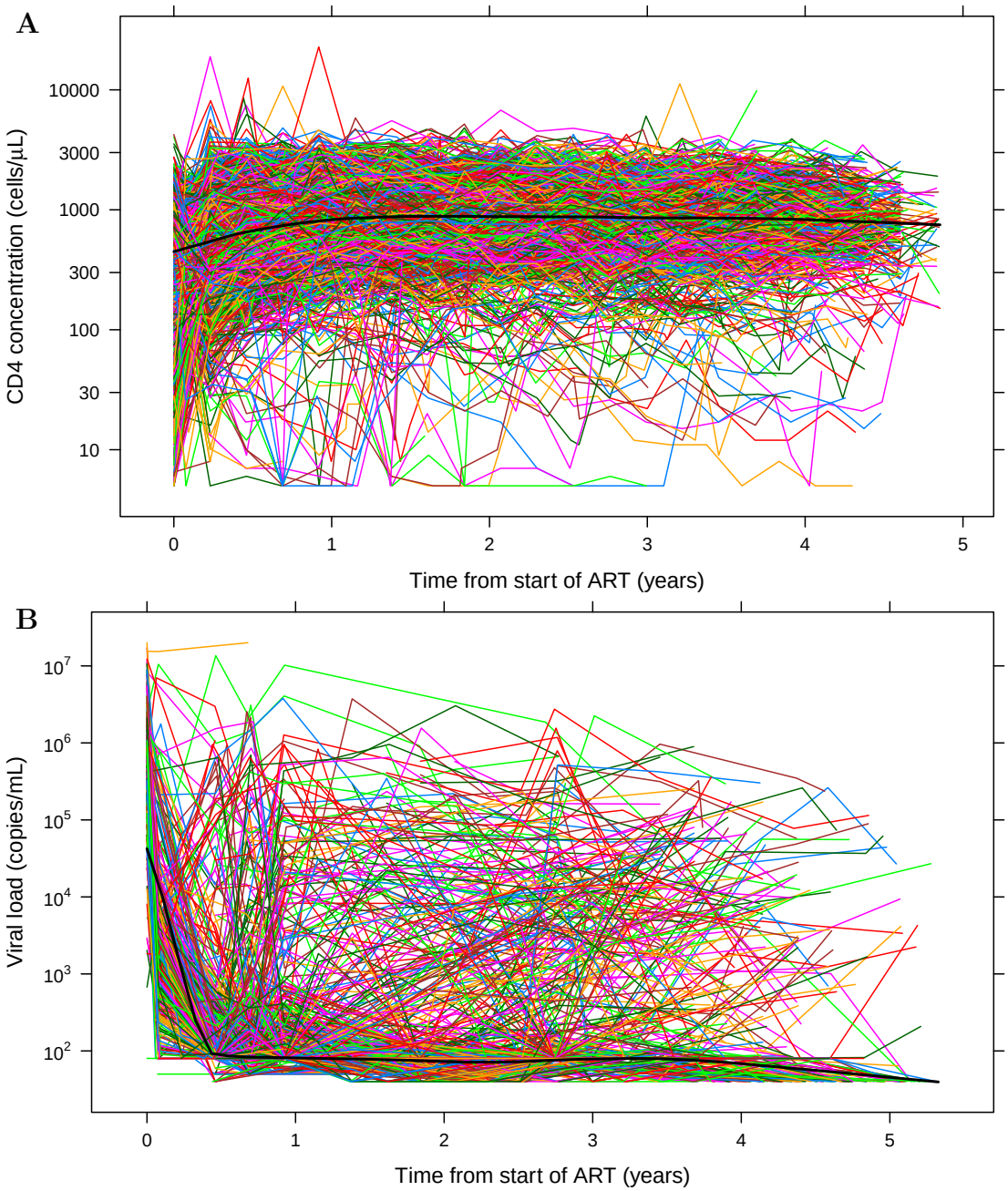

Figure 4.3: ARROW data for (**A**) CD4 T cell reconstitution and (**B**) viral load for children commencing ART. The coloured lines give the responses of individual children to ART. The black line gives a local regression curve.

term impact of starting ART with four drugs from two classes, three NRTIs and one NNRTIs, followed by maintenance with three drugs as opposed to continuous treatment with the standard three drug regimen.

Before this trial, for children starting ART, standard practice was to perform laboratory tests every three to six months assessing viral load, immune

Table 4.2: Breakdown of the demographics and the drugs used for the patients in the ARROW dataset for HIV-infected children.

	No	%		No	%
Age			**Sex**		
0→2	239	20	Male	596	49
2→4	222	18	Female	610	51
4→7	219	18	**Life Status**		
7→10	277	23	Alive	1150	95
10→	249	21	Dead	56	4.6
Centre			**Cause of death**		
Entebbe, Uganda	188	16	Def/prob HIV related	44	3.6
JCRC, Kampala, Uganda	318	26	Def/prob drug related	2	0.16
Harare, Zimbabwe	400	33	Uncertain HIV/drug related	2	0.16
PIDC, Mulago, Uganda	300	25	Unlikely HIV/drug related	2	0.16
			Unknown	6	0.50

Randomisation			
	Drug Arm		
	A	B	C
Laboratory and clinical monitoring	198	201	201
Clinical monitoring	199	203	204

Abbreviations: JCRC, Joint Clinical Research Centre; PIDC, Paediatric Infectious Diseases Centre.

markers and ART side effects. This is an expensive process, and the benefits of these tests to HIV-infected children had not been examined. The first part of this study compared two arms of patients, one where patients received clinically driven monitoring (CDM) and the other where patients received laboratory plus clinical monitoring (LCM). In both arms, haematology and liver function tests and measures of CD4 and CD8 T cell concentrations were performed regularly, but the results were withheld in the CDM arm unless requested by the treating physician. Meanwhile the results were independently monitored by the Data Monitoring Committee. The exception to this was haemoglobin results, which were returned on all children at week 8.

For the second part of the trial, children were independently randomised to three arms. The control arm A received three drugs, two NRTIs (lamivudine and abacavir) and one NNRTI (nevirapine or efavirenz). Arms B and C received a

fourth drug, the NRTI zidovudine for the first 36 weeks, after which they dropped one of the drugs for the remainder of their first-line ART. Arm B stopped taking zidovudine, while Arm C stopped taking the NNRTI.

While HIV viral loads were not assayed in real-time, specimens were stored allowing retrospective testing. Clinical markers such as height, weight, WHO disease staging and any adverse events were also recorded at clinic visits.

The trial found that there were no significant differences between the groups, either for initiating treatment with four drugs rather than three or between clinically driven monitoring and laboratory plus clinical monitoring [202]. The authors conclude that given the expensive laboratory testing does not improve outcome, it is unnecessary and so more of the limited resources should be spent on ART drugs, which would allow longer treatment in more individuals [202].

Because there were no significant differences found, in this analysis I pooled the data from all the groups into one cohort for modelling. In the dataset there are 20,989 measurements of CD4 concentration and 4795 measurements of viral load of which 2760 are below the limit of quantification. A breakdown of the demographics is given in Table 4.2, and the data is shown in Figure 4.3.

4.3 Methods

4.3.1 Model building

In order to model the responses of HIV-infected children to ART and its effect on CD4 T cell reconstitution, I start with the basic dynamic model described above by Nowak *et al* [179,180] and shown in Figure 4.4, as a simple model for viral load and CD4 concentration co-dependence that had been applied to paediatric data [176]. This model was then modified by removing the explicit compartment for infected T cells in a similar manner to that described by Perelson *et al* [182], by including the mechanistic elements for age and competition for resources as

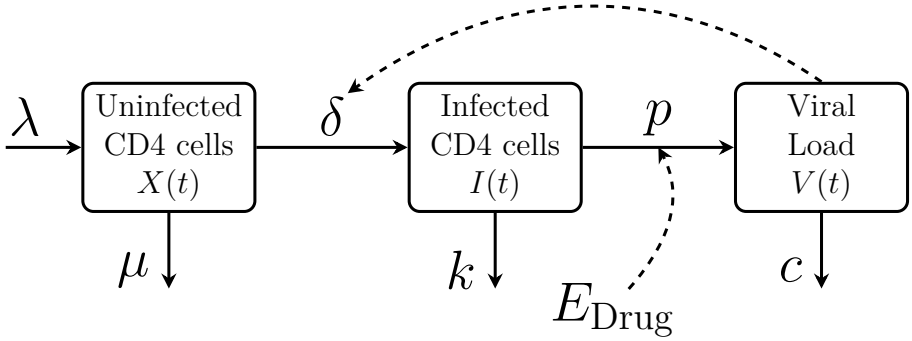

Figure 4.4: Schematic of the basic dynamic model for CD4 concentration and HIV dynamics as developed by Nowak *et al* [179,180] and as applied to a paediatric dataset by Bouzza *et al* [176]. Uninfected CD4 cells are produced at zero-order rate λ. These cells die at rate μ and are infected by virus at rate δV to become infected cells. Infected cells die at rate k and produce virus at rate p. Virus then dies at rate c. The effect of the ART drugs is to reduce the production of new virus p.

in the model for CD4 reconstitution in paediatric HSCT in Chapter 3, and by allowing for multi-phasic decline in viral load. This process is described in more detail below.

Basic dynamic model properties

The basic dynamic model as described above and shown in Figure 4.4 has three compartments representing uninfected T cells $X(t)$, infected T cells $I(t)$ and viral load $V(t)$. In this model, uninfected T cells are produced at a zero-order rate λ and then die at a first-order rate μ, where μ gives the difference between cell death and cell proliferation. Uninfected cells can become infected at a virus dependent first-order rate δ. These infected cells can then die at first order rate k. Virus is produced by infected cells at a first-order rate p, dependent on the concentration of infected cells, and then dies at first-order rate c. The equations

for the dynamics are given by,

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \lambda - \mu\, X - \delta\, V\, X \tag{4.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I = \delta\, V\, X - k\, I \tag{4.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V = p\, I - c\, V\,. \tag{4.3}$$

When using this model it is assumed that the system is in steady state before any ART is given. As such, the model parameters are inter-related with,

$$\delta = \frac{\lambda - \mu\, X_0}{X_0\, V_0} \tag{4.4}$$

$$k = \frac{\lambda - \mu\, X_0}{I_0} \tag{4.5}$$

$$p = \frac{c\, V_0}{I_0}\,, \tag{4.6}$$

where T_0, I_0 and V_0 are the uninfected CD4, infected CD4 and viral load respectively before the start of ART. This means that the model depends on six parameters, $\lambda, \mu, c, X_0, I_0, V_0$.

Dynamics with ART

The effect of the ART drugs is to prevent HIV virus being formed. As such, they can be modelled on the viral load compartment as,

$$\frac{\mathrm{d}}{\mathrm{d}t}V = p\, I(1 - E_{\mathrm{Drug}}) - c\, V\,. \tag{4.7}$$

where E_{Drug} is between 0 and 1 and would be 1 for a fully effective drug and 0 for a fully ineffective drug. The pharmacodynamic models of Section 2.2.3 in Chapter 2 are often used to define E_{Drug} based on drug concentration. In this analysis, because ART is given continuously, it is possible to assume drug concentration is continuous and as such treat E_{Drug} as a parameter confined to between 0 and 1.

In the case of a fully effective drug, the dynamics become very simple. The viral load no longer depends on the infected CD4 cells and can be solved to give $V = V_0 \mathrm{e}^{-ct}$, and correspondingly the dynamics of the uninfected cells no longer depend on the infected cells. The dynamics of infected cells therefore become isolated, depending only on the initial concentration, I_0.

Simplifying the basic dynamic model

If E_{Drug} is less than one, the system becomes more complicated, with inter-dependence of all three compartments. However, the proportion of T cells infected at any one time is very small ($< 1\%$ [198]). This means that the observed data for CD4 T cell concentrations will largely be uninfected CD4 cells. Furthermore, because viral dynamics are very fast in comparison to CD4 dynamics, the changes in viral load are largely driven by changes to the concentration of infected cells. As a result the quasi-steady state approximation can be used to simplify the model, removing the infected CD4 cell compartment and leaving a two compartment version of the model, using a similar method to that of Perelson *et al* [182].

Because the dynamics of viral load are so dependant on the concentration of infected CD4 cells, we can say that the viral load is in quasi-steady state with infected CD4 cells. This means that in the time-frame of virus, infected cells are at any instant effectively constant and the change in viral load with time with constant infected CD4 cell concentration is 0,

$$\left.\frac{\mathrm{d}V}{\mathrm{d}t}\right|_I = 0. \tag{4.8}$$

As a result, the (unmeasured) infected CD4 concentration can be defined in terms

of the (measured) viral load,

$$I(t) = \frac{c}{p}V(t) = \frac{I_0}{V_0}V(t). \tag{4.9}$$

Substituting (4.9) into (4.2) gives,

$$\frac{I_0}{V_0}\frac{\mathrm{d}V}{\mathrm{d}t} = \delta\, X\, V - k\frac{I_0}{V_0}V, \tag{4.10}$$

and by rearranging and using the relations for δ, I_0 and p in (4.4) – (4.6), we get,

$$\frac{\mathrm{d}}{\mathrm{d}t}V = \delta\frac{V_0}{I_0}X\, V - k\, V \tag{4.11}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V = \frac{(\lambda - \mu\, X_0)}{X_0\, V_0}\frac{k\, V_0}{(\lambda - \mu\, X_0)}X\, V - k\, V \tag{4.12}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V = \frac{X}{X_0}k\, V - k\, V. \tag{4.13}$$

Including the effects of ART drugs gives us the model for viral dynamics,

$$\frac{\mathrm{d}}{\mathrm{d}t}V = k\frac{X}{X_0}V(1 - E_{\mathrm{Drug}}) - k\, V\,, \tag{4.14}$$

where the rate of the viral load dynamics is now dependent on k, the loss rate of infected CD4 T cells.

Allowing for multi-phasic viral load decline

The decline in viral load on ART is thought to be multi-phasic, with the loss rates of viral load decreasing as the infected CD4 cells that contain virus increasingly become the longer-lived cell subsets [199]. This effect was modelled with a sigmoidal function on the loss term, $\frac{V}{V+V_{50}}$ such that viral load loss rates decline with viral load. This also has the effect of allowing viral load decline to stop with some residual viral load remaining. So that the parameter V_0 retains the same meaning as the steady-state viral load before ART, we reparameterise (4.14) to

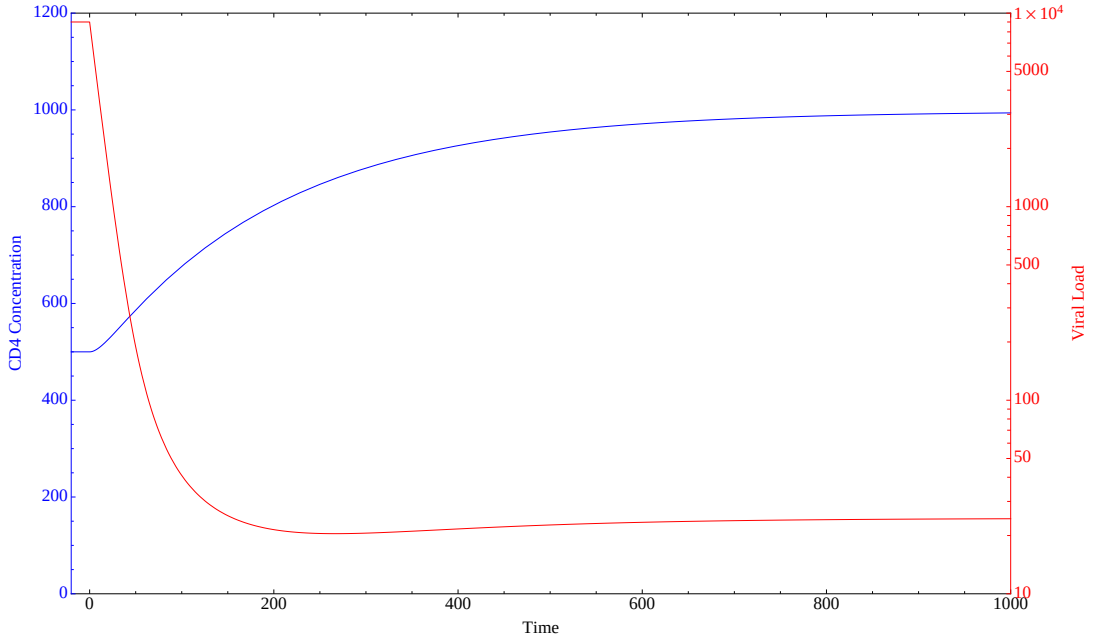

Figure 4.5: Output from the simplified basic dynamic model, including allowing for multi-phasic viral load loss. The model starts in steady state before the start of ART on day 0, after which viral load falls rapidly and CD4 concentration reconstitutes slowly. $X_0 = 500, \lambda = 5, \mu = 0.05, V_0 = 9000, k = 0.1, V_{50} = 100, E_{\text{Drug}} = 0.9$.

give,

$$\frac{\mathrm{d}}{\mathrm{d}t}V = \frac{k\,V_0}{V_0 + V_{50}}\,\frac{X}{X_0}V(1 - E_{\text{Drug}}) - k\,V\frac{V}{V + V_{50}}. \tag{4.15}$$

The equation for T cell concentration is

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \lambda - \mu\,X - \delta\,XV, \qquad \text{where} \quad \delta = \frac{\lambda - \mu\,X_0}{X_0\,V_0}. \tag{4.16}$$

The output from this model in Figure 4.5 shows that viral load declines to a non-zero steady state and CD4 concentration increases at a much slower rate to a steady state.

Development of resistance

As mentioned in Section 4.1, because the RNA transcription in HIV replication is inexact, there is a very high mutation rate. As such, HIV often becomes resistant to ART. Mechanisms for the development of resistance are complicated, but here

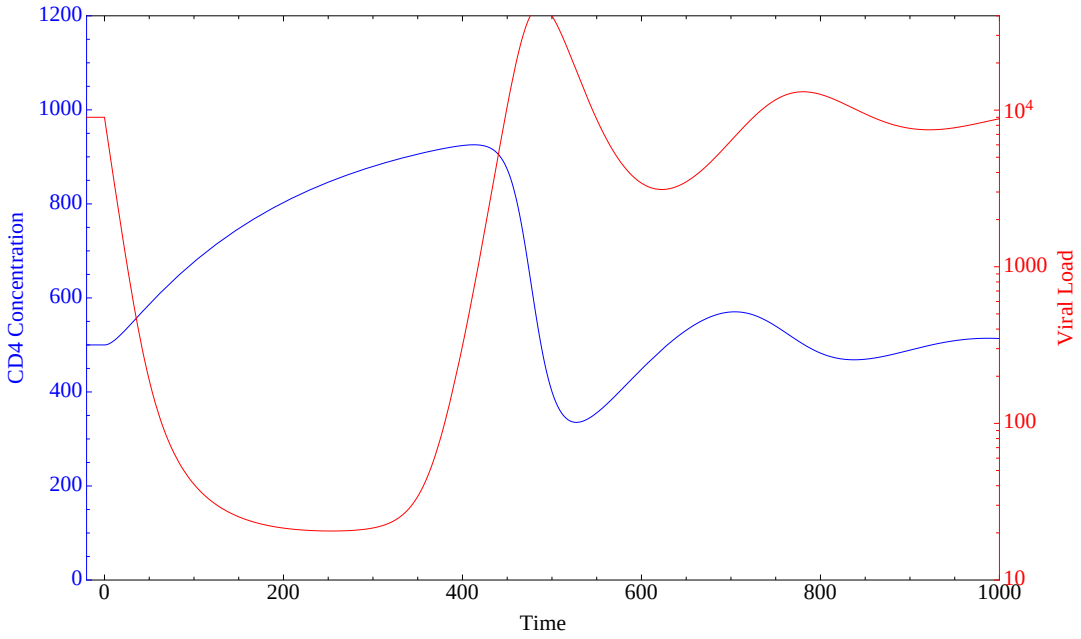

Figure 4.6: Effects of the development of resistance in the simplified basic dynamic model. The model starts in steady state before the start of ART on day 0, after which viral load falls rapidly and CD4 concentration reconstitutes slowly. After a certain amount of time, resistance to ART develops and E_{Drug} decreases. Viral load therefore increases fast and CD4 declines. With the model in this form, CD4 declines rapidly and oscillations form. $X_0 = 500, \lambda = 5, \mu = 0.05, V_0 = 9000, k = 0.1, V_{50} = 100, E_{\text{Drug}} = 0.9, T_{\text{Drug}} = 300, \rho = 20$.

we use a very simple model for the development of resistance with the intention of demonstrating the effects on the model dynamics.

In this model, the effects of resistance are simply to reduce the drug effect, E_{Drug} to zero at time t_R with a time-dependent step function to produce the following for E_{Drug} the drug effect including resistance,

$$E_{\text{Drug}} = E_{\max}\left(E_{\text{R}} + \frac{1 - E_{\text{R}}}{\exp[\frac{t-t_R}{\rho}] + 1}\right), \tag{4.17}$$

where t_R is the time at which resistance to the drug with be $\frac{1}{2}$, $E_{\max}$ is the maximum drug effect before resistance, E_{R} is the residual drug effect after the development of resistance, and ρ is the rate at which resistance develops.

The development of resistance has some unexpected effects in the model, as demonstrated in Figure 4.6. The inter-dependence of viral load and CD4

concentration causes oscillatory behaviour, as high CD4 concentrations cause large growth in viral load, which in turn causes high death in CD4 concentration, resulting in a decline in production of viral load and so on. Furthermore, viral load spikes to quantities far higher than those observed before the commencement of ART, and CD4 concentrations decline very sharply on the emergence of resistance. These effects are observed in many of the previously proposed models [179,180, 182].

There is little evidence that either these oscillations or the sharp decline in CD4 concentration exist *in vivo* [183,184,193–195] and the data available for this analysis do not seem to support this model. While viral load can increase very sharply on the emergence of resistance, it does not seem to rise to loads far higher than that observed before the start of ART. Meanwhile, CD4 concentration then tends to decline steadily rather than sharply as modelled here. Furthermore, oscillatory dynamics do not seem to be observed, although the resolution in the data may not be high enough to be certain that these do not exist.

I therefore alter the equation for viral load (4.15) to remove dependence on CD4 concentration. A mechanistic explanation for the independence of virus production from CD4 concentration is that the proportion of CD4 T cells that are infected is so low, uninfected cells are in excess and viral production is therefore not limited by the CD4 T cell concentration. The resulting model's output is a more realistic representation of the data. This gives the following equation for viral load dynamics,

$$\frac{\mathrm{d}}{\mathrm{d}t}V = \frac{k\,V_0}{V_0 + V_{50}}\,V(1 - E_{\mathrm{Drug}}) - k\,V\frac{V}{V + V_{50}}. \tag{4.18}$$

while the equation for CD4 concentration remains,

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \lambda - \mu\,X - \delta\,XV, \qquad \text{where} \quad \delta = \frac{\lambda - \mu\,X_0}{X_0\,V_0}, \tag{4.19}$$

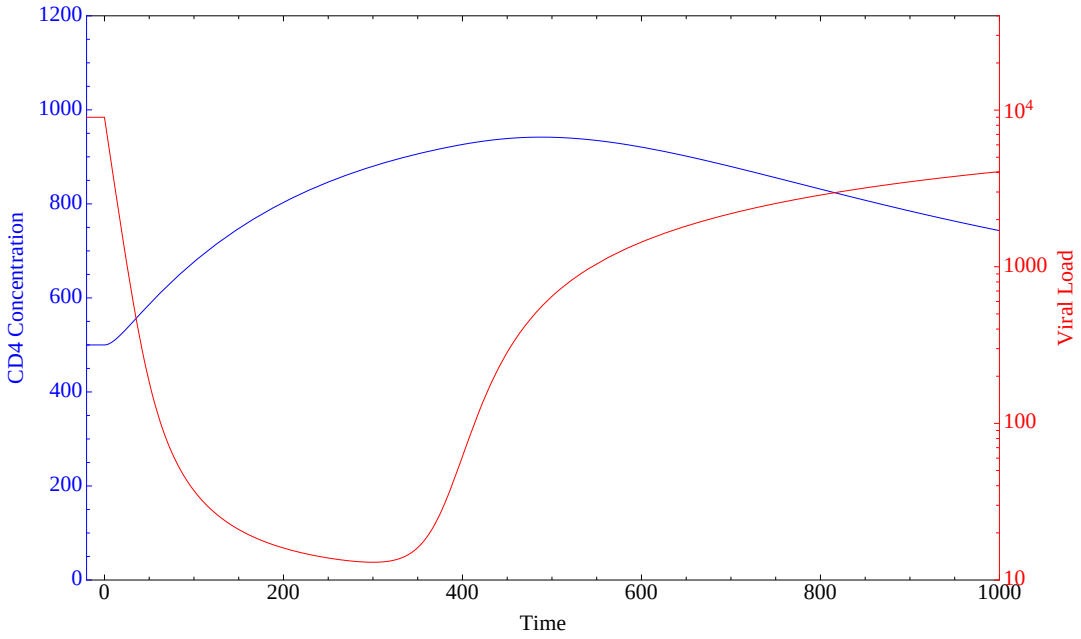

Figure 4.7: Effects of the development of resistance following removal of the dependence of viral load production of CD4 concentration. The model starts in steady state before the start of ART on day 0, after which viral load falls rapidly and CD4 concentration reconstitutes slowly. After a certain amount of time, resistance to ART develops and E_{Drug} decreases. Viral load therefore increases fast and CD4 declines. By removing the dependance of viral load production on T cell concentration, oscillations are removed and CD4 declines slowly. $X_0 = 500, \lambda = 5, \mu = 0.05, V_0 = 9000, k = 0.1, V_{50} = 100, E_{\text{Drug}} = 0.9, T_{\text{Drug}} = 300, \rho = 20$.

for which the dynamics are given in Figure 4.7. As can be seen, this produces more realistic dynamics for the emergence of HIV resistance to ART.

Incorporation of age effects from HSCT model

In order to fit the model to paediatric data spanning multiple years, it is necessary to account for the expected changes in the immune system with age. As mentioned in Section 4.1, previous methods have involved accounting for age by adjusting the data either to z-scores [171] or to the ratio of the measured to the expected [172]. In this chapter, I model the CD4 concentration without adjusting it by using the mechanistic model developed in Chapter 3 to account for age-related changes. Furthermore, by including the mechanistic elements for competition for resources it is possible to identify both proliferation and loss rather than just net

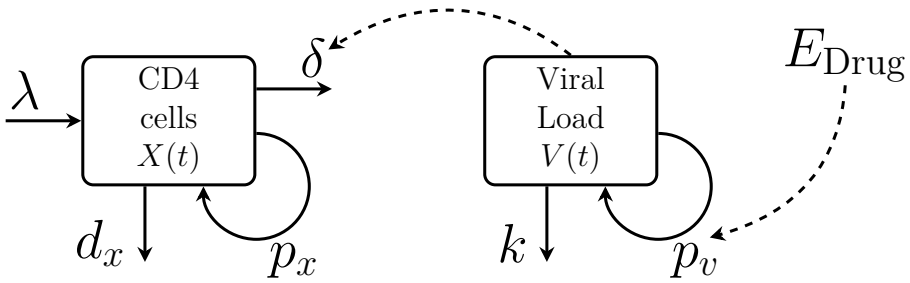

Figure 4.8: Schematic of the final model for CD4 concentration and HIV dynamics. CD4 cells are produced at rate λ, proliferate at rate p_x and die at rate d_x in the absence of virus. In the presence of HIV, CD4 loss rate increased by δV. Meanwhile, viral load is produced at rate p_v and declines at rate k. The ART affects the rate of production of virus p_v.

loss.

The equations for the CD4 T cell concentration dynamics therefore become,

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \lambda - \mathrm{d}_x\, X + p_x\, X - \delta\, V\, X\,, \tag{4.20}$$

where

$$\lambda = \lambda_0 \frac{y(\tau)\, V(\tau)\, \gamma}{\eta(c-\gamma)} \tag{4.21}$$

$$p_x = y(\tau) p_0 \mathrm{e}^{c_p\left(1-\frac{X(t)}{V(\tau)}\right)} \tag{4.22}$$

$$d_x = y(\tau) d_0 \mathrm{e}^{c_d\left(\frac{X(t)}{V(\tau)}-1\right)}\,, \tag{4.23}$$

with as before, $y(\tau)$ the proportion of CD4 cells expressing Ki67 with age (Figure 3.4); $V(\tau)$ the expected CD4 T cell concentration with age (Figure 1.2); $\eta = 0.52$ the duration of Ki67 expression; and $c = 0.25$ and $\gamma = 0.08$ constants related to the TREC content of CD4 cells as they leave the thymus. The parameter λ_0 is included such that the proportion of the expected thymic output for age is estimated but the shape of the changes with age are maintained. d_0 and p_0 give the rates of loss and proliferation when the CD4 concentration is that expected of a healthy child of the same age, while c_d and c_p give the strength of the effects of competition for resources on loss and proliferation respectively.

The viral dynamics remain unchanged as,

$$\frac{\mathrm{d}}{\mathrm{d}t}V = \frac{k\,V_0}{V_0 + V_{50}}\,V(1 - E_{\mathrm{Drug}}) - k\,V\frac{V}{V + V_{50}}. \tag{4.24}$$

leaving ten parameters to be estimated, six for CD4 concentration: X_0 the initial concentration of T cells, λ_0 proportion of the expected thymic output with age, p_0 and d_0 the proliferation and loss rates respectively when the concentration is that of a healthy child, c_p and c_d the strength of the effects of competition for resources on proliferation and loss respectively; and four for viral load: V_0 the initial viral load, k the rate of decline in viral load giving the loss rate of infected CD4 T cells, V_{50} the viral load at which the decline in viral loss is half, and E_{Drug} the strength of the drug effect. A schematic of the final model is shown in Figure 4.8.

4.3.2 Model fitting

Nonlinear mixed-effects (NLME) modelling, as outlined in Section 2.3.2 on Page 48, was used to fit the model to the two datasets. The model-fitting was carried out using NONMEM 7.3 [85], using the importance sampling expectation-maximisation algorithm and the ADVAN13 (general nonlinear kinetics) subroutine [123]. The M3 method described in Section 2.3.4 was used to cope with viral load data below the limit of quantification.

4.4 Results

4.4.1 Patients with full viral suppression

The first aim was to see whether it was possible to fit a model for CD4 reconstitution to data that have not been pre-adjusted for age through building the effects of age directly into the model as described in Chapter 3. In order to sim-

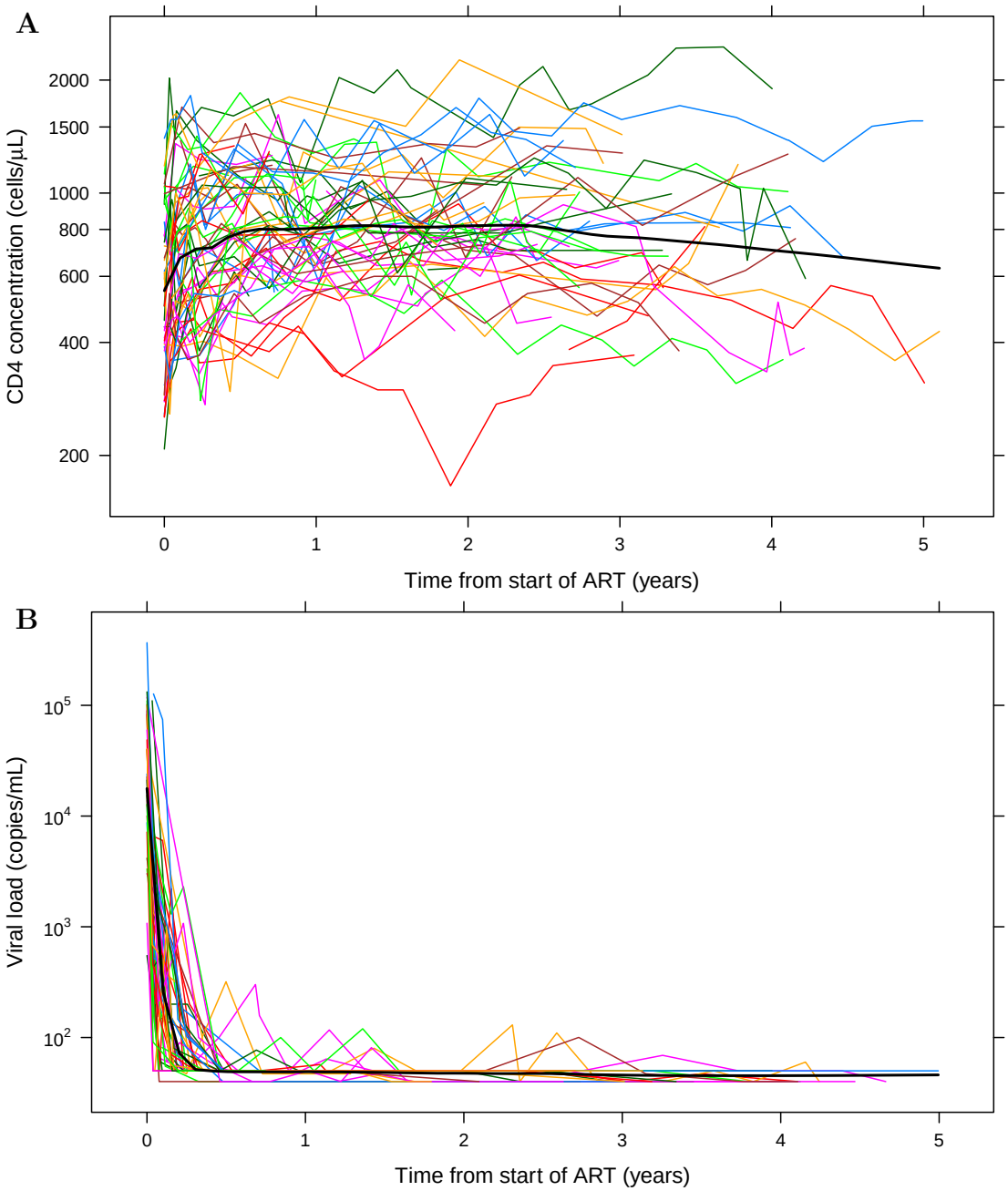

Figure 4.9: PENTA data for children restarting ART after PTI, (**A**) CD4 T cell reconstitution and (**B**) viral load excluding patients without complete viral suppression. Coloured lines give the responses of individual patients to ART. The black line gives a local regression curve.

plify model-fitting in this initial analysis, viral load trajectories where the viral load was not fully suppressed were excluded. This was defined as patients with a measurement of viral load greater than 1000 copies/mL more than 100 days after the start of ART. These excluded patients were a combination of patients

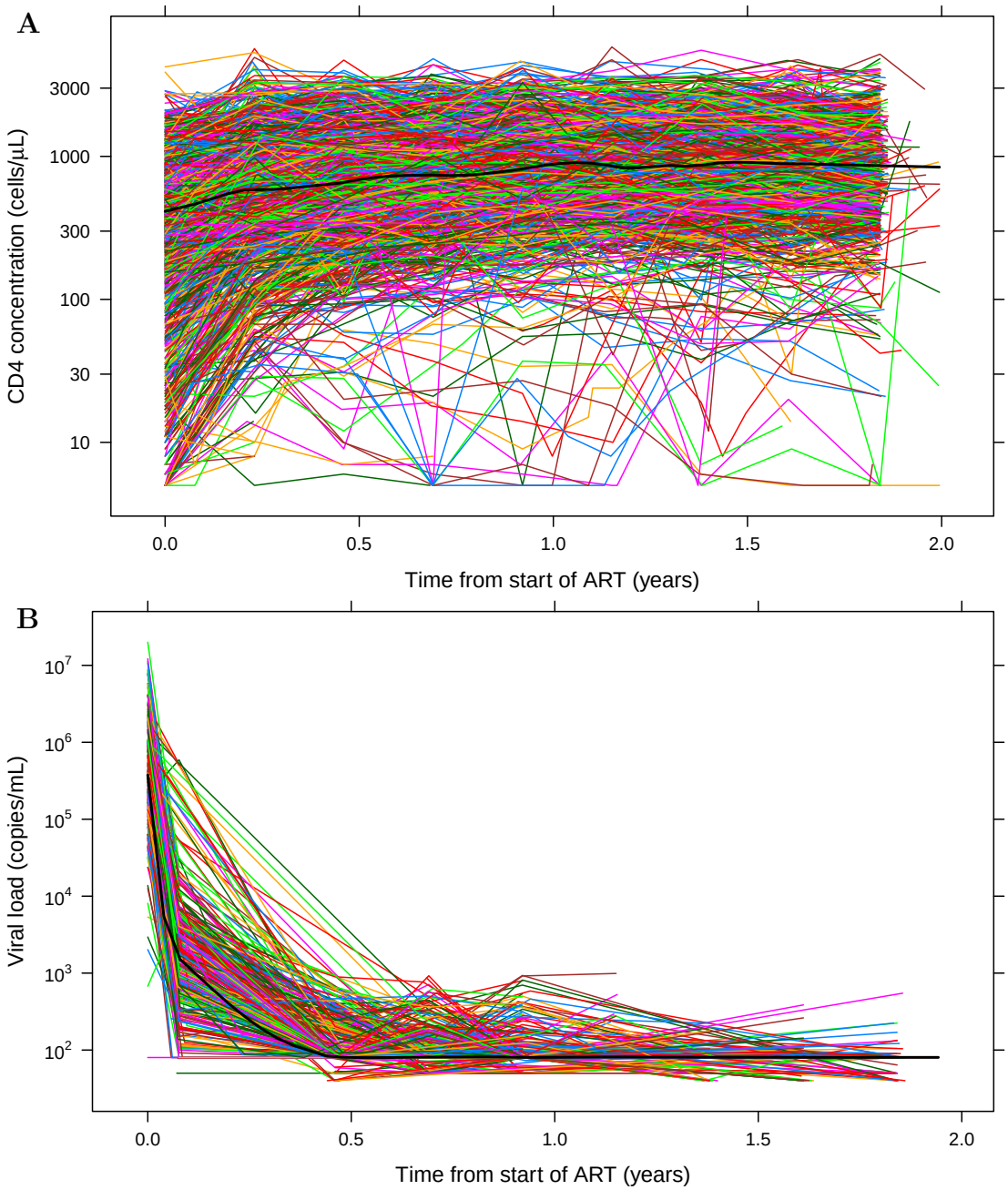

Figure 4.10: ARROW data cut to first two years post ART commencement for (**A**) CD4 T cell reconstitution and (**B**) viral load excluding patients without complete viral suppression. Coloured lines give the responses of individual patients to ART. The black line gives a local regression curve.

in whom the viral load did not decline, or it declined but only very slowly, or the viral load rebounded at some later date due either to non-adherence or the development of resistance to ART.

The resultant PENTA data after the exclusion of viral load trajectories without

Table 4.3: Parameter estimates from fitting the model for children starting ART to PENTA and ARROW datasets.

		PENTA		ARROW	
Parameter		Estimate	Ω	Estimate	Ω
λ_0	Proportion theoretical thymic output [33] (cells/day)	0.178	0.267	0.104	2.48
d_0	Proportion expected loss (/day)	0.545	2.81	0.462	2.17
p_0	Proportion expected proliferation (/day)	0.166	5.68	0.176	0.879
X_0	Initial concentration of T cells (cells/μL)	559	0.168	288	1.66
c_d	Strength of competition loss	1.21	1.09	1.18	0.631
c_p	Strength of competition proliferation	2.57	1.05	1.02	0.420
V_0	Initial viral load (copies/mL)	22,200	0.937	337,000	2.15
k	Rate of loss for viral load (/day)	0.299	0.585	0.259	0.948
V_{50}	Viral load at decrease of cell loss (copies/mL)	277	1.41	751	1.39
E_{Drug}	Effect of drug	0.989	1.91	0.976	1.28
σ_x	Variance of the residual error CD4 concentration	0.0420	—	0.153	—
σ_v	Variance of the residual error viral load	1.48	—	0.661	—

Ωs are the variances of random effects for that parameter.

full suppression are shown in Figure 4.9. In this dataset there are 721 measurements of CD4 concentration and 525 measurements of HIV viral load, of which 388 are below the limit of quantification (BLQ).

Because the ARROW data are so rich, it was possible to cut the data to just the first two years post ART commencement, which is the time period in which there are the greatest changes in viral load and CD4 concentration. In addition, there were 11 measurements of CD4 concentration that were out of the biologically feasible range ($>$ 6000 cells/μL) and as such are likely to be reporting errors; these were therefore removed. The resultant ARROW data used for model-fitting are shown in Figure 4.10. In these data there are 10,479 measurements of CD4 concentration and 1904 measurements of HIV viral load, of which 1118 are BLQ.

Parameter estimates from fitting the model to the PENTA and ARROW datasets are given in Table 4.3. As can be seen, parameter estimates are mostly consistent across the two datasets. There are however some differences, most notably between the estimates for initial CD4 concentration and viral load. Initial

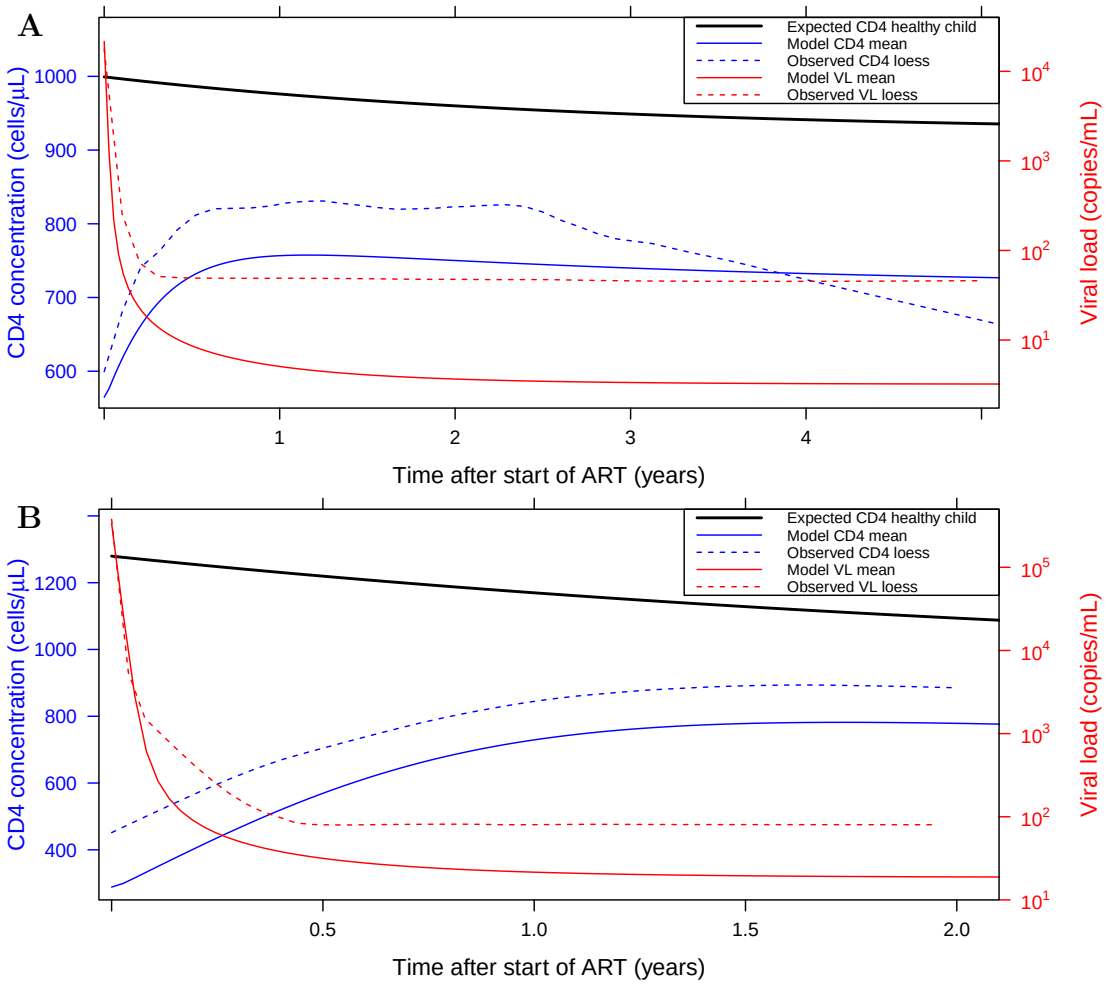

Figure 4.11: Comparison of model output to the expected CD4 concentration of a healthy child and to a local regression curve (LOESS) of the observed data for PENTA (**A**) and ARROW (**B**).

CD4 concentrations are estimated to be much lower for the ARROW dataset and the corresponding viral load is higher. The parameter defining the viral load at which viral load decline is half also has a higher estimate for the ARROW data over the PENTA data. The differences in initial viral load and CD4 concentration are supported by the observed data.

The estimates for the thymic output were λ_0 of 18% and 10% of the previously predicted thymic output [33] for PENTA and ARROW respectively. With this model it was possible to estimate the strength of the competition for resources unlike in the model in Chapter 3 where the effects of these parameters were inter-related with the effects of the parameters for the delay to thymic output recovery

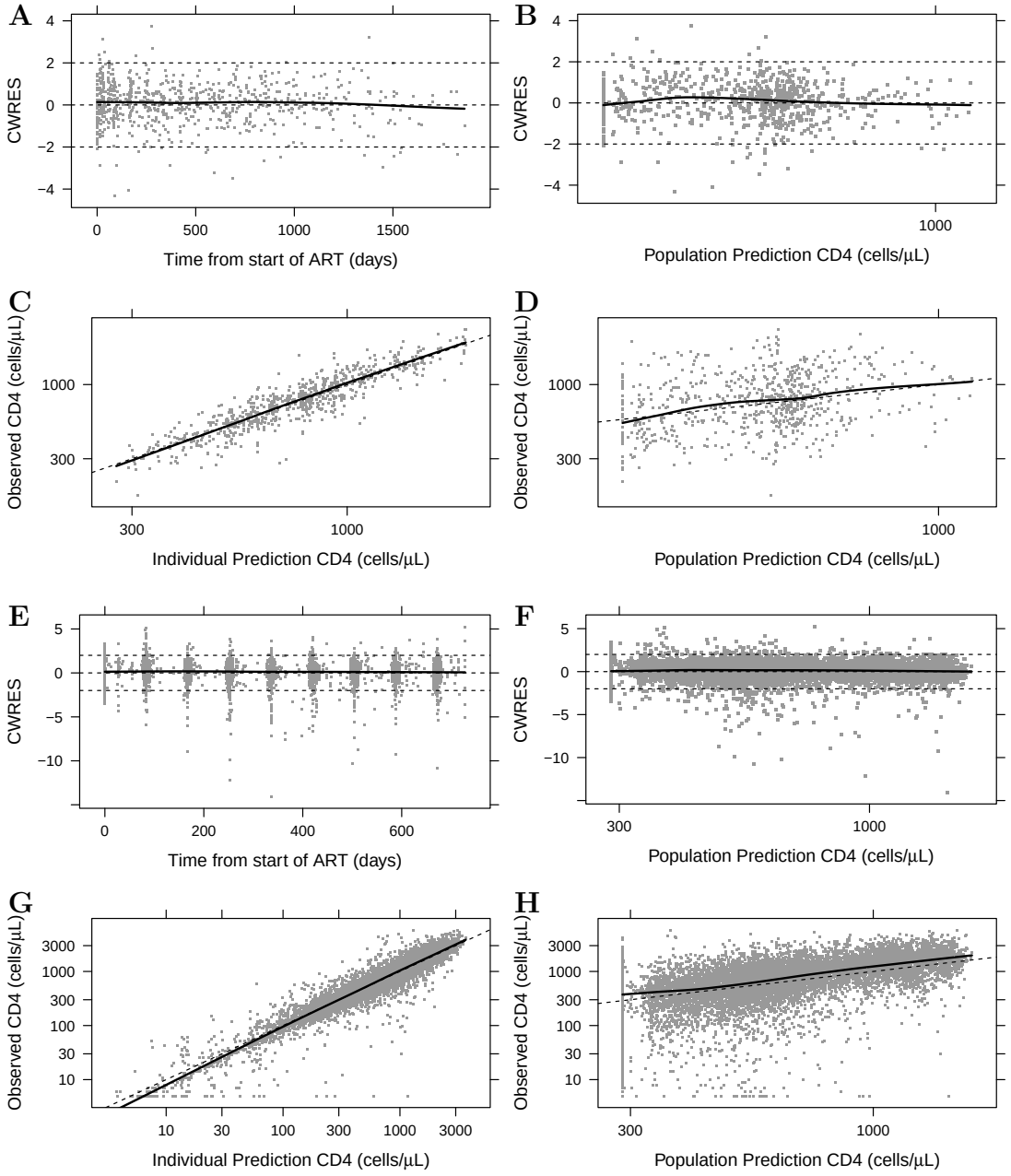

Figure 4.12: Diagnostic plots for the PENTA and ARROW datasets. **A** and **B** give the CWRES against time and population prediction respectively for PENTA, **C** and **D** give the observed data against the individual-level and population-level predictions respectively for PENTA. **E** and **F** give the CWRES against time and population prediction respectively for ARROW, **G** and **H** give the observed data against the individual-level and population-level predictions respectively for ARROW.

following HSCT. The parameter estimates found by the HIV dynamics model were close to the value at which they were fixed for paediatric HSCT.

Comparisons of the model output for a child of median age at start of ART

(9.6 years for PENTA and 6.0 years for ARROW) and a local regression curve of the observed data are shown in Figure 4.11. As can be seen, for viral loads, the model follows the observed data well at first, but as increasing amounts of the data are BLQ, the model diverges. For CD4 concentrations, the predicted CD4 concentration is below that of the local regression curve.

This could be caused by biases in the observed data in which there is more data for younger patients for whom the expected CD4 concentrations would be higher. The decrease in observed CD4 concentrations at later times in the PENTA data is likely to be due to a similar effect: only a few patients have data trajectories that continue that long after the start of ART.

Diagnostic plots for PENTA and ARROW are shown in Figure 4.12. Because of the high proportion of BLQ data in the viral load data, only the residuals for CD4 concentration are informative and are shown here. As can be seen from the plots of the residuals, for both PENTA and ARROW the residuals are roughly normally distributed with mean 0, and there are no apparent biases with time or population prediction demonstrating that they are independent. There are however a few very low residuals for the ARROW data. This could be due to further mis-recorded observations – mistaken observations of low CD4 concentration are hard to distinguish from genuine observations. The plots of observed against individual and population predictions also demonstrate that there are no obvious biases in the data.

The visual predictive checks (VPCs) for the PENTA data in Figures 4.13A and B demonstrate that data simulated from the model matches the observed data well. In particular, the model captures the dynamics of the reconstitution as well as its variability for CD4 reconstitution, and for the viral load it captures the expected proportions of measurements that are BLQ.

The VPCs for the ARROW data in Figure 4.13C and D also mostly match the observed data well. The median of the simulated data for CD4 concentration

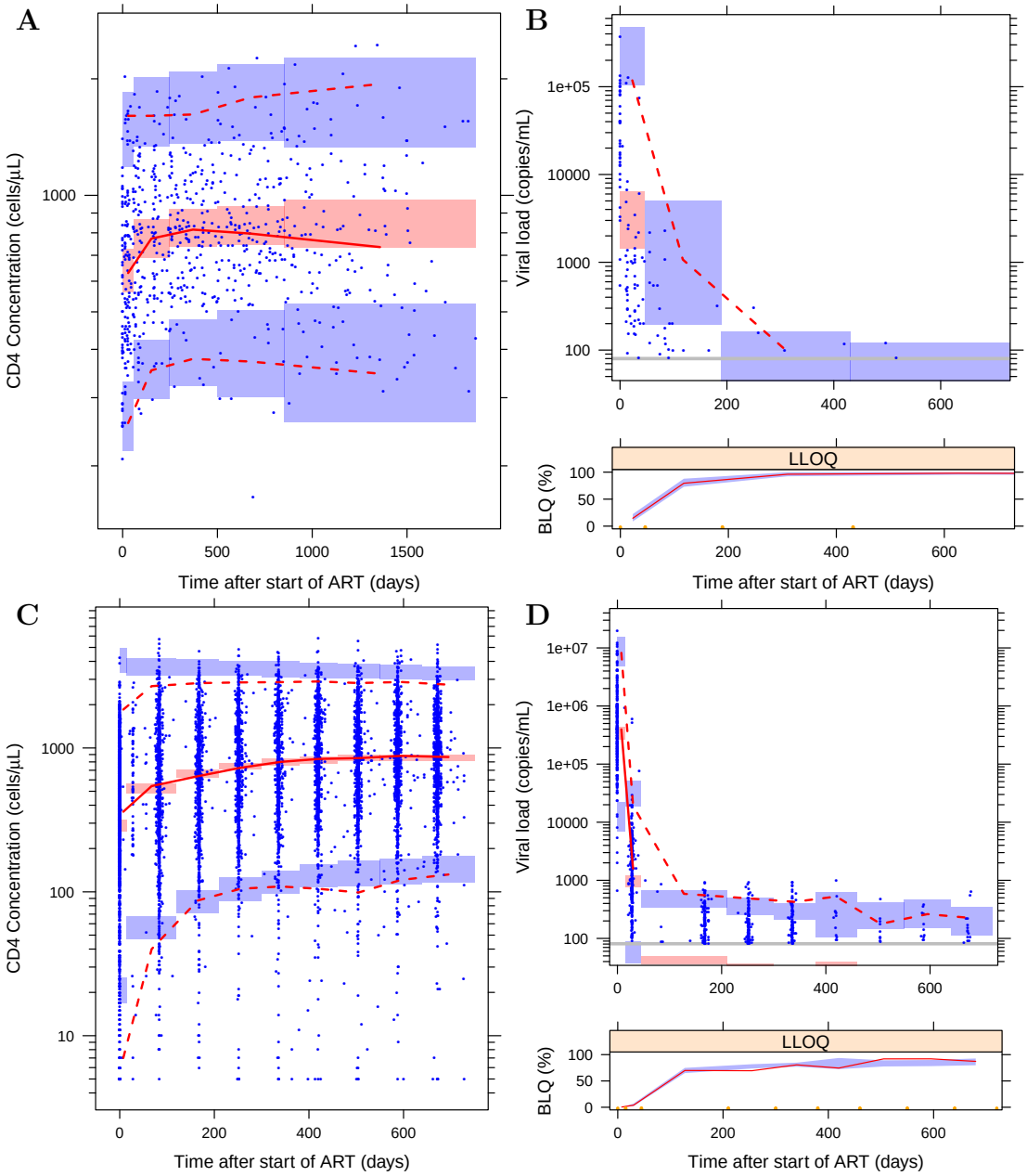

Figure 4.13: Visual predictive checks. **A** PENTA CD4 concentration; **B** PENTA viral load. **C** ARROW CD4 concentration; **D** ARROW viral load. The blue dots give the observed data, with the corresponding median, 2.5th and 97.5th percentiles given by the filled and dashed red dashed lines respectively. The 95% confidence intervals around the model-simulated median and percentiles are given by the red and blue shaded regions respectively. In the BLQ plot, the observed proportion of measurements that were BLQ is given by the red line, while the blue shaded area gives the 95% confidence interval of the model-simulated BLQ data.

matches the observed data, and while the long-term concentrations match, the model does not capture the variability in the initial concentrations correctly, overpredicting both the minimum and the maximum of the initial concentrations.

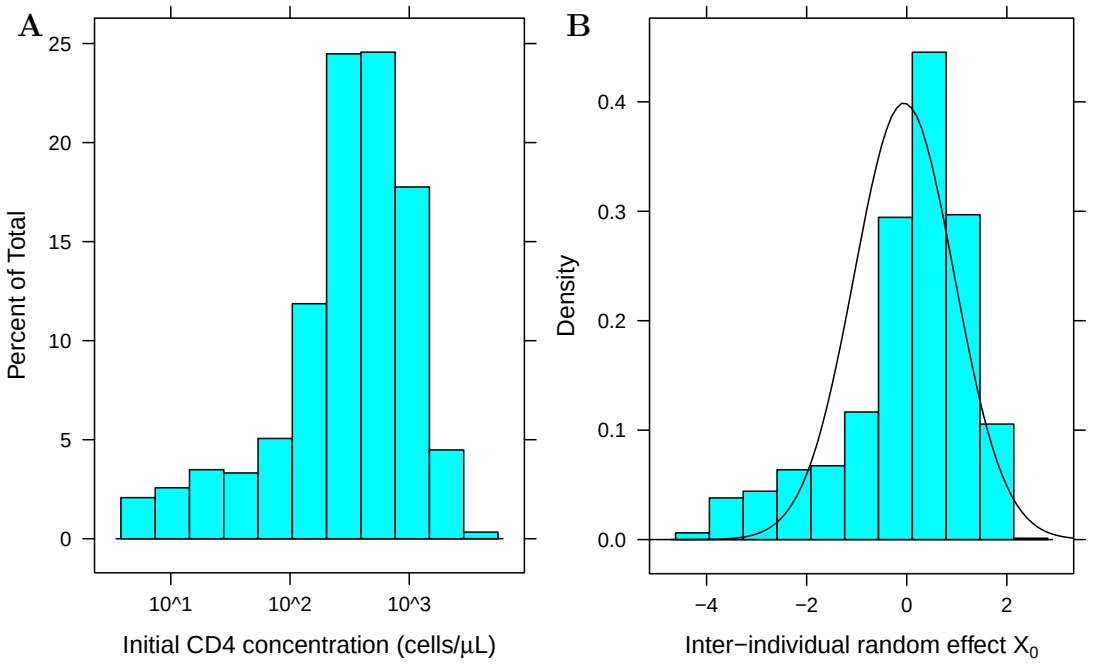

Figure 4.14: Histograms of (**A**) the initial CD4 concentration in the observed data on a log scale and (**B**) the model-output random effect of X_0, the initial CD4 concentration compared to a normal distribution, demonstrating that a lognormal assumption was inaccurate.

This could be due to the initial concentrations not being log-nomally distributed, as is assumed in the model (see Figure 4.14). However, the model-simulated data for viral load captures the median of the observed data and the variability as well as the proportion of measurements that were BLQ.

4.4.2 Patients with viral load rebound

Viral loads rebound either when a patient develops resistance to ART or due to non-adherence to the ART. The patient's regimen may then be changed or the patient may adhere again and viral load may fall as a result, although if the CD4 concentration is largely unaffected, then the regimen may not be altered.

Out of the ARROW dataset, 261 patients did not maintain full suppression and had a measurement of viral load exceeding 1000 copies/mL more than 100 days after the start of ART. Because the aim of this section is to model the development of resistance, only patients for whom there were more than four

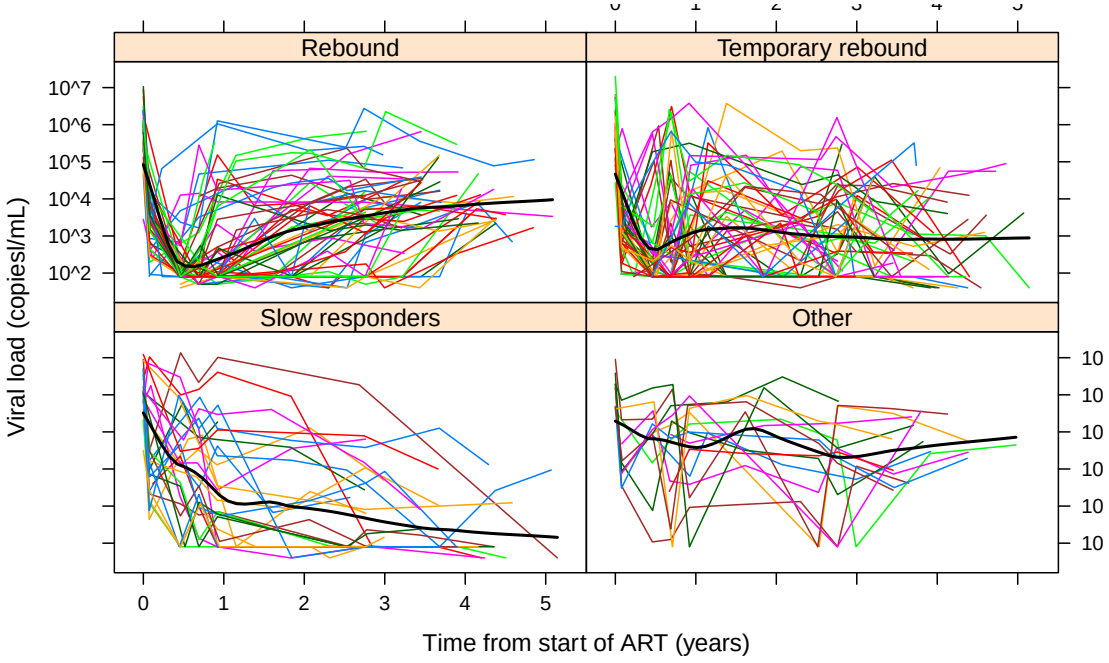

Figure 4.15: Viral load profiles of patients that do not have full suppression of viral load, split by visual inspection into categories of viral load suppression.

measurements of viral load in the dataset were considered in order to allow categorisation. This left 148 patients who were then categorised according to the shape of their viral load profile (see Figure 4.15). For 57 patients, the viral load after initially falling rebounded and stayed high; for 48 patients, the viral load after initially falling rebounded temporarily, with some patients having a second rebound as well; for 28 patients their viral load decreased only slowly under ART, in contrast to the fast suppression seen in the majority of patients. The remaining 15 patients had viral load profiles that could not be categorised.

In order to assess the ability of the model to reflect the rebound in viral load due to the development of resistance to ART, the 57 patients were modelled who had a viral load profile where the viral load initially fell but then rebounded and remained high for the rest of the time course. In order to capture the development of resistance rather than just the initial reconstitution, patients were modelled for the full time-course of data available, up to five years after the start of ART. In this subset of data, there were 1011 observations of CD4 concentration, and 413

measurements for viral load of which 114 were BLQ. The patients had a mean of 18 measurements of CD4 concentration and 7 measurements for viral load in this time-course. The children in this subset of the data had a median age at start of ART of three years old.

The model for the development of resistance in Section 4.3.1 was used,

$$E_{\text{Drug}} = E_{\max}\left(E_{\text{R}} + \frac{1 - E_{\text{R}}}{\exp[\frac{t-t_R}{\rho}] + 1}\right), \tag{4.25}$$

where $E_{\max}$ is the drug effect before the development of resistance, E_{R} is the residual drug effect after the development of resistance, t_R is the time at which the drug effect is 50% of the way between $E_{\max}$ and E_{Drug}, and ρ is the rate at which resistance develops. In order to fit the model, parameter estimates from fitting the model to the whole dataset for viral load and CD4 dynamics were used and fixed. The three parameters related to the development of resistance were then estimated along with their random effects using the importance sampling algorithm.

The parameter estimates of these three parameters and their random effect variances are given in Table 4.4. For those patients who did develop resistance, the mean time to the development of resistance was 13 months after the start of ART, and according to the model the development of resistance was on average fast, taking 1 month for the ART to go from 90% to 10% effective. When resistance developed, there was very little residual drug effect, with on average just a 4.5%

Table 4.4: Parameter estimates for the development of resistance in a subset of 57 ARROW patients.

		ARROW	
Parameter		Estimate	Ω
t_R	Time after ART start at development of resistance (days)	406	0.798
ρ	Rate of the development of resistance (days)	5.92	2.54
E_{R}	Residual drug effect after development of resistance	0.0447	0.764

Ωs are the variances of random effects for that parameter.

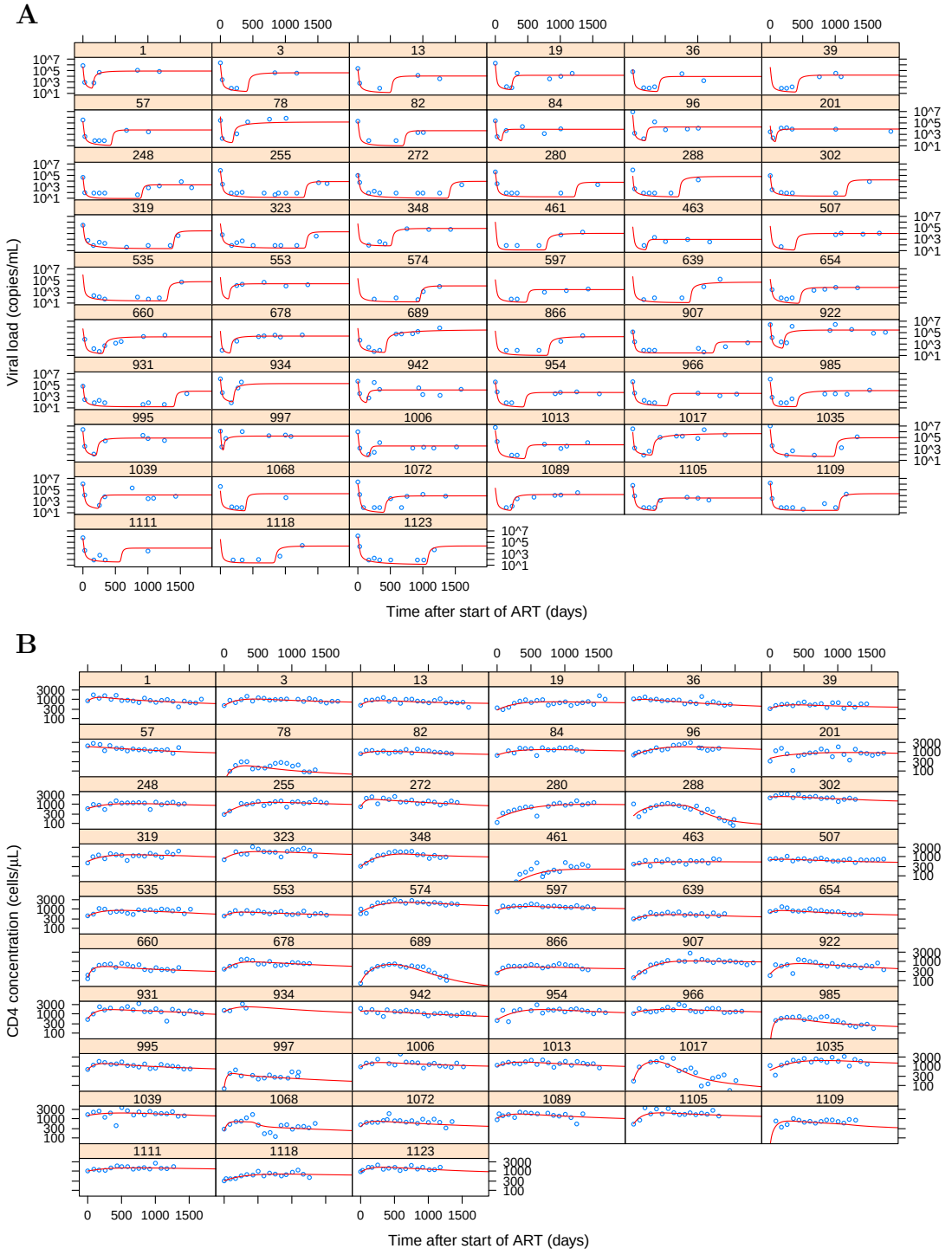

Figure 4.16: Comparison of observed data and model output for patients who develop resistance in the ARROW dataset. Red lines give the model output, while blue circles give the observed data. **A** gives the plots for viral load, **B** gives the corresponding plots for CD4 concentration.

reduction from the pre-ART rate of virus production expected.

Plots in Figure 4.16 give a comparison for each individual of their observed

data and the model output for both viral load and CD4 concentration. As can be seen, the model reflects the data well. For viral load, it manages to pick up the differences in the trajectories from individuals who develop resistance early to those that develop it late. Because of the relatively low resolution in the data, with measurements taken at large time gaps, it was not possible for many patients to determine the exact time at which resistance developed. For the corresponding CD4 concentrations, the model captures first the reconstitution in CD4 concentration and then the decrease in CD4 concentration as resistance to ART develops and viral load increases.

4.4.3 Covariate analysis

In order to find the factors that affect both viral load decline and the CD4 reconstitution, covariate analysis was performed using the stepwise covariate model-building (SCM) procedure. In the forward search for testing which covariates to include in the model a p-value of $p_1 < 0.01$ was used, while in the back-

Table 4.5: Covariates included in the model for PENTA and ARROW datasets.

	Parameter	Covariate	Effect size	p-value
PENTA	X_0	ART age	-0.805	$\ll 0.001$
	X_0	Zidovudine	0.195	0.002
	p_0	Nevirapine	3.00	$\ll 0.001$
	c_p	Lamivudine	0.411	$\ll 0.001$
	k	Efavirenz	1.04	$\ll 0.001$
	V_{50}	Zidovudine	-0.836	$\ll 0.001$
ARROW	X_0	ART age	-1.10	$\ll 0.001$
	X_0	Treatment centre:		$\ll 0.001$
		Harare, Zimbabwe	$0^\dagger$	
		JCRC, Kampala, Uganda	0.221	
		PIDC, Mulago,Uganda	0.0937	
		Entebbe, Uganda	0.599	
	λ_0	Died by end of trial	-0.983	$\ll 0.001$

For categorical covariates, the parameter for patients who had the respective covariate is multiplied by (1 + Effect size). So an effect below 0 decreases the parameter and an effect greater than 0 increases the parameter for that covariate. For continuous covariates, the parameter is multiplied by $(1 + \text{Effect size}) \times \frac{\text{ART age - Median ART age}}{\text{Maximum ART age}}$. The null hypothesis is then that the effect size is zero. † Typical individual.

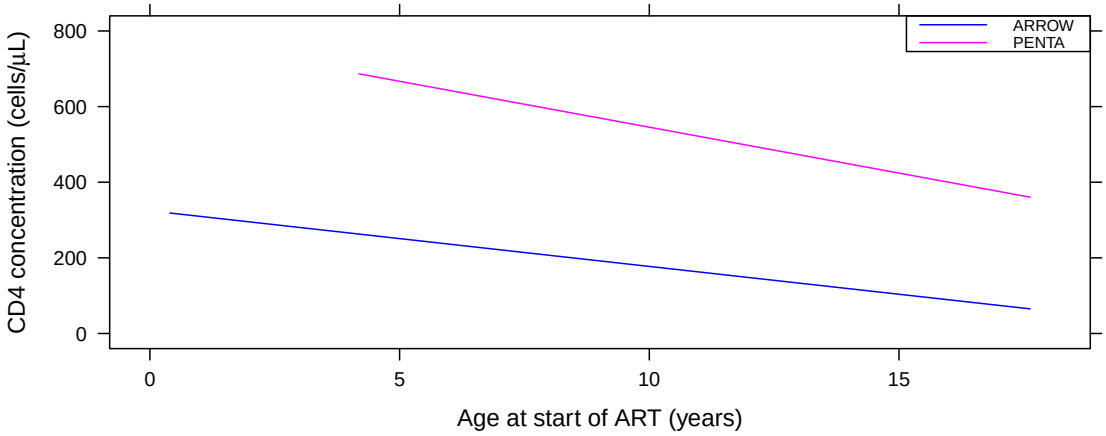

Figure 4.17: Age effects on the initial concentration of CD4 T cells X_0, plotted for the range of ages for which there were data in each dataset. The differences between the ARROW and PENTA data are due to differences in the parameter estimates for X_0 between the datasets.

wards search for testing which of these included covariates to retain, a p-value of $p_2 < 0.005$ was used. The covariates tested are listed in Table 4.1 for the PENTA data and in Table 4.2 for the ARROW data and these covariates were tested on each of the parameters that were estimated in the model. The covariates that were included are shown in Table 4.5.

Age effects on X_0

For both models, the initial concentration of T cells was found to be negatively correlated with age (see Figure 4.17). This is not surprising, given that CD4 concentrations are expected to fall across childhood. In both the PENTA and the ARROW data, age did not correlate with any other parameter.

PENTA data

For the PENTA data, the four factors that were associated with changes to the dynamics were ART drugs (see Figure 4.18 and Table 4.5). Two were ART drugs given before the beginning of the PTI: zidovudine (ZDV) and nevirapine (NVP). The others were those given after the PTI, during viral load decline and CD4 reconstitution: efavirenz (EFZ) and lamivudine (3TC). Zidovidine is

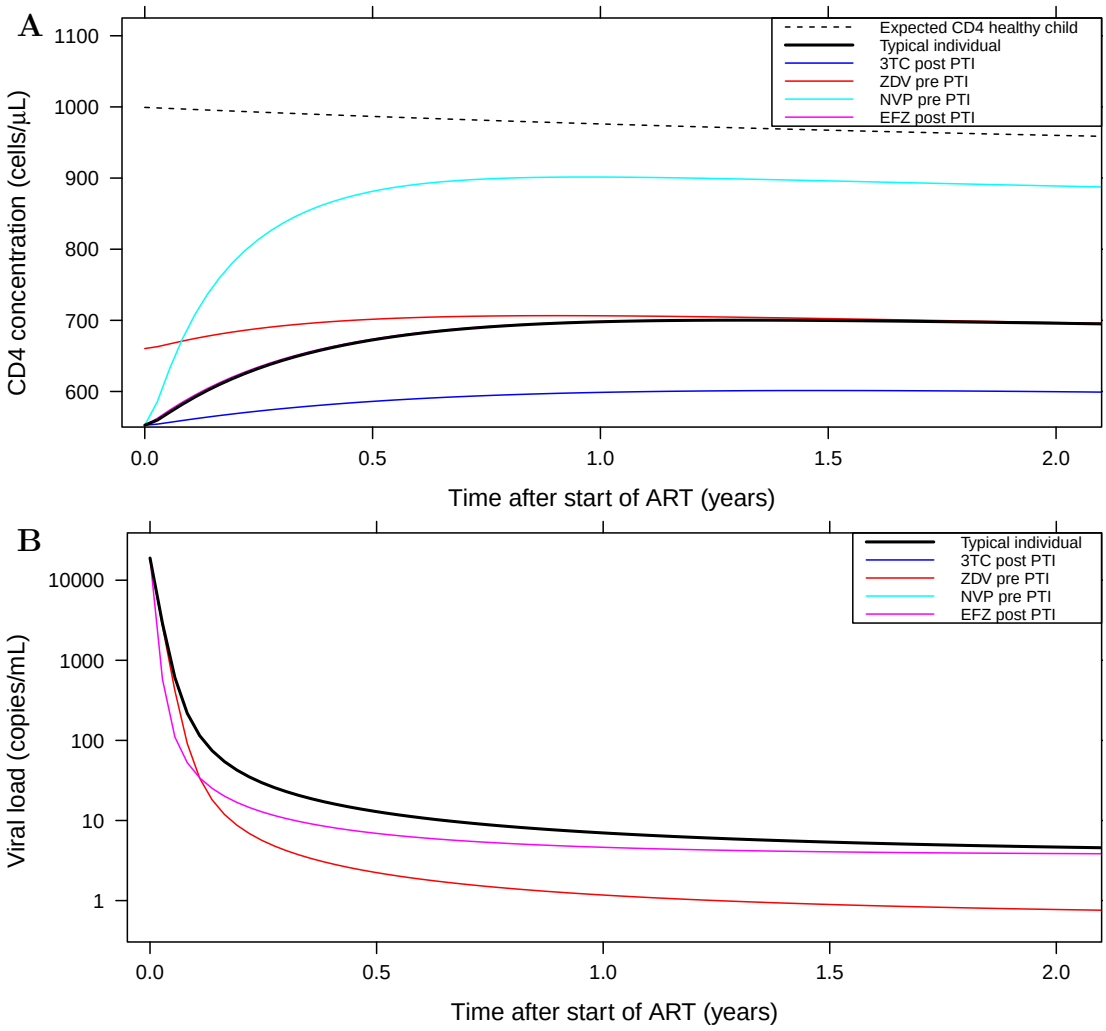

Figure 4.18: Effects of the factors that affect (**A**) CD4 concentration and (**B**) viral load on restarting ART in the PENTA data. *Abbreviations*: 3TC: lamivudine; ZDV: zidovudine; NVP: nevirapine; EFZ: efavirenz.

an NRTI, and was associated with increased CD4 at the end of the PTI and a small decrease in viral load after the restart of ART. Nevirapine is an NNRTI and was associated with better CD4 reconstitution following the restart of ART. Efavirenz is an NNRTI and was asoicatied with faster viral load decline in the weeks following the start of ART. Lamivudine is an NRTI and was associated with poorer CD4 reconstitution following the restart of ART.

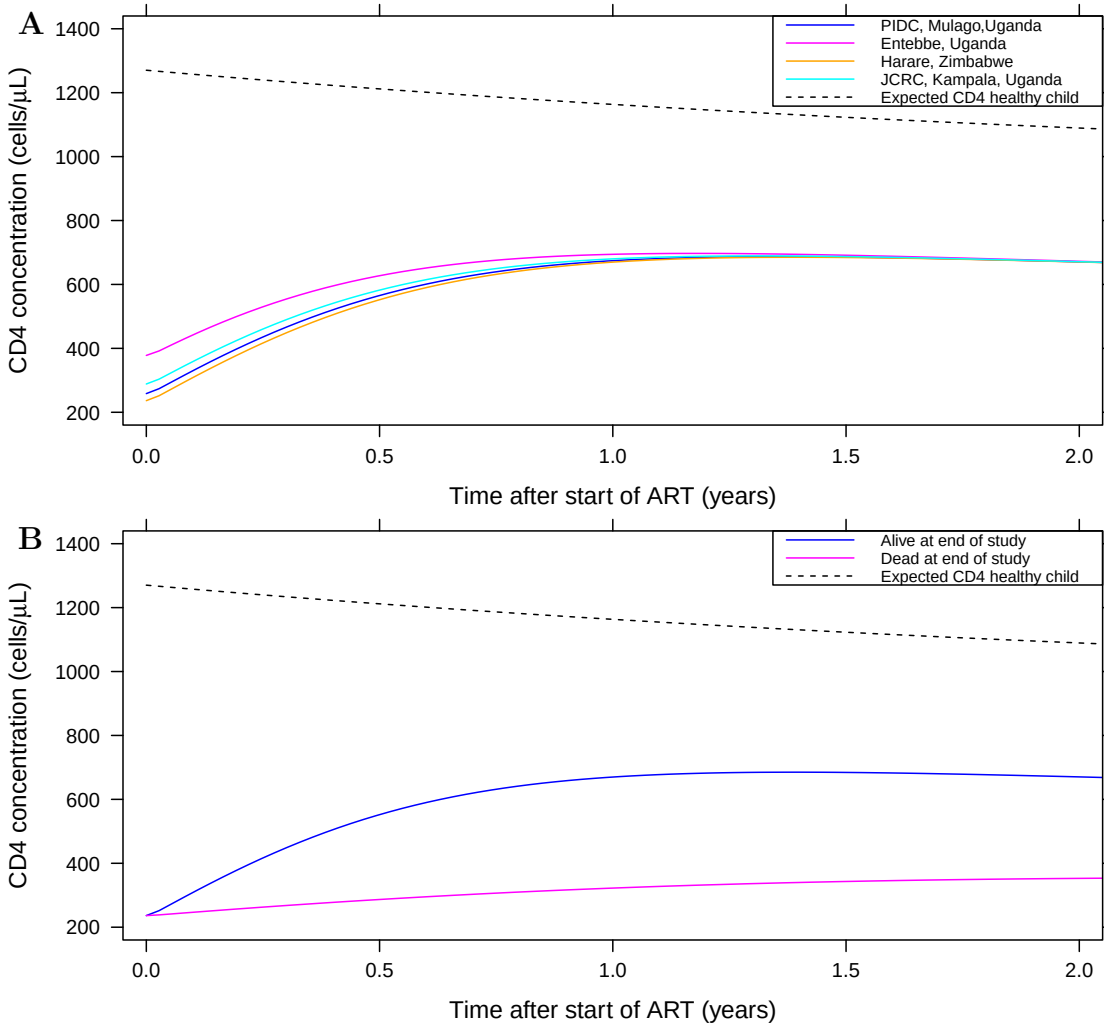

Figure 4.19: Effets of the factors that affect CD4 concentration on starting ART in the ARROW data. (**A**) The centre at which they were enrolled and (**B**) whether they were alive at the end of the trial.

ARROW data

For the ARROW data, the two factors that affected the dynamics were the centre at which the child was enrolled and whether or not they were alive at the end of the study (see Figure 4.19 and Table 4.5). The model found an association between the initial concentration of CD4 T cells and the centre at which they were enrolled, with patients at the centre in Entebbe, Uganda having a higher concentration than those at the other three centres. The patients that had died by the end of the trial (n=56) were found by the model to have on average a much reduced thymic output, resulting in a very low long-term CD4 concentration, with

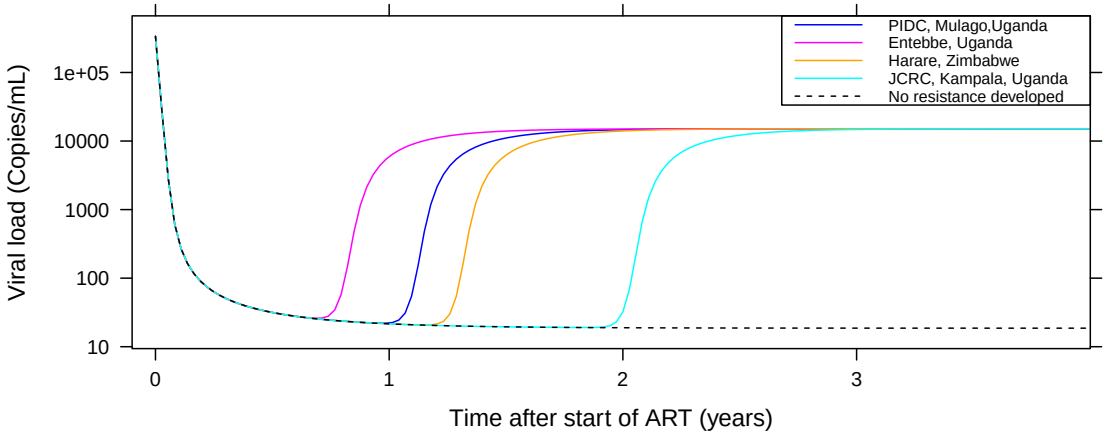

Figure 4.20: Effects of the treatment centre on time at which viral load rebounds in the ARROW data due to either the development of resistance to ART or non-adherence.

the concentration barely increasing from that when they started ART.

Development of resistance in ARROW data

Looking at the development of resistance in the ARROW data, the centre at which the child was treated and seen was found to affect the time at which the viral load rebounded (see Figure 4.20). As can be seen, the viral load declines and increases at the same rate for each centre, but it increases at a different time for each centre.

4.5 Discussion

In this chapter, I have developed a new mechanistic model for the dynamics of CD4 reconstitution and viral load decline in HIV-infected children starting ART. The model combines a simplified version of a previous model for HIV dynamics with the age and competition effects from the model in Chapter 3 for long-term CD4 reconstitution following paediatric HSCT. This allows the model to be fitted to datasets of paediatric patients without having to adjust the data to account for age.

The model was fitted to two datasets, from two clinical trials, PENTA and

ARROW. While most of the parameter estimates obtained from fitting the models were similar between the datasets, some parameters differed substantially. In particular, the viral loads were found to be higher and CD4 concentrations lower at the start of ART in the ARROW dataset. The differences between the parameter estimates for the datasets are sensible both biologically and due to the different structures of the two trials. Biologically, patients with higher viral loads would be expected to have lower CD4 concentrations. Structurally PENTA is a planned treatment interruption trial with children closely monitored for the re-start of ART whereas the children in ARROW are receiving treatment for the first time. Furthermore, the children in ARROW are from less developed countries and as such potentially have not had access to the same standard of care as the patients in PENTA.

The inverse of the rate of loss of CD4 concentration can be used to find the cell lifespan. The parameter estimates implied a CD4 T cell lifetime of 196 days for a child of the median age at ART for the ARROW dataset (6 years old), with a range from 120 days for a 1 year old to 620 days for an 18 year old. Similarly to the model of CD4 reconstitution following HSCT, these estimates agree well with the experimental evidence from the deuterium and BrdU labelling studies that have taken into account kinetic heterogeneity in the T cell population, finding lifetime estimates between 222 and 611 days (range 167 to 1245) [21,22,126,127].

Because the viral load is in quasi-steady state with infected CD4 T cells, the rate of decline of viral load gives the rate of decline of infected CD4 T cells. The estimates for the rate of loss of viral load k therefore implies an average lifetime of infected CD4 T cells of 3.2 days and 3.9 days for PENTA and ARROW respectively. Previous studies have estimated the lifetime of infected CD4 T cells at around 3 days [184,185,196,197], in agreement with the parameter estimates from this model.

The inclusion of a viral-load-dependant sigmoidal function on the rate of loss

of viral load allows the decline in viral load to be multi-phasic. It also means that viral load plateaus at a residual low level rather than continuing asymptotically towards zero. This means that resistance can develop more quickly than would otherwise be possible from the model. The residual low level is below the limit of quantification, so it is not possible to determine experimentally where this level might actually be.

The model found that the CD4 concentration for age does not on average return to the level of a healthy child, with the long-term concentration on average 75% of that expected. This is in agreement with other analyses of both the datasets [172,201] and other datasets [171], and could have major implications for the health of the children's immune systems.

The model has been applied to 57 patients from the ARROW dataset that developed resistance to ART. These patients were chosen as having had an initial decrease in viral load that then increased and remained high. The model was able to match the data well for the combination of viral load and CD4 concentration, capturing the time at which resistance developed. The accuracy of this time estimate however is dependent on the resolution in the data. From the observed data, resistance appears to develop rapidly with very few observations made where resistance is partial. Resistance can therefore usually only be estimated to have developed between two time points and as such the accuracy of the time at which resistance develops depends on the frequency of the measurements of viral load. The model for resistance allows for further analysis of the factors that affect the timing of the development of resistance.

Covariate analysis was performed to find what factors significantly alter either the viral load or the CD4 reconstitution profiles following the start of ART. For both datasets, age at the start of ART affected the initial concentration of CD4 T cells. This is to be expected partly because the CD4 concentration is expected to decline with age in healthy children and partly because older children will have

been infected with HIV for longer. In this analysis, a linear correlation was tested between age at ART and CD4 concentration. While an exponential decay could have better extrapolation properties, it was not tested during covariate analysis in order to simplify the covariate analysis. Age was not found to be a significant covariate with any other parameter in the model, which implies that the model has managed to account for age-related changes of CD4 concentration. Analysis of this ARROW data [172] and of a different cohort of children [171] using CD4 z-scores to account for age related differences found that age affected not only the initial CD4 concentration in agreement with this analysis but also the long-term concentration, an effect we did not observe here.

For the PENTA data, significant differences in CD4 reconstitution and viral load were found with four of the ART drugs that were given: two NRTIs, zidovudine and lamivudine, and two NNRTIs, nevirapine and efavirenz. Patients that had received zidovudine before the PTI appeared to fare much better over the PTI, with a higher CD4 concentration at the re-start of ART. Their viral load also fell to a lower concentration in the ensuing weeks. Patients that received nevirapine pre-PTI also fared better after the PTI, with increased CD4 concentrations. In contrast, those patients that received lamivudine post PTI had decreased CD4 reconstitution in the weeks that followed the start of ART. Efavirenz was associated with an increased rate of loss in viral load, which resulted in a faster viral load decline following the start of ART. There were however only 56 trajectories for patients following PTI in this analysis, so while this demonstrates that the model can be used to find what affects reconstitution, these results should be taken with caution and it would require a larger dataset to have full confidence in them.

For the ARROW data, significant differences were found between the initial CD4 concentrations of patients that were enrolled at different centres. There was no difference found between the viral loads of these patients. In the covariate

analysis presented here, an association was also found between decreased thymic output, and a resultant significant decrease in long-term CD4 concentration in the patients that died before the end of the study. This effect is supported by the observed data (mean CD4 concentration at 2 yrs post start of ART for patients that survive to end of trial is 900 cells/μL in comparison to 200 cells/μL for those that do not), and agrees with the previous analysis of this data, which found that patients who had a non-asymptotic (i.e. flat) profile following the start of ART were more likely to have worse disease progression than those who displayed an asymptotic recovery [172].

Analysis of the 57 patients that had some sort of rebound in their viral load following an initial increase found that the centre at which the patient was treated appeared to significantly affect the time at which this rebound happened. The rebound could happen for three major reasons: the development of resistance to the drug combination that the patient is receiving, non-adherence to the drug protocol that the patient has been set, or a combination of the two with non-adherence raising the risk of the development of resistance. Given that the patients in the centres received similar combinations of drugs and that the drug combinations do not seem to have affected outcome, it seems that the likeliest cause of these differences is differences in adherence levels of the patients. It would be interesting to look into this further to see whether the centres have different strategies to promote adherence.

In this chapter, I have successfully simplified and adapted a previous model for the dynamics of HIV and CD4 concentration by removing the infected CD4 cells compartment, allowing for the multi-phasic nature of viral load decline, including the effects of age-related immune development, and including the effects of competition for resources. This model was successfully applied to two different datasets, finding parameter values that were consistent across the datasets and with the previous applications of similar models. The model was then further

extended to model the viral load rebound due to either non-adherence or the development of resistance. The model was used to perform covariate analysis to find which factors affect the viral load decline or the CD4 reconstitution in both datasets or the development of resistance in the ARROW dataset. While more work is required to ascertain the nature of these covariate effects, it is clear that the model can be used to interrogate a dataset and find useful information about what affects the recovery of children following the start of ART.

Chapter 5

A global competition model for T cell homeostasis

5.1 Introduction

In this chapter, I present a stochastic global clonal competition model for T cell homeostasis, on which I have been collaborating with Grant Lythe, University of Leeds.

T cell receptors (TCRs) are antigen specific, and it is this specificity that allows T cells to fight pathogens effectively. Each T cell will express one type of TCR, with about 30,000 TCRs per cell [203]. As discussed in Section 1.5, T cell receptors have a vast potential diversity of forms. Gene rearrangement in the thymus has the potential to produce more than 10^{15} different TCRs [204], although it is reckoned that the human T cell receptor repertoire actually has fewer than 10^8 different TCR clones in the body at any one time [205]. Because of specificity, this high diversity is required to enable T cells to recognise as many different pathogens as possible. TCRs, however, recognise peptides up to around 14 amino acids long when presented on MHC. As such, given there are 20 amino acids involved in protein formation, there are more than 10^{16} such peptides that

could be presented to T cells by antigen presenting cells (APCs) [204,205]. Thus in order to be able to recognise all peptides, TCRs must be cross-reactive and able to recognise multiple peptide-MHC combinations [206,207].

T cells with the same TCR can be grouped together into clonotypes. Whereas all T cells interactions with cytokines are general and non-specific, interactions with different sp-MHC combinations are specific to particular TCR clonotypes according to some interaction matrix. T cells compete with each other for these interactions with resources to survive and proliferate [29] and the resulting inter-clonotype competition drives the observed diversity of the T cell repertoire [27,28]. While a more complex model has been developed that considers competition for resources in the presence of a full sp-MHC interaction matrix [208], in this chapter we consider competition for non-specific resources such as cytokines.

This "Global" stochastic competition model can be used to study clonotype dynamics and the diversity and survival of TCR clonotypes. With the model in this form, all T cells compete equally for the same resources, which has two advantages. Firstly, it is computationally much less expensive, which means numerical results can be simulated for much greater time spans and with many more clonotypes, thus modelling a system much closer in scale to a human repertoire and offering greater insights. Secondly because it is simpler, it is mathematically tractable, which means analytical results can be found. This in turn allows the results to be scaled to a full-sized system. Some results from this work were also applicable to the more complex model [208].

5.2 The model

All T cells compete globally for a central stimulus which represents the total pool of resources. It is assumed that this pool of resources is of finite size and that the amount of resource does not change either with time or with the T cell

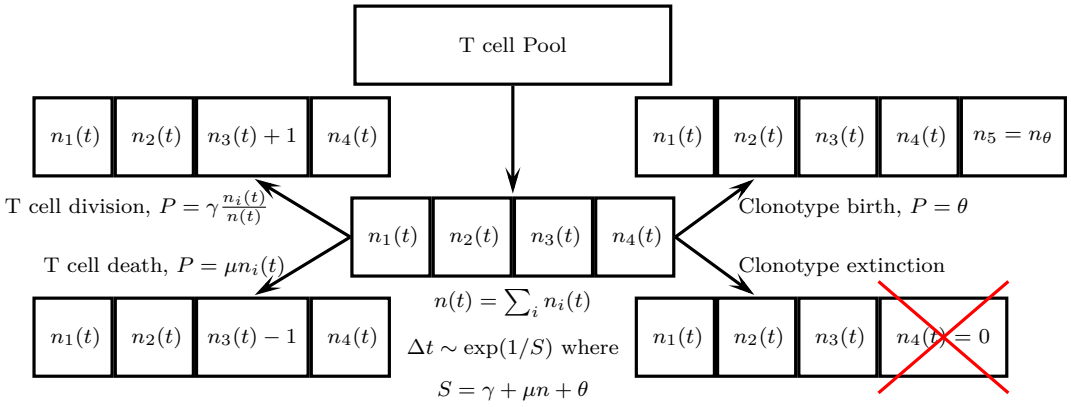

Figure 5.1: Schematic of the model for global competition between TCR clonotypes. The T cell pool with $n(t)$ T cells is split into N clonotypes with clonotype i having n_i T cells. A cell can then die or compete for resource to proliferate so that the number of cells in a clonotype can increase by one with a probability $\gamma \frac{n_i}{n}$ or decrease by one with probability μn_i. New clonotypes with n_θ cells per clonotype are be produced at rate θ. If the number of cells in a clonotype reaches zero, that clonoype is extinct. One such event happens per time step, and time between steps in incremented by Δt, where Δt is exponentially distributed according to the rates of the dynamics.

population. T cells then compete equally for this constant stimulus γ in order to divide. The death rate is equal for each T cell in the population and is given by μ, the probability per unit time of dying. For the whole population of cells, the birth rate is then γ and the death rate is $\mu n(t)$ where $n(t)$ is the total number of T cells with time t. This can be written as,

$$\lim_{\Delta t \to 0} \Delta t^{-1} \mathcal{P}[n(t + \Delta t) - n(t) = 1] = \gamma \qquad \text{and,} \tag{5.1}$$

$$\lim_{\Delta t \to 0} \Delta t^{-1} \mathcal{P}[n(t + \Delta t) - n(t) = -1] = \mu n(t). \tag{5.2}$$

The T cell pool $n(t)$ is split into $N(t)$ clonotypes, labelled by i, where each clonotype represents a group of cells with the same TCR (see Figure 5.1 for a model schematic). The number of T cells in each clonotype $n_i(t)$ is an integer ≥ 0. The system starts with N_0 clonotypes with n_0 cells per clonotype. Each T cell has equal probability per unit time of dying μ and thus the death rate for the cells in one clonotype $\mu_i(t) = \mu n_i(t)$. T cells compete equally for resource γ such that the birth rate for each cell is $\frac{\gamma}{n(t)}$ and the birthrate for cells in one clonotype

$\lambda_i(t) = \frac{\gamma n_i(t)}{n(t)}$, or,

$$\lambda_i(t) = \lim_{\Delta t \to 0} \Delta t^{-1} \mathcal{P}[n_i(t + \Delta t) - n_i(t) = 1] = \gamma \frac{n_i(t)}{n(t)} \qquad \text{and,} \tag{5.3}$$

$$\mu_i(t) = \lim_{\Delta t \to 0} \Delta t^{-1} \mathcal{P}[n_i(t + \Delta t) - n_i(t) = -1] = \mu n_i(t). \tag{5.4}$$

If the number of cells in a clonotype reaches zero that clonotype becomes extinct, so if $n_i(t_1) = 0$ then $n_i(t > t_1) = 0$ for all $t > t_1$.

Furthermore, the activity of the thymus is modelled through new clonotypes entering the system with probability per unit time θ. When new T cells enter the circulation from the thymus they have usually undergone a few rounds of division before leaving the thymus. The new clonotypes that enter the system therefore have n_θ cells per clonotype, where n_θ is usually in the range of two to ten cells.

Numerical simulations from the model use the Gillespie algorithm to track the integer number of T cells in each clonotype as they decrease or increase through cell death or division, and new clonotypes are produced by the thymus. At each step of the algorithm, one of three events can happen: a cell in a clonotype divides, a cell in a clonotype dies or a new clonotype is produced by the thymus. Time is then increased before the next step by an amount Δt distributed $\Delta t \sim \exp(1/S)$ where S is the sum of the rates of the dynamics $S = \gamma + \mu n + \theta$. Numerical simulations have also been carried out using a τ-leaping algorithm as an approximation of the Gillespie algorithm. In the τ-leaping algorithm, time is incremented by Δt and the number of events (cell death, cell division, clonotype birth) is Poisson distributed according to the rates μ, γ and θ. The τ-leaping algorithm has the advantage of being less computationally expensive, allowing the simulation of larger systems in shorter periods of time. Numerical simulations are then compared to analytical solutions.

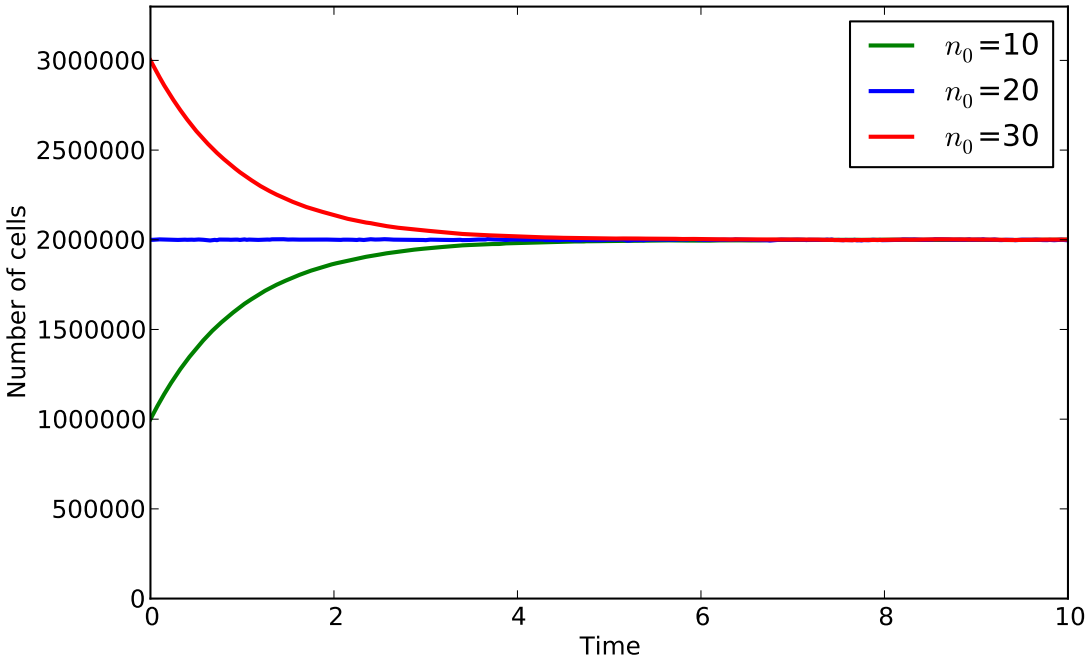

Figure 5.2: Total number of T cells with time from numerical simulation, demonstrating that the system heads towards an equilibrium from three different starting points. The equilibrium number of T cells is given by Eq (5.6). $\mu = 1.0$, $\gamma = 2{,}000{,}000$, $N_0 = 100{,}000$.

5.3 Results

5.3.1 Without thymic output

In the first instance, we look at the results from the model in the simplest situation, where there is no thymic output of new clonotypes so that $\theta = 0$. This can be thought of as the situation where the thymus is damaged, such as following HSCT, or where it has been removed. The model then depends on four parameters, the stimulus γ the death rate μ, and the two parameters giving the initial set-up of the system: the initial number of clonotypes N_0 and the initial number of cells per clonotype n_0.

Number of T cells

Although the numbers of T cells will fluctuate randomly due to the stochasticity in the model, the mean total number of T cells will reach a homeostatic steady

state, with this steady state dependent on the stimulus γ and the death rate μ. If we define the mean total number of T cells as $x = \mathbb{E}(n(t))$ then in the absence of any T cell production from the thymus,

$$\frac{\mathrm{d}}{\mathrm{d}t}x = \gamma - \mu x, \tag{5.5}$$

and thus in steady state, as $t \to \infty$, the mean total number of T cells is constant and given by the ratio of stimulus to death rate,

$$x \to \frac{\gamma}{\mu}. \tag{5.6}$$

Figure 5.2 demonstrates that an equilibrium level of T cells is reached from three different starting positions and that this steady state is indeed independent of the initial number of T cells.

Number of clonotypes

The number of clonotypes however does not have a steady-state number. Even when the total number of T cells is in steady state, the number of T cells in each clonotype will be fluctuating randomly, and as a result, the number of T cells in a clonotype can reach zero, resulting in the clonotype becoming extinct. Because of this extinction, the number of clonotypes $N(t)$ can only be a non-increasing function with time.

In order to find an analytic solution for the number of clonotypes with time, we look at a single clonotype and we take the situation where the rate of stimulus is much greater than the death rate, $\gamma \gg \mu$. We can therefore make the approximation that the system is close to its steady state, such that $n(t) \sim x(t) = \frac{\gamma}{\mu}$, and so,

$$\lambda_i(t) = \gamma\frac{n_i(t)}{n(t)} = \mu n_i(t). \tag{5.7}$$

Thus for each clonotype we have a birth-death process with no net drift.

Furthermore, with $\gamma \gg \mu$, because there will be large numbers of cells, we can treat changes in the numbers of cells as continuous, which allows the use of the diffusion approximation. We therefore approximate $n_i(t)$ as a diffusion process, $n_i(t) \sim \mathbf{X}_t$. In order to write down the corresponding stochastic differential equation (SDE) for the system, we define the mean and mean square for the change in $\mathbf{X}_t$, with time through comparison to the birth-death process,

$$\mathbb{E}(n_i(t+\Delta t) - n_i(t)) = \lambda_i(t)\Delta t - \mu n_i(t)\Delta t = 0, \tag{5.8}$$

where we have used the fact that $\lambda_i(t) = \mu n_i(t)$ and that there is therefore no net drift. The mean for the diffusion process is therefore given by $\mathbb{E}(\mathbf{X}_{t+\Delta t} - \mathbf{X}_t) = 0$. Similarly, to find the mean square,

$$\mathbb{E}((n_i(t+\Delta t) - n_i(t))^2) = \lambda_i(t)\Delta t + \mu n_i(t)\Delta t \tag{5.9}$$

$$= 2\mu n_i(t)\Delta t. \tag{5.10}$$

The mean square for the diffusion process is therefore given by $\mathbb{E}((\mathbf{X}_{t+\Delta t} - \mathbf{X}_t)^2) = 2\mu \mathbf{X}_t \Delta t$. The SDE can thus be written as,

$$\mathrm{d}\mathbf{X}_t = \sqrt{2\mu \mathbf{X}_t}\mathrm{d}\mathbf{W}_t. \tag{5.11}$$

This SDE is relatively well characterised, with known solutions [209,210]. In particular, if the probability of reaching $\mathbf{X} = 0$ before time t starting with $\mathbf{X}_0 = b$ cells is $F(t, b)$, then $F(t, b)$ satisfies the following partial differential equation,

$$\frac{\partial}{\partial t}F(t,b) = \frac{1}{2}\mu x_0 \frac{\partial^2}{\partial x_0^2}F(t,b), \tag{5.12}$$

where $F(t, 0) = 1$. This is found from the Kolmogorov backwards equation [210].

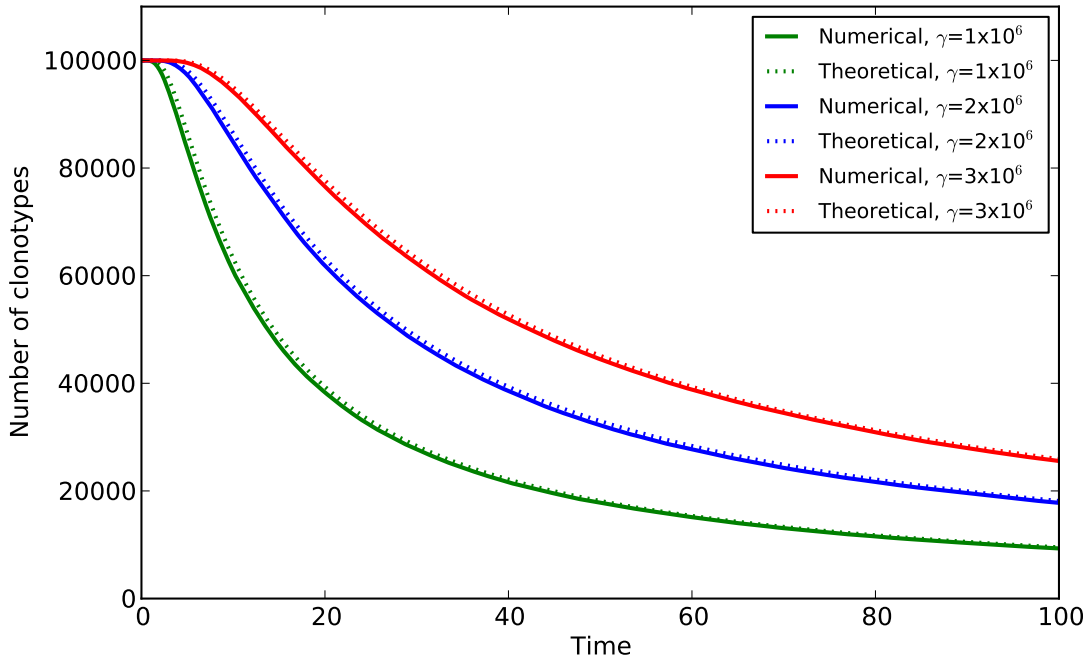

Figure 5.3: Comparison of number of clonotypes with time $N(t)$ from numerical simulation and theoretical prediction. The theoretical prediction is given by Eq (5.15). $\mu = 1.0$, $N_0 = 100{,}000$, n_0 varies according to γ to ensure the system starts with the steady state number of T cells, with $n_0 = 10$ for the green line, $n_0 = 20$ for the blue line and $n_0 = 10$ for the red line.

The value of b is the number of cells necessary in each clonotype at time $t = 0$ such that there are the expected mean number of cells in the system. In the case of this model, $b = \frac{\gamma}{\mu N_0}$, where N_0 is the number of clonotypes at $t = 0$. The solution to (5.12) is,

$$F(t, b) = 1 - \exp\left(-\frac{b}{\mu t}\right), \tag{5.13}$$

giving,

$$\mathcal{P}[\mathbf{X}_t = 0 | \mathbf{X}_0 = b] = 1 - \exp\left(-\frac{b}{\mu t}\right). \tag{5.14}$$

And so $N(t)$ the expected number of clonotypes with time is given by,

$$N(t) = N_0 \left(1 - \exp\left(-\frac{b}{\mu t}\right)\right) \tag{5.15}$$

$$= N_0 \left(1 - \exp\left(-\frac{\gamma}{\mu^2 N_0 t}\right)\right). \tag{5.16}$$

A comparison between this theoretical prediction and numerical simulations

of the model are shown in Figure 5.3 for three different values of γ. As can be seen, the theoretical predictions match the numerical simulations well. It can also be seen that the number of clonotypes falls fast at first and then the decrease slows down. Because the number of T cells remains constant (Figure 5.2), the average size of a clonotype increases as the number of clonotypes falls, and so the probability of extinction for each clonotype decreases. The number of clonotypes does however continually fall, with the probability of full extinction as $t \to \infty$ equal to 1. There is therefore no steady-state number of clonotypes or steady-state distribution of clonotypes sizes.

5.3.2 Including thymic output

Including the release of new clonotoypes into the periphery from the thymus makes the model a closer representation of the actual system. By including thymic output in the model, there will be a point at which the rate of production of new clonotypes equals the extinction rate of clonotypes and an equilibrium number of clonotypes can be reached. This allows long-term numbers of clonotypes and distributions of clonotypes sizes to be found from the model.

Number of cells

After including thymic output, similarly to the scenario without thymic output, the total number of cells will fluctuate randomly. The mean total number of T cells in the population does however still have a steady state, but this steady state now depends on the extent of thymic output given by the rate of production of new clonotypes θ and the number of cells per new clonotpye n_θ, as well as the stimulus γ and the death rate μ. Taking the mean total number of T cells again as $x = \mathbb{E}(n(t))$, in the presence of thymic output of new clonotypes then,

$$\frac{\mathrm{d}}{\mathrm{d}t}x = \gamma - \mu x + \theta n_\theta, \tag{5.17}$$

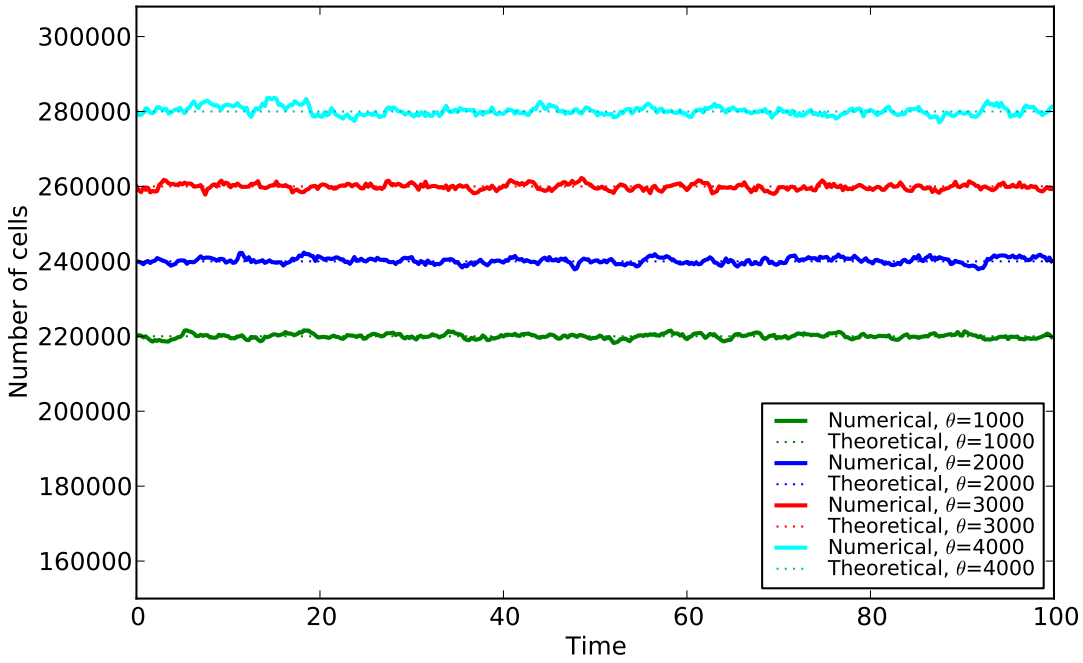

Figure 5.4: Total number of T cells with time from numerical simulation including thymic output, demonstrating the effects of different levels of thymic output θ on the homeostatic set point. Theoretical values from Eq (5.18). $\mu = 1.0$, $\gamma = 200{,}000$, $n_0 = 20$, $N_0 = 10{,}000$, $n_\theta = 20$.

and thus as $t \to \infty$, the mean total number of T cells still reaches a steady state number, given by,

$$x(t) \to \frac{\gamma + \theta n_\theta}{\mu}. \tag{5.18}$$

The effect of changing θ on this homeostatic equilibrium value is demonstrated in Figure 5.4, with the number of T cells reaching a dynamic steady state which increases with the level of thymic output.

Clonotype lifetime and the number of clonotypes

In order to find the clonotype lifetimes and the number of clonotypes, similarly to the case with no thymic output, we first need to define the stochastic differential equation for the system. Again we look at a single clonotype in the case that $\gamma \gg \mu$ allowing us to assume that the number of cells in a clonotype is continuous. By including thymic output, the deathrate is now roughly equal to the total production of cells from both proliferation and thymic output at steady state or

$\mu n(t) \simeq \gamma + n_\theta \theta$. The birthrate is therefore given by,

$$\lambda_i(t) = \gamma \frac{n_i}{n(t)} = \gamma \frac{\mu n_i}{\gamma + n_\theta \theta} = \frac{\mu n_i}{1+\beta} \tag{5.19}$$

where

$$\beta = \frac{n_\theta \theta}{\gamma} \tag{5.20}$$

is the ratio of the thymic output of T cells to the production of T cells through peripheral division. Thus we now need to consider a birth-death process [211], where the death rate is slightly higher than the birthrate and there is thus net drift within a single clonotype. The expected change in the birth-death process is now,

$$\begin{aligned}
\mathbb{E}(n_i(t+\Delta t) - n_i(t)) &= (\lambda_i(t) - \mu n_i(t))\Delta t && (5.21)\\
&= \mu n_i(t) \left(\frac{1}{1+\beta} - 1 \right) \Delta t && (5.22)\\
&= \mu n_i(t) \left(\frac{-\beta}{1+\beta} \right) \Delta t && (5.23)\\
&= -\alpha \mu n_i(t) \Delta t, && (5.24)
\end{aligned}$$

where $\alpha = \frac{\beta}{1+\beta}$. Similarly, for the mean square,

$$\begin{aligned}
\mathbb{E}((n_i(t+\Delta t) - n_i(t))^2) &= (\lambda_i(t) + \mu n_i(t))\Delta t && (5.25)\\
&= \mu n_i(t) \left(\frac{1}{1+\beta} + 1 \right) \Delta t && (5.26)\\
&= \mu n_i(t) \left(\frac{2+\beta}{1+\beta} \right) \Delta t && (5.27)\\
&= (2-\alpha) \mu n_i(t) \Delta t. && (5.28)
\end{aligned}$$

The mean and mean square for the diffusion process are therefore given by,

$$\mathbb{E}(\mathbf{X}_{t+\Delta t} - \mathbf{X}_t) = -\alpha \mathbf{X}_t \Delta t \qquad \mathbb{E}((\mathbf{X}_{t+\Delta t} - \mathbf{X}_t)^2) = (2-\alpha)\mu \mathbf{X}_t \Delta t. \tag{5.29}$$

With $0 < \alpha \ll 1$, so in the scenario where most of the T cells are produced through peripheral devision rather than from the thymus, the SDE for the system can be approximated as,

$$\mathrm{d}\mathbf{X}_t = -\alpha\mu\mathbf{X}_t\mathrm{d}\mathbf{X}_t + \sqrt{2\mu\mathbf{X}_t}\mathrm{d}\mathbf{W}_t. \tag{5.30}$$

Similarly to the case where there was no thymic output, this equation is relatively well characterised with known solutions [209,211]. The mean clonotype lifetime is given by the time to extinction for clonotype, or the time it takes on average for a clonotype to have zero T cells through random fluctuations. If $T(x, \alpha)$ is the time to extinction of a typical clonotype starting with x cells in the clonotype, then $T(x, \alpha)$ must a solution to the following general ODE for time to reach a boundary from a starting point in between [211],

$$\frac{1}{2}\bar{\sigma}^2(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2}T(x,\alpha) + \bar{\mu}(x)\frac{\mathrm{d}}{\mathrm{d}x}T(x,\alpha) = -1 \;\text{ for } c < x < d\,. \tag{5.31}$$

In the situation presented here, the values for $\bar{\sigma}$ and $\bar{\mu}$ are given by $\bar{\sigma}(x) = \sqrt{2\mu x}$ and $\bar{\mu}(x) = -\alpha\mu x$. We are also only interested in the time to reach 0 and so set the boundary conditions as $c = 0$ and $d \to \infty$. Hence we need to find a solution for $T(x, \alpha)$ for the following ODE,

$$\mu x\frac{\mathrm{d}^2}{\mathrm{d}x^2}T(x,\alpha) - \alpha\mu x\frac{\mathrm{d}}{\mathrm{d}x}T(x,\alpha) = -1 \;\text{ for } 0 < x < \infty\,, \tag{5.32}$$

with boundary conditions $T(0, \alpha) = 0$ and $T(\infty, \alpha) = 0$. Mathematica version 9.0 [212] and the analytical equation solver `DSolve` was used to find a solution for $T(x, \alpha)$ giving,

$$T(x,\alpha) = \frac{1}{\alpha\mu}\Big(\gamma_{\mathrm{E}} - e^{\alpha x}\mathrm{Ei}(-\alpha x) + \ln(\alpha x)\Big) \tag{5.33}$$

where $\mathrm{Ei}(x)$ is the Exponential Integral and γ_{E} is the Euler–Mascheroni constant,

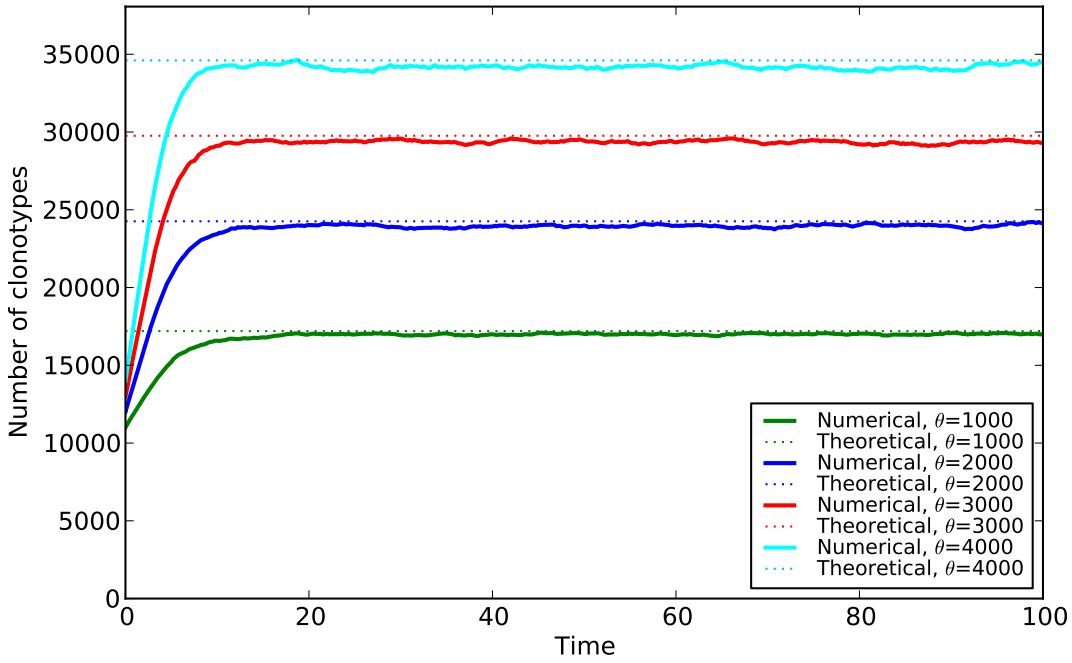

Figure 5.5: Total number of clonotypes with time from numerical simulation including thymic output, demonstrating the effects of different levels of thymic output θ on the homeostatic set point. The theoretical values come from Eq (5.36). The system is started such that there are the homeostatic number of T cells. $\mu = 1.0$, $\gamma = 200{,}000$, $n_0 = 20$, $N_0 = 10{,}000$, $n_\theta = 20$.

given by

$$\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t}\,\mathrm{d}t \qquad \text{and} \tag{5.34}$$

$$\gamma_{\mathrm{E}} = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln(n)\right) \simeq 0.57721\ldots\ . \tag{5.35}$$

Therefore, if the initial number if cells in a clonotype in n_θ and the mean lifetime of a clonotype is $T(n_\theta, \alpha)$, then the mean number of surviving clonotypes, as $t \to \infty$, is

$$\bar{N} = \theta\, T(n_\theta, \alpha) = \frac{\theta}{\alpha\mu}\Big(\gamma_{\mathrm{E}} - e^{\alpha n_\theta}\mathrm{Ei}(-\alpha n_\theta) + \ln(\alpha n_\theta)\Big)\ . \tag{5.36}$$

The number of clonotypes therefore reaches steady state (see Figure 5.5). The steady state number is dependent on α, related to the ratio of cells produced by thymic output to cells produced by proliferation, and n_θ, the number of cells per

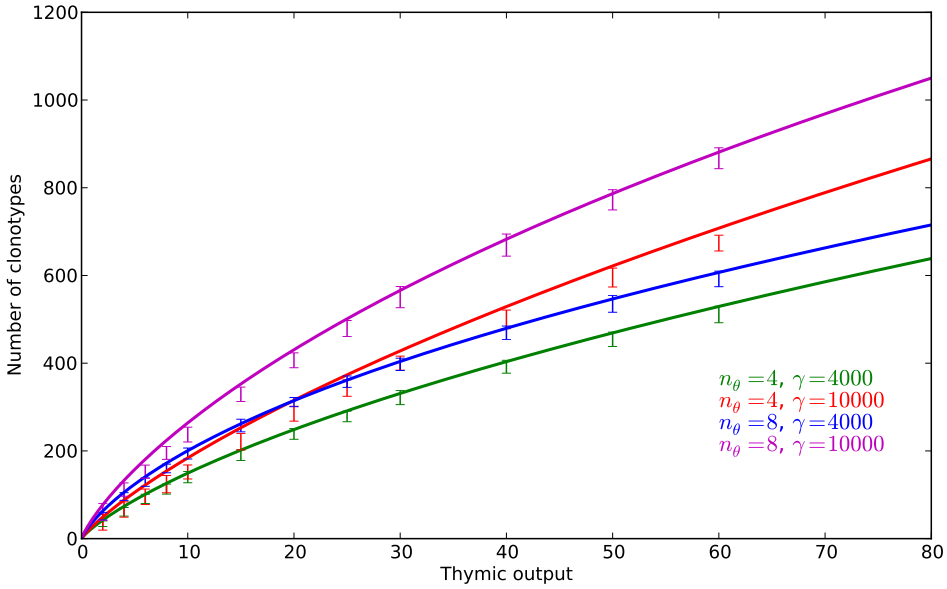

Figure 5.6: Mean long-term number of clonotypes with thymic output for four different combinations of γ and n_θ. The lines give the theoretical number of clonotypes $\bar{N}$ from formula (5.36), while the error bars give the results from numerical simulation. $\mu = 1.0$.

new clonotype. Comparisons between the results of numerical simulations of the model for $\bar{N}$ with thymic output and formula (5.36) can be seen in Figure 5.6, demonstrating the relationships between the number of clonotypes and the values for γ, θ and n_θ.

Distribution of clonotype sizes

Because the number of clonotypes reaches a steady state, the sizes of the clonotypes will form a steady state distribution. In order to find this long-term distribution, we first assume each clonotype can be approximated as an independent realisation of the birth-death process, obeying SDE (5.30). From this can be calculated $G(y, x)$, the occupation density at y (the number of clonotypes with y cells) given that the process starts at a source x (each clonotype has x cells in when it enters the population) and is absorbed at 0 (clonotypes go extinct when

they have 0 cells). $G(y,x)$ is known to satisfy the following ODE,

$$\mu x \frac{\mathrm{d}^2}{\mathrm{d}x^2} G(y,x) - \alpha\mu x \frac{\mathrm{d}}{\mathrm{d}x} G(y,x) = -\delta(x-y) \text{ for } 0 < x < \infty \tag{5.37}$$

with boundary condition $G(y,0) = 0$. The solution can be constructed from combinations of a constant and $\mathrm{e}^{\alpha x}$ as,

$$G(y,x) = \begin{cases} c(\mathrm{e}^{\alpha x} - 1) & y < x \\ c(\mathrm{e}^{\alpha y} - 1) & y \geq x \ . \end{cases} \tag{5.38}$$

By using the δ–function property that

$$\lim_{\epsilon \to 0} \int_{y-\epsilon}^{y+\epsilon} \delta(x-y)dx = 1 \ , \tag{5.39}$$

and by integrating (5.37) with respect to x between the limits $x = y - \epsilon$ and $x = y + \epsilon$ and taking the limit $\epsilon \to 0$, it can be shown that the step change in $G'(y,x)$ at $x = y$ must be equal to $-1/\mu y$. Thus by differentiating the two parts of (5.38), and taking the difference we find,

$$\frac{c\alpha}{\mathrm{e}^{-\alpha y}} = \frac{1}{\mu y} \tag{5.40}$$

and thus,

$$c = \frac{1}{\alpha\mu} \frac{\mathrm{e}^{-\alpha y}}{y}. \tag{5.41}$$

In this case, the source x is the number of cells per clonotype output from the thymus, n_θ. Thus with n_θ fixed we find the number of clonotypes with y cells in them to be,

$$G(y,n_\theta) = \begin{cases} \dfrac{1}{\alpha\mu} \dfrac{1 - \mathrm{e}^{-\alpha y}}{y} & y < n_\theta \\ \dfrac{\mathrm{e}^{\alpha n_\theta} - 1}{\alpha\mu} \dfrac{\mathrm{e}^{-\alpha y}}{y} & y \geq n_\theta \ . \end{cases} \tag{5.42}$$

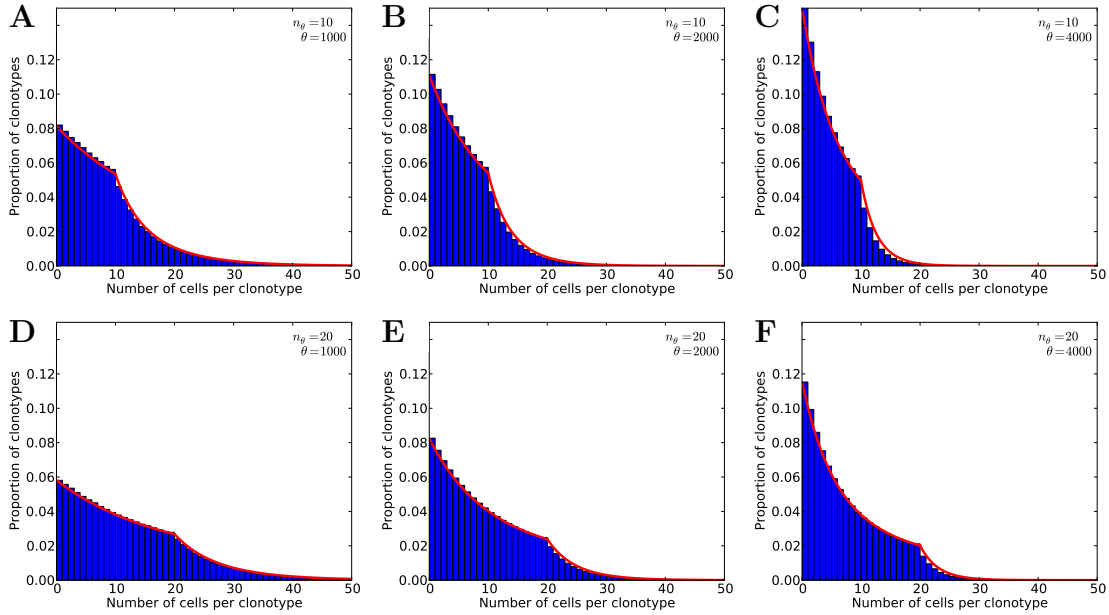

Figure 5.7: Histograms of numerical simulations of the number of T cells in each clonotype (blue bars) is compared with the theoretical distribution of clonotype sizes (red lines) given by $G(y,b)/T(b,\alpha)$ where $G(y,b)$ is given by formula (5.42). Multiple values of θ and n_θ are shown to demosntrate the differences in their resultant distributions. $\gamma = 200{,}000$, $\mu = 1.0$, $n_0 = n_\theta$

Comparisons between numerical simulations and the formula given in (5.42) are shown in Figure 5.7. As can be seen, this formula appears to fit the simulations well with these parameter values. Increased levels of thymic output give increased drift for each individual clonotype towards zero. Thus the distributions become increasingly pushed towards many clonotypes with small numbers of cells in them.

Gini coefficient

The Gini coefficient is often used as a metric for the dispersion and inequality of a population, vaying between 0 and 1. A Gini coefficient of 1 means complete inequality and a Gini coefficient of 0 means complete equality in that every instance of the population has the same size, so in this case every clonotype contains an equal number of cells.

The Lorenz curve is used to define the Gini coefficient. In this case, the Lorenz curve is the cumulative proportion of the total number of T cells with increasing

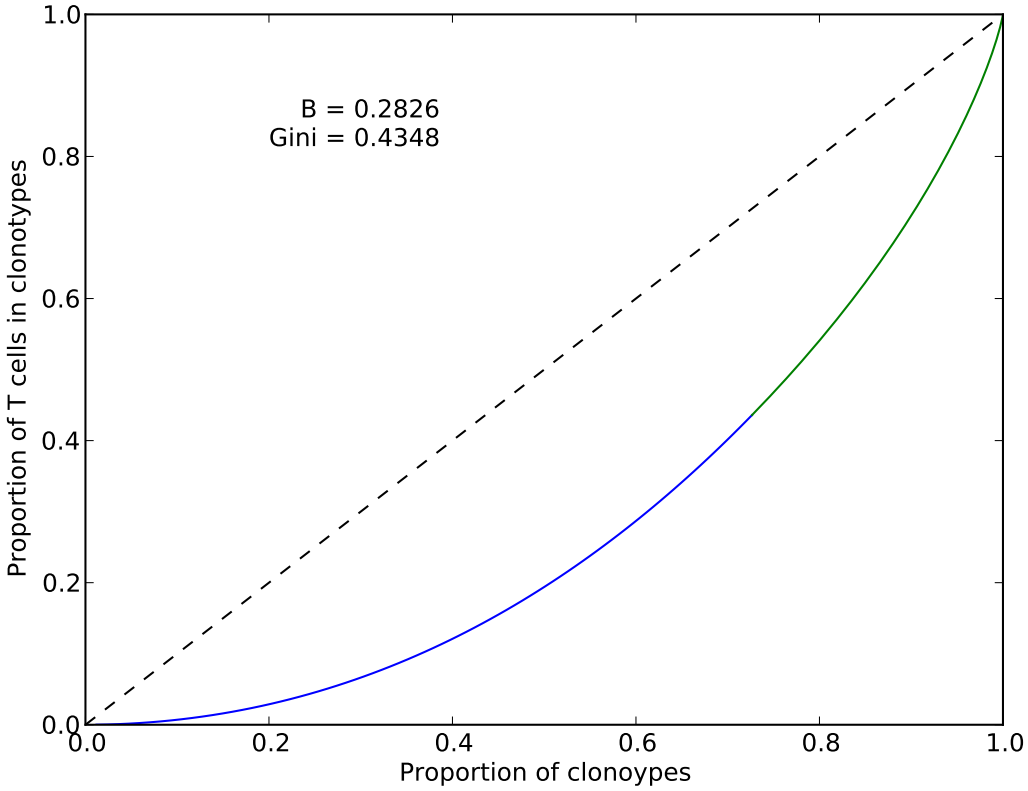

Figure 5.8: A Lorenz curve from global competition model by plotting $F(z)$ given by (5.43) against $H(z)$ given by (5.44) where z is the number of T cells in a clonotype. The blue section is for $z < n_\theta$, and the green section for $z > n_\theta$. B and the Gini coefficient were calculated using (5.46). $\gamma = 4000$, $\theta = 80$, $\mu = 1.0$, $n_\theta = 8$.

clonotype size plotted against the cumulative proportion of total clonotypes. The Gini coefficient is then twice the area between this curve and the line of equality.

The Lorenz curve in this case, as demonstrated in Figure 5.8, is given by a plot of the cumulative total number of T cells in z clonotypes, $F(z) = \int_0^z zG(z,b)dz$ on the y-axis against the cumulative total number of clonotypes, $H(z) = \int_0^z G(z,b)dz$ as a parametric plot with increasing z. Where,

$$F(z) = \begin{cases} \frac{1}{\alpha b}\left(\alpha z - 1 + \mathrm{e}^{-\alpha z}\right) & z < b \\ \frac{1}{\alpha b}\left(\alpha b + (1 - \mathrm{e}^{\alpha b})\mathrm{e}^{-\alpha z}\right) & z \geq b \end{cases} \tag{5.43}$$

$$H(z) = \begin{cases} \frac{1}{\alpha\mu}\left(\gamma_E + \log(\alpha z) - \mathrm{Ei}(\alpha z)\right) & z < b \\ \frac{1}{\alpha\mu}\left(\gamma_E + \log(\alpha b) - \mathrm{e}^{\alpha b}\mathrm{Ei}(-\alpha b) + (\mathrm{e}^{\alpha b} - 1)\mathrm{Ei}(-\alpha b)\right) & z \geq b. \end{cases} \tag{5.44}$$

From these, B the area under the Lorenz curve is given by,

$$B = \int_0^\infty F(z)H'(z)\mathrm{d}z \tag{5.45}$$

where $H'(z) = G(z, b)$. Thus,

$$B = \left\{ \frac{1}{\alpha^2 b\mu T}\left(\alpha b + \mathrm{e}^{-\alpha b} - 1 - \gamma_E - \log\left(\frac{\alpha b}{2}\right) + 2\mathrm{Ei}(-\alpha b) - \mathrm{Ei}(-2\alpha b)\right) + \frac{1-\mathrm{e}^{\alpha b}}{\alpha^2 \mu b T}\left(\alpha b\mathrm{Ei}(-\alpha b) + (1-\mathrm{e}^{\alpha b})\mathrm{Ei}(-2\alpha b)\right)\right\}, \tag{5.46}$$

where b is the number of cells per new clonotype, n_θ. The Gini coefficient is then given by $\mathrm{Gini} = 1 - 2B$.

5.3.3 Application to murine CD8 memory data

The experimental setup

The model has been applied to data from an experiment conducted by Thea Hogan and Benedict Seddon at the Royal Free London NHS Foundation Trust. In this experiment, a series of mice were given busulfan at 8 weeks of age to ablate the bone marrow whilst leaving the circulating immune cells intact. New bone marrow that is genetically identical except for a change to a single marker which allows it and its daughter cells to be identified is then grafted into the mouse. The replacement of the original host T cells in the body by the new donor T cells is then measured over the following months.

The experimenters expected to find that the donor T cells replaced the host T cells until the proportions in the circulations were equal to the proportion of donor bone marrow to host bone marrow after the engraftment. With the application of mathematical modelling in conjunction with Andrew Yates and Graeme Gossel, however, they discovered that the T cells did not achieve complete replacement for any T cell subset. This effect was more marked with memory T cells and

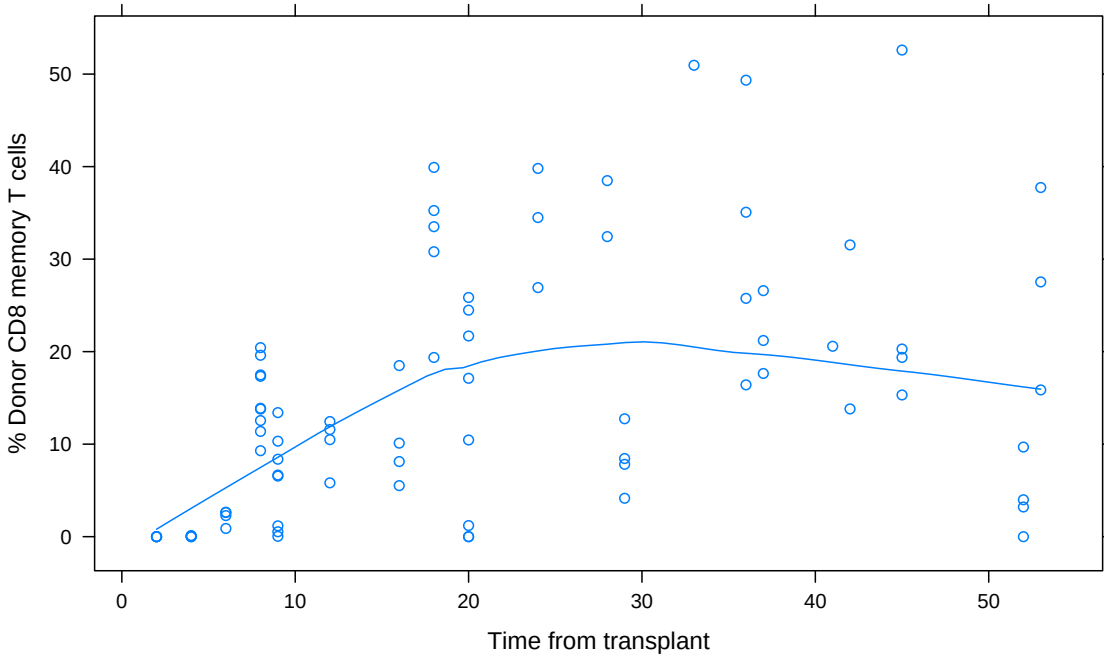

Figure 5.9: Results of the mouse experiment for CD8 memory T cell replacement from the experiment conducted by Thea Hogan at the Royal Free London NHS Foundation. Each datapoint is a different mouse. It gives the percentage of the cells that are from the donor bone marrow out of the CD8 memory cells, normalised to the percentage of the bone marrow that had been replaced for each mouse. The line gives a local regression curve for the data. As can be seen, there is large variability in the data, but no mouse manages more than 55% replacement, with the mean long-term replacement in the region of 20%.

particularly with CD8 memory compartment where the donor T cells only managed to replace approximately 20% of the compartment, after normalising for the percentage of the bone marrow that had been replaced (see Figure 5.9).

While naïve T cell populations require a combination of signalling from sp-MHC and cytokines for survival and proliferation, memory T cells mostly only require interactions with cytokines such as IL-7 and IL-15 [19] and are largley sp-MHC independent. As such, these T cells compete only for global resources. The situation is simpler in CD8 memory, where most subsets are reckoned to compete only for IL-15 cytokines [19], and as such the cells are competing for a single global resource. Therefore the model presented in this chapter could be appropriate to look at the dynamics of the CD8 memory T cell subset.

The cell surface marker CD122 is the β-chain of the IL-2 receptor and is also a

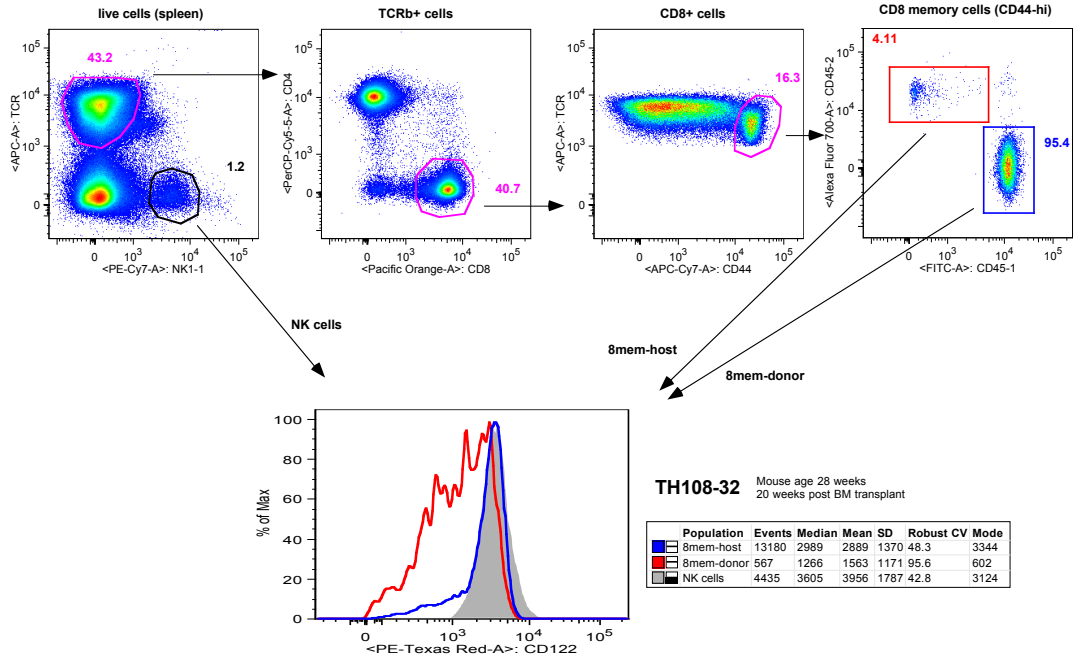

Figure 5.10: Gating strategy for CD8 memory and CD122 expression levels. The cells in the spleen were split into NK cells and T cells using the TCR. These were subsetted into CD8+ CD4- and then further into CD44 positive to give the memory CD8 cells. The two different markers for CD45 were used to differentiate the donor and host populations of these cells. Expression levels of CD122 could then be compared to those of the NK cells to give their relative expression level. With thanks to Thea Hogan.

component of the IL-15 receptor. As such, expression levels of CD122 can be used as a proxy for IL-15 cytokine receptor levels for a cell. Figure 5.10 gives the gating strategy for looking at the CD8 memory subset, and at their CD122 expression levels. As can be seen, the CD122 expression level for donor memory cells is lower than that for host memory cells in this example. The complete distributions of median CD122 expresion levels of host and donor cells is shown in Figure 5.11. Because of flow-cytometry inaccuracies, data collected on different days cannot be compared directly. NK cell CD122 distributions however are not expected to change and as such were used to normalise the CD8 memory CD122 expression levels for both host and donor. It can be seen that CD122 expression levels are significantly higher for host CD8 memory cells than donor CD8 memory cells. This implies that IL-15 receptor levels are also correspondingly higher for host CD8 memory T cells over donor CD8 memory T cells, giving them an advantage

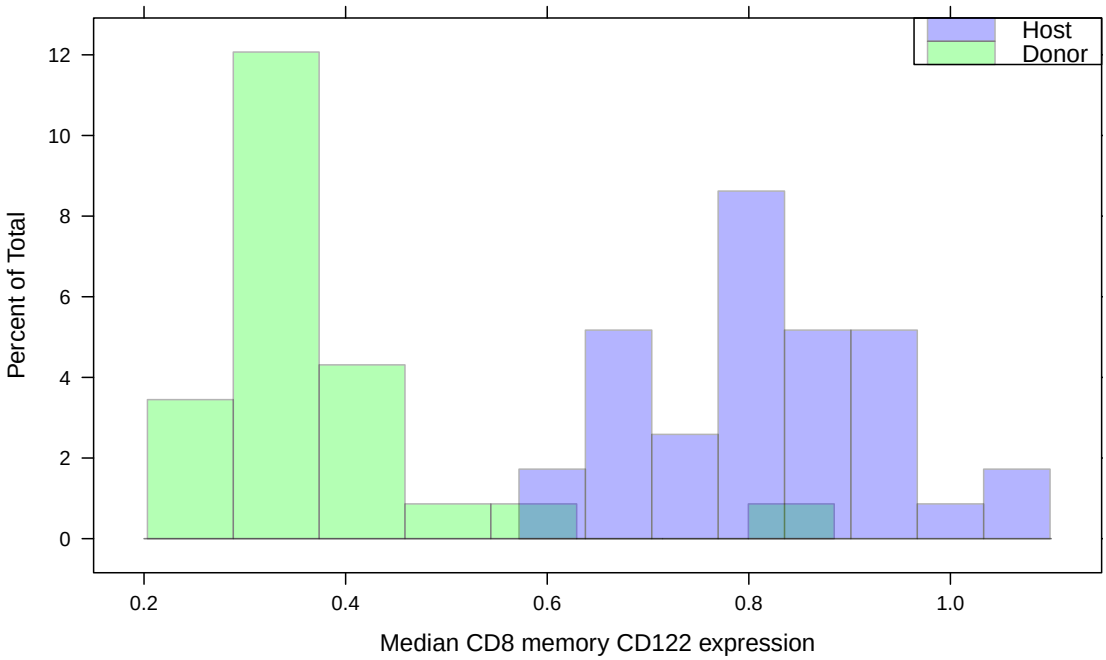

Figure 5.11: CD122 expression level distributions for host and donor CD8 memory cells. The CD122 expression is reported as relative to the CD122 expression of the NK cells from the same mouse to normalise between differences from the flow cytometry being carried out on different days.

in competing for resources.

Modelling

In order to model the situation represented in these data, the global competition model was updated such that, while the total level of resource remained constant, cells from each clonotype have an affinity for resource drawn from a truncated normal distribution. It is assumed that this would in some way be related to the number of receptors for cytokines on their cell surface. Cells from each clonotype then compete for the resource. In this situation, the mechanism that was described as thymic output becomes clonotype entry into the memory pool from the naïve pool. As a new clonotype enters, its affinity for resource is drawn randomly from the same truncated normal distribution.

The system starts with a large number of clonotypes in the memory pool simulating the effects of birth and the immediate recognition of many pathogens.

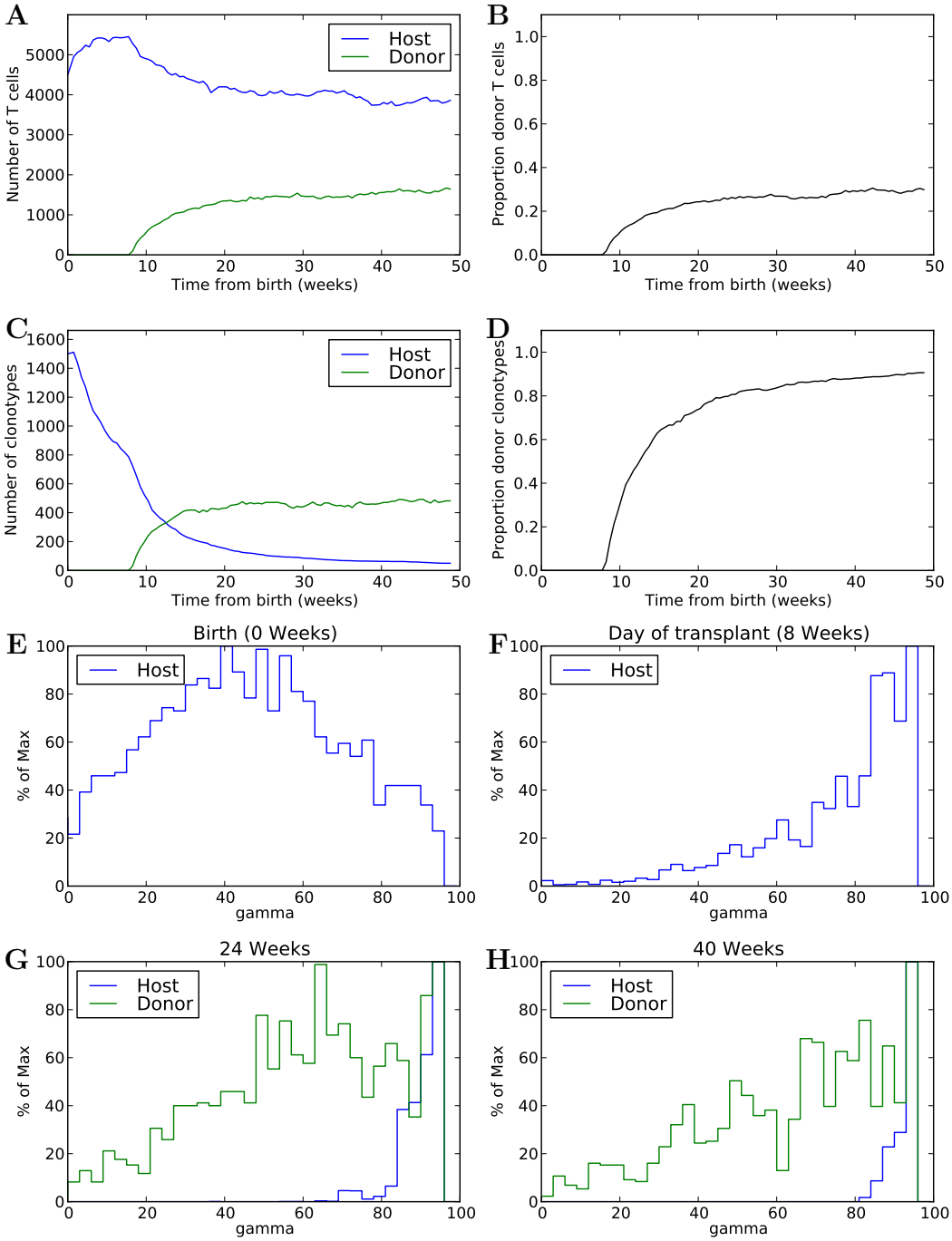

Figure 5.12: The impact of competition for resources as an explanation for the incomplete replacement of CD8 memory cell in mice. The system starts with an influx of memory clonotypes (**C**) with affinity for resource drawn from the distribution shown at birth (**E**). After 8 weeks of competition, by the day of transplant, the distribution has skewed with the clonotypes with higher affinity out-competing those with lower affinities (**F**). The donor clonotypes, which are drawn from the same distribution as at birth therefore do not compete for resource equally (**G** and **H**) and cannot displace these incumbent cells high-affinity cells and therefore, while the proportion of donor clonotypes becomes high (**D**), the proportion of donor T cells remains low (**B**).

These clonotypes each have an affinity for resource drawn randomly from the truncated normal distribution. The cells in the clonotypes then compete for 8 weeks, during which time, new host clonotypes enter the memory pool at a fixed rate, also with their affinity for resource also drawn randomly from the same distribution. At 8 weeks, the mouse receives busulfan ablating the host bone marrow and new donor bone marrow is injected. In the model, after this bone marrow transplant, the new clonotypes entering the memory pool are now donor CD8 memory T cells, with their affinity for resource again drawn from the same distribution as the host clonotypes.

The results from the model are shown in Figure 5.12. In the first 8 weeks, the number of clonotypes and T cells falls towards a homeostatic level. Because of competition for resources, the host clonotypes from the original distribution (Figure 5.12E) with a high affinity for resource are selected and survive and come to dominate the pool so that the distribution of affinities is heavily weighted towards high-affinity (Figure 5.12F). When donor clonotypes then attempt to enter the pool, because their affinity is drawn from the original truncated normal distribution (the same as that in Figure 5.12E), on average each clonotype will have a lower affinity than the established host memory cells. As a result, they struggle to out-compete the host cells, which causes incomplete replacement in line with the experimental data (Figure 5.12B). The model predicts that there will be almost complete replacement of the T cell clonotypes (Figure 5.12D), with only those with high affinity for resource surviving (Figure 5.12H). Over a long enough period of time, the number of host T cells will decline as clonotypes randomly die out and are not replaced.

5.4 Discussion

A stochastic model for global competition between T cell clonotypes has been developed. Because the fundamentals of the model are not too complex, it is not too computationally expensive to run the model with large numbers of clonotypes or for long time scales. This allows greater numerical insight into the properties of the system. Furthermore, the model is simple enough to characterise analytically, allowing the development of analytic solutions that are generalisable to all scales, allowing the results to be applied at the level of mice and humans.

Without a constant input of new clonotypes from the thymus, the stochastic fluctuations in the number of T cells in each clonotype causes the number of T cells to reach zero and some clonotypes to become extinct. Hence the number of distinct clonotpes, and thus diversity, is expected to fall (Figure 5.3). This has implications in the months after an HSCT when the thymus is not producing any T cells; whilst T cell numbers might start to recover through homeostatic mechanisms, until the thymus starts producing new T cells, diversity can be expected not only to remain low but also to get worse.

The model reaches a steady state number of T cells (Figures 5.2 and 5.4), with the number of T cells resulting from a balance of cell death rates and cell production rates from both thymic output and cell division. Without thymic output, the number of T cells still remains constant. Accordingly, as the number of clonotypes decreases, the number of T cells in each clonotype must increase. This means, measuring just the concentrations of T cells following interventions, such as thymectomies, may be misleading in terms of the damage done.

In the presence of thymic output, the number of clonotypes reaches a steady state, with the number dependent on the level of thymic output (Eq. (5.36) and Figures 5.5 and 5.6). A prediction from this model is therefore that as thymic output falls in a healthy child, the number of distinct T cell clonotypes can be expected to fall as well.

The mean lifetime of a clonotype in the perihery is dependent on thymic output (Eq (5.33)). As the thymic output increases, the net drift effect on the number of T cells in a clonotype increases, driving the clonotypes towards extinction more rapidly, shortening clonotype lifetime. This implies that in young children where there is higher thymic output, clonotypes and to an extent immune memory may have a shorter lifetime than in older children. This might in part explain why children who are more likely to become atopic, develop an immune response to dust mites at a later age [213], and may also explain the decreased ability of children to form immunity to Plasmodium falciparum in comparison to adults [214]. The number of clonotypes in the T cell pool is also dependent on thymic output, with increased thymic output resulting in more, smaller, clonotypes.

The model output for the distribution of clonotype sizes is highly dependent on both thymic output of new clonotypes (θ) and the number of cells in the clonotypes (n_θ) that are produced by the thymus. On the one hand, the higher the thymic output of new clonotypes, the more heavily the distribution is skewed towards low numbers of cells per clonotype. On the other hand, the more cells per new clonotype, the flatter the distribution and the more clonotypes with more cells per clonotype. Thymic output is usually just measured as cells/day. Figures 5.7B and D both have the same thymic output in terms of numbers of cells, as do and Figures 5.7C and E. The stark differences in their distributions therefore imply that measuring thymic output in this way misses key information related to the drivers of clonotype diversity.

The model output for the distribution of the number of T cells in each clonotype could be compared to data from sequencing T cell receptors. Comparing these distributions would allow validation of the model output and the estimation of some parameter values. Because the analytical solutions for the model are scalable, it might then be possible to form predictions of distributions in different circumstances, such as following thymectomy, following HSCT or in old age. Fur-

thermore, the model's scalability allows comparisons to be made between humans and mice.

The model has been able to reproduce the observed effects from replacing the bone marrow in a mouse and the lack of full replacement. In this solution, the surviving host cells and invading donor cells are from identical populations but because of the time advantage that the host cells have had to select clonotypes with higher affinity, on average they effectively have a competitive advantage over the invading donor cells. This demonstrates that a biologically simple solution is a possible explanation for the data, using purely stochastic dynamics and competition for resources. This represents a key advantage of stochastic modelling over deterministic modelling, with it being possible to create dynamics that incorporate the stochastic nature of elements of the system, in this case the affinity for cytokines. Over a long enough period of time, enough donor clonotypes with high affinity would enter the pool that eventually the donor cells would be expected to completely replace the host cells. This period of time is however likely to be longer than the lifetime of the mouse.

The model does however have limitations. Naïve T cells compete not only for cytokines but also for sp-MHC, with the affinity of different clonotypes for different sp-MHC governed by an interaction matrix. As such, the situation is much more complex for these cell subsets and results from this model can only be applied as an explanation for one part of their clonotype dynamics. Furthermore, the results for clonotype size distributions and lifetimes have assumed that all clonotypes compete equally. Again the situation is more complex, with cells likely to have affinities for resource drawn from some sort of distribution, in a similar manner to that used in modelling the mouse CD8 memory data.

In this chapter, we have developed a stochastic model for global competition of clonotypes in T cell homeostasis. The model was simple enough that it was possible to use it for large numbers of T cells and clonotypes and for long time-

periods. It was also simple enough that it was possible to find analytical solutions that can recreate the numerical simulations, allowing the model results to be scaled to the size of a human. The model was successfully adapted to be applied to a system from a murine experiment, demonstrating that a simple solution reliant on dynamics from competition for global resources could explain the apparent complications in the data.

Contributions

Grant Lythe proposed the model and did the analysis in the case with no thymus in Section 5.3.1 and found the solution for long-term clonotype distributions. I performed the analysis of the case with a thymus in Section 5.3.2 to find the solutions for long-term clonotype lifetimes, long-term clonotype numbers and the Gini coefficient. I also produced all the numerical solutions throughout the chapter and carried out all the analysis for the CD8 memory T cells mouse experiment in Section 5.3.3.

Chapter 6

Conclusions

Understanding the reconstitution of the haematopoietic system following paediatric haematopoietic stem cell transplantation is important because of the high mortality and morbidity associated with the intervention. Similarly, in HIV-infected children starting ART, the immune system does recover but its recovery is incomplete and patients can develop resistance. The aim of this thesis was to adapt and construct mechanistic mathematical models to better understand the underlying dynamics that govern these systems. While one part of this work looked at a stochastic model for global competition, studying its effects on T cell clonotype dynamics, the majority of this work used deterministic models to look at populations of cells, applying them to data.

Following paediatric HSCT neutrophils and platelets recover quickly, and so age-related effects can be ignored in the structural model. It was therefore possible to apply previous adult models for neutrophil and platelet dynamics. For neutrophils, the model of Friberg *et al* [52] was mechanistically altered to account for the transient release of extra neutrophils on the administration of steroids [60–63] and to allow the elimination rate to differ from the transfer and synthesis rates. For platelets, the model of Friberg *et al* [52] was found to provide the better fit than the model of Hayes *et al* [72], demonstrating that the feedback

term was necessary to account for the oscillatory reconstitution profile. These platelet oscillations had not been observed in previous uses of these models because they did not have either the resolution in the data or the experimental setup [67–71], although fluctuations on a similar time-scale have been observed in healthy adults [96], with the negative feedback thought to be due to levels of TGF-beta [97]. Covariate analysis found that long-term platelet concentration was lowered in patients that received cord blood stem cells and raised in patients that received peripheral blood stem cells in comparison to bone marrow stem cells. This potentially implies that early platelet reconstitution is derived from precursor cells contained in the graft, rather than bone marrow derived. It was notable that no inter-individual age differences were observed for either neutrophils or platelets, implying that not only are concentrations relatively stable through childhood, as expected [98,99], but also the dynamics are stable as well.

In contrast, CD4 T cells recover slowly following paediatric HSCT and so age-related changes will manifest themselves within an individual's data [31]. In this work, a novel mechanistic model has been constructed that accounted for these age-related changes as effects on thymic output, loss and proliferation. Further mechanistic elements were included to account for the delay to thymic output in the months following HSCT [100,117–119] and the effects on proliferation and loss of competition for resources [14–19]. Testing these mechanistic components demonstrated that they were necessary to achieve a good fit to the data. In the resulting model, younger patients recovered more quickly than older patients, in line with the data. Covariate analysis identified that alemtuzumab and anti-thymocyte globulin (ATG), two anti-lymphocyte antibodies, reduced the initial concentration of CD4 T cells. Whilst this effect is to be expected, with the model it was possible to quantify the size of the effect. It was also possible to demonstrate that, even though reconstitution was delayed because it was starting from a lower concentration, the rate of reconstitution was unaffected by these drugs.

Patients that received cord blood stem cells had been noticed to have improved early reconstitution. These patients were however younger on average and were less likely to have received either alemtuzumab or ATG and the model was able to demonstrate that these two effects were enough to explain the differences found in the observed data. The model was used to form predictions of reconstitution on an individualised basis using just early data and the relevant covariates. The model's predictive ability was found to be high, capturing the trajectory in 81% of the patients in a new dataset of 75 patients that had not been looked at before.

In HIV-infected children starting ART, viral loads decline sharply but CD4 reconstitution is slow enough that age effects need to be included in the model. A previous model for HIV viral load and CD4 concentration dynamics [179,180] was successfully simplified using a quasi-steady state approximation to remove the infected CD4 T cells compartment because they represent such a small fraction of the total CD4 T cell population [198]. A further adaption was made in order to remove the oscillatory dynamics that the original model produced because the data did not support these dynamics. The mechanistic elements from the model of CD4 reconstitution following HSCT were then included to account for age-related changes in the immune system and competition for resources. A mechanism was also included to account for the multi-phasic nature of viral load decline [199]. The model was then successfully applied to two datasets for HIV-infected children starting ART. Covariate analysis showed that in both datasets, the initial concentration of CD4 T cells was inversely affected by age, as would be expected as CD4 concentrations are expected to fall through childhood and also older children have been HIV-infected for longer. In the ARROW dataset, lower thymic output and thus poor reconstitution was associated with the patients that died before the end of the trial, in agreement with another analysis of this data [172]. The model was extended to assess viral load rebound due to either the development of resistance or non-adherence to the ART. The mean time following

the start of ART that the virus rebounded was different for the four treatment centres. It would be interesting to look further into whether the adherence-promoting programs at the centres could explain these differences.

The effects of competition for resources on T cell clonotype dynamics was explored using the stochastic global competition model. The model was simple enough that not only could long run-times and large cell populations be simulated numerically but also the model was analytically tractable. This meant it was possible to produce equations for mean clonotype lifetimes, the mean long-term number of clonotypes and for the clonotype size distributions. These equations are then general and so can be scaled to systems larger than can be simulated numerically, such as a human. It is notable that the purely random dynamics can produce distributions of clonotype sizes. The model demonstrated that the shapes of the distributions are highly dependent on not just the quantity of T cells produced by the thymus, but also the number of T cells in each clonotype. The stochastic model was extended to allow clonotypes to have affinities for resource drawn randomly from a truncated normal distribution. Numerical simulations were then compared to CD8 memory T cell data from a bone marrow replacement experiment in mice, demonstrating that the model can be used to explain the observed complex effects whilst relying on only a small number of biological assumptions that would otherwise require complicated explanations.

In this work, mechanistic models have been applied to paediatric data for reconstitution following HSCT to our knowledge for the first time. This is also the first time a paediatric-specific model for HIV and CD4 concentration dynamics has been applied to data for HIV-infected children. This work has demonstrated that mechanistic models can be used to assess the fundamental drivers of the system. Furthermore, the deterministic mechanistic models have allowed the identification of factors that potentially would not otherwise have been identifiable across a whole dataset of diverse children, treatments and conditions.

6.1 Recommendations for developing mechanistic models of immune reconstitution

Mechanistic mathematical models aim to abstract the fundamental biology of the system in a mathematical context. As such, the first part of developing mechanistic models is to ascertain the biology that drives the system. Close collaboration is therefore required with immunologists and clinicians to understand the relevant biology. The biology is however complex and the aim is to apply these models to data. As such, the more the models can be simplified, the more likely it is to be possible to fit to data. One key to this is time-scales. Events that happen on similar time-scales need to be considered together, whereas events that happen on largely different time scales can be considered separately. In this modelling, this meant that age had to be accounted for in the context of CD4 reconstitution, but could be ignored for platelets and neutrophils. It also allowed the removal of the infected CD4 cells compartment from the HIV model.

Mechanistic elements can then be added into the model and tested for their necessity. Whilst the mechanistic elements that provide the highest quality of fit are the most appealing, an eye should be kept in the interpretability of the resulting parameters from the model. This maximises the possibility of further collaboration with clinicians and immunologists to interpret the results. Furthermore, there is a temptation to over-test the mechanisms in the model as a form of over-fitting, which may not improve the model's performance on a different dataset.

Whilst modelling assumptions can be used, often more specific data would add more information and allow better models to be developed. Proliferation and loss rates are both first-order interactions and so are mathematically difficult to separate. In this analysis, we have used a model for competition for resources to separate these parameters for CD4 reconstitution, which can work because

there are large changes in CD4 concentration. A better method however would be to have the proportion of cells expressing Ki67, a marker for proliferation, to split proliferation from loss. The model also relies on a sigmoidal function to model the impairment of thymic output of CD4 cells immediately post HSCT. Better would be to have data for the TREC content of the CD4 population that could give a proxy for the actual thymic output for each individual.

Fitting these nonlinear mixed effects models to data is a time consuming process, particularly during the process of model-development when many different iterations and small changes to the model need to be tested. Parameter estimation then becomes a balance between the time required to achieve parameter estimates that are accurate enough and the accuracy of the estimates that is required. Once a model has been developed, then more accurate parameter estimates can be sought. Parallelisation and the use of clusters can help to speed runtimes, but there is a cost involved with setting up the run and often with queueing to use the cluster.

6.2 Future work

The lasso or the least absolute shrinkage and selection separator needs further exploration for the performance of multivariate analysis. In this work, multivariate analysis was carried out using stepwise covariate model-building (SCM). SCM however has the capability to produce covariate models that have selected the wrong covariates [215], and requires the covariate model to be validated which can be computationally intensive [216]. Hence care must be taken with the results of SCM. Another option for multivariate analysis that has been applied in the nonlinear mixed-effects modelling context is the lasso [217–219]. In lasso, all covariates are normalised and included in the model at once. They are tested with a restriction on the total of the coefficients, which forces some coefficients

towards zero, leaving just significant covariates. This method was however problematic with these data. The method struggles when some covariates have strong effects such as were the effects of conditioning by alemtuzumab or ATG or no conditioning for the model of CD4 reconstitution following HSCT. Furthermore, the number of covariates that needed to be tested was high and involved the inclusion and estimation of very many parameters, which proved difficult. The use of lasso does however offer the chance to validate the covariate model.

Many developments are possible to the models. Potential developments to the short-term reconstitution following HSCT modelling include testing different numbers of transit compartments, different kinetic-pharmacodynamic models and different forms for the feedback loop. Many developments to the model for long-term reconstitution following HSCT have already been tested, including altered forms for the competition functions and the function delaying thymic recovery post HSCT. One further development could be modelling the whole time course including CD4 decline with conditioning for patients. This was not possible with the present data because there is not enough CD4 concentration data in the time period around the HSCT.

In this work, models have been fitted using maximum likelihood methods. For nonlinear mixed effects modelling these have an advantage in that there are many algorithms and much software that have been developed to perform parameter estimation. It would however be interesting to apply Bayesian methods to these models and data. Bayesian methods would have the advantage of producing distributions around parameter estimates that would be more realistic outputs from the model. They also allow the incorporation of priors, which could be particularly useful when applying these models to new data where parameter estimates and variance-covariance matrices could be used as priors. The requirement for priors can however lead to error propagation if the original priors were wrong and the data uninformative.

Having data on the concentrations of the drugs used in the interventions could improve the characteristics of the models. The drug effect identified in the short-term reconstitution following HSCT had a half-life that was substantially longer than the half lives of the drugs involved, implying that there is some model mis-specification. The most likely area for this mis-specification is the proliferating compartment. It is assumed that there is a resident population of proliferating cells, when in fact biologically, it will be a population in dynamic equilibrium with its bone marrow precursors. The drugs will also act to stop proliferation in these compartments, resulting in a delay to production in the proliferating compartment in the model. As such, an extension to the model could be to include more compartments before the proliferating compartment on which the drugs would be expected to act in tandem. To identify parameters surrounding these extra compartments, it would however be necessary to have higher resolution data than is currently available for drug concentrations and the times at which the drugs were given. This would enable the construction of a full pharmacokinetic-pharmacodynamic model that might be able to explain these effects and tease out the full relationship between the drug effects and the extra bone marrow compartments. Similarly, extra information on dosing and drug concentrations could be used to improve the accuracy and performance in the long-term reconstitution of CD4 T cells. By including more information on the mode of therapy, some of the inter-individual differences in the initial concentration of T cells might be explained.

Extra data could also help with the modelling of viral load rebound in HIV-infected children. For this work, the data only contained viral loads, while in future longitudinal data for HIV sequences will also be available. This information could be used to assess the susceptibility of different HIV strains to drugs, allowing the explanation and prediction of viral rebound. Furthermore, some clinical trials now monitor ART adherence with pill counters or more sophisticated methods

such as microchips in bottles or bottle caps that record usage. Adding this adherence data might also explain some of the viral rebounds.

Whilst predictions for individual patients have only been produced in one context in this work, it demonstrates that this type of mechanistic model can be used in this manner. With the method that has been developed here to make predictions, predictions become more accurate when larger datasets improve parameter estimates and make the identification of more factors possible. Furthermore the method allows for the predictions to be updated when new observations are taken for the individual patient, with confidence intervals narrowing as more information becomes available. With the increased use of electronic hospital records there is therefore the opportunity to produce predictions of reconstitution that could be updated immediately following a new observation, providing a graphical output that could be given to the clinician, thus providing a useful clinical tool. Predictions of reconstitution have the ability to help clinicians by giving them more information on what to expect in the months and years following the intervention or start of treatment and as such whether any changes to treatment protocols should be made. For short-term reconstitution following paediatric HSCT, this could include administration of GCSF, transfusion with extra platelets or a longer stay in isolation in hospital. For CD4 reconstitution, this could include prophylactic drugs or repeat HSCT. Predictions also allow the clinicians to receive early warnings of problems following the development of a new conditioning regimen.

More understanding is needed about changes to T cell receptor diversity following HSCT and in HIV-infected patients starting ART. For full functional reconstitution of CD4 T cells, not only does the number of T cells have to reconstitute but also the diversity that allows it to respond to so many pathogens in a specific manner. In the months following HSCT, there is little or no thymic output of new CD4 T cells. As a result, no new clonotypes can be produced after the transplant and the only T cell clonotypes in circulation are those that

were in the graft or survived the pre-HSCT conditioning regimen. Thus with no new clonotypes, the stochastic model of Chapter 5 predicts that the diversity can then be expected to fall over the ensuing months as clonotypes go extinct by randomness. Diversity could then be expected to increase as thymic output of new clonotypes increases. It would be very interesting to look at these effects using data for T cell receptor diversity to see whether the model's predictions are valid and if so to find parameter estimates. These estimates could then be used to predict how clonotype diversity is expected to recover as the concentration reconstitutes, giving estimates of the time to reach fully-functional diversity.

Models for diversity can be incorporated into the dynamics of CD4 T cell concentration reconstitution. CD4 T cells compete for both cytokines and sp-MHC. Cytokine competition is global and all T cells compete equally but sp-MHC competition is specific with different clonotypes competing for different sp-MHC. Intuitively, it can be imagined that when there are very few clonotypes, the cytokines will be in excess and the competition will mostly be intra-clonotype competition for sp-MHC whereas when there are many clonotypes, the competition is likely to be for a mixture of cytokines and sp-MHC. In situations where the diversity is changing it might therefore be possible to identify the different components of competition for sp-MHC and cytokines. Ciupe *et al* [28] attempt this with patients that have had thymus transplants for DiGeorge anomaly, using assays of T cell receptor vBeta family usage and CDR3 length as a measure of diversity. An attempt to perform a similar analysis using just vBeta data for long-term reconstitution following HSCT proved not to be possible because of the quality and quantity of data available. This is however a very important area for research and one in which there is much of interest if data for TCR diversity across time following HSCT were to become available. Furthermore, if the stochastic global competition model's results could be validated with data and parameter estimates found, it could be used to estimate diversity and thus

to allow the separation of the two forms of competition.

This work aimed to investigate reconstitution following paediatric HSCT and in HIV-infected children starting ART by using mechanistic mathematical modelling. By adapting previous work and constructing new mathematical models, I have explored both these problems, improving the understanding of the processes that govern reconstitution, finding factors that affect reconstitution and finally predicting reconstitution trajectories. This is the first time that some of these systems have been studied using these techniques and as such there is great scope for further investigation, both of the systems presented here and by applying these techniques to other systems.

References

[1] Liu, K., and M. C. Nussenzweig, 2010. Origin and development of dendritic cells. *Immunological Reviews* 234:45–54.

[2] Murphy, K., 2012. *Janeway's Immunobiology*. Garland Science.

[3] Yip, J., Y. Shen, M. C. Berndt, and R. K. Andrews, 2005. Primary platelet adhesion receptors. *IUBMB life* 57:103–108.

[4] Harker, L. A., L. K. Roskos, U. M. Marzec, R. A. Carter, J. K. Cherry, B. Sundell, E. N. Cheung, D. Terry, and W. Sheridan, 2000. Effects of megakaryocyte growth and development factor on platelet production, platelet life span, and platelet function in healthy human volunteers. *Blood* 95:2514–2522.

[5] Bainton, D. F., J. L. Ullyot, and M. G. Farquhar, 1971. The development of neutrophilic polymorphonuclear leukocytes in human bone marrow origin and content of azurophil and specific granules. *Journal of Experimental Medicine* 134:907–934.

[6] Kennedy, A. D., and F. R. DeLeo, 2009. Neutrophil apoptosis and the resolution of infection. *Immunologic Research* 43:25–61.

[7] Hickey, M. J., and P. Kubes, 2009. Intravascular immunity: the host–pathogen encounter in blood vessels. *Nature Reviews Immunology* 9:364–375.

[8] Athens, J., O. Haab, S. Raab, A. Mauer, H. Ashenbrucker, G. Cartwright, and M. Wintrobe, 1961. Leukokinetic studies. IV. The total blood, circulating and marginal granulocyte pools and the granulocyte turnover rate in normal subjects. *Journal of Clinical Investigation* 40:989.

[9] Tofts, P. S., T. Chevassut, M. Cutajar, N. G. Dowell, and A. M. Peters, 2011. Doubts concerning the recently reported human neutrophil lifespan of 5.4 days. *Blood* 117:6050–6052.

[10] Tak, T., K. Tesselaar, J. Pillay, J. A. Borghans, and L. Koenderman, 2013. What's your age again? Determination of human neutrophil half-lives revisited. *Journal of Leukocyte Biology* 94:595–601.

[11] Cox, G., 1995. Glucocorticoid treatment inhibits apoptosis in human neutrophils. Separation of survival and activation outcomes. *Journal of Immunology* 154:4719–4725.

[12] Yun, T. J., and M. J. Bevan, 2001. The Goldilocks conditions applied to T cell development. *Nature Immunology* 2:13–14.

[13] Trepel, F., 1974. Number and distribution of lymphocytes in man. A critical analysis. *Klinische Wochenschrift* 52:511–515.

[14] Vivien, L., C. Benoist, and D. Mathis, 2001. T lymphocytes need IL-7 but not IL-4 or IL-6 to survive in vivo. *International Immunology* 13:763–768.

[15] Kondrack, R. M., J. Harbertson, J. T. Tan, M. E. McBreen, C. D. Surh, and L. M. Bradley, 2003. Interleukin 7 regulates the survival and generation of memory CD4 cells. *Journal of Experimental Medicine* 198:1797–1806.

[16] Martin, B., C. Bécourt, B. Bienvenu, and B. Lucas, 2006. Self-recognition is crucial for maintaining the peripheral CD4+ T-cell pool in a nonlymphopenic environment. *Blood* 108:270–277.

[17] Reynolds, J., M. Coles, G. Lythe, and C. Molina-Paris, 2013. Mathematical model of

naive T cell division and survival IL-7 thresholds. *Frontiers in Immunology* 4:434.

[18] Pearson, C., A. Silva, M. Saini, and B. Seddon, 2011. IL-7 determines the homeostatic fitness of T cells by distinct mechanisms at different signalling thresholds in vivo. *European Journal of Immunology* 41:3656–3666.

[19] Surh, C. D., and J. Sprent, 2008. Homeostasis of naive and memory T cells. *Immunity* 29:848–862.

[20] Yates, A., M. Saini, A. Mathiot, and B. Seddon, 2008. Mathematical modeling reveals the biological program regulating lymphopenia-induced proliferation. *Journal of Immunology* 180:1414–1422.

[21] De Boer, R. J., A. S. Perelson, and R. M. Ribeiro, 2012. Modelling deuterium labelling of lymphocytes with temporal and/or kinetic heterogeneity. *Journal of the Royal Society Interface* 9:2191–2200.

[22] De Boer, R. J., and A. S. Perelson, 2013. Quantifying T lymphocyte turnover. *Journal of Theoretical Biology* 327:45–87.

[23] Palmer, M. J., V. S. Mahajan, J. Chen, D. J. Irvine, and D. A. Lauffenburger, 2011. Signaling thresholds govern heterogeneity in IL-7-receptor-mediated responses of naïve CD8+ T cells. *Immunology Cell Biology* 89:581–594.

[24] Seddon, B., P. Tomlinson, and R. Zamoyska, 2003. Interleukin 7 and T cell receptor signals regulate homeostasis of CD4 memory cells. *Nature Immunology* 4:680–686.

[25] Seddon, B., and R. Zamoyska, 2003. Regulation of peripheral T-cell homeostasis by receptor signalling. *Current Opinion in Immunology* 15:321–324.

[26] Tan, J. T., B. Ernst, W. C. Kieper, E. LeRoy, J. Sprent, and C. D. Surh, 2002. Interleukin (IL)-15 and IL-7 jointly regulate homeostatic proliferation of memory phenotype CD8+ cells but are not required for memory phenotype CD4+ cells. *Journal of Experimental Medicine* 195:1523–1532.

[27] Mahajan, V. S., I. B. Leskov, and J. Z. Chen, 2005. Homeostasis of T cell diversity. *Cellular Molecular Immunology* 2:1–10.

[28] Ciupe, S. M., B. H. Devlin, M. L. Markert, and T. B. Kepler, 2009. The dynamics of T-cell receptor repertoire diversity following thymus transplantation for DiGeorge anomaly. *PLoS Computational Biology* 5:e1000 396.

[29] Troy, A. E., and H. Shen, 2003. Cutting edge: Homeostatic proliferation of peripheral T lymphocytes is regulated by clonal competition. *Journal of Immunology* 170:672–676.

[30] Hao, Y., N. Legrand, and A. A. Freitas, 2006. The clone size of peripheral CD8 T cells is regulated by TCR promiscuity. *Journal of Experimental Medicine* 203:1643–1649.

[31] Huenecke, S., M. Behl, C. Fadler, S. Y. Zimmermann, K. Bochennek, L. Tramsen, R. Esser, D. Klarmann, M. Kamper, A. Sattler, et al., 2008. Age-matched lymphocyte subpopulation reference values in childhood and adolescence: application of exponential regression analysis. *European Journal of Haematology* 80:532–539.

[32] Steinmann, G. G., B. Klaus, and H. K. Müller-Hermelink, 1985. The involution of the ageing human thymic epithelium is independent of puberty. *Scandinavian Journal of Immunology* 22:563–575.

[33] Bains, I., R. Thiébaut, A. J. Yates, and R. Callard, 2009. Quantifying thymic export: combining models of naive T cell proliferation and TCR excision circle dynamics gives an explicit measure of thymic output. *Journal of Immunology* 183:4329–4336.

[34] Douek, D. C., R. D. McFarland, P. H. Keiser, E. A. Gage, J. M. Massey, B. F. Haynes, M. A. Polis, A. T. Haase, M. B. Feinberg, J. L. Sullivan, et al., 1998. Changes in thymic function with age and during the treatment of HIV infection. *Nature* 396:690–695.

[35] Rufer, N., T. H. Brümmendorf, S. Kolvraa, C. Bischoff, K. Christensen, L. Wadsworth, M. Schulzer, and P. M. Lansdorp, 1999. Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. *Journal of Experimental Medicine* 190:157–168.

[36] Bains, I., R. Antia, R. Callard, and A. J. Yates, 2009. Quantifying the development of the peripheral naive CD4+ T-cell pool in humans. *Blood* 113:5480–5487.

[37] Barker, C. I., E. Germovsek, R. L. Hoare, J. M. Lestner, J. Lewis, and J. F. Standing, 2014. Pharmacokinetic/pharmacodynamic modelling approaches in paediatric infectious diseases and immunology. *Advanced Drug Delivery Reviews* 73:127–139.

[38] Gratwohl, A., H. Baldomero, M. Aljurf, M. C. Pasquini, L. F. Bouzas, A. Yoshimi, J. Szer, J. Lipton, A. Schwendener, M. Gratwohl, et al., 2010. Hematopoietic stem cell transplantation: A global perspective. *JAMA* 303:1617–1624.

[39] British Society for Blood and Marrow Transplantation, 2013. UK and ROI transplant activity, accessed 31/07/15. *http://bsbmt.org/2013-activity/* .

[40] Miano, M., M. Labopin, O. Hartmann, E. Angelucci, J. Cornish, E. Gluckman, F. Locatelli, A. Fischer, R. Egeler, R. Or, et al., 2007. Haematopoietic stem cell transplantation trends in children over the last three decades: a survey by the paediatric diseases working party of the European Group for Blood and Marrow Transplantation. *Bone Marrow Transplant* 39:89–99.

[41] Almyroudis, N., A. Fuller, A. Jakubowski, K. Sepkowitz, D. Jaffe, T. Small, T. Kiehn, E. Pamer, and G. Papanicolaou, 2005. Pre-and post-engraftment bloodstream infection rates and associated mortality in allogeneic hematopoietic stem cell transplant recipients. *Transplant Infectious Disease* 7:11–17.

[42] Massaro, K., R. Macedo, B. de Castro, F. Dulley, M. Oliveira, M. Yasuda, A. Levin, and S. Costa, 2014. Risk factor for death in hematopoietic stem cell transplantation: are biomarkers useful to foresee the prognosis in this population of patients? *Infection* 42:1023–1032.

[43] Nash, R. A., T. Gooley, C. Davis, and F. R. Appelbaum, 1996. The problem of thrombocytopenia after hematopoietic stem cell transplantation. *Oncologist* 1:371–380.

[44] Han, T., L. Xu, D. Liu, K. Liu, F. Wang, Y. Wang, C. Yan, Y. Chen, Y. Sun, Y. Ji, et al., 2015. Recombinant human thrombopoietin promotes platelet engraftment after haploidentical hematopoietic stem cell transplantation: a prospective randomized controlled trial. *Annals of Hematology* 94:117–128.

[45] Ramirez, P., C. Brunstein, B. Miller, T. Defor, and D. Weisdorf, 2011. Delayed platelet recovery after allogeneic transplantation: a predictor of increased treatment-related mortality and poorer survival. *Bone Marrow Transplant* 46:981–986.

[46] Friberg, L. E., and M. O. Karlsson, 2003. Mechanistic models for myelosuppression. *Investigational New Drugs* 21:183–194.

[47] Bodey, G. P., M. Buckley, Y. Sathe, and E. J. Freireich, 1966. Quantitative relationships between circulating leukocytes and infection in patients with acute leukemia. *Annals of Internal Medicine* 64:328–340.

[48] Moore, M., and J. Theissen, 1992. *Cytotoxics and irreversible effects. In:C.J. van Boxtel, N.H.G. Holford, M. Danhof M (eds) The in vivo study of drug action. pp 337–400*, volume 377. Elsevier Science Publisher BV, Amsterdam.

[49] Crawford, J., H. Ozer, R. Stoller, D. Johnson, G. Lyman, I. Tabbara, M. Kris, J. Grous, V. Picozzi, G. Rausch, et al., 1991. Reduction by granulocyte colony-stimulating factor of fever and neutropenia induced by chemotherapy in patients with small-cell lung cancer. *New England Journal of Medicine* 325:164–170.

[50] Rosner, G. L., and P. Müller, 1994. Pharmacodynamic analysis of hematologic profiles. *Journal of Pharmacokinetics and Biopharmaceutics* 22:499–524.

[51] Karlsson, M. O., V. Molnar, J. Bergh, A. Freijs, and R. Larsson, 1998. A general model for time-dissociated pharmacokinetic-pharmacodynamic relationships exemplified by paclitaxel myelosuppression. *Clinical Pharmacology & Therapeutics* 63:11–25.

[52] Friberg, L. E., A. Henningsson, H. Maas, L. Nguyen, and M. O. Karlsson, 2002. Model of chemotherapy-induced myelosuppression with parameter consistency across drugs. *Journal of Clinical Oncology* 20:4713–4721.

[53] Minami, H., Y. Sasaki, N. Saijo, T. Ohtsu, H. Fujii, T. Igarashi, and K. Itoh, 1998. Indirect-response model for the time course of leukopenia with anticancer drugs. *Clinical Pharmacology & Therapeutics* 64:511–521.

[54] Friberg, L., C. Brindley, M. Karlsson, and A. Devlin, 2000. Models of schedule dependent haematological toxicity of 2′-deoxy-2′-methylidenecytidine (DMDC). *European Journal of Clinical Pharmacology* 56:567–574.

[55] Friberg, L. E., A. Freijs, M. Sandström, and M. O. Karlsson, 2000. Semiphysiological model for the time course of leukocytes after varying schedules of 5-fluorouracil in rats. *Journal of Pharmacology and Experimental Therapeutics* 295:734–740.

[56] Zamboni, W. C., D. Z. D'Argenio, C. F. Stewart, T. MacVittie, B. J. Delauter, A. M. Farese, D. M. Potter, N. M. Kubat, D. Tubergen, and M. J. Egorin, 2001. Pharmacodynamic model of topotecan-induced time course of neutropenia. *Clinical Cancer Research* 7:2301–2308.

[57] Panetta, J. C., M. N. Kirstein, A. Gajjar, G. Nair, M. Fouladi, and C. F. Stewart, 2003. A mechanistic mathematical model of temozolomide myelosuppression in children with high-grade gliomas. *Mathematical Biosciences* 186:29–41.

[58] Bulitta, J. B., P. Zhao, R. D. Arnold, D. R. Kessler, R. Daifuku, J. Pratt, G. Luciano, A.-R. Hanauske, H. Gelderblom, A. Awada, et al., 2009. Multiple-pool cell lifespan models for neutropenia to assess the population pharmacodynamics of unbound paclitaxel from two formulations in cancer patients. *Cancer Chemotherapy and Pharmacology* 63:1035–1048.

[59] Soto, E., A. Staab, C. Doege, M. Freiwald, G. Munzert, and I. F. Trocóniz, 2011. Comparison of different semi-mechanistic models for chemotherapy-related neutropenia: application to BI 2536 a Plk-1 inhibitor. *Cancer Chemotherapy and Pharmacology* 68:1517–1527.

[60] Cream, J., 1968. Prednisolone-induced granulocytosis. *British Journal of Haematology* 15:259–268.

[61] Mishler, J. M., and P. M. Emerson, 1977. Development of neutrophilia by serially increasing doses of dexamethasone. *British Journal of Haematology* 36:249–257.

[62] Denison, F. C., C. L. Elliott, and E. M. Wallace, 1997. Dexamethasone-induced leucocytosis in pregnancy. *BJOG: An International Journal of Obstetrics & Gynaecology* 104:851–853.

[63] John, T. J., 1966. Leukocytosis during steroid therapy. *American Journal of Diseases of Children* 111:68–70.

[64] Ozawa, K., H. Minami, and H. Sato, 2007. Population pharmacokinetic and pharmacodynamic analysis for time courses of docetaxel-induced neutropenia in Japanese cancer patients. *Cancer Science* 98:1985–1992.

[65] Soto, E., A. Staab, M. Freiwald, G. Munzert, H. Fritsch, C. Döge, and I. Troconiz, 2010. Prediction of neutropenia-related effects of a new combination therapy with the anticancer drugs BI 2536 (a Plk1 inhibitor) and pemetrexed. *Clinical Pharmacology & Therapeutics* 88:660–667.

[66] Quartino, A. L., M. O. Karlsson, H. Lindman, and L. E. Friberg, 2014. Characterization of endogenous G-CSF and the inverse correlation to chemotherapy-induced neutropenia in patients with breast cancer using population modeling. *Pharmaceutical Research* 31:3390–3403.

[67] van Kesteren, C., A. S. Zandvliet, M. O. Karlsson, R. A. Mathôt, C. J. Punt, J.-P. Armand, E. Raymond, A. D. Huitema, C. Dittrich, H. Dumez, et al., 2005. Semiphysiological model describing the hematological toxicity of the anti-cancer agent indisulam. *Investigational New Drugs* 23:225–234.

[68] Joerger, M., A. D. Huitema, D. J. Richel, C. Dittrich, N. Pavlidis, E. Briasoulis, J. B. Vermorken, E. Strocchi, A. Martoni, R. Sorio, et al., 2007. Population pharmacokinetics and pharmacodynamics of paclitaxel and carboplatin in ovarian cancer patients: a study by the European organization for research and treatment of cancer-pharmacology and molecular mechanisms group and new drug development group. *Clinical Cancer Research*

13:6410–6418.

[69] Schmitt, A., L. Gladieff, C. M. Laffont, A. Evrard, J.-C. Boyer, A. Lansiaux, C. Bobin-Dubigeon, M.-C. Etienne-Grimaldi, M. Boisdron-Celle, M. Mousseau, et al., 2010. Factors for hematopoietic toxicity of carboplatin: refining the targeting of carboplatin systemic exposure. *Journal of Clinical Oncology* 28:4568–4574.

[70] Bender, B. C., F. Schaedeli-Stark, R. Koch, A. Joshi, Y.-W. Chu, H. Rugo, I. E. Krop, S. Girish, L. E. Friberg, and M. Gupta, 2012. A population pharmacokinetic/pharmacodynamic model of thrombocytopenia characterizing the effect of trastuzumab emtansine (T-DM1) on platelet counts in patients with HER2-positive metastatic breast cancer. *Cancer Chemotherapy and Pharmacology* 70:591–601.

[71] Kaefer, A., J. Yang, P. Noertersheuser, S. Mensing, R. Humerickhouse, W. Awni, and H. Xiong, 2014. Mechanism-based pharmacokinetic/pharmacodynamic meta-analysis of navitoclax (ABT-263) induced thrombocytopenia. *Cancer Chemotherapy and Pharmacology* 74:593–602.

[72] Hayes, S., D. Ouellet, J. Zhang, M. B. Wire, and E. Gibiansky, 2011. Population PK/PD modeling of eltrombopag in healthy volunteers and patients with immune thrombocytopenic purpura and optimization of response-guided dosing. *Journal of Clinical Pharmacology* 51:1403–1417.

[73] Hayes, S., P. N. Mudd Jr, D. Ouellet, B. M. Johnson, D. Williams, and E. Gibiansky, 2013. Population PK/PD modeling of eltrombopag in subjects with advanced solid tumors with chemotherapy-induced thrombocytopenia. *Cancer Chemotherapy and Pharmacology* 71:1507–1520.

[74] Farrell, C., S. C. Hayes, M. Wire, and J. Zhang, 2014. Population pharmacokinetic/pharmacodynamic modelling of eltrombopag in healthy volunteers and subjects with chronic liver disease. *British Journal of Clinical Pharmacology* 77:532–544.

[75] Zhang, J., M. Thapar, C. Farrell, and M. B. Wire, 2014. Modeling and simulation support eltrombopag dosing in thrombocytopenic patients with chronic HCV infection. *Pharmaceutical Research* pages 1–14.

[76] Wu, K., M. Thapar, C. Farrell, S. Hayes, H. Guo, M. Hou, and J. Zhang, 2015. Population pharmacokinetic and pharmacodynamic modeling and effects on platelet counts of different dosages of eltrombopag in Chinese patients with chronic primary immune thrombocytopenia. *Clinical Therapeutics* 37:1382–1395.

[77] Upton, R., and D. Mould, 2014. Basic concepts in population modeling, simulation, and model-based drug development: Part 3 — Introduction to pharmacodynamic modeling methods. *CPT: Pharmacometrics & Systems Pharmacology* 3:1–16.

[78] Jacqmin, P., R. Gieschke, P. Jordan, J.-L. Steimer, T. Goggin, G. Pillai, E. Snoeck, and P. Girard, 2001. Modeling drug induced changes in biomarkers without using drug concentrations: introducing the K-PD model. *10th Population Approach Group Conference, Basel, Switzerland, http://www. page-meeting. org.*

[79] Vassal, G., A. Gouyette, O. Hartmann, J. Pico, and J. Lemerle, 1989. Pharmacokinetics of high-dose busulfan in children. *Cancer Chemotherapy and Pharmacology* 24:386–390.

[80] De Jonge, M. E., A. D. Huitema, S. Rodenhuis, and J. H. Beijnen, 2005. Clinical pharmacokinetics of cyclophosphamide. *Clinical Pharmacokinetics* 44:1135–1164.

[81] Gandhi, V., and W. Plunkett, 2002. Cellular and clinical pharmacology of fludarabine. *Clinical Pharmacokinetics* 41:93–103.

[82] Gouyette, A., O. Hartmann, and J.-L. Pico, 1986. Pharmacokinetics of high-dose melphalan in children and adults. *Cancer Chemotherapy and Pharmacology* 16:184–189.

[83] Główka, F., M. Karaźniewicz-Łada, G. Grund, T. Wrobel, and J. Wachowiak, 2008. Pharmacokinetics of high-dose iv treosulfan in children undergoing treosulfan-based preparative regimen for allogeneic haematopoietic sct. *Bone Marrow Transplant* 42:S67–S70.

[84] Pinheiro, J. C., and D. M. Bates, 2000. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

[85] Sheiner, L. B., and S. L. Beal, 1983. Evaluation of methods for estimating population pharmacokinetic parameters. III. Monoexponential model: Routine clinical pharmacokinetic data. *Journal of Pharmacometrics and Biopharmaceutics* 11:303–319.

[86] Jonsson, E. N., and M. O. Karlsson, 1998. Automated covariate model building within NONMEM. *Pharmaceutical Research* 15:1463–1468.

[87] Beal, S. L., 2001. Ways to fit a PK model with some data below the quantification limit. *Journal of Pharmacokinetics and Pharmacodynamics* 28:481–504.

[88] Ahn, J. E., M. O. Karlsson, A. Dunne, and T. M. Ludden, 2008. Likelihood based approaches to handling data below the quantification limit using NONMEM VI. *Journal of Pharmacokinetics and Pharmacodynamics* 35:401–421.

[89] Bergstrand, M., and M. O. Karlsson, 2009. Handling data below the limit of quantification in mixed effect models. *AAPS Journal* 11:371–380.

[90] Bergstrand, M., E. Plan, M. Kjellsson, and M. O. Karlsson, 2007. A comparison of methods for handling of data below the limit of quantification in NONMEM VI. *Annual Meeting of the Population Approach Group in Europe*, page 16.

[91] Karlsson, M., and R. Savic, 2007. Diagnosing model diagnostics. *Clinical Pharmacology & Therapeutics* 82:17–20.

[92] Hooker, A. C., C. E. Staatz, and M. O. Karlsson, 2007. Conditional weighted residuals (CWRES): a model diagnostic for the FOCE method. *Pharmaceutical Research* 24:2187–2197.

[93] Holford, N., 2005. The visual predictive check—superiority to standard diagnostic (Rorschach) plots. *Abstr*, volume 738, page 14.

[94] Karlsson, M. O., and N. Holford, 2008. A tutorial on visual predictive checks. *17th meeting of the Population Approach Group in Europe, Marseille, France, page Abstr*, volume 1434.

[95] Bergstrand, M., A. C. Hooker, J. E. Wallin, and M. O. Karlsson, 2011. Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. *AAPS Journal* 13:143–151.

[96] Von Schulthess, G., and U. Gessner, 1986. Oscillating platelet counts in healthy individuals: experimental investigation and quantitative evaluation of thrombocytopoietic feedback control. *Scandinavian Journal of Haematology* 36:473–479.

[97] Hirayama, Y., S. Sakamaki, Y. Tsuji, T. Matsunaga, and Y. Niitsu, 2003. Cyclic platelet and leukocyte count oscillation in chronic myelocytic leukemia regulated by the negative feedback of transforming growth factor β. *International Journal of Hematology* 77:71–74.

[98] Segel, G. B., and J. S. Halterman, 2008. Neutropenia in pediatric practice. *Pediatrics in Review* 29:12.

[99] Soldin, S. J., C. Brugnara, and E. C. Wong, 2003. *Pediatric reference ranges*. Amer. Assoc. for Clinical Chemistry.

[100] Charrier, E., P. Cordeiro, R.-M. Brito, S. Mezziani, S. Herblot, F. Le Deist, and M. Duval, 2013. Reconstitution of maturating and regulatory lymphocyte subsets after cord blood and BMT in children. *Bone Marrow Transplant* 48:376–382.

[101] Bartelink, I. H., S. V. Belitser, C. A. J. Knibbe, M. Danhof, A. J. de Pagter, T. C. G. Egberts, and J. J. Boelens, 2013. Immune reconstitution kinetics as an early predictor for mortality using various hematopoietic stem cell sources in children. *Biology of Blood and Marrow Transplantation* 19:305–313.

[102] Berger, M., O. Figari, B. Bruno, A. Raiola, A. Dominietto, M. Fiorone, M. Podesta, E. Tedone, S. Pozzi, F. Fagioli, et al., 2008. Lymphocyte subsets recovery following allogeneic bone marrow transplantation (BMT): CD4+ cell count and transplant-related mortality. *Bone Marrow Transplant* 41:55–62.

[103] Schwinger, W., D. Weber-Mzell, B. Zois, T. Rojacher, M. Benesch, H. Lackner, H. J. Dornbusch, P. Sovinz, A. Moser, G. Lanzer, et al., 2006. Immune reconstitution after

purified autologous and allogeneic blood stem cell transplantation compared with unmanipulated bone marrow transplantation in children. *British Journal of Haematology* 135:76–84.

[104] Kim, H. O., H. J. Oh, J. W. Lee, P. Jang, N. Chung, B. Cho, and H. Kim, 2013. Immune reconstitution after allogeneic hematopoietic stem cell transplantation in children: a single institution study of 59 patients. *Korean Journal of Pediatrics* 56:26–31.

[105] Chen, X., G. A. Hale, R. Barfield, E. Benaim, W. H. Leung, J. Knowles, E. M. Horwitz, P. Woodard, K. Kasow, U. Yusuf, et al., 2006. Rapid immune reconstitution after a reduced-intensity conditioning regimen and a CD3-depleted haploidentical stem cell graft for paediatric refractory haematological malignancies. *British Journal of Haematology* 135:524–532.

[106] Fedele, R., M. Martino, C. Garreffa, G. Messina, G. Console, D. Princi, A. Dattola, T. Moscato, E. Massara, E. Spiniello, et al., 2012. The impact of early CD4+ lymphocyte recovery on the outcome of patients who undergo allogeneic bone marrow or peripheral blood stem cell transplantation. *Blood Transfusions* 10:174–180.

[107] Tedeschi, S. K., M. Jagasia, B. G. Engelhardt, J. Domm, A. A. Kassim, W. Chinratanalab, S. L. Greenhut, S. Goodman, J. P. Greer, F. Schuening, et al., 2011. Early lymphocyte reconstitution is associated with improved transplant outcome after cord blood transplantation. *Cytotherapy* 13:78–82.

[108] Koenig, M., S. Huenecke, E. Salzmann-Manrique, R. Esser, R. Quaritsch, D. Steinhilber, H. H. Radeke, H. Martin, P. Bader, T. Klingebiel, et al., 2010. Multivariate analyses of immune reconstitution in children after allo-SCT: risk-estimation based on age-matched leukocyte sub-populations. *Bone Marrow Transplant* 45:613–621.

[109] Rénard, C., V. Barlogis, V. Mialou, C. Galambrun, D. Bernoux, M. P. Goutagny, L. Glasman, A. D. Loundou, F. Poitevin-Later, F. Dignat-George, et al., 2011. Lymphocyte subset reconstitution after unrelated cord blood or bone marrow transplantation in children. *British Journal of Haematology* 152:322–330.

[110] Koehl, U., K. Bochennek, S. Y. Zimmermann, T. Lehrnbecher, J. Sörensen, R. Esser, C. Andreas, C. Kramm, H. P. Grüttner, E. Falkenberg, et al., 2007. Immune recovery in children undergoing allogeneic stem cell transplantation: absolute CD8+ CD3+ count reconstitution is associated with survival. *Bone Marrow Transplant* 39:269–278.

[111] Barlogis, V., L. Glasman, C. Brunet, A. D. Loundou, C. Lemarie, C. Galambrun, I. Thuret, C. Curtillet, M. Le Meignen, F. Bernard, et al., 2011. Impact of viable CD45 cells infused on lymphocyte subset recovery after unrelated cord blood transplantation in children. *Biology of Blood and Marrow Transplantation* 17:109–116.

[112] Duval, M., B. Pédron, P. Rohrlich, F. Legrand, A. Faye, B. Lescoeur, P. Bensaid, R. Larchee, G. Sterkers, and E. Vilmer, 2002. Immune reconstitution after haematopoietic transplantation with two different doses of pre-graft antithymocyte globulin. *Bone Marrow Transplant* 30:421–426.

[113] Giannelli, R., M. Bulleri, M. Menconi, B. Casazza, D. Focosi, S. Bernasconi, and C. Favre, 2012. Reconstitution rate of absolute CD8+ T lymphocyte counts affects overall survival after pediatric allogeneic hematopoietic stem cell transplantation. *Journal of Pediatric Hematology and Oncology* 34:29–34.

[114] Wade, A., and A. Ades, 1994. Age-related reference ranges: significance tests for models and confidence intervals for centiles. *Statistics in Medicine* 13:2359–2367.

[115] van Gent, R., A. W. Schadenberg, S. A. Otto, R. A. Nievelstein, G. T. Sieswerda, F. Haas, F. Miedema, K. Tesselaar, N. J. Jansen, and J. A. Borghans, 2011. Long-term restoration of the human T-cell compartment after thymectomy during infancy: a role for thymic regeneration? *Blood* 118:627–634.

[116] Hapuarachchi, T., J. Lewis, and R. E. Callard, 2013. A mechanistic model for naive CD4 T cell homeostasis in healthy adults and children. *Frontiers in Immunology* 4:366.

[117] Fallen, P. R., L. McGreavey, J. A. Madrigal, M. Potter, M. Ethell, H. G. Prentice,

A. Guimarães, and P. J. Travers, 2003. Factors affecting reconstitution of the T cell compartment in allogeneic haematopoietic cell transplant recipients. *Bone Marrow Transplant* 32:1001–1014.

[118] Douek, D. C., R. A. Vescio, M. R. Betts, J. M. Brenchley, B. J. Hill, L. Zhang, J. R. Berenson, R. H. Collins, and R. A. Koup, 2000. Assessment of thymic output in adults after haematopoietic stem-cell transplantation and prediction of T-cell reconstitution. *Lancet* 355:1875–1881.

[119] Clave, E., D. Lisini, C. Douay, G. Giorgiani, M. Busson, M. Zecca, F. Moretta, G. Acquafredda, L. P. Brescia, F. Locatelli, et al., 2013. Thymic function recovery after unrelated donor cord blood or T-cell depleted HLA-haploidentical stem cell transplantation correlates with leukemia relapse. *Frontiers in Immunology* 4:54.

[120] Soetaert, K., and T. Petzoldt, 2010. Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME. *Journal of Statistical Software* 33:1–28.

[121] R Development Core Team, 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[122] Omlin, M., R. Brun, and P. Reichert, 2001. Biogeochemical model of Lake Zürich: sensitivity, identifiability and uncertainty analysis. *Ecological Modelling* 141:105–123.

[123] Beal, S., L. B. Sheiner, A. Boeckmann, and R. J. Bauer, 2013. NONMEM User's Guides. (1989-2013). *Icon Development Solutions, Ellicott City, MD, USA* .

[124] Lindbom, L., P. Pihlgren, E. N. Jonsson, and N. Jonsson, 2005. PsN-Toolkit – a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Computational Methods and Programs in Biomedicine* 79:241–257.

[125] Booth, C., and P. Veys, 2013. T cell depletion in paediatric stem cell transplantation. *Clinical Experimental Immunology* 172:139–147.

[126] Hellerstein, M. K., R. A. Hoh, M. B. Hanley, D. Cesar, D. Lee, R. A. Neese, J. M. McCune, et al., 2003. Subpopulations of long-lived and short-lived T cells in advanced HIV-1 infection. *Journal of Clinical Investigation* 112:956–966.

[127] Vrisekoop, N., I. den Braber, A. B. de Boer, A. F. Ruiter, M. T. Ackermans, S. N. van der Crabben, E. H. Schrijver, G. Spierenburg, H. P. Sauerwein, M. D. Hazenberg, et al., 2008. Sparse production but preferential incorporation of recently produced naive T cells in the human peripheral pool. *Proceedings of the National Academy of Sciences* 105:6115–6120.

[128] Mohri, H., A. S. Perelson, K. Tung, R. M. Ribeiro, B. Ramratnam, M. Markowitz, R. Kost, L. Weinberger, D. Cesar, M. K. Hellerstein, et al., 2001. Increased turnover of T lymphocytes in HIV-1 infection and its reduction by antiretroviral therapy. *Journal of Experimental Medicine* 194:1277–1288.

[129] Rebello, P., K. Cwynarski, M. Varughese, A. Eades, J. Apperley, and G. Hale, 2001. Pharmacokinetics of CAMPATH-1H in BMT patients. *Cytotherapy* 3:261–267.

[130] Bunn, D., C. Lea, D. Bevan, R. Higgins, and B. Hendry, 1996. The pharmacokinetics of anti-thymocyte globulin (ATG) following intravenous infusion in man. *Clinical Nephrology* 45:29–32.

[131] Lane, J. P., P. T. Evans, Z. Nademi, D. Barge, A. Jackson, S. Hambleton, T. J. Flood, A. J. Cant, M. Abinun, M. A. Slatter, et al., 2014. Low-dose serotherapy improves early immune reconstitution after cord blood transplantation for primary immunodeficiencies. *Biology of Blood and Marrow Transplantation* 20:243–249.

[132] Ballen, K. K., 2014. ATG for cord blood transplant: yes or no? *Blood* 123:7–8.

[133] Lindemans, C. A., R. Chiesa, P. J. Amrolia, K. Rao, O. Nikolajeva, A. de Wildt, C. E. Gerhardt, K. C. Gilmour, M. B. Bierings, P. Veys, et al., 2014. Impact of thymoglobulin prior to pediatric unrelated umbilical cord blood transplantation on immune reconstitution and clinical outcome. *Blood* 123:126–132.

[134] Shah, A. J., N. Kapoor, G. M. Crooks, K. I. Weinberg, H. A. Azim, R. Killen, L. Kuo, T. Rushing, D. B. Kohn, and R. Parkman, 2007. The effects of Campath 1H upon graft-

versus-host disease, infection, relapse, and immune reconstitution in recipients of pediatric unrelated transplants. *Biology of Blood and Marrow Transplantation* 13:584–593.

[135] Hale, G., and H. Waldmann, 1994. Control of graft-versus-host disease and graft rejection by T cell depletion of donor and recipient with Campath-1 antibodies. Results of matched sibling transplants for malignant diseases. *Bone Marrow Transplant* 13:597–611.

[136] Kottaridis, P., D. W. Milligan, R. Chopra, R. Chakraverty, S. Chakrabarti, S. Robinson, K. Peggs, S. Verfuerth, R. Pettengell, J. C. Marsh, et al., 2001. In vivo Campath-1H prevents GvHD following nonmyeloablative stem-cell transplantation. *Cytotherapy* 3:197–201.

[137] Chakraverty, R., K. Peggs, R. Chopra, D. W. Milligan, P. D. Kottaridis, S. Verfuerth, J. Geary, D. Thuraisundaram, K. Branson, S. Chakrabarti, et al., 2002. Limiting transplantation-related mortality following unrelated donor stem cell transplantation by using a nonmyeloablative conditioning regimen. *Blood* 99:1071–1078.

[138] Hale, G., P. Jacobs, L. Wood, W. Fibbe, R. Barge, N. Novitzky, C. d. Toit, L. Abrahams, V. Thomas, D. Bunjes, et al., 2000. CD52 antibodies for prevention of graft-versus-host disease and graft rejection following transplantation of allogeneic peripheral blood stem cells. *Bone Marrow Transplant* 26:69–76.

[139] Fernandes, J. F., V. Rocha, M. Labopin, B. Neven, D. Moshous, A. R. Gennery, W. Friedrich, F. Porta, C. D. de Heredia, D. Wall, et al., 2012. Transplantation in patients with SCID: mismatched related stem cells or unrelated cord blood? *Blood* 119:2949–2955.

[140] Chiesa, R., K. Gilmour, W. Qasim, S. Adams, A. J. Worth, H. Zhan, C. A. Montiel-Equihua, S. Derniame, C. Cale, K. Rao, et al., 2012. Omission of in vivo T-cell depletion promotes rapid expansion of naïve CD4+ cord blood lymphocytes and restores adaptive immunity within 2 months after unrelated cord blood transplant. *British Journal of Haematology* 156:656–666.

[141] McCune, J. M., 2001. The dynamics of CD4+ T-cell depletion in HIV disease. *Nature* 410:974–979.

[142] De Rossi, A., A. S. Walker, N. Klein, D. De Forni, D. King, and D. M. Gibb, 2002. Increased thymic output after initiation of antiretroviral therapy in human immunodeficiency virus type 1–infected children in the Paediatric European Network for Treatment of AIDS (PENTA) 5 trial. *Journal of Infectious Diseases* 186:312–320.

[143] Douek, D. C., R. A. Koup, R. D. McFarland, J. L. Sullivan, and K. Luzuriaga, 2000. Effect of HIV on thymic function before and after antiretroviral therapy in children. *Journal of Infectious Diseases* 181:1479–1482.

[144] Ometto, L., D. De Forni, F. Patiri, V. Trouplin, F. Mammano, V. Giacomet, C. Giaquinto, D. Douek, R. Koup, and A. De Rossi, 2002. Immune reconstitution in HIV-1-infected children on antiretroviral therapy: role of thymic output and viral fitness. *AIDS* 16:839–849.

[145] Sandgaard, K. S., J. Lewis, S. Adams, N. Klein, and R. Callard, 2014. Antiretroviral therapy increases thymic output in children with HIV. *AIDS* 28:209–214.

[146] Hazenberg, M. D., D. Hamann, H. Schuitemaker, and F. Miedema, 2000. T cell depletion in HIV-1 infection: how CD4+ T cells go out of stock. *Nature immunology* 1:285–289.

[147] Giorgi, J. V., L. E. Hultin, J. A. McKeating, T. D. Johnson, B. Owens, L. P. Jacobson, R. Shih, J. Lewis, D. J. Wiley, J. P. Phair, et al., 1999. Shorter survival in advanced human immunodeficiency virus type 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine coreceptor usage. *Journal of Infectious Diseases* 179:859–870.

[148] Borkow, G., Q. Leng, Z. Weisman, M. Stein, N. Galai, A. Kalinkovich, and Z. Bentwich, 2000. Chronic immune activation associated with intestinal helminth infections results in impaired signal transduction and anergy. *Journal of Clinical Investigation* 106:1053.

[149] Schacker, T. W., P. L. Nguyen, G. J. Beilman, S. Wolinsky, M. Larson, C. Reilly, and

A. T. Haase, 2002. Collagen deposition in HIV-1 infected lymphatic tissues and T cell homeostasis. *Journal of Clinical Investigation* 110:1133–1139.

[150] Zeng, M., P. J. Southern, C. S. Reilly, G. J. Beilman, J. G. Chipman, T. W. Schacker, and A. T. Haase, 2012. Lymphoid tissue damage in HIV-1 infection depletes naive T cells and limits T cell reconstitution after antiretroviral therapy. *PLoS Pathog* 8:e1002 437.

[151] Zeng, M., M. Paiardini, J. C. Engram, G. J. Beilman, J. G. Chipman, T. W. Schacker, G. Silvestri, and A. T. Haase, 2012. Critical role of CD4 T cells in maintaining lymphoid tissue structure for immune cell homeostasis and reconstitution. *Blood* 120:1856–1867.

[152] Bujdoso, R., P. Young, J. Hopkins, D. Allen, and I. Mcconnell, 1989. Non-random migration of CD4 and CD8 T cells: changes in the CD4: CD8 ratio and Interleukin 2 responsiveness of efferent lymph cells following in vivo antigen challenge. *European Journal of Immunology* 19:1779–1784.

[153] Deeks, S. G., 2012. HIV infection, lymphoid fibrosis, and disease. *Blood* 120:1753–1754.

[154] Zhang, Z.-Q., D. W. Notermans, G. Sedgewick, W. Cavert, S. Wietgrefe, M. Zupancic, K. Gebhard, K. Henry, L. Boies, Z. Chen, et al., 1998. Kinetics of CD4+ T cell repopulation of lymphoid tissues after treatment of HIV-1 infection. *Proceedings of the National Academy of Sciences* 95:1154–1159.

[155] Pakker, N. G., D. W. Notermans, R. J. De Boer, M. Roos, F. d. Wolf, A. Hill, J. M. Leonard, S. A. Danner, F. Miedema, and P. T. A. Schellekens, 1998. Biphasic kinetics of peripheral blood T cells after triple combination therapy in HIV-1 infection: a composite of redistribution and proliferation. *Nature Medicine* 4:208–214.

[156] Svarovskaia, E. S., S. R. Cheslock, W.-H. Zhang, W.-S. Hu, and V. K. Pathak, 2003. Retroviral mutation rates and reverse transcriptase fidelity. *Frontiers in Bioscience* 8:d117–d134.

[157] World Health Organisation, 2013. Estimate percentage of pregnant women living with HIV who receive antiretrovirals (ART) for preventing mother-to-child transmission (PMTCT), accessed 10/09/15. *http://gamapserver.who.int/gho/interactive_charts/hiv/pmtct/atlas.html* .

[158] Msellati, P., M. Newell, and F. Dabis, 1995. Rates of mother-to-child transmission of HIV-1 in Africa, America and Europe: results from 13 perinatal studies. *Journal of Acquired of Immune Deficiency Syndromes and Human Retrovirology* 8:506–510.

[159] Townsend, C. L., M. Cortina-Borja, C. S. Peckham, A. de Ruiter, H. Lyall, and P. A. Tookey, 2008. Low rates of mother-to-child transmission of HIV following effective pregnancy interventions in the United Kingdom and Ireland, 2000–2006. *AIDS* 22:973–981.

[160] Scarlatti, G., 1996. Paediatric HIV infection. *Lancet* 348:863–868.

[161] European Collaborative Study, 1991. Children born to women with HIV-1 infection: natural history and risk of transmission. *Lancet* 337:253–260.

[162] Barnhart, H. X., M. B. Caldwell, P. Thomas, L. Mascola, I. Ortiz, H.-W. Hsu, J. Schulte, R. Parrott, Y. Maldonado, R. Byers, et al., 1996. Natural history of human immunodeficiency virus disease in perinatally infected children: an analysis from the pediatric spectrum of disease project. *Pediatrics* 97:710–716.

[163] De Beaudrap, P., F. Rouet, P. Fassinou, A. Kouakoussui, S. Mercier, R. Ecochard, and P. Msellati, 2008. CD4 cell response before and after HAART initiation according to viral load and growth indicators in HIV-1-infected children in Abidjan, Cote d'Ivoire. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 49:70–76.

[164] Newell, M., D. Patel, T. Goetghebuer, and C. Thorne, 2006. CD4 cell response to antiretroviral therapy in children with vertically acquired HIV infection: is it associated with age at initiation? *Journal of Infectious Diseases* 193:954–962.

[165] Goetghebuer, T., J. Le Chenadec, E. Haelterman, L. Galli, C. Dollfus, C. Thorne, A. Judd, O. Keiser, J. T. Ramos, J. Levy, et al., 2012. Short-and long-term immunological and virological outcome in HIV-infected infants according to the age at antiretroviral treatment initiation. *Clinical Infectious Diseases* 54:878–881.

[166] Hainaut, M., M. Ducarme, L. Schandené, C. A. Peltier, D. Marissens, G. Zissis, F. Mascart, and J. Levy, 2003. Age-related immune reconstitution during highly active antiretroviral therapy in human immunodeficiency virus type 1-infected children. *Pediatric Infectious Disease Journal* 22:62–69.

[167] Kekitiinwa, A., K. J. Lee, A. S. Walker, A. Maganda, K. Doerholt, S. B. Kitaka, A. Asiimwe, A. Judd, P. Musoke, D. M. Gibb, et al., 2008. Differences in factors associated with initial growth, CD4, and viral load responses to ART in HIV-infected children in Kampala, Uganda, and the United Kingdom/Ireland. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 49:384–392.

[168] Paediatric European Network for Treatment of AIDS and others, 2010. Response to planned treatment interruptions in HIV infection varies across childhood. *AIDS* 24:231–241.

[169] Puthanakit, T., S. J. Kerr, J. Ananworanich, T. Bunupuradah, P. Boonrak, and V. Sirisanthana, 2009. Pattern and predictors of immunologic recovery in human immunodeficiency virus-infected children receiving non-nucleoside reverse transcriptase inhibitor-based highly active antiretroviral therapy. *Pediatric Infectious Disease Journal* 28:488–492.

[170] Walker, A. S., K. Doerholt, M. Sharland, D. M. Gibb, et al., 2004. Response to highly active antiretroviral therapy varies with age: the UK and Ireland Collaborative HIV Paediatric Study. *AIDS* 18:1915–1924.

[171] Lewis, J., A. S. Walker, H. Castro, A. de Rossi, D. M. Gibb, C. Giaquinto, N. Klein, and R. Callard, 2012. Age and CD4 count at initiation of antiretroviral therapy in HIV-infected children: Effects on long-term T-cell reconstitution. *Journal of Infectious Diseases* 205:548–556.

[172] Picat, M.-Q., J. Lewis, V. Musiime, A. Prendergast, K. Nathoo, A. Kekitiinwa, P. Nahirya Ntege, D. M. Gibb, R. Thiebaut, A. S. Walker, et al., 2013. Predicting patterns of long-term CD4 reconstitution in HIV-infected children starting antiretroviral therapy in sub-Saharan Africa: a cohort-based modelling study. *PLoS Medicine* 10:e1001 542.

[173] Guiguet, M., F. Boué, J. Cadranel, J.-M. Lang, E. Rosenthal, D. Costagliola, C. E. G. of the FHDH-ANRS CO4 Cohort, et al., 2009. Effect of immunodeficiency, HIV viral load, and antiretroviral therapy on the risk of individual malignancies (FHDH-ANRS CO4): a prospective cohort study. *Lancet Oncology* 10:1152–1159.

[174] Lichtenstein, K. A., C. Armon, K. Buchacz, J. S. Chmiel, K. Buckner, E. Tedaldi, K. Wood, S. D. Holmberg, and J. T. Brooks, 2010. Low CD4+ T cell count is a risk factor for cardiovascular disease events in the HIV outpatient study. *Clinical Infectious Diseases* 51:435–447.

[175] Triant, V. A., S. Regan, H. Lee, P. E. Sax, J. B. Meigs, and S. K. Grinspoon, 2010. Association of immunologic and virologic factors with myocardial infarction rates in a US health care system. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 55:615.

[176] Bouazza, N., J.-M. Tréluyer, P. Msellati, P. Van de Perre, S. Diagbouga, B. Nacro, H. Hien, E. Zoure, F. Rouet, A. Ouiminga, et al., 2013. A novel pharmacokinetic approach to predict virologic failure in HIV-1-infected paediatric patients. *AIDS* 27:761–768.

[177] Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, M. Markowitz, et al., 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373:123–126.

[178] Wei, X., S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, et al., 1995. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* 373:117–122.

[179] Nowak, M. A., S. Bonhoeffer, G. M. Shaw, and R. M. May, 1997. Anti-viral drug treatment: dynamics of resistance in free virus and infected cell populations. *Journal of Theoretical Biology* 184:203–217.

[180] Bonhoeffer, S., R. M. May, G. M. Shaw, and M. A. Nowak, 1997. Virus dynamics and

drug therapy. *Proceedings of the National Academy of Sciences* 94:6971–6976.

[181] McLean, A. R., V. C. Emery, A. Webster, and P. D. Griffiths, 1991. Population dynamics of HIV within an individual after treatment with zidovudine. *AIDS* 5:485–490.

[182] Perelson, A. S., D. E. Kirschner, and R. De Boer, 1993. Dynamics of HIV infection of CD4+ T cells. *Mathematical Biosciences* 114:81–125.

[183] Funk, G. A., M. Fischer, M. Opravil, H. F. Günthard, B. Ledergerber, S. Bonhoeffer, et al., 2001. Quantification of in vivo replicative capacity of HIV-1 in different compartments of infected cells. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 26:397–404.

[184] Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho, 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 271:1582–1586.

[185] Perelson, A. S., and P. W. Nelson, 1999. Mathematical analysis of HIV-1 dynamics in vivo. *SIAM Review* 41:3–44.

[186] Perelson, A. S., 2002. Modelling viral and immune system dynamics. *Nature Reviews Immunology* 2:28–36.

[187] De Boer, R. J., and A. S. Perelson, 1998. Target cell limited and immune control models of HIV infection: a comparison. *Journal of Theoretical Biology* 190:201–214.

[188] Ribeiro, R. M., H. Mohri, D. D. Ho, and A. S. Perelson, 2002. In vivo dynamics of T cell activation, proliferation, and death in HIV-1 infection: why are CD4+ but not CD8+ T cells depleted? *Proceedings of the National Academy of Sciences* 99:15 572–15 577.

[189] Yates, A., J. Stark, N. Klein, R. Antia, and R. Callard, 2007. Understanding the slow depletion of memory CD4+ T cells in HIV infection. *PLoS Medicine* 4:e177.

[190] Callard, R. E., J. Stark, and A. J. Yates, 2003. Fratricide: a mechanism for T memory-cell homeostasis. *Trends in immunology* 24:370–375.

[191] Putter, H., S. Heisterkamp, J. Lange, and F. De Wolf, 2002. A Bayesian approach to parameter estimation in HIV dynamical models. *Statistics in Medicine* 21:2199–2214.

[192] Huang, Y., D. Liu, and H. Wu, 2006. Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* 62:413–423.

[193] Guedj, J., R. Thiébaut, and D. Commenges, 2007. Maximum likelihood estimation in dynamical models of HIV. *Biometrics* 63:1198–1206.

[194] Prague, M., D. Commenges, J. Drylewicz, and R. Thiébaut, 2012. Treatment monitoring of HIV-infected patients based on mechanistic models. *Biometrics* 68:902–911.

[195] Prague, M., D. Commenges, and R. Thiébaut, 2013. Dynamical models of biomarkers and clinical progression for personalized medicine: The HIV context. *Advanced Drug Delivery Reviews* 65:954–965.

[196] Drylewicz, J., J. Guedj, D. Commenges, and R. Thiébaut, 2010. Modeling the dynamics of biomarkers during primary HIV infection taking into account the uncertainty of infection date. *Annals of Applied Statistics* pages 1847–1870.

[197] Lavielle, M., A. Samson, A. Karina Fermin, and F. Mentré, 2011. Maximum likelihood estimation of long-term HIV dynamic models and antiviral response. *Biometrics* 67:250–259.

[198] Josefsson, L., M. S. King, B. Makitalo, J. Brännström, W. Shao, F. Maldarelli, M. F. Kearney, W.-S. Hu, J. Chen, H. Gaines, et al., 2011. Majority of CD4+ T cells from peripheral blood of HIV-1–infected individuals contain only one HIV DNA molecule. *Proceedings of the National Academy of Sciences* 108:11 199–11 204.

[199] Palmer, S., L. Josefsson, and J. Coffin, 2011. HIV reservoirs and the possibility of a cure for HIV infection. *Journal of Internal Medicine* 270:550–560.

[200] Gibb, D. M., A. Compagnucci, H. Green, M. Lallemant, Y. Saidi, N. Ngo-Giang-Huong, C. Taylor, L. Mofenson, F. Monpoux, M. Tomé, et al., 2008. Treatment interruption in children with chronic HIV-infection: the results of the paediatric European network for

treatment of AIDS (PENTA) 11 trial. *Journal of the International AIDS Society* 11:1–2.

[201] Bunupuradah, T., T. Duong, A. Compagnucci, P. McMaster, S. Bernardi, S. Kanjanavanit, O. Rampon, A. Faye, Y. Saïdi, Y. Riault, et al., 2013. Outcomes after reinitiating antiretroviral therapy in children randomized to planned treatment interruptions. *AIDS* 27:579–589.

[202] ARROW Trial Team and others, 2013. Routine versus clinically driven laboratory monitoring and first-line antiretroviral therapy strategies in African children with HIV (ARROW): a 5-year open-label randomised factorial trial. *Lancet* 381:1391–1403.

[203] Varma, R., 2008. TCR triggering by the pMHC complex: Valency, affinity, and dynamics. *Science Signaling* 1:pe21.

[204] Sewell, A. K., 2012. Why must T cells be cross-reactive? *Nature Reviews Immunology* 12:669–677.

[205] Mason, D., 1998. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology Today* 19:395–404.

[206] Newell, E. W., L. K. Ely, A. C. Kruse, P. A. Reay, S. N. Rodriguez, A. E. Lin, M. S. Kuhns, K. C. Garcia, and M. M. Davis, 2011. Structural basis of specificity and cross-reactivity in T cell receptors specific for cytochrome c–I-Ek. *Journal of Immunology* 186:5823–5832.

[207] Calis, J. J. A., R. J. de Boer, and C. Keşmir, 2012. Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire. *PLoS Computational Biology* 8:e1002 412.

[208] Lythe, G., R. Callard, R. L. Hoare, and C. Molina-París, 2016. How many TCR clonotypes does a body maintain? *Journal of Theoretical Biology* 389:214–224.

[209] Revuz, D., and M. Yor, 1999. *Continuous Martingales and Brownian Motion*. Springer.

[210] Voss, A., and J. Voss, 2008. A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology* 52:1–9.

[211] Karlin, S., and H. M. Taylor, 1981. *A Second Course in Stochastic Processes*. Gulf Professional Publishing.

[212] Wolfram Research, Inc., 2012. *Mathematica*. Version 9.0, Champaign, Illinois.

[213] Prescott, S., C. Macaubas, T. Smallacombe, B. Holt, P. Sly, R. Loh, and P. Holt, 1998. Reciprocal age-related patterns of allergen-specific T-cell immunity in normal vs. atopic infants. *Clinical and Experimental Allergy* 28:39–44.

[214] Baird, J., 1998. Age dependent characteristics of protection v. susceptibility to plasmodium falciparum. *Annals of Tropical Medicine and Parasitology* 92:367–390.

[215] Ribbing, J., and E. N. Jonsson, 2004. Power, selection bias and predictive performance of the population pharmacokinetic covariate model. *Journal of Pharmacokinetics and Pharmacodynamics* 31:109–134.

[216] Sauerbrei, W., 1999. The use of resampling methods to simplify regression models in medical statistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 48:313–329.

[217] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 267–288.

[218] Ribbing, J., J. Nyberg, O. Caster, and E. N. Jonsson, 2007. The lasso – a novel method for predictive covariate model building in nonlinear mixed effects models. *Journal of Pharmacokinetics and Pharmacodynamics* 34:485–517.

[219] Tessier, A., J. Bertrand, M. Chenel, and E. Comets, 2015. Comparison of nonlinear mixed effects models and noncompartmental approaches in detecting pharmacogenetic covariates. *AAPS Journal* 17:597–608.

Appendix A

NONMEM model files

A.1 Short-term reconstitution of neutrophils following paediatric HSCT

```
$PROBLEM   Paediatric HSCT Neutrophil Reconstitution
;____ Data Input and setup of problem ___________________________________
$INPUT     ID TIME DV EVID CMT AMT DRUG STEROIDS DENDT DRUGT AGE BMTAGE
     ALEM ATG DONORTYPE LEUK CORD
$DATA      nonmemdata_HSCT_neutrophils_log.csv IGNORE=#
$SUBROUTINE ADVAN13 TOL=6
$MODEL     COMP(STEM) COMP(TRN1) COMP(TRN2) COMP(TRN3) COMP(CIRC,
   DEFOBS) COMP(INPUT)
$PK

;____ Initiate population level parameters for the model _______________
TVBASE = THETA(1)
TVSET  = THETA(2)
TVMTT  = THETA(3)
TVGAM  = THETA(4)
TVEDRUG= THETA(5)
TVKIN  = THETA(6)
TVIP0  = THETA(7)
TVK    = THETA(8)
TVTDRUG= THETA(9)

;____ Modelling covariate effects on population level parameters ______
IF(DONORTYPE.EQ.2) MTTDONORTYPE = 1
IF(DONORTYPE.EQ.3) MTTDONORTYPE = (1 + THETA(12))
IF(DONORTYPE.EQ.1) MTTDONORTYPE = (1 + THETA(13))
IF(DONORTYPE.EQ.0) MTTDONORTYPE = (1 + THETA(14))
IF(LEUK.EQ.0)    BASELEUK    =  1
IF(LEUK.EQ.1)    BASELEUK    = (1 + THETA(11))
IF(CORD.EQ.0)    BASECORD    =  1
IF(CORD.EQ.1)    BASECORD    = (1 + THETA(10))
```

```
TVBASE = TVBASE * BASECORD * BASELEUK
TVMTT  = TVMTT * MTTDONORTYPE

;____ Mu modelling to improve model run-times ____________________________
MU_1 = LOG(TVBASE)
MU_2 = LOG(TVSET)
MU_3 = LOG(TVMTT)
MU_4 = LOG(TVGAM)
MU_5 = TVEDRUG
MU_6 = LOG(TVKIN)
MU_7 = LOG(TVIP0)
MU_8 = LOG(TVK)
MU_9 = LOG(TVTDRUG)

;____ Including individual random effects on parameters _____________
BASE = EXP(MU_1+ETA(1))
SET  = EXP(MU_2+ETA(2))
MTT  = EXP(MU_3+ETA(3))
GAM  = EXP(MU_4+ETA(4))
EDRUG= MU_5+ETA(5)
KIN  = EXP(MU_6+ETA(6))
IP0  = EXP(MU_7+ETA(7))
K    = EXP(MU_8+ETA(8))
TDRUG= EXP(MU_9+ETA(9))

;____ Define initial concentration for each compartment _____________
A_0(1) = BASE/K
A_0(2) = BASE/K
A_0(3) = BASE/K
A_0(4) = BASE/K
A_0(5) = BASE
A_0(6) = IP0

;____ Define rates from MTT _____________________________________________
KELM = 4/(MTT*K)
KSYN = 4/MTT
KTR  = 4/MTT

;____ Change in asymptote at time of HSCT ______________________________
$DES
IF (TIME.LT.14) THEN
   ASY=BASE
ELSE
   ASY=SET
ENDIF
RBD=(ASY/A(5))**GAM

;____ Linear K-PD model ________________________________________________
IF (TIME.LE.DENDT) THEN
```

```
   DRUG1 = DRUG/(1+EXP(-EDRUG))
ELSE
   DRUG1 = DRUGT/(1+EXP(-EDRUG))*EXP(-(T-DENDT)/TDRUG)
ENDIF

;____ Effects of steroids ____________________________________________
STER=0
IF (ALEM==1) THEN
   IF (TIME.GE.5.AND.TIME.LT.9) THEN
      STER = 1
   ENDIF
ENDIF
IF (ATG==1) THEN
   IF (TIME.GE.8.AND.TIME.LT.11) THEN
      STER = 1
   ENDIF
ENDIF

;____ Define the differential equations for the model ________________
DADT(1) = KSYN*A(1)*RBD*(1 - DRUG1) - KTR*A(1)
DADT(2) = KTR*A(1) - KTR*A(2)
DADT(3) = KTR*A(2) - KTR*A(3)
DADT(4) = KTR*A(3) - KTR*A(4)
DADT(5) = KTR*A(4) - KELM*A(5) + STER*KIN*A(6)
DADT(6) =-KIN*A(6)*STER

;____ Compare the model output (IPRED) to the observed data (Y) _______
$ERROR
IPRED=A(5)
IPRED=LOG(IPRED)
Y=IPRED+EPS(1)

;____ Initial estimates for the population level parameters ___________
$THETA
 (0,1039.250) ; 1. BASE, initial neutrophil concentration
 (0,2570)     ; 2. SET, long-term neutrophil concentration
 (0,5.0)      ; 3. MTT, transfer time through transit
 (0,0.084)    ; 4. GAM, strength of feedback
 (-20,2,20)   ; 5. EDRUG, maximum drug effect
 (0,0.8)      ; 6. KIN, neutrophil input rate from steroids
 (0,1200)     ; 7. IP0, initial neutrophil conc in input comp
 (0,1)        ; 8. K, difference between KE, KS and KT
 (0,5)        ; 9. TDRUG, Drug effect lifetime from last dose
 (-1,-0.001)  ; 10. BASECORD1, effects of CORD on BASE
 (-1,-0.001)  ; 11. BASELEUK1, effects of LEUK on BASE
 (-1,-0.001)  ; 12. MTTDONORTYPE3, effects of DT3 on MTT
 (-1,-0.001)  ; 13. MTTDONORTYPE1, effects of DT1 on MTT
 (-1,-0.001)  ; 14. MTTDONORTYPE0, effects of DT0 on MTT
```

```
;____ Initial estimates for the random effects var-covar matrix _______
$OMEGA BLOCK(9)
 0.1
 0.01 0.1
 0.01 0.01 0.1
 0.01 0.01 0.01 0.1
 0.01 0.01 0.01 0.01 0.1
 0.01 0.01 0.01 0.01 0.01 0.1
 0.01 0.01 0.01 0.01 0.01 0.01 0.1
 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.1
 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.1
;____ Initial estimate for the residual error variance _____________
$SIGMA 0.5

;____ Estimation algorithm to be used (EONLY=1 Expectation Only) ______
$ESTIMATION METHOD=SAEM INTERACTION NBURN=800 NITER=1000 PRINT=10 CTYPE
   =3
$ESTIMATION METHOD=IMP EONLY=1 ISAMPLE=10000 IACCEPT=0.1 NITER=12 PRINT
   =1 DF=0 ISAMPEND=10000 STDOBJ=2 MAPITER=0 PRINT=1 RANMETHOD=3S2

;____ Output results of the model-fitting into a table _________________
$TABLE     ID TIME DV PRED IPRED CWRES EVID ETA1 ETA2 ETA3 ETA4 ETA5
   ETA6 ETA7 ETA8 ETA9 OBJI NOAPPEND ONEHEADER NOPRINT FILE=
   sdtab_HSCT_neutro
```

A.2 Short-term reconstitution of platelets following paediatric HSCT

```
$PROBLEM   Paediatric HSCT Platelet Reconstitution
;____ Data Input and setup of problem _____________________________
$INPUT     ID TIME DV EVID CMT AMT DRUG STEROIDS DENDT DRUGT AGE BMTAGE
     ALEM ATG DONORCELLS BUSU CYCL IMDEF
$DATA      nonmemdata_HSCT_platelets_log.csv IGNORE=#
$SUBROUTINE ADVAN13 TOL=6
$MODEL     COMP(STEM) COMP(TRN1) COMP(TRN2) COMP(TRN3) COMP(CIRC,
   DEFOBS)
$PK

;____ Initiate population level parameters for the model _______________
TVBASE = THETA(1)
TVSET  = THETA(2)
TVMTT  = THETA(3)
TVGAM  = THETA(4)
TVEDRUG= THETA(5)
TVTDRUG= THETA(6)

;____ Modelling covariate effects on population level parameters ______
IF(IMDEF.EQ.0)     SETIMDEF    =  1
```

```
IF(IMDEF.EQ.1)     SETIMDEF    = (1 + THETA(11))
IF(DONORCELLS.EQ.0) SETDONORCELLS = 1
IF(DONORCELLS.EQ.2) SETDONORCELLS = (1 + THETA(9))
IF(DONORCELLS.EQ.1) SETDONORCELLS = (1 + THETA(10))
IF(CYCL.EQ.0)      BASECYCL     =  1
IF(CYCL.EQ.1)      BASECYCL     = (1 + THETA(8))
IF(BUSU.EQ.0)      BASEBUSU     =  1
IF(BUSU.EQ.1)      BASEBUSU     = (1 + THETA(7))

TVBASE = TVBASE * BASEBUSU * BASECYCL
TVSET  = TVSET * SETDONORCELLS * SETIMDEF

;____ Mu modelling to improve model run-times ___________________________
MU_1 = LOG(TVBASE)
MU_2 = LOG(TVSET)
MU_3 = LOG(TVMTT)
MU_4 = LOG(TVGAM)
MU_5 = TVEDRUG
MU_6 = LOG(TVTDRUG)

;____ Including individual random effects on parameters _____________
BASE = EXP(MU_1+ETA(1))
SET  = EXP(MU_2+ETA(2))
MTT  = EXP(MU_3+ETA(3))
GAM  = EXP(MU_4+ETA(4))
EDRUG= MU_5+ETA(5)
TDRUG= EXP(MU_6+ETA(6))

;____ Define initial concentration for each compartment _____________
A_0(1) = BASE
A_0(2) = BASE
A_0(3) = BASE
A_0(4) = BASE
A_0(5) = BASE

;____ Define rates from MTT _____________________________________________
KELM = 4/MTT
KSYN = 4/MTT
KTR  = 4/MTT

;____ Change in asymptote at time of HSCT _______________________________
$DES
IF (TIME.LT.14) THEN
   ASY=BASE
ELSE
   ASY=SET
ENDIF
RBD=(ASY/A(5))**GAM
```

```
;____ Linear K-PD model _____________________________________________
IF (TIME.LE.DENDT) THEN
   DRUG1 = DRUG/(1+EXP(-EDRUG))
ELSE
   DRUG1 = DRUGT/(1+EXP(-EDRUG))*EXP(-(T-DENDT)/TDRUG)
ENDIF

;____ Define the differential equations for the model ________________
DADT(1)=KSYN*A(1)*RBD*(1 - DRUG1) - KTR*A(1)
DADT(2)=KTR*A(1) - KTR*A(2)
DADT(3)=KTR*A(2) - KTR*A(3)
DADT(4)=KTR*A(3) - KTR*A(4)
DADT(5)=KTR*A(4) - KELM*A(5)

;____ Compare the model output (IPRED) to the observed data (Y) _______
$ERROR
IPRED=A(5)
IPRED=LOG(IPRED)
Y=IPRED+EPS(1)

;____ Initial estimates for the population level parameters ___________
$THETA
 (0,176791)    ; 1. BASE, initial platelet concentration
 (0,149943)    ; 2. SET, long-term platelet concentration
 (0,5.16337)   ; 3. MTT, transfer time through transit
 (0,0.116843)  ; 4. GAM, strength of feedback
 (-20,-1.00,20) ; 5. EDRUG, maximum drug effect
 (0,6.9009)    ; 6. TDRUG, drug effect lifetime from last dose
 (-1,0.698179) ; 7. BASEBUSU1, effects of BUSU on BASE
 (-1,-0.437215) ; 8. BASECYCL1, effects of CYCL on BASE
 (-1,0.373364) ; 9. SETIMDEF1, effects of IMDEF of SET
 (-1,-0.001)   ; 10. SETDONORCELLS2, effects of DC2 on SET
 (-1,-0.001)   ; 11. SETDONORCELLS1, effects of DC1 on SET

;____ Initial estimates for the random effects var-covar matrix _______
$OMEGA BLOCK(6)
 0.77513
 0.246018 0.834123
 -0.34828 -0.126483 0.577297
 -0.230623 -0.480143 0.302266 1.11016
 0.186967 -0.0423519 0.576467 0.277966 1.4537
 -0.205738 0.847174 -0.73557 0.0774342 -2.02161 5.63053
;____ Initial estimate for the residual error variance _______________
$SIGMA 0.267354

;____ Estimation algorithm to be used (EONLY=1 Expectation Only) ______
$ESTIMATION METHOD=SAEM INTERACTION NBURN=1000 NITER=1000 PRINT=10
$ESTIMATION METHOD=IMP EONLY=1 ISAMPLE=10000 IACCEPT=0.1 NITER=12 PRINT
   =1 DF=0 ISAMPEND=10000 STDOBJ=2 MAPITER=0 PRINT=1 RANMETHOD=3S2
```

```
;____ Output results of the model-fitting into a table _________________
$TABLE     ID TIME DV PRED IPRED CWRES EVID ETA1 ETA2 ETA3 ETA4 ETA5
   ETA6 OBJI NOAPPEND ONEHEADER NOPRINT FILE=sdtab_HSCT_platelets
```

A.3 Long-term reconstitution of CD4 concentration following paediatric HSCT

```
$PROBLEM   Paediatric HSCT CD4 T cell Reconstitution
;____ Data Input and setup of problem ___________________________________
$INPUT     ID TIME DV EVID AGE GVHYN COND ALEM ATG LEUK
$DATA      nonmemdata_HSCT_cd4_log.csv IGNORE=@
$SUBROUTINE ADVAN13 TOL=9
$MODEL     COMP=(CENTRAL)
$PK

;____ Initiate population level parameters for the model _______________
TVL   = THETA(1)
TVD   = THETA(2)
TVPR  = THETA(3)
TVINT = THETA(4)
TVLH  = THETA(5)
TVLR  = THETA(6)

;____ Modelling covariate effects on population level parameters ______
IF(ALEM.EQ.0) INTALEM = 1
IF(ALEM.EQ.1) INTALEM = (1 + THETA(7))
IF(ATG.EQ.0)  INTATG  =  1
IF(ATG.EQ.1)  INTATG  = (1 + THETA(8))
IF(GVHYN.EQ.0) INTGVHYN = 1
IF(GVHYN.EQ.1) INTGVHYN = (1 + THETA(9))
IF(LEUK.EQ.0) LLEUK   =  1
IF(LEUK.EQ.1) LLEUK   = (1 + THETA(10))
IF(COND.EQ.0) PRCOND  =  1
IF(COND.EQ.1) PRCOND  = (1 + THETA(11))

TVINT = TVINT * INTALEM * INTATG * INTGVHYN
TVL   = TVL * LLEUK
TVPR  = TVPR * PRCOND

;____ Mu modelling to improve model run-times ___________________________
MU_1 = LOG(TVL)
MU_2 = LOG(TVD)
MU_3 = LOG(TVPR)
MU_4 = LOG(TVINT)
MU_5 = LOG(TVLH)
MU_6 = LOG(TVLR)
```

```
;____ Naive (V) and total (VT) CD4 T cell concentration with age ______
VT = 924 + 2354*EXP(-0.001012*AGE)
V = 496.5 + 2074*EXP(-0.000869*AGE)

;____ Including Age effects and individual random effects on params ___
L   = EXP(MU_1 + ETA(1)) * V * 0.9 * 0.02*EXP(-0.00027*AGE)
D   = EXP(MU_2 + ETA(2)) * 0.02*EXP(-0.00027*AGE)
PR  = EXP(MU_3 + ETA(3)) * 0.02*EXP(-0.00027*AGE)
INT = EXP(MU_4 + ETA(4))
LH  = EXP(MU_5 + ETA(5))
LR  = EXP(MU_6 + ETA(6))
CD  = 1
CP  = 1
A_0(1) = INT

;____ Define the differential equation for the model ___________________
$DES
DADT(1) = L * (1-EXP(-2*T/LH))/(1+EXP(LR*(1-T/LH))) - A(1)*(D*EXP(CD*(A
    (1)/VT-1)) - PR * EXP(CP*(-A(1)/VT+1)))

;____ Compare the model output (IPRED) to the observed data (Y) _______
$ERROR
IPRED = A(1)
IPRED = LOG(IPRED)
Y = IPRED + EPS(1)

;____ Initial estimates for the population level parameters ___________
$THETA
 (0,0.222489) ; 1. L0, thymic output
 (0,0.454536) ; 2. D0, loss
 (0,0.195179) ; 3. P0, proliferation
 (0,164.2850) ; 4. X0, Initial concentration of cells
 (0,133.1450) ; 5. LH, time to recovery in thymic output
 (0,9.757030) ; 6. LR, rate of recovery of thymic output
 (-1,-0.83722) ; 7. INTALEM, effects of alemtuzumab on A_0
 (-1,-0.93890) ; 8. INTATG, effects of ATG on A_0
 (-1,0.331920) ; 9. INTGVHYN, effects of acute GvHD on A_0
 (-1,1.289310) ; 10. LLEUK, effects of leukaemia on thymic output
 (-1,-0.88522) ; 11. PRCOND, effects of no conditioning on
    proliferation

;____ Initial estimates for the random effects var-covar matrix _______
$OMEGA BLOCK(6)
 1.553010
 0.522870 1.662840
 0.161760 0.320213 0.247498
 0.405366 0.415663 0.218925 1.286390
 0.534060 -0.44269 -0.06167 0.766879 1.214760
 0.083104 0.340042 -0.13533 -0.89795 -0.77866 1.259670
```

```
;____ Initial estimate for the residual error variance _____________
$SIGMA 0.218784

;____ Estimation algorithm to be used (EONLY=1 Expectation Only) ______
$ESTIMATION METHOD=IMP INTERACTION ISAMPLE=1000 NITER=80 PRINT=10 CTYPE
   =1
$ESTIMATION METHOD=IMP INTERACTION EONLY=1 ISAMPLE=10000 IACCEPT=0.1
   NITER=8 PRINT=1 DF=0 ISAMPEND=10000 STDOBJ=2 MAPITER=0 RANMETHOD=3S2

;____ Output results of the model-fitting into a table _____________
$TABLE     ID TIME DV PRED IPRED CWRES EVID ETA1 ETA2 ETA3 ETA4 ETA5
   ETA6 LAMBDA DELTA INT OBJI NOPRINT NOAPPEND ONEHEADER FILE=
   sdtab_HSCT_cd4
```

A.4 CD4 concentration and viral load in HIV-infected children restarting ART, PENTA data

```
$PROBLEM   HIV infected children starting ART, PENTA data
;____ Data Input and setup of problem _____________________________
$INPUT     ID TIME DV EVID FLAG VLBLQ VLLOQ L2 AGE PTINO ARTAGE NVP0
   ZDV0 D3TC EFZ CART
$DATA      nonmemdata_HIV_penta_log.csv IGNORE=@
$SUBROUTINE ADVAN13 TOL=9
$MODEL     COMP(CD4)
$MODEL     COMP(VL)
$PK

;____ Initiate population level parameters for the model ______________
;; CD4 Parameters
TVL0  = THETA(1)
TVD0  = THETA(2)
TVP0  = THETA(3)
TVINT = THETA(4)
TVCD  = THETA(5)
TVCP  = THETA(6)
;; Viral Load parameters
TVVL0  = THETA(7)
TVVK   = THETA(8)
TVV50  = THETA(9)
TVEDRUG = THETA(10)

;____ Modelling covariate effects on population level parameters ______
IF(EFZ.EQ.0) VKEFZ  =  1
IF(EFZ.EQ.1) VKEFZ  = (1 + THETA(16))
IF(ZDV0.EQ.0) V50ZDV0 = 1
IF(ZDV0.EQ.1) V50ZDV0 = (1 + THETA(15))
```

```
IF(NVP0.EQ.0) PONVP0 = 1
IF(NVP0.EQ.1) PONVP0 = (1 + THETA(14))
IF(ZDV0.EQ.0) INTZDV0 = 1
IF(ZDV0.EQ.1) INTZDV0 = (1 + THETA(13))
IF(D3TC.EQ.1) CPD3TC = 1
IF(D3TC.EQ.0) CPD3TC = (1 + THETA(12))
INTCART = ( 1 + THETA(11)*(CART - 0.53))

TVINT = TVINT * INTCART * INTZDV0
TVVK  = TVVK * VKEFZ
TVV50 = TVV50 * V50ZDV0
TVP0  = TVP0 * PONVP0
TVCP  = TVCP * CPD3TC

;____ Mu modelling to improve model run-times ___________________________
MU_1  = LOG(TVL0)
MU_2  = LOG(TVD0)
MU_3  = LOG(TVP0)
MU_4  = LOG(TVINT)
MU_5  = LOG(TVCD)
MU_6  = LOG(TVCP)
MU_7  = LOG(TVVL0)
MU_8  = LOG(TVVK)
MU_9  = LOG(TVV50)
MU_10 = TVEDRUG

;____ Including individual random effects on parameters _______________
L     = EXP(MU_1 + ETA(1))
D     = EXP(MU_2 + ETA(2))
PR    = EXP(MU_3 + ETA(3))
INT   = EXP(MU_4 + ETA(4))
CD    = EXP(MU_5 + ETA(5))
CP    = EXP(MU_6 + ETA(6))
VL0   = EXP(MU_7 + ETA(7))
VK    = EXP(MU_8 + ETA(8))
V50   = EXP(MU_9 + ETA(9))
EDRUG =   MU_10 + ETA(10)

;____ Define initial concentration for each compartment _______________
A_0(1) = INT
A_0(2) = VL0

;____ Naive (V) and total (VT) CD4 T cell concentration with age
    ------------
VT    = 924 + 2354*EXP(-0.001012*AGE)
V     = 496.5 + 2074*EXP(-0.000869*AGE)
VTART = 924 + 2354*EXP(-0.001012*ARTAGE)
VART  = 496.5 + 2074*EXP(-0.000869*ARTAGE)
```

```
;____ Including Age effects on parameters ______________________________
LAMBDA = L * 0.02*EXP(-0.00027*AGE) * V * 0.905
LOSS   = D * 0.02*EXP(-0.00027*AGE)
PROL   = PR * 0.02*EXP(-0.00027*AGE)
DELTA  = (0.02*EXP(-0.00027*ARTAGE)/VL0) *(VART*0.905*L/INT - D*EXP(CD*(
    INT/VTART-1)) + PR*EXP(CP*(1-INT/VTART)))
DRUG   = 1/(1+EXP(-EDRUG))

;____ Define the differential equations for the model ___________________
$DES
DADT(1) = LAMBDA - A(1)*(LOSS * EXP(CD*(A(1)/VT-1)) - PROL * EXP(CP*(-A
    (1)/VT+1))) - DELTA*A(1)*A(2)
DADT(2) = (1-DRUG)*VK*A(2)*VL0/(VL0+V50) - VK*A(2) *A(2)/(A(2) + V50)

;____ Compare the model output (IPRED) to the observed data (Y) _______
$ERROR
IPRED=0
; VIRUS
IF(FLAG==2) IPRED = LOG(A(2))
IF(VLBLQ==0.AND.FLAG==2) F_FLAG = 0
IF(VLBLQ==0.AND.FLAG==2) Y = IPRED + EPS(1)
IF(VLBLQ==1.AND.FLAG==2) F_FLAG = 1
IF(VLBLQ==1.AND.FLAG==2) SD = SQRT(SIGMA(1,1))
IF(VLBLQ==1.AND.FLAG==2) Y = PHI((VLLOQ-IPRED)/SD)+1.0E-30
IF(VLBLQ==1.AND.FLAG==2) MDVRES=1
; CD4
IF(FLAG==1) IPRED = LOG(A(1))
IF(FLAG==1) Y = IPRED + EPS(2)

;____ Initial estimates for the population level parameters ___________
$THETA
 (0,0.141795)      ; 1. L0, thymic output
 (0,0.380595)      ; 2. D0, loss
 (0,0.0475488)     ; 3. P0, proliferation
 (0,551.052)       ; 4. X0, initial concentration CD4
 (0,0.974182)      ; 5. CD, strength competition loss
 (0,4.75739)       ; 6. CP, strength competition proliferation
 (0,20654.8)       ; 7. VL0, initial viral load
 (0,0.208448)      ; 8. VK, rate viral load decline
 (0,363.183)       ; 9. V50, viral load multiphasic
 (-20,4.42277,20)  ; 10. EDRUG, drug effect
 (-2.127,-0.800645) ; 11. INTCART1, ART age effects on X0
 (-1,-0.381881)    ; 12. CPD3TC1, 3TC effects on CP
 (-1,0.192567)     ; 13. INTZDV01, ZDV effects on X0
 (-1,2.95009)      ; 14. P0NVP01, NVP effects on P0
 (-1,-0.839658)    ; 15. V50ZDV01, ZDV effects on V50
 (-1,0.986536)     ; 16. VKEFZ1, EFZ effects on VK
```

```
;____ Initial estimates for the random effects var-covar matrix _______
$OMEGA BLOCK(10)
 0.27265
 0.706376 3.77097
 0.975012 5.08962 9.92586
 0.0182628 0.190841 0.882143 0.144507
 0.432257 2.1399 2.38794 0.00361615 1.33157
 -0.193464 -1.53942 -2.46365 -0.170988 -0.779164 0.75726
 -0.0121435 -1.31401 -1.50906 -0.0675266 -0.763466 0.653639 1.06044
 0.199772 1.00175 1.33651 0.0438143 0.572176 -0.401686 -0.30425
    0.280002
 0.202649 0.776869 2.30924 0.271685 0.20535 -0.47313 0.104938 0.238378
    0.942788
 0.0921026 0.990572 3.30598 0.44775 0.16056 -0.773472 -0.144805
    0.277563 1.28223 2.04978
;____ Initial estimate for the residual error variances ______________
$SIGMA BLOCK(2)
 1.40524
 -0.023082 0.0411791

;____ Estimation algorithm to be used (EONLY=1 Expectation Only) ______
$ESTIMATION METHOD=IMP INTER LAPLACE GRD=SN(1) NITER=80 PRINT=10
   ISAMPLE=1000 CTYPE=1 NOABORT
$ESTIMATION METHOD=IMP INTER LAPLACE GRD=SN(1) EONLY=1 NITER=4 PRINT=1
   ISAMPLE=10000 IACCEPT=0.1 DF=0 ISAMPEND=10000 STDOBJ=2 MAPITER=0
   RANMETHOD=3S2 NOABORT

;____ Output results of the model-fitting into a table _________________
$TABLE     ID TIME DV EVID PRED IPRED FLAG VLBLQ CWRES ETA1 ETA2 ETA3
   ETA4 ETA5 ETA6 ETA7 ETA8 ETA9 ETA10 OBJI NOPRINT NOAPPEND ONEHEADER
   FILE=sdtab_HIV_penta
```

A.5 CD4 concentration and viral load in HIV-infected children starting ART, ARROW data

```
$PROBLEM   HIV infected children starting ART, ARROW data
;____ Data Input and setup of problem _________________________________
$INPUT     ID TIME DV FLAG EVID VLBLQ VLLOQ L2 AGE ARTAGE CTR CHART
   MART DIED RXINT RXBDOD CART
$DATA      nonmemdata_HIV_arrow_log.csv IGNORE=@
$SUBROUTINE ADVAN13 TOL=9
$MODEL     COMP(CD4) COMP(VL)
$PK

;____ Initiate population level parameters for the model ______________
;; CD4 Parameters
```

```
TVL0  = THETA(1)
TVD0  = THETA(2)
TVP0  = THETA(3)
TVINT = THETA(4)
TVCD  = THETA(5)
TVCP  = THETA(6)
;; Viral Load parameters
TVVL0  = THETA(7)
TVVK   = THETA(8)
TVV50  = THETA(9)
TVEDRUG = THETA(10)

;____ Modelling covariate effects on population level parameters ______
IF(DIED.EQ.0) L0DIED = 1
IF(DIED.EQ.1) L0DIED = (1 + THETA(15))
IF(CTR.EQ.1) INTCTR = 1
IF(CTR.EQ.2) INTCTR = (1 + THETA(14))
IF(CTR.EQ.3) INTCTR = (1 + THETA(13))
IF(CTR.EQ.0) INTCTR = (1 + THETA(12))
INTCART = ( 1 + THETA(11)*(CART - 0.34))

TVL0  = TVL0 * L0DIED
TVINT = TVINT * INTCART * INTCTR

;____ Mu modelling to improve model run-times ___________________________
MU_1  = LOG(TVL0)
MU_2  = LOG(TVD0)
MU_3  = LOG(TVP0)
MU_4  = LOG(TVINT)
MU_5  = LOG(TVCD)
MU_6  = LOG(TVCP)
MU_7  = LOG(TVVL0)
MU_8  = LOG(TVVK)
MU_9  = LOG(TVV50)
MU_10 = TVEDRUG

;____ Including individual random effects on parameters _______________
L     = EXP(MU_1 + ETA(1))
D     = EXP(MU_2 + ETA(2))
PR    = EXP(MU_3 + ETA(3))
INT   = EXP(MU_4 + ETA(4))
CD    = EXP(MU_5 + ETA(5))
CP    = EXP(MU_6 + ETA(6))
VL0   = EXP(MU_7 + ETA(7))
VK    = EXP(MU_8 + ETA(8))
V50   = EXP(MU_9 + ETA(9))
EDRUG =   MU_10 + ETA(10)

;____ Define initial concentration for each compartment _______________
```

```
A_0(1) = INT
A_0(2) = VL0

;____ Naive (V) and total (VT) CD4 T cell concentration with age ______
VT    = 924 + 2354*EXP(-0.001012*AGE)
V     = 496.5 + 2074*EXP(-0.000869*AGE)
VTART = 924 + 2354*EXP(-0.001012*ARTAGE)
VART  = 496.5 + 2074*EXP(-0.000869*ARTAGE)

;____ Including Age effects on parameters ______________________________
LAMBDA = L * 0.02*EXP(-0.00027*AGE) * V * 0.905
LOSS   = D * 0.02*EXP(-0.00027*AGE)
PROL   = PR * 0.02*EXP(-0.00027*AGE)
DELTA  = (0.02*EXP(-0.00027*ARTAGE)/VL0) *(VART*0.905*L/INT - D*EXP(CD*(
    INT/VTART-1)) + PR*EXP(CP*(1-INT/VTART)))
DRUG   = 1/(1+EXP(-EDRUG))

;____ Define the differential equations for the model ________________
$DES
DADT(1) = LAMBDA - A(1)*(LOSS * EXP(CD*(A(1)/VT-1)) - PROL * EXP(CP*(-A
    (1)/VT+1))) - DELTA*A(1)*A(2)
DADT(2) = (1-DRUG)*VK*A(2)*VL0/(VL0+V50) - VK*A(2) *A(2)/(A(2) + V50)

;____ Compare the model output (IPRED) to the observed data (Y) _______
$ERROR
IPRED=0
; VIRUS
IF(FLAG==2) IPRED = LOG(A(2))
IF(VLBLQ==0.AND.FLAG==2) F_FLAG = 0
IF(VLBLQ==0.AND.FLAG==2) Y = IPRED + EPS(1)
IF(VLBLQ==1.AND.FLAG==2) F_FLAG = 1
IF(VLBLQ==1.AND.FLAG==2) SD = SQRT(SIGMA(1,1))
IF(VLBLQ==1.AND.FLAG==2) Y = PHI((VLLOQ-IPRED)/SD)+1.0E-30
IF(VLBLQ==1.AND.FLAG==2) MDVRES=1
; CD4
IF(FLAG==1) IPRED = LOG(A(1))
IF(FLAG==1) Y = IPRED + EPS(2)

;____ Initial estimates for the population level parameters ___________
$THETA
 (0,0.25)          ; 1. L0, thymic output
 (0,0.5)           ; 2. D0, loss
 (0,0.3)           ; 3. P0, proliferation
 (0,460)           ; 4. X0, initial concentration CD4
 (0,1)             ; 5. CD, strength competition loss
 (0,1)             ; 6. CP, strength competition proliferation
 (0,200000)        ; 7. VL0, initial viral load
 (0,0.33)          ; 8. VK, rate viral load decline
 (0,400)           ; 9. V50, viral load multiphasic
```

```
 (-20,3,20)       ; 10. EDRUG, drug effect
 (-1.515,-0.001515) ; 11. INTCART1, ART age effects on X0
 (-1,-0.001)      ; 12. INTCTR2, Centre effects on X0
 (-1,-0.001,5)    ; 13. INTCTR3, Centre effects on X0
 (-1,-0.001,5)    ; 14. INTCTR0, Centre effects on X0
 (-1,-0.001, 10.00) ; 15. L0DIED1, Effects of DIED on L0

;____ Initial estimates for the random effects var-covar matrix _______
$OMEGA BLOCK(10)
 1
 0.1 1
 0.1 0.1 1
 0.1 0.1 0.1 1
 0.1 0.1 0.1 0.1 1
 0.1 0.1 0.1 0.1 0.1 1
 0.1 0.1 0.1 0.1 0.1 0.1 1
 0.1 0.1 0.1 0.1 0.1 0.1 0.1 1
 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 1
 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 1
;____ Initial estimate for the residual error variances ______________
$SIGMA BLOCK(2)
 1
 0.1 1

;____ Estimation algorithm to be used (EONLY=1 Expectation Only) ______
$ESTIMATION METHOD=IMP INTER LAPLACE GRD=SN(1) NITER=80 PRINT=10
   ISAMPLE=1000 CTYPE=1 NOABORT
$ESTIMATION METHOD=IMP INTER LAPLACE GRD=SN(1) EONLY=1 NITER=4 PRINT=1
   ISAMPLE=10000 IACCEPT=0.1 DF=0 ISAMPEND=10000 STDOBJ=2 MAPITER=0
   RANMETHOD=3S2 NOABORT

;____ Output results of the model-fitting into a table _________________
$TABLE     ID TIME DV EVID PRED IPRED FLAG VLBLQ CWRES ETA1 ETA2 ETA3
   ETA4 ETA5 ETA6 ETA7 ETA8 ETA9 ETA10 OBJI NOPRINT NOAPPEND ONEHEADER
   FILE=sdtab_HIV_arrow
```

A.6 CD4 concentration and viral load in HIV-infected children starting ART, viral load rebound in ARROW data

```
$PROBLEM   HIV infected children starting ART, ARROW data, viral load
   rebound
;____ Data Input and setup of problem _________________________________
$INPUT     ID TIME DV FLAG EVID VLBLQ VLLOQ L2 AGE ARTAGE CTR
$DATA      nonmemdata_HIV_arrow_viral_rebound_log.csv IGNORE=@
$SUBROUTINE ADVAN13 TOL=9
$MODEL     COMP(CD4) COMP(VL)
```

```
$PK

;____ Initiate population level parameters for the model ______________
;; CD4 Parameters
TVL0  = THETA(1)
TVD0  = THETA(2)
TVP0  = THETA(3)
TVINT = THETA(4)
TVCD  = THETA(5)
TVCP  = THETA(6)
;; Viral Load parameters
TVVL0  = THETA(7)
TVVK   = THETA(8)
TVV50  = THETA(9)
;; Drug effect parameters
TVEDRUG = THETA(10)
TVDRUGT = THETA(11)
TVEDRES = THETA(12)
TVDRATE = THETA(13)

;____ Modelling covariate effects on population level parameters ______
IF(CTR.EQ.3) DRUGTCTR = 1 ; Most common
IF(CTR.EQ.2) DRUGTCTR = ( 1 + THETA(15))
IF(CTR.EQ.0) DRUGTCTR = ( 1 + THETA(16))
IF(CTR.EQ.1) DRUGTCTR = ( 1 + THETA(17))

TVDRUGT = TVDRUGT * DRUGTCTR

;____ Mu modelling to improve model run-times ___________________________
MU_1  = LOG(TVL0)
MU_2  = LOG(TVD0)
MU_3  = LOG(TVP0)
MU_4  = LOG(TVINT)
MU_5  = LOG(TVCD)
MU_6  = LOG(TVCP)
MU_7  = LOG(TVVL0)
MU_8  = LOG(TVVK)
MU_9  = LOG(TVV50)
MU_10 =    TVEDRUG
MU_11 = LOG(TVDRUGT)
MU_12 =    TVEDRES
MU_13 = LOG(TVDRATE)

;____ Including individual random effects on parameters _______________
L     = EXP(MU_1 + ETA(1))
D     = EXP(MU_2 + ETA(2))
PR    = EXP(MU_3 + ETA(3))
INT   = EXP(MU_4 + ETA(4))
CD    = EXP(MU_5 + ETA(5))
```

```
CP    = EXP(MU_6 + ETA(6))
VL0   = EXP(MU_7 + ETA(7))
VK    = EXP(MU_8 + ETA(8))
V50   = EXP(MU_9 + ETA(9))
EDRUG =    MU_10+ ETA(10)
DRUGT = EXP(MU_11+ ETA(11))
EDRES =    MU_12+ ETA(12)
DRATE = EXP(MU_13+ ETA(13))

;____ Define initial concentration for each compartment _____________
A_0(1) = INT
A_0(2) = VL0

;____ Naive (V) and total (VT) CD4 T cell concentration with age ______
VT    = 924 + 2354*EXP(-0.001012*AGE)
V     = 496.5 + 2074*EXP(-0.000869*AGE)
VTART = 924 + 2354*EXP(-0.001012*ARTAGE)
VART  = 496.5 + 2074*EXP(-0.000869*ARTAGE)

;____ Including Age effects on parameters ____________________________
LAMBDA = L * 0.02*EXP(-0.00027*AGE) * V * 0.905
LOSS   = D * 0.02*EXP(-0.00027*AGE)
PROL   = PR * 0.02*EXP(-0.00027*AGE)
DELTA  = (0.02*EXP(-0.00027*ARTAGE)/VL0) *(VART*0.905*L/INT - D*EXP(CD*(
    INT/VTART-1)) + PR*EXP(CP*(1-INT/VTART)))
DRUG   = 1/(1+EXP(-EDRUG))
DRES   = 1/(1+EXP(-EDRES))


;____ Define the drug resistance equation ____________________________
$DES
DRUGRES = DRUG*(DRES + (1-DRES)/(1+EXP((T-DRUGT)/DRATE)))

;____ Define the differential equations for the model _________________
DADT(1) = LAMBDA - A(1)*(LOSS * EXP(CD*(A(1)/VT-1)) - PROL * EXP(CP*(-A
    (1)/VT+1))) - DELTA*A(1)*A(2)
DADT(2) = (1-DRUG*(DRES + (1-DRES)/(1+EXP((T-DRUGT)/DRATE-4))))*VK*A(2)
    *VL0/(VL0+V50) - VK*A(2) *A(2)/(A(2) + V50)


;____ Compare the model output (IPRED) to the observed data (Y) _______
$ERROR
IPRED=0
; VIRUS
IF(FLAG==2) IPRED = LOG(A(2))
IF(VLBLQ==0.AND.FLAG==2) F_FLAG = 0
IF(VLBLQ==0.AND.FLAG==2) Y = IPRED + EPS(1)
IF(VLBLQ==1.AND.FLAG==2) F_FLAG = 1
IF(VLBLQ==1.AND.FLAG==2) SD = SQRT(SIGMA(1,1))
```

```
IF(VLBLQ==1.AND.FLAG==2) Y = PHI((VLLOQ-IPRED)/SD)+1.0E-30
IF(VLBLQ==1.AND.FLAG==2) MDVRES=1
; CD4
IF(FLAG==1) IPRED = LOG(A(1))
IF(FLAG==1) Y = IPRED + EPS(2)

;____ Initial estimates for the population level parameters __________
$THETA
 (0,0.103837)  FIX ; 1. L0, thymic output
 (0,0.461765)  FIX ; 2. D0, loss
 (0,0.175663)  FIX ; 3. P0, proliferation
 (0,288.4140)  FIX ; 4. X0, initial concentration CD4
 (0,1.178360)  FIX ; 5. CD, strength competition loss
 (0,1.019090)  FIX ; 6. CP, strength competition proliferation
 (0,337370.0)  FIX ; 7. VL0, initial viral load
 (0,0.258776)  FIX ; 8. VK, rate viral load decline
 (0,750.5140)  FIX ; 9. V50, viral load mulitphasic
 (-20,3.693170) FIX ; 10. EDRUG, drug effect
 (0,363.666)       ; 11. DRUGT, time at which resistance develops
 (-20,-2.99475,20) ; 12. EDRES, residual drug effect
 (0,9.41068)       ; 13. DRATE, rate of development of resistance
 (-1,0.930066)     ; 14. DRUGTCTR2, Centre effects on rebound time
 (-1,-0.297463)    ; 15. DRUGTCTR0, Centre effects on rebound time
 (-1,0.198859)     ; 16. DRUGTCTR1, Centre effects on rebound time

;____ Initial estimates for the random effects var-covar matrix _______
BLOCK(10) FIX
 2.484630
 1.258550 2.173280
 0.287249 1.092420 0.878814
 0.385209 -1.08255 -0.86133 1.661450
 -0.30083 -0.03684 -0.03540 -0.28335 0.630546
 0.210169 -0.07209 -0.19530 0.229530 -0.16358 0.420285
 -0.11873 -0.00676 0.024708 -0.22191 0.158112 -0.07459 2.146130
 -0.04519 0.001880 0.008743 -0.06100 0.005354 -0.01568 0.053634
    0.094765
 0.260986 -0.46731 -0.26785 0.528847 -0.05784 0.031735 0.901993
    0.080490 1.393940
 0.324752 -0.07808 -0.16514 0.461696 0.048112 0.079234 0.460435
    -0.03956 0.861229 1.281940
$OMEGA 0.67719
$OMEGA 2.46763
$OMEGA 0.01  FIX

;____ Initial estimate for the residual error variances ______________
$SIGMA BLOCK(2)
 1.45644
 0.00113859 0.0972418
```

```
;____ Estimation algorithm to be used (EONLY=1 Expectation Only) ______
$ESTIMATION METHOD=IMP INTER LAPLACE GRD=SN(1) NITER=80 PRINT=10
   ISAMPLE=1000 CTYPE=1 NOABORT
$ESTIMATION METHOD=IMP INTER LAPLACE GRD=SN(1) EONLY=1 NITER=4 PRINT=1
   ISAMPLE=10000 IACCEPT=0.1 DF=0 ISAMPEND=10000 STDOBJ=2 MAPITER=0
   RANMETHOD=3S2 NOABORT

;____ Output results of the model-fitting into a table _________________
$TABLE     ID TIME DV EVID PRED IPRED FLAG VLBLQ CWRES DRUGRES NOPRINT
   NOAPPEND ONEHEADER FILE= sdtab_HIV_arrow_viral_rebound
```

Appendix B

R and NONMEM scripts for predicting long-term CD4 reconstitution following paediatric HSCT

B.1 R script

```
library(lattice)
library(FME)
library(latticeExtra)
rm(list=ls())

#########################################################################
##### Input Patient Characteristics ####################################
# Patient Hostpital Number or another numeric identifier
Patient_ID=1004

# The Age (in days) on the day of the HSCT
Age_at_HSCT = 4676

## Input diagnosis:
# 1: Leukaemia (any form)
# 2. Any other diagnosis
Diagnosis = 2

## Conditioning details:
# 1: Alemtuzumab
# 2. ATG
# 3. any other form of conditioning (e.g. TBI, ACD45, other drugs etc)
# 4. No conditioning at all
Conditioning = 1

## Acute GvHD Status
```

```
# 1: Never had acute GvHD
# 2: Had acute GvHD
Acute_GvHD_status = 1


##############################################################################
##### Input CD4 concentration data ###########################################
# Formats:
# Times in days following HSCT (with day 0 as the day of HSCT)
# Concentrations in cells/uL (i.e. should be roughly in the range of 10
    to 4000)
Times = c(20,176,204,239,281,323)
Concentrations = c(10,120,160,340,680,1010)


##############################################################################
##### Simulation of data characteristics #####################################
# The number of sample parameter values formed:
#   - More samples gives a more accurate representation
#   - Fewer samples is faster
#   - Somewhere between 300 and 1000 is recomended
samples=500

# The size of the confidence interval of the trajectory to plot (%):
Confidence_Interval = 68


##############################################################################
##### Model File Name ########################################################
Model = "CD4_individual_prediction"


##############################################################################
##### Set directory for modelling ############################################
Directory = "~/Documents/R/CD4/Predictions"
# Model file should be saved in this directory in format: runModel.mod


##############################################################################
##### Create directory for individual patient for prediction #########
Individual_Directory = paste(Directory,"/Patient_",Patient_ID,sep="")
system(paste("mkdir ",Individual_Directory,sep=""))
setwd(Individual_Directory)



##############################################################################
##############################################################################
##### Carry out parameter estimation and simulation ##################
```

```
#########################################################################

#########################################################################
##### Create data set given information above ###########################
nonmem_data <- data.frame(
ID = Patient_ID,
TIME = c(0,Times),
DV = c(0,log(Concentrations)),
EVID = c(2,rep(0,length(Times))),
AGE = c(0,Times) + Age_at_HSCT,
GVHYN = 0,
COND=0,
ALEM=0,
ATG=0,
LEUK=0)
if (Conditioning == 1){nonmem_data$ALEM <- 1}
if (Conditioning == 2){nonmem_data$ATG <- 1}
if (Conditioning == 4){nonmem_data$COND<-1}
if (Diagnosis == 1){nonmem_data$LEUK <- 1}
if (Acute_GvHD_status == 2){nonmem_data$GVHYN <- 1}


names(nonmem_data)[1] <- "#ID"
write.csv(nonmem_data,"nonmemdata_log_predictions.csv",row.names=FALSE,
    quote=FALSE)


#########################################################################
##### Carry out model fitting in NONMEM #################################
system(paste("cp ",Directory,"/run",Model,".mod run",Model,".mod",sep
    =""))
system(paste("execute run",Model,".mod -directory=NM_run",Model,sep="")
    )
system(paste("rm -r NM_run",Model,sep=""))


#########################################################################
##### Read back in NONMEM results #######################################
phi <- read.table(paste("run",Model,".phi",sep=""),header=T,skip=1)
sigma <- read.table(paste("run",Model,".ext",sep=""),header=T,skip=1)
    [1,"SIGMA.1.1."]

#########################################################################
##### Create sample parameter sets from NONMEM results ##################
nparam=6 # Number of parameters with random effects
varcovar_list <- list()
for (i in 1:nparam){
       for (j in 1:nparam){
              if (j>i) {name <- paste("PHC",j,i,"",sep=".")}
```

```
                else {name <- paste("PHC",i,j,"",sep=".")}
                varcovar_list <- append(varcovar_list,name)}}
varcovar_list <- unlist(varcovar_list)

mean_list <- list()
for (i in 1:nparam){
       mean_name <- paste("PHI",i,"",sep=".")
       mean_list=append(mean_list,mean_name)}
mean_list <- unlist(mean_list)

parMean <- unlist(phi[,mean_list])
parCovar <- matrix(data=unlist(phi[,varcovar_list]),nrow=nparam)

sample_params <- data.frame(Norm(parMean=parMean,parCovar=parCovar,num=
   samples))
names(sample_params) <- c("lambda","d","p","X0","LH","LR")
real_sample_params <- exp(sample_params)


#########################################################################
##### Simulate data from sample parameter sets ##########################

sample_param <- real_sample_params[1,]
pars <- list(lambda = sample_param$l,
                            d   = sample_param$d,
                            p   = sample_param$p,
                            X0  = sample_param$X0,
                            LH  = sample_param$LH,
                            LR  = sample_param$LR)

#Define model to solve, including time range, differential equations,
   starting estimates for compartments
solveCD4 <- function(pars, times=seq(0,3000,by=10)){
       derivs <- function(t, state, pars) {
              with(as.list(c(state,pars)),{
                     Age = Age_at_HSCT
                     V = (497+2070*exp(-0.00087*(t+Age)))
                     V2 <- (924+2354*exp(-0.001012*(t+Age)))
                     L = lambda*exp(-0.00027*(t+Age))*V*0.02*(1-exp(-2*
                         t/LH))/(1+exp(LR*(1-t/LH)))*0.02/0.0221
                     D = d*0.018*exp(-0.00027*(t+Age))*exp(1*(X/V2-1))
                     P = p*0.02 *exp(-0.00027*(t+Age))*exp(1*(1-X/V2))
                     dX <- L - X*(D-P)
                     return(list(c(dX)))
              })
       }
       state <- c(X = pars$X0)
       return(ode(y=state,times=times,func=derivs,parms=pars))
}
```

```
# Carry out estimation from model
out <- data.frame(solveCD4(real_sample_params[1,]))

for (i in 2:samples){
       out_new <- data.frame(solveCD4(real_sample_params[i,]))
       out[,i+1] <- out_new$X
       }

#########################################################################
#########################################################################
##### Plot Results ######################################################
#########################################################################


#########################################################################
##### Calculate confidence intervals for plots ##########################
CI.lower= (50 - Confidence_Interval*0.5)/100.0
CI.higher=(50 + Confidence_Interval*0.5)/100.0

#xyplot(X~time,out,pch=".",scales=list(y=list(log=T)))
out$lower <- apply(out[,2:ncol(out)],1,quantile,probs=c(CI.lower))*exp
   (-sigma)
out$median <- apply(out[,2:ncol(out)],1,quantile,probs=c(0.5))
out$higher <- apply(out[,2:ncol(out)],1,quantile,probs=c(CI.higher))*
   exp(sigma)
out <- as.data.frame(out[,c("time","lower","median","higher")])

write.csv(out,paste("Predicted Trajectories",Patient_ID,".csv",sep=""))
#########################################################################
##### Include CD4 concentration with age ################################
CD4_expected <- data.frame(
AGE = 1:8000)
CD4_expected$CD4 <- 924+2354*exp(-0.001012*CD4_expected$AGE)
CD4_expected$CI05 <- 0.469*CD4_expected$CD4
CD4_expected$CI95 <- 1.687*CD4_expected$CD4
out$AGE <- Age_at_HSCT + out$time
out <- merge(out,CD4_expected)
out$min <- do.call(pmin,data.frame(out$lower,out$CI05))
out$time <- out$time/30.5

#########################################################################
##### Plot results and CD4 concentration for age and save
   ##############
pdf(file=paste("Predicted trajectory ",Patient_ID,".pdf",sep=""),width
   =8, height=5)
xyplot(CD4~time, out,type="l", scales=list(x=list(limit=c(0,39), at=c
   (6,12,18,24,30,36)), y=list(log=T,limit=c(3,6000), at=c
   (10,30,100,300,1000,3000))), xlab=list(label="Time after HSCT (
```

```
   months)", cex=1.5), ylab = list(label=expression(paste("CD4
   concentration (cells/",mu,"L)")),cex=1.5)) +
       as.layer(xyplot(CI95~time,out,type="l",lty=2, scales=list(y=list
          (log=T)), panel=panel.xyarea, border=0,origin=0,aspect='xy',
          col=rgb(0,0,1,0.2))) +
       as.layer(xyplot(CI05~time,out,type="l",lty=2, scales=list(y=list
          (log=T)), panel=panel.xyarea, border=0,origin=0,aspect='xy',
          col="white", col.line="white")) +
       as.layer(xyplot(higher~time,out,type="l",lty=2, scales=list(y=
          list(log=T)), panel=panel.xyarea, border=0,origin=0,aspect='
          xy',col=rgb(0.1,1,0,0.3))) +
       as.layer(xyplot(lower~time,out,type="l",lty=2, scales=list(y=
          list(log=T)), panel=panel.xyarea, border=0,origin=0,aspect='
          xy',col="white"))+
       as.layer(xyplot(lower~time,out,type="l",lty=2, scales=list(y=
          list(log=T)), panel=panel.xyarea, border=0,origin=0,aspect='
          xy',col=rgb(0,0,1,0.2))) +
       as.layer(xyplot(min~time,out, scales=list(y=list(log=T)), panel=
          panel.xyarea,border=0, origin=0,aspect='xy',col="white", col.
          line="white"))+
       as.layer(xyplot(higher~time,out,type="l",lty=2, col="black",
          scales=list(y=list(log=T)))) +
       as.layer(xyplot(lower~time,out,type="l",lty=2,col="black",
          scales=list(y=list(log=T)))) +
       as.layer(xyplot(median~time,out,type="l",lty=1,col="black",
          scales=list(y=list(log=T)))) +
       as.layer(xyplot(exp(DV)~TIME/30.5,nonmem_data,cex=0.7, type="p",
           col="black", scales=list(y=list(log=T))))
dev.off()

#########################################################################
#########################################################################
```

B.2 NONMEM script

```
$PROBLEM   Paediatric HSCT CD4 T cell Reconstitution Individual
   Predictions
;____ Data Input and setup of problem _____________________________
$INPUT     ID TIME DV EVID AGE GVHYN COND ALEM ATG LEUK
$DATA      nonmemdata_log_predictions.csv IGNORE=@
$SUBROUTINE ADVAN13 TOL=9
$MODEL     COMP=(CENTRAL)
$PK

;____ Initiate population level parameters for the model ______________
TVL   = THETA(1)
TVD   = THETA(2)
TVPR  = THETA(3)
```

```
TVINT = THETA(4)
TVLH  = THETA(5)
TVLR  = THETA(6)

;____ Modelling covariate effects on population level parameters ______
IF(ALEM.EQ.0) INTALEM = 1
IF(ALEM.EQ.1) INTALEM = (1 + THETA(7))
IF(ATG.EQ.0)  INTATG  =  1
IF(ATG.EQ.1)  INTATG  = (1 + THETA(8))
IF(GVHYN.EQ.0) INTGVHYN = 1
IF(GVHYN.EQ.1) INTGVHYN = (1 + THETA(9))
IF(LEUK.EQ.0) LLEUK   =  1
IF(LEUK.EQ.1) LLEUK   = (1 + THETA(10))
IF(COND.EQ.0) PRCOND  =  1
IF(COND.EQ.1) PRCOND  = (1 + THETA(11))

TVINT = TVINT * INTALEM * INTATG * INTGVHYN
TVL   = TVL * LLEUK
TVPR  = TVPR * PRCOND

;____ Mu modelling to improve model run-times ___________________________
MU_1 = LOG(TVL)
MU_2 = LOG(TVD)
MU_3 = LOG(TVPR)
MU_4 = LOG(TVINT)
MU_5 = LOG(TVLH)
MU_6 = LOG(TVLR)

;____ Naive (V) and total (VT) CD4 T cell concentration with age ______
VT = 924 + 2354*EXP(-0.001012*AGE)
V = 496.5 + 2074*EXP(-0.000869*AGE)

;____ Including Age effects and individual random effects on params ___
L   = EXP(MU_1 + ETA(1)) * V * 0.9 * 0.02*EXP(-0.00027*AGE)
D   = EXP(MU_2 + ETA(2)) * 0.02*EXP(-0.00027*AGE)
PR  = EXP(MU_3 + ETA(3)) * 0.02*EXP(-0.00027*AGE)
INT = EXP(MU_4 + ETA(4))
LH  = EXP(MU_5 + ETA(5))
LR  = EXP(MU_6 + ETA(6))
CD  = 1
CP  = 1
A_0(1) = INT

;____ Define the differential equation for the model ___________________
$DES
DADT(1) = L * (1-EXP(-2*T/LH))/(1+EXP(LR*(1-T/LH))) - A(1)*(D*EXP(CD*(A
    (1)/VT-1)) - PR * EXP(CP*(-A(1)/VT+1)))

;____ Compare the model output (IPRED) to the observed data (Y) _______
```

```
$ERROR
IPRED = A(1)
IPRED = LOG(IPRED)
Y = IPRED + EPS(1)

;____ Initial estimates for the population level parameters __________
$THETA
 (0,0.222489) ; 1. L, thymic output
 (0,0.454536) ; 2. D, loss
 (0,0.195179) ; 3. PR, proliferation
 (0,164.2850) ; 4. A_0, Initial concentration of cells
 (0,133.1450) ; 5. LH, time to recovery in thymic output
 (0,9.757030) ; 6. LR, rate of recovery of thymic output
 (-1,-0.83722) ; 7. INTALEM, effects of alemtuzumab on A_0
 (-1,-0.93890) ; 8. INTATG, effects of ATG on A_0
 (-1,0.331920) ; 9. INTGVHYN, effects of acute GvHD on A_0
 (-1,1.289310) ; 10. LLEUK, effects of leukaemia on thymic output
 (-1,-0.88522) ; 11. PRCOND, effects of no conditioning on
    proliferation

;____ Initial estimates for the random effects var-covar matrix _______
$OMEGA BLOCK(6)
 1.553010
 0.522870 1.662840
 0.161760 0.320213 0.247498
 0.405366 0.415663 0.218925 1.286390
 0.534060 -0.44269 -0.06167 0.766879 1.214760
 0.083104 0.340042 -0.13533 -0.89795 -0.77866 1.259670

;____ Initial estimate for the residual error variance _____________
$SIGMA 0.218784

;____ Estimation algorithm to be used (EONLY=1 Expectation Only) ______
$ESTIMATION METHOD=IMP INTERACTION EONLY=1 ISAMPLE=10000 IACCEPT=0.1
   NITER=10 PRINT=1

;____ Output results of the model-fitting into a table _________________
$TABLE     ID TIME DV PRED IPRED CWRES EVID ETA1 ETA2 ETA3 ETA4 ETA5
   ETA6 OBJI NOPRINT NOAPPEND ONEHEADER FILE=
   sdtab_individual_predictions
```

Appendix C

Python code

C.1 Global competition model

```
## Global competition model with thymus
## File output of lifetimes and clonotype distributions
## And with automatic graphs of the distributions

from numpy import random,array,log,exp,sqrt,mean,arange,cumsum,
    searchsorted,where,histogram,concatenate,loadtxt,transpose
from pylab import show,savefig,ion,arange,histogram,step,draw,figure,
    plot,ylabel,zeros,hist,xlabel,ylim,title,text,bar,subplot,
    tight_layout
import datetime,time,sys
from scipy.special import *

#########################################################
# Give folder name for this run
runname = 'Global_thymus'
#########################################################
### Set variables, or read in from command line
#########################################################
if len(sys.argv)==1:
    n0 = 8
    delta = 80.0
    gamma = 4000.0
    folder_name = 'Global_thymus'
else:
    n0 = int(sys.argv[1])
    delta = float(sys.argv[2])
    gamma= float(sys.argv[3])
    folder_name= str(sys.argv[4])
    today_start = str(sys.argv[5])
mu = 1.0
N = int(round(gamma/n0 + delta))
tmax = 2000.0
tint = tmax/50.0
```

```
TcellNumb = [n0]*N
t=0.0
tt=-0.5*tint
birthtime = [0.0]*N
NNow=N

############################################################
### Find start time/date and open file for output
############################################################
todayraw = datetime.datetime.today()
today = todayraw.strftime("%b%d%y %Hh%M.%S")
life_t_file = open(folder_name+'/lifetimes_'+runname+'_'+str(gamma)+'_
    '+str(n0)+'_'+str(delta)+'.txt', 'w')
life_t_file.write( 'N=%d n0=%d delta=%.2f gamma=%.2f tmax=%.2f\n' %(N,
    n0,delta,gamma,tmax))
life_t_file.write( ' expected mean number of cells=%.2f\n' % ((gamma+
    delta*n0)/mu))
clono_file = open(folder_name+'/clonotype_distributions_'+runname+'_'+
    str(gamma)+'_'+str(n0)+'_'+str(delta)+'.txt', 'w')
clono_file.write('t\t')
for i in range(99):
    clono_file.write('%d\t'%i)
print 'N=',N,'n0=',n0,'delta=',delta,'gamma=',gamma,'tmax=',tmax
print 'expected mean number of cells=' ,(gamma+delta*n0)/mu

############################################################
### Gillespie Algorithm for the dynamics
############################################################
def Gillespiestep(gamma,mu,delta):
    '''One step of the Gillespie algorith'''
    TcellSum = sum(TcellNumb)
    Lamb = [x*gamma/TcellSum for x in TcellNumb]
    deathrate = [x*mu for x in TcellNumb]
    rates = deathrate+Lamb+[delta] # concatenates arrays
    ratesum = sum(rates)
    urv = random.uniform()
    i = searchsorted(cumsum(rates)/ratesum,urv)
    idead=NNow
    if i < NNow:
        TcellNumb[i]-=1 # Loss
        if TcellNumb[i] == 0:
            idead = i
            TcellNumb.pop(i) # Clonotype extinction
    elif i==2*NNow:
        TcellNumb.append(n0) # New clonotype
    else:
        i = i-NNow
        TcellNumb[i] += 1 # Division
    return i,idead,ratesum
```

```
############################################################
### Carry out the actual dynamics
############################################################
start=time.clock()
print start
while t < tmax:
    i,idead,ratesum = Gillespiestep(gamma,mu,delta)
    tinc = -log(random.uniform())/ratesum
    if idead != NNow:
        NNow-=1
        life_t_file.write("%.3f\t%.3f\n" %(birthtime[i],t-birthtime[i]))
        birthtime.pop(i)
    elif i==2*NNow:
        birthtime.append(t)
        NNow+=1
    if tt>tint:
        tt=0
        celldens = histogram(TcellNumb, bins=(arange(100)+1))
        clono_file.write('\n%.3f\t'%t)
        for tcells in celldens[0]:
            clono_file.write('%d\t'%tcells)
        life_t_file.flush()
        clono_file.flush()
        print t,NNow,sum(TcellNumb)
        timenow = time.clock()
        timetaken=timenow-start
        print timetaken
        start=timenow
    t+=tinc
    tt+=tinc

############################################################
### Output data into files and close files
############################################################
clono_file.write('\n%.3f\t'%t)
for tcells in celldens[0]:
    clono_file.write('%d\t'%tcells)
clono_file.close()
life_t_file.close()

############################################################
### Create histogram of distribution of clonotype lifetimes
############################################################
alpha = delta*n0/gamma
gamma_E = 0.57721566490
T = 1.0/(alpha*mu)*(gamma_E - exp(alpha*n0)*expi(-alpha*n0) + log(alpha
    *n0))
mybins=arange(101)
```

```
lifetime = loadtxt(folder_name+'/lifetimes_'+runname+'_'+str(gamma)+'_
    '+str(n0)+'_'+str(delta)+'.txt', skiprows=2)
lifetime = transpose([lifetime for lifetime in lifetime if lifetime
    [0]!=0])
subplot(2,1,1)
ydens=histogram(lifetime[1], bins=mybins, normed=True)[0]
hist(lifetime[1], mybins, normed=True)
x = (arange(1000)+1)*0.1
pred=(n0/x**2)*exp(-n0/x)
l=plot(x,pred, linewidth=2)
xlabel('Clonotype lifetime')
title(r'Histogram of clonotype lifetimes')
text(75,0.02,'$N=%d$ \n $n_0=%d$ \n $\delta =%.0f$ \n $\gamma=%.0f$ \n
    $\mu=%.0f$ \n $t_{max}=%.0f$'%(N,n0,delta,gamma,mu,tmax), fontsize
    =14)
############################################################
### Create plot of clonotype sizes
############################################################
subplot(2,1,2)
clonotypes = loadtxt(folder_name+'/clonotype_distributions_'+runname+'_
    '+str(gamma)+'_'+str(n0)+'_'+str(delta)+'.txt', skiprows=1).T
clonotypes_t = transpose(clonotypes[1:])
mean_clonos = sum(clonotypes_t)/sum(sum(clonotypes_t))
ind=arange(len(mean_clonos))
width=1
p1 = bar(ind,mean_clonos,width,color='b')
x1=((arange(n0*10)+1)*0.1)
pred1=(1/(T*alpha))*(1-exp(-alpha*x1))/x1
l1=plot(x1,pred1,linewidth=2,color='g')
x2=n0+arange(900)*0.1
pred2=(1/T)*(exp(alpha*n0)-1)*exp(-alpha*x2)/(alpha*x2)
l2=plot(x2,pred2,linewidth=2,color='g')
xlabel('T cells per clonotype')
title(r'Histogram of clonotype sizes')
tight_layout(True)

############################################################
### Save figures
############################################################
today_date = todayraw.strftime("%b%d%y")
savefig(folder_name+'/'+str(today_date)+'_lifetime_size_%.0f_%.0f_%.0f.
    pdf'%(gamma,n0,delta))

clono_n_file_1 = open(folder_name+'/clono_totals_'+str(n0)+'_'+str(
    gamma)+'.txt', 'a')
clono_totals = sum(clonotypes[1:])
for clono_total in clono_totals:
    clono_n_file_1.write('%d\t'%clono_total)
clono_n_file_1.close()
```

C.2 Global competition model, including distributions of resource affinity, as applied to mouse data for CD8 memory T cells

```
## Global competition model with thymus
## Set up as for mice experiments for CD8 memory T cells donor & host
## With gamma ditributed

from numpy import random,array,log,exp,sqrt,mean,arange,cumsum,
    searchsorted,where,histogram,concatenate,loadtxt,transpose,append
from pylab import show,savefig,ion,arange,histogram,step,draw,figure,
    plot,ylabel,zeros,hist,xlabel, xlim,ylim,title,text,bar,subplot,
    tight_layout,legend,close,ion,ioff, rcParams
import datetime,time,sys
from scipy.special import *

############################################################
# Give folder name for this run
runname = 'Mice_global'
############################################################
### Set variables, or read in from command line
############################################################
if len(sys.argv)==1:
    n0=3
    N = 1500
    n_th = 3
    delta = 120.0
    gamma = 4000.0
    gamma_donor=800.0
    folder_name = 'Mice'

else:
    N = int(sys.argv[1])
    delta = float(sys.argv[2])
    gamma= float(sys.argv[3])
    folder_name= str(sys.argv[4])
    today_start = str(sys.argv[5])

mu = 0.8
tmax = 49.0
tint = 0.5
ttrans = 8.0
TcellNumb = [n0]*N
t=0.0
tt=-0.5*tint
birthtime = [0.0]*N
Ninit=N
NNow=N
```

```
NHost=N
NDonor=0
Host_or_donor = [1]*N
t_list=[0]
NDonor_list=[0]
NHost_list=[N]
T_donor_list=[0]
T_host_list=[sum(TcellNumb)]
gamma_i = [-1]*N
for i in range(N):
    while gamma_i[i]<=0.0 or gamma_i[i]>=100.0:
        gamma_i[i] = random.normal(50,30)
mybins=3*arange(34)

#############################################################
### Find start time/date and open file for output
#############################################################
todayraw = datetime.datetime.today()
today = todayraw.strftime("%b%d%y %Hh%M.%S")
print 'N=',N,'n0=',n0,'n_th=',n_th,'delta=',delta,'gamma=',gamma,'tmax
    =',tmax
print 'expected mean No cells=' ,(gamma+delta*n_th)/mu, "Init no cells
    =",sum(TcellNumb)

#############################################################
### Gillespie Algorithm for the dynamics
#############################################################
def Gillespiestep(gamma,mu,delta):
    '''One step of the Gillespie algorith'''
    gammaNormaliser = sum([gamma_i[i]*TcellNumb[i] for i in range(NNow)
        ])
    Lamb = [TcellNumb[i]*gamma_i[i]*gamma/(gammaNormaliser) for i in
        range(NNow)]
    deathrate = [x*mu for x in TcellNumb]
    rates = deathrate+Lamb+[delta] # concatenates arrays
    ratesum = sum(rates)
    urv = random.uniform()
    i = searchsorted(cumsum(rates)/ratesum,urv)
    idead=NNow
    if i < NNow:
        TcellNumb[i]-=1 # Loss
        if TcellNumb[i] == 0:
            idead = i
            TcellNumb.pop(i) # Clonotype extinction
    elif i==2*NNow:
        TcellNumb.append(n_th) # New clonotype
    else:
        i = i-NNow
        TcellNumb[i] += 1 # Division
```

```
    return i,idead,ratesum

Tcell_gamma_dens = sum([[gamma_i[i]]*TcellNumb[i] for i in range(N)
    ],[])
gamma_dens = histogram(Tcell_gamma_dens, bins=mybins)[0]
gamma_dens = append(gamma_dens,[0])
gamma_list = [gamma_dens]

##########################################################
### Carry out the actual dynamics
##########################################################
start=time.clock()
print start
while t < ttrans:
    i,idead,ratesum = Gillespiestep(gamma,mu,delta)
    tinc = -log(random.uniform())/ratesum
    if idead != NNow:
        NNow-=1
        NHost -= 1
        birthtime.pop(i)
        Host_or_donor.pop(i)
        gamma_i.pop(i)
    elif i==2*NNow:
        birthtime.append(t)
        Host_or_donor.append(1)
        new_gamma=-1
        while new_gamma<=0 or new_gamma > 100:
            new_gamma = random.normal(50,30)
        gamma_i.append(new_gamma)
        NNow +=1
        NHost += 1
    if tt>tint:
        tt=0
        TcellNumb_host=0
        TcellNumb_donor=0
        TcellNumb_host=sum(TcellNumb)
        print TcellNumb_host, TcellNumb_donor
        T_donor_list.append(TcellNumb_donor)
        T_host_list.append(TcellNumb_host)
        t_list.append(t)
        NDonor_list.append(NDonor)
        NHost_list.append(NHost)
        print t,NNow,sum(TcellNumb)
        timenow = time.clock()
        timetaken = timenow-start
        Tcell_gamma_dens = sum([[gamma_i[i]]*TcellNumb[i] for i in range
            (NNow)],[])
        print len(Tcell_gamma_dens)
        gamma_dens = histogram(Tcell_gamma_dens, bins=mybins)[0]
```

```
        gamma_dens = append(gamma_dens,[0])
        start=timenow
    t+=tinc
    tt+=tinc

gamma_list.append(gamma_dens)
gamma_dens_donor = [0]*len(mybins)
gamma_list_donor = [gamma_dens_donor]
ttt=0.0

while t < tmax:
    i,idead,ratesum = Gillespiestep(gamma,mu,delta)
    tinc = -log(random.uniform())/ratesum
    if idead != NNow:
        NNow-=1
        if Host_or_donor[idead] == 0:
            NDonor -= 1
        if Host_or_donor[idead] == 1:
            NHost -= 1
        birthtime.pop(i)
        Host_or_donor.pop(i)
        gamma_i.pop(i)
    elif i==2*NNow:
        birthtime.append(t)
        Host_or_donor.append(0)
        new_gamma=-1
        while new_gamma<=0 or new_gamma > 100:
#           new_gamma = random.uniform(0,100)
            new_gamma = random.normal(50,30)
        gamma_i.append(new_gamma)
        NNow +=1
        NDonor += 1
    if tt>tint:
        tt=0
        TcellNumb_host=0
        TcellNumb_donor=0
        for i in range(len(Host_or_donor)):
            if Host_or_donor[i]==1:
                TcellNumb_host += TcellNumb[i]
            if Host_or_donor[i]==0:
                TcellNumb_donor += TcellNumb[i]
        print TcellNumb_host, TcellNumb_donor
        T_donor_list.append(TcellNumb_donor)
        T_host_list.append(TcellNumb_host)
        t_list.append(t)
        NDonor_list.append(NDonor)
        NHost_list.append(NHost)
        print t,NNow,sum(TcellNumb)
        Tcell_gamma_dens = [[gamma_i[i]]*TcellNumb[i] for i in range(
```

```
            NNow)]
        #print len(Tcell_gamma_dens)
        gamma_dens = histogram(sum([Tcell_gamma_dens[i] for i in range(
            NNow) if Host_or_donor[i] ==1],[]),bins=mybins)[0]
        gamma_dens = append(gamma_dens,[0])
        gamma_dens_donor = histogram(sum([Tcell_gamma_dens[i] for i in
            range(NNow) if Host_or_donor[i] ==0],[]),bins=mybins)[0]
        gamma_dens_donor = append(gamma_dens_donor,[0])
        timenow = time.clock()
        timetaken = timenow-start
        start=timenow
    if ttt>ttrans*2:
        ttt=0.0
        gamma_list.append(gamma_dens)
        gamma_list_donor.append(gamma_dens_donor)
    t+=tinc
    tt+=tinc
    ttt+=tinc
gamma_list.append(gamma_dens)

rcParams['figure.figsize'] = 10.5, 13.5

subplot(4,2,1)
plot(t_list,T_host_list, label='Host')
plot(t_list,T_donor_list, label='Donor')
xlabel('Time from birth (weeks)')
ylabel('Number of T cells')
ylim(0,1.1*max(T_host_list))
legend(labelspacing=0.02)

subplot(4,2,3)
plot(t_list,NHost_list, label='Host')
plot(t_list,NDonor_list, label='Donor')
xlabel('Time from birth (weeks)')
ylabel('Number of clonotypes')
ylim(0,1.1*max(NHost_list))
legend(labelspacing=0.02)

proportion_T = [1.0*T_donor_list[i]/(1.0*T_host_list[i]+1.0*
    T_donor_list[i]) for i in range(len(T_donor_list))]

subplot(4,2,2)
plot(t_list,proportion_T,c='black', label='Proportion')
xlabel('Time from birth (weeks)')
ylabel('Proportion donor T cells')
ylim(0,1.1)

proportion_C = [1.0*NDonor_list[i]/(1.0*NHost_list[i]+1.0*NDonor_list[i
    ]) for i in range(len(NDonor_list))]
```

```
subplot(4,2,4)
plot(t_list,proportion_C,c='black', label='Proportion')
xlabel('Time from birth (weeks)')
ylabel('Proportion donor clonotypes')
ylim(0,1.1)

subplot(4,2,5)
title('Birth (0 Weeks)')
step(mybins,gamma_list[0]/float(max(gamma_list[0]))*100, label='Host')
xlabel('gamma')
ylabel('% of Max')
legend(loc=2,labelspacing=0.02)

subplot(4,2,6)
title('Day of transplant (%.0f Weeks)' % ttrans)
step(mybins,gamma_list[1]/float(max(gamma_list[1]))*100, label='Host')
ylabel('% of Max')
xlabel('gamma')
legend(loc=2,labelspacing=0.02)

subplot(4,2,7)
title('%.0f Weeks' % (3*ttrans))
step(mybins,gamma_list[2]/float(max(gamma_list[2]))*100, label='Host')
step(mybins,gamma_list_donor[1]/float(max(gamma_list_donor[1]))*100,
    label='Donor')
xlabel('gamma')
ylabel('% of Max')
legend(loc=2,labelspacing=0.02)

subplot(4,2,8)
title('%.0f Weeks' % (5*ttrans))
step(mybins,gamma_list[3]/float(max(gamma_list[3]))*100, label='Host')
step(mybins,gamma_list_donor[2]/float(max(gamma_list_donor[2]))*100,
    label='Donor')
xlabel('gamma')
ylabel('% of Max')
legend(loc=2, labelspacing=0.02)

tight_layout()
savefig('Mice/Gamma distributions with donor nth='+str(n_th)+' quick mu
    =0.8.pdf')
show(block=True)
```