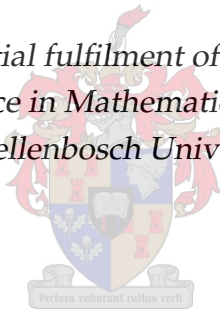


Immune Biomarker Reference Range Estimation for Healthy Paediatric Patients in South Africa

by

Steve Bicko Cygu

*Thesis presented in partial fulfilment of the requirements for the
degree of Master of Science in Mathematics in the Faculty of Science
at Stellenbosch University*



Department of Mathematical Sciences,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Supervisor: Prof. Martin. Nieuwoudt and Prof. Cang. Hui

March 2017

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: Steve Bicko Cygu

Date: March 2017

Copyright © 2017 Stellenbosch University
All rights reserved.

Abstract

Immune Biomarker Reference Range Estimation for Healthy Paediatric Patients in South Africa

Steve Bicko Cygu

*Department of Mathematical Sciences,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.*

Thesis: MSc. (Mathematics)

March 2017

Understanding and quantifying human peripheral blood T-lymphocyte immunophenotypes is necessary for the diagnosis and treatment of immune and haematological disorders. This is only possible if comparison to control-data from healthy subjects is available for all the biomarkers of interest. Historical empirical studies in industrialized countries have described normal reference ranges for such biomarkers in children by grouping particular age ranges into ‘age-blocks’. Since such markers change with age this has resulted in a loss of precision in determining whether patients that lie close to the limits of age-ranges are normal or not. Previous studies have relied on fitting single exponential models to such data, which makes the simple assumption of an exponential decline in cell markers with age. However, the counts of such markers have been observed to increase from birth to between 6 months to 12 months from birth and then decrease continuously with age. There is a dearth of reference range estimation methods which are age-continuous and incorporate biologically *mechanistic* models. A more ideal solution would be the development of appropriate mathematical models and model-based, age-continuous reference range estimation methods that describe such changes in a continuous manner. Such models may then be used to investigate the influence of population covariates on age-related changes in the biomarkers of interest. In this study, we employ paediatric data from a cohort of 381 healthy South African children. This is

cross-sectional in design and the biomarkers described include: CD3+, CD19+, CD8+, CD4+, ratio of CD4+ naive/memory, CD18+CD56+ and CD3-CD56+. Using weighted generalized nonlinear least squares, we fit and compare single and double exponential semi-mechanistic models. An ideal model is selected based on the Akaike's Information Criterion (AIC). The double exponential model is found to be the best fit for age-related changes in such biomarkers. This predicts that cell counts rise after birth to a maximum at approximately 12 months of age and decline in an exponential manner towards an asymptote in adulthood. This is in agreement with prior empirical and mechanistic studies. We extend the double exponential model to investigate the influence of particular covariates. The type of feeding in the first 6 months following birth is found to be the covariate with the greatest influence on age-related changes in the majority of the biomarkers investigated. A model-based method to estimate age-continuous reference ranges is then proposed. This assumes that particular reference ranges are a specified shift of the 'running' standard deviations of residuals away from a fitted central model function by a Z-score. We compare this method to reference ranges calculated using traditional centile curves. Centile curves demonstrate a simpler negative single-exponential decline as age advances and enable no *mechanistic* interpretation for this pattern. The models employed in this study may lead to the development of a laboratory tool by which individual cell-marker values may be compared to healthy age-continuous reference ranges.

Keywords: immunological cell biomarkers, healthy children, model-based, age-continuous reference ranges.

Opsomming

Die skatting van Immuun-Biomarker Verwysings Waardes vir gesonde Suid-Afrikaanse kinders

("Immune Biomarker Reference Range Estimation for Healthy Paediatric Patients in South Africa ")

Steve Bicko Cygu

*Departement Wiskundige Wetenskappe,
Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.*

Tesis: MSc. (Wiskunde)

Maart 2017

Die diagnose en behandeling van immuun en hematologiese afwykings benodig die begrip van en kwantifisering van menslike perifere bloed T-limfosiet immuun-fenotipes. Dit is slegs moontlik indien vergelyking met data van gesonde pasiente beskikbaar is vir al die biomerkers van belang. Historiese empiriese studies in geïndustrialiseerde lande het al sulke normale verwysings waardes vir kinders beskryf, maar deur groepering van die data in spesifieke 'ouderdoms-blokke'. Aangesien sulke biomerkers aanhoudend verander met ouderdom, het dit gelei tot 'n verlies aan akkuraatheid in die bepaling van normaliteit van pasiënte wat naby aan die grense van sulke ouderdoms-groepe lê. Ander studies het staatgemaak op die pas van enkel-eksponensiële modelle op sodanige data, wat die eenvoudige aanname maak van 'n eksponensiële afname in sel merkers met ouderdom. Dit was wel al waargeneem dat sulke merkers verhoog vanaf geboorte tot tussen 6 en 12 maande na geboorte en dan voortdurend daal met vergrotende ouderdom. Daar is 'n gebrek aan verwysings waarde skattings metodes wat ouderdom-aaneenlopend is en biologies meganistiese modelle inkooporeer. 'N Meer ideale oplossing sou die ontwikkeling wees van toepaslike wiskundige verwysings waarde skattings metodes wat model-gebaseer en ouderdoms-aaneenlopend is. Sulke modelle kan dan gebruik word om die invloed van koveranderlikes wat ouderdoms-verwante veran-

deringe in die biomerkers van belang veroorsaak te ondersoek. In hierdie studie gebruik ons die ouderdoms-deursnee data van 'n groep van 381 gesonde Suid-Afrikaanse kinders. Wat die volgende biomerkers insluit: CD3+, CD19+, CD8+, CD4+, die verhouding van CD4+ naïef/geheue selle, CD18+, CD56+ en CD3-CD56+. Ons gebruik dan nie-lineêre kleinste kwadrate metodes om enkel en dubbel eksponensiële semi-meganistiese modelle te pas en vergelyk. Die ideale model is gekies op grond van die Akaike se inligtings Maatstaf. Sodoende bepaal ons dat die dubbel eksponensiële model die mees geskik is om ouderdoms-verwante veranderinge in sulke biomerkers aan te dui. Hierdie model voorspel dat seltellings styg na geboorte tot 'n maksimum by ongeveer 12 maande en daarna eksponensiël daal tot 'n asimptoot in volwassenheid. Hierdie resultaat stem ooreen met vorige empiriese en meganistiese studies. Daarna gebruik ons die dubbele eksponensiële model om die invloed van koveranderlikes te ondersoek. Die koveranderlike met die meeste invloed was die tipe voeding in die eerste 6 maande na geboorte. Model gebaseerde metode wat ouderdoms-aaneenlopende verwysings waardes skat is daarna voorgestel. Hierdie metode veronderstel dat die verwysings waardes 'n gespesifiseerde Z-verskuiwing van die 'lopende' standaardafwykings van die residue vanaf die sentraal gepaste model funksie is. Ons vergelyk hierdie metode met tradisionele persentiel kurwes. Persentiel kurwes toon 'n eenvoudiger negatiewe enkel-eksponensiële daling met ouderdom, en is nie in staat om meganistiese interpretasies te maak vir hierdie patroon nie. Die modelle wat in hierdie studie voorgestel is mag lei tot die ontwikkeling van 'n laboratorium instrument waarmee individuele sel-merker waardes kan vergelyk word met gesonde ouderdoms aaneenlopende verwysings waardes.

Sleutelwoorde: immunologiese sel biomerkers, gesonde kinders, model gebaseer, ouderdoms-aaneenlopende verwysings waardes.

Acknowledgements

I would first like to wholeheartedly thank the Almighty God for His guidance and infinitely abundant grace upon me throughout this research period, He has been a great source of strength.

Next, I would like to express my deep and profound gratitude to my supervisor, Prof. Martin Nieuwoudt; for his effort, patient guidance, enthusiastic encouragement and useful critiques throughout the course of this research work. I would not manage to list everything I have learnt during my duration of working with you. Thank you also for introducing me to immune biomarkers and helping me realize over this research period the intricate art of communication in Science. Moreover, I extend special thanks to my co-supervisor, Prof. Cang Hui. Your tireless support and zeal in helping me discuss my research question and findings has gone a long way in aiding me communicate the results well.

Sincere thanks to all the wonderful people I have been working with at DST-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), whose friendship and wit has not only kept me motivated but smiling too throughout my research work. Special thanks to Wanja, Zinhle, Christianah, James, Tokpa, Mahasa, Trust and Fanuel who quickly became not only colleagues but family too.

My great appreciation to SACEMA for the financial support during my study period.

Lastly, I thank all my family members who have not only borne the distance but have been with me both emotionally and spiritually. Special thanks to Dani Nyalum, Perez Nyaka-gwa, Dickson & Rhoda Nyambori, Philip Odhiambo and Hezron Osodo. To my girlfriend, Verah, thanks a lot for the understanding, encouragement and support. I appreciate you all.

Dedications

I dedicate this thesis to the memory of my parents and grandparents, Charles Sigu, Sarah Akinyi, Damar Aseto, Okeyo Migunga and Otigo Nyaranga, who believed in me and without whose unconditional sacrifices and hard work I would not be where I am today.

Publications

The following are titles of publications in progress and are extracts from this thesis. These will be submitted to international journals after this thesis is complete.

1. An Age-continuous, Model-based Biomarker Reference Range Estimation.
2. Comparison of lymphocyte subset populations in children from South Africa, US and Europe: Effect of Evolution or Environment on the Paediatric immune system?
3. The influence of covariate factors on Age-related changes in lymphocyte cell markers for South African healthy children.

Contents

Declaration	i
Abstract	ii
Opsomming	iv
Publications	viii
List of Figures	xii
List of Tables	xv
1 Introduction	1
1.1 Background	1
1.2 Biological background	5
1.2.1 T cell development, function and activation	5
1.2.2 Maintenance and regulation of the T cell population	10
1.2.3 Age-related changes	11
1.3 Problem statements	11
1.4 Objectives of the study	12
1.5 Project outline	12
2 Literature Review	13
2.1 Mathematical and statistical models for lymphocyte cell markers	13
2.1.1 Ordinary differential equation models (ODEs)	13
2.1.2 Statistical models	17
2.2 Factors influencing age-related changes	19
2.2.1 Age	19
2.2.2 Sex	19

2.2.3	Race	20
2.2.4	Food in the first 6 months following birth	21
2.2.5	History of exposure to illness	21
2.3	Methods for constructing age-related reference ranges	23
2.3.1	The clinical and laboratory standards institute (CLSI) approved guidelines	23
2.3.2	The LMS (lambda-mu-sigma) method and centile curves	24
2.3.3	Other approaches	25
3	Appropriate Models for Age-related changes in lymphocyte cell markers	26
3.1	Introduction	26
3.2	Study data	27
3.2.1	Participants	27
3.2.2	Laboratory testing	27
3.3	Methods	29
3.3.1	Formulation of models	29
3.3.2	Parameter estimation	30
3.3.3	Model comparison and selection	32
3.4	Results	34
3.4.1	Parameter estimates	35
3.4.2	Models-fit comparison and assessment	38
3.5	Summary	41
4	The influence of covariate factors on Age-related changes in lymphocyte cell markers	42
4.1	Introduction	42
4.2	Methods	43
4.2.1	Generalized nonlinear least squares	43
4.2.2	Extending the double exponential model	45
4.3	Results	47
4.3.1	Descriptives	47
4.3.2	Model prediction comparisons and extended model estimates	48
4.4	Summary	69
5	An Age-continuous Model-based Biomarker Reference Range estimation method	70
5.1	Introduction	70
5.2	Methods	71

5.3 Results	77
5.4 Summary	82
6 Discussion and Conclusion	83
6.1 Discussion	83
6.2 Conclusion	85
Appendix	87
List of references	87

List of Figures

1.1	The three models for T cell differentiation. The diagrams are adapted from Bains et al. [9].	9
2.1	A simple model of naive CD4+ T cell dynamics. The model divides the naive population into two components. Cell loss is caused by death, differentiation and migration.	15
2.2	Estimated number of naive CD4+ T cells generated by peripheral division per day. Mean (solid line), 2.5% and 97.5% quantiles (dashed lines).	16
3.1	A comparison of the formulated models. Model 1 (the single exponential) showed an exponential decay immediately from birth, Model 2 (the double exponential) demonstrates an exponential growth from birth followed by an exponential decline.	30
3.2	The current data represented using the Comans-Bitter et al. [24] and Shearer et al. [95] 'age-blocks'. For CD19+ and CD8+ the markers clearly increase in the first 12 months and then decline thereafter. In the case of CD3+ the trend is also present but less pronounced.	34
3.3	Plots showing changes in absolute counts of CD3+, CD4+, CD8+ and CD19+ over time for healthy South African children. For CD3+, CD8+ and CD19+, the double exponential model predicts a rise in the counts of cell makers from birth to about 12 months followed by an exponential decline to older ages. This pattern is less apparent for CD4+.	36

3.4	Plots showing changes in absolute counts of ratio of CD4+ naive to memory T cells, CD16+CD56+ and CD3-CD56+ over time for healthy South African children. For the ratio of CD4+ naive to memory T cells, the double exponential model predicts an initial rise in the counts from birth to about 12 months followed by an exponential decline. For CD16+CD56+ and CD3-CD56+, the prediction for the double exponential model is similar to that of the single exponential model and thus either of the models is preferred.	37
3.5	Plots of standardized residuals against the fitted values to assess heteroscedasticity using <i>gnls</i> for the double exponential.	39
3.6	Plot of empirical autocorrelation function for standardized residuals to assess the correlation of the errors for the double exponential model using <i>gnls</i>	40
4.1	A plot showing age-related changes in CD3+ T cells across the different covariates.	49
4.2	A plot showing age-related changes in CD4+ T cells across the different covariates.	52
4.3	A plot showing age-related changes in CD8+ across different covariates. . . .	55
4.4	A plot showing age-related changes in CD19+ across different covariates. . . .	58
4.5	plot showing age-related changes in Ratio CD4+ naive/memory T cells across the different covariates.	61
4.6	A plot showing age-related changes in CD16+CD56+ across the covariates. . . .	64
4.7	A plot showing age-related changes in CD3-CD56+ across the covariates. . . .	67
5.1	An illustration of how deviation from the mean accounts for different percentages of the sample. For the normal distribution, the values less than one standard deviation away from the mean account for 68.27% of the set; while two standard deviations from the mean account for 95.45%; and three standard deviations account for 99.73%. The values within two standard deviations are the normal reference range.	71
5.2	Flowchart of the model-based estimation procedure.	75
5.3	Plots showing model-based, age-continuous reference ranges for absolute counts of CD3+, CD4+, CD8+ and CD19+.	79
5.4	Plots showing model-based, age-continuous reference ranges for absolute counts of CD16+CD56+, CD3-CD56+ and ratio of CD4+ naive to memory T cells for healthy South African children aged 12.5 years and below. The fits assume similar mechanistic properties of the described model.	80

5.5	A comparison between centile curves (black lines) and model-based fits for absolute counts of CD3+, CD4+, CD8+ and CD19+. At early ages, centile curves (black lines) estimate wider, and simpler single-exponential-like, reference ranges than the double exponential model (red lines).	81
-----	--	----

List of Tables

3.1	Explanation of cell-markers investigated in this study.	28
3.2	Comparison of parameter estimates for the two models.	35
3.3	Log-likelihoods and AIC for the two fitted models	38
4.1	Characteristics of the study population.	47
4.2	Parameter estimates for the extended model with CD3+.	50
4.3	Parameter estimates for the extended model for Age-related changes in CD4+.	53
4.4	Parameter estimates for the extended model for CD8+.	56
4.5	Parameter estimates for the extended model for CD19+ B lymphocytes.	59
4.6	Parameter estimates for the extended model for Age-related changes in the ratio of CD4+ naive/memory T cells.	62
4.7	Parameter estimates for the extended model for Age-related changes in CD16+CD56+ T cells.	65
4.8	Parameter estimates for the extended model for Age-related changes in CD3–CD56+ T cells.	68
5.1	Parameter estimates for the central (mean), lower and upper fits for a 95% model-based reference ranges.	78

Chapter 1

Introduction

1.1 Background

Understanding and quantifying human blood lymphocyte immunophenotypes (immune biomarkers), i.e. cell-surface markers or clusters of differentiation (CD) is necessary for the diagnosis and treatment of immunological and haematological disorders. This includes autoimmune and infectious diseases, such as HIV, which may give rise to selective marker proliferation or deficiency. CD4+ antigen provides a primary binding site for the HIV virus. Acquired Immunodeficiency Syndrome (AIDS) results in the decline in the counts of CD4+ T cells with consequent invasion by opportunistic infections. CD4+ T cell counts are commonly used to determine when to commence and monitor HIV patients in highly-active antiretroviral treatment (HAART) programs, assess progression of immunological diseases and the prophylaxis of opportunistic infections [50]. Since CD4+ T cells are the only routinely monitored markers in HAART, historical studies of cell-surface markers have mainly focused on only this one. Few studies have systematically investigated changes in cell-markers from birth through adulthood for the blood lymphocyte immunophenotypes of any of the other main lymphocyte subgroups [51]. A proper understanding of the changes in all blood lymphocyte subsets can only be achieved if comparison to control-data from normal (healthy) subjects is available for the markers of interest [24].

Homeostatic proliferation in the thymus ensures maintenance and production of cell-surface markers. In paediatric patients, the immature immune system produces quantitatively different numbers of peripheral blood cell-surface markers compared to adults [29]. For instance, an adult has approximately 3000 lymphocyte *cell/ml* in the periph-

eral blood, while in children, the number of cell-surface markers rises from birth to a maximum (approximately 4000 *cell/ml*) between six months [50] to one year of age [7], it then follows an exponential decline as the child grows to adulthood. This is due to a number of inter-related factors. For example, the progressive involution of the thymus in the first 20 years of life, exposure to antigens and the alteration of naive to memory cell ratios associated with immunological ‘learning’. Other factors include the rapid change in body size and blood volume associated with growth and the progressive age-related replacement of primary thymic production by peripheral cell division [29]. As a result, both the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) have developed age specific CD4+ cell count systems for use in classifying children into appropriate ‘immune-categories’ [50].

Although considerable progress has been made in immunological research over the last decade, quantifying lymphocyte dynamics remains a challenge. There are divergent predictions of blood lymphocyte population estimates and age-related changes in them. These include varying estimates of thymic production, cell division rates and T cell lymphocyte longevity. As a result, quantification and understanding of homeostatic processes and the naive lymphocyte repertoire remains challenging [29]. Various experimental methods and techniques have been employed to empirically quantify peripheral blood cell surface markers. The requisite antibodies and laboratory techniques for flow cytometric immunophenotyping are generally expensive and time-consuming, and further, the use of T-cell receptor excision circle (TRECs) is problematic as these change in quantity over time [29]. Lymphocyte cell marker measurements are also characterized by large inter-individual variation with measured residuals which are not usually distributed in a Gaussian normal pattern (personal experience).

Reference ranges are commonly used in medicine to compare individual clinical measurements to population values. There have been relatively few studies describing normal reference ranges for all the major cell surface markers [24, 33, 95] and of these, all were conducted in first-world industrialized countries. The comparability of such ranges to those for people in resource-limited settings, such as Sub-Saharan Africa (SSA), is questionable given the different environmental and immunological milieu. Historical studies in industrialized countries have described normal reference ranges for such biomarkers in children by grouping particular age ranges into ‘age-blocks’ [24, 95]. Since such markers change continuously with age this has resulted in a loss of precision in determining whether patients that lie close to the limits of age-ranges are normal or not.

Other studies have fitted single-exponential models to such data, simply assuming an exponential decline in cell markers with age [48]. However, the counts of such markers have been empirically observed to increase from birth to between 6 months to 12 months and then decrease continuously with age. The current paediatric immunological reference intervals in use in the South African National Health Laboratory Service (NHLS) have been adopted from international publications [33, 65, 95].

Although limited paediatric immunohaematological reference ranges have been published from Africa [50, 67, 69, 71, 93, 105], the values vary considerably between countries and regions. The variations in the international and African reference data have been attributed to, among other factors, ethnicity, diet, endemic infections, altitude and the testing methodologies used. Other factors may include inadequate number of samples used in the analysis and the failure to provide confidence intervals around the predicted central values.

Studies that are longitudinal in design are relatively expensive and are prone to difficulties in following-up the respondents. Cost cutting in the form of cross-sectional study designs excludes the effects of longitudinal individual aging but does show 'age snapshot' population behaviour. This in turn necessitates large sample sizes within particular age ranges which then *hopefully* represent the biological dynamics within. The large sample sizes also tend to negate the cost-cutting efforts. In view of the challenges associated with empirically determining biomarker reference ranges, a working group, The Clinical and Laboratory Standards Institute Approved Guideline (3rd edition) for defining and establishing reference intervals [21].

An alternative to purely empirical methods for determining cell marker reference ranges lies in fitting non-linear *mechanistic* or *semi-mechanistic* models to such data [7, 8, 29, 48, 110]. Such models vary in terms of the number of parameters to be estimated and the assumptions made regarding the underlying age-related dynamics. In the case of semi-mechanistic models, they make only the assumption that the change in the cell marker with increasing age is an exponential which proceeds to an asymptote in adulthood [48]. However, the simplicity of such an assumption somewhat detaches it from the underlying biological mechanism. Fully *mechanistic* models, on the other hand, may require difficult or even impossible-to-measure parameters in order for them to be fitted to peripheral blood biomarker data [29]. The benefits of models as compared to purely empirical methods lies in their requiring relatively fewer subjects to enable a fit

to data along with the ready ability to calculate descriptive statistics, such as the median and 95% confidence intervals. Further, they represent age as a continuous rather than an age-range-limited variable [48, 110]. Thus, the data are not split into arbitrary ‘blocks’ but represent the continuous change in the biomarker over time. The application of such models does hold challenges. For example, the requisite specification of the assumptions regarding the underlying biological dynamics, statistical procedures which require normality assumptions, and the absence of model-based statistical methods to determine age-continuous reference ranges [48]. The main difference between *mechanistic* or *semi-mechanistic* models and *purely empirical methods* is that *mechanistic* or *semi-mechanistic* models require an understanding and incorporation of the underlying biological mechanisms while *purely empirical methods* simply involves the calculation of descriptive statistics such as mean, median and standard deviation of the cell markers.

In the current study we review appropriate background immunobiological information. However, the focus is primarily on a mathematical investigation of age-related changes in paediatric peripheral blood immune biomarkers (CD positives). We compare single and double exponential models and extend the double exponential model to investigate the influence of covariates. We also develop a robust model-based approach to estimate age-continuous reference ranges. The latter method is compared with commonly used centile curves.

From a technical perspective it is desirable to have robust diagnostic mathematical tools and methods by which to compare the quality of the model fits and the calculation of reference ranges for this type of data. Such methods may lead to the development of easily employed tools for the laboratory that enable the determination of whether individual cell marker values are normal or not, in a precise and age-continuous manner. The age-continuous, model-based reference estimation algorithm has resulted in development an R software package which will be available for public release. The resolution of these issues provides a methodical foundation by which similar data may be analyzed in the future, and may also provide insights into the biological questions that remain, regarding age-related changes in immune biomarkers.

1.2 Biological background

Lymphocyte cell markers form an important part of the immune system which allows the body to initiate responses to invading pathogens. T cell lymphocytes express receptors on their surfaces which, together with major histocompatibility (MHC) molecules, enables them to identify particular antigen-peptides. As part of the adaptive immune system this offers long-term protection against reinfection by a particular antigen [68]. The development process of the peripheral blood lymphocytes involves the production of T cell precursors in the bone marrow, development in the thymus and entry into the circulating pool as mature T cells. Mature T cell lymphocytes are identified by their expression of CD3+, CD4+, CD8+ or other antigens. This involves extensive cell differentiation, division and selection, as well as T cell receptor (TCR) gene rearrangement. However, only a small percentage of T cell lymphocytes live through this process to enter the circulating pool [9]. Lymphocytes are composed of T-lymphocytes, B-lymphocytes and natural-killer (NK) cells. The development of T-lymphocytes takes place in the thymus whereas B-lymphocytes develop in the bone marrow in adults and in the foetal liver.

T cell receptors (TCR) determine the binding affinity of T cell clones bound to major histocompatibility complex (MHC) molecule. In other words, the receptors expressed by individual T cells on their surfaces help in recognizing specific antigen-peptides in association with the MHC molecules. T cell clones are very specific to peptides bound to a single MHC (pMHC) and activates the appropriate T cell clone. During T cell development in the thymus, the T cells that are specific for self-pMHC complexes are deactivated. This ensures that T cells only respond to foreign antigens. T cell clones are subdivided into naive, activated, effector, and memory cells [9, 68].

1.2.1 T cell development, function and activation

The development of peripheral T lymphocytes takes place in the thymus. The production of T cell precursors begins in the bone marrow; lymphocyte development, 'thymopoiesis', occurs in the thymus and these then enter the peripheral circulating pool as mature T cells [9, 75]. T cell development in the thymus is characterized by cell differentiation, cell division, T cell receptor (TCR) gene rearrangement and cell selection. The thymus is composed of the following: non-lymphatic tissues, thymic epithelial space (TES) and lymphocytic perivascular space. Thymic epithelial space is divided into cortex and medulla. Only a small percentage of thymic lymphocytes survive and enter the

circulating pool [9, 52]. Thymopoiesis takes place mainly in the TES, which contains approximately 90% of the thymocytes at any particular time [42, 59]. Before cell division and maturation, cell proliferation which involves TCR β chain rearrangement takes place in the cortex. Just before this rearrangement, the CD4 positive and CD8 positive subpopulations are upregulated by the thymocytes to become double positive, i.e. positive to both markers. Thereafter, the thymocytes move to the cortex for a vigorous selection process which leads to the formation of two major subpopulations, namely: CD4+ helper T cells and CD8+ cytotoxic T cells [29]. In this process cells that are not capable of binding to self-peptides bound to MHC molecules die by neglect (a process which eliminates those T cells which would be non-functional due to an inability to bind to MHC), which accounts for 80% – 90% of thymocytes loss [109]. Single positive thymocytes move to the medulla through the expression of chemokine receptor, CCR7. After this, the thymocytes (referred to as mature naive T cells) then enter the peripheral circulating population [9]. Immature cells that express very high affinity to self-peptides on MHCs are lost through apoptosis, which refers to a normal, genetically regulated process leading to the death of cells and triggered by the presence or absence of certain stimuli. The final stages of thymic development involves sequential changes in the expression of surface markers and the functional activation of thymocytes for various stimuli [54, 112].

The principal function of T cells is to recognise the foreign antigens presented by antigen presenting cells (APCs) and to elicit appropriate responses. This process is catalysed by MHC molecules and takes place in the secondary lymphoid organs. T cell activation occurs as a result of expansion of T cell receptor clones, followed by activation of naive T cells which undergo vigorous cell proliferation and cell differentiation to produce effector and memory T cells. The effector T cells produced at this stage are capable of destroying infected cells and producing cytokines that are able to stimulate B cells. Large numbers of effector T cells are believed to die through apoptosis, leaving only a small minority to move to the memory T cell pool [3, 80]. The factors affecting T cell function and activation include: TCR activation strength, presence of environmental stimuli, binding extent, MHC peptide exposure and other co-stimulation factors [9].

1.2.1.1 Naive T cell function and activation

At any given time, only 2% of lymphocytes are found in the circulating peripheral blood. The other 98% of lymphocytes are distributed throughout the other tissues of the body. However, almost all the lymphocytes pass through the blood every day [10, 98]. Naive

T cells are exported from the thymus and are in continuous circulation between peripheral blood and the secondary lymphoid organs. Secondary lymphoid organs, which include lymph nodes and the spleen, maintain mature naive lymphocytes and initiate an adaptive immune response. Naive T cell peptides are presented by the APCs, B lymphocytes or dendritic cells (DCs), while still in the secondary lymphoid organs. This enables them to identify specific antigens through a high speed random process. The half-life of a naive T cell within a given lymph vessel is approximately a quarter of an hour. Cells which are not able to recognise their specific antigen return to lymphoid circulation through different lymph nodes. Chemokine receptors (CCR7, CXCR5 and CD31, among others) regulate the recirculation of naive T cells into the lymph nodes [87, 108].

Immune responses and TCR signals in the activated naive T cells are initiated by MHC peptides binding to the antigen specific TCR/CD3 complex [9]. The activation process of naive T cells is determined by the level of TCR affinity, co-stimulation and inflammation. Co-stimulation is not only necessary for proliferation and cell survival but is also important in the production of cytokines and trafficking of synapses to appropriate antigen recognition. Naive CD4⁺ T cells are activated by MHC molecules binding with the TCR and the CD28 co-receptor [90]. Once successfully activated, the naive T cells produce effector T cells through clonal expansion and cell differentiation.

1.2.1.2 Effector T cell function and activation

Effector T cells respond to a variety of invading pathogens and are consequently considered to represent a heterogeneous cell population. The availability of co-stimulatory complexes, antigen loads and cytokines determines the activation of T cells. Due to deficiency in CCR7 and CD27, effector CD4⁺ T cells do not home in on lymph nodes. Instead, they have other molecules which cause them to move to sites of infection. CD4⁺ effector T cells contain type 17 T helper (Th17) which produces interleukin (IL), i.e., IL-12, IL-22 & IL-23, all of which induce immune responses against bacteria, fungi and tissue inflammatory processes [18]. The regulation of effector T cells responses is necessary, as overproduction of either Th1 or Th17 may result in organ-specific auto-immunity. Allergies and asthma may also occur as a result of unregulated Th2 responses [118].

1.2.1.3 Memory T cell function and activation

The adaptive immune system retains small numbers of long-lived memory T cells after an infection. This constitutes ‘immune-memory’ for pathogens enabling prolonged pathogen immunity. Memory T cells respond with greater sensitivity and rapidity to antigens than naive T cells. This is due to the higher proportion of antigen-specific T cells as compared to varied naive T cells [56]. Memory T cells are subdivided into two main subpopulations: central and effector memory T cells, (TCM and TEM respectively). TCM which express CD62L and CCR7 are activated in response to secondary antigen and move through the blood and secondary lymph nodes. TEMs do not express CD62L and CCR7 and circulate through the blood and peripheral organs. TEMs are responsible for the secretion of cytokines and other cytolytic activities [9, 72, 113].

Various models have been proposed which describe the development of memory T cells. These include: Linear differentiation, bifurcation and self-renewal models. First we consider the linear differentiation model: This proposes that activated naive T cells produce a large pool of effector T cells. Only a small proportion of these cells survive and differentiate as long-lived memory T cells [56, 114]. The bifurcation model proposes that the naive T cells divide ‘asymmetrically’, i.e. one daughter cell produces effector cells, while the other regulates the production of memory T cells [9, 17]. Studies have also demonstrated that memory cell production can take place without antigen stimulation, implying that naive T cells do not necessarily need to go through the activation phase to become memory T cells [41, 79]. The self-renewal model proposes that naive T cells differentiate directly into central memory T cells (TCM) and effector memory T cells (TEM). TEM cells move to the inflammation sites and differentiate to become effector T cells. These are illustrated in Figures 1.1(a, b and c).

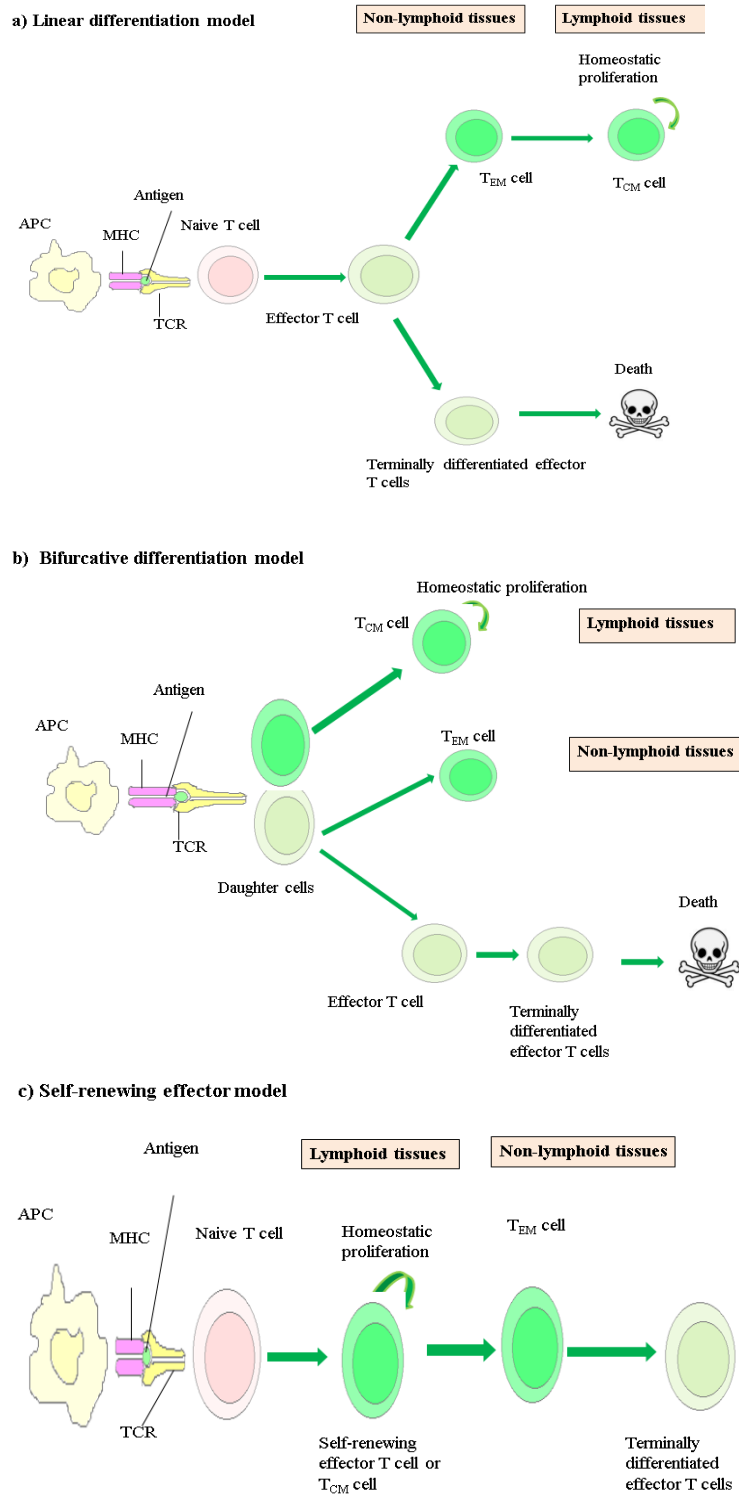


Figure 1.1: The three models for T cell differentiation. The diagrams are adapted from Bains et al. [9].

1.2.2 Maintenance and regulation of the T cell population

The T cell population is regulated and maintained through homeostasis which is responsible for the addition or removal of thymic tissues [9]. Additionally, the expansion and survival of the T cell population is due to its ability to reconstitute following immunological depletion [91]. Cell division, differentiation and survival are all homeostatically regulated. However, the homeostatic effect is subtle, hindering our understanding of T cells. The rate of thymic production and cells exported per day is not known. In *in vivo* thymectomy experiments have been used to investigate the dynamics of T cell populations in healthy human subjects [9].

It is well known that thymic involution occurs with advancement in age [7], although the majority of prior studies have not shown any change in the overall thymic volume with advancing age [9]. However, an age-related decline in the volume of TES, and growth in adipose tissues, has previously been observed [99]. T cell receptor excision circles (TRECs) are considered the principal thymic output marker. TREC numbers have been shown to be significantly higher in children but decline with advances in age, and they are also lower in HIV-infected people [7]. This is further described in Section 2.1.

Mathematical models recommend the use of TRECs/*ml* to measure blood thymic output as opposed to TREC content per cell. This follows the notion that the decline in the TREC population cannot be directly attributed to cell division. TRECs also influence thymic output through cell proliferation and cell loss through death or phenotypic changes [32, 34].

1.2.2.1 Naive T cell homeostasis

Two processes are involved in naive T cell homeostasis, the stimulation of cytokines and the interaction between self-peptide & MHC-TCR. These begin with the expansion and release of large quantities of effector cells following activation by foreign antigens. Only a small number of effector T cells survive this to become memory T cells [9, 101].

1.2.2.2 Memory T cell homeostasis

Studies have shown that the memory group develops from the naive T cell group, although this process is not well understood [70]. Memory T cells undergo periodic self-

replacement, supported by continuous induced expansion of antigens. The volume of the memory T cell pool tends to remain constant in adults as opposed to children which seems to increase from birth to a maximum during early year of life. Memory T cells acquire new members through differentiation, while the pool is regulated through homeostasis. The division of memory T cells is faster than that of naive T cells in both healthy children and adults [70, 94]. The survival and turnover of memory T cells is largely influenced by MHC interactions. However, CD8+ T cells are capable of undergoing homeostatic proliferation in the absence of MHC I. The memory T cell population is regulated through cell death, and this is compensated through production of new cells [9].

1.2.3 Age-related changes

The immune system grows from childhood to a relatively constant size in adulthood. In longitudinal studies of T cell populations, it has been shown that CD4+ T cell numbers decline exponentially with advancement in age from birth to adolescence [38, 48]. At birth, CD4+ and CD8+ T cells constitute approximately 90% of the total naive T cell population. However, with advances in age, the naive T cell population declines and the ratio of naive to memory T cells approaches 1 : 1 at approximately 20 years. Thereafter, the population of both naive and memory T cells remains relatively constant into old age [36, 45, 48, 82]. Despite the reduction of the newly produced naive T cells in the thymus, studies have shown that, on average, the number of T cell receptor (TCR) sequences within naive T cells for individuals aged between 20 to 65 years are relatively constant, but significantly higher in individuals older than 65 [9, 58].

1.3 Problem statements

In this study, we seek to address the following problems:

1. The loss of precision resulting from the practice of grouping age ranges into 'blocks' in cross-sectional studies that investigate changes in blood immunophenotypes.
2. The fitting of appropriate semi-mechanistic models to quantify blood lymphocyte immunophenotypes in healthy children aged 2 weeks to 12.5 years.
3. Variation in the estimation of age-related changes in lymphocyte cell markers due to population characteristics (covariate factors) of the study subjects.

4. ‘Standardization’ of age-continuous, model-based methods for estimating reference ranges for lymphocyte cell markers.
5. The absence of a statistical package in R software to estimate age-continuous, model-based reference ranges.

1.4 Objectives of the study

The fundamental aim of this study is to estimate model-based, age-continuous immune biomarker reference ranges for healthy paediatric patients in South Africa. Specifically, the study seeks to achieve the following objectives;

1. To fit single and double exponential models to the data and compare them in order to produce improved models.
2. To investigate the influence of covariate factors on age-related changes in lymphocyte cell markers.
3. To propose and implement a more robust model-based, age-continuous method for estimating reference ranges for lymphocyte cell markers. This will lead to development of an R package for estimating model-based, age-continuous reference ranges.

1.5 Project outline

In Chapter 2 we explore existing literature on mathematical and statistical models for age-related changes in lymphocyte cell markers. We review available methods for reference range estimation and the possible biological factors which may influence such changes. In Chapter 3 we present the comparison between double and single exponential models, including their formulation, parameter-estimation and interpretation of the results. In Chapter 4 we review the generalized nonlinear least squares estimation method and extend the double exponential model to incorporate covariates. In Chapter 5 we present the new model-based, age-continuous estimation method, its implementation and results. We also compare this method with traditional centile curves. In Chapter 6 we discuss and conclude this study, provide recommendations and areas for future research.

Chapter 2

Literature Review

To construct reference ranges for immune biomarkers, the following are required; appropriate mathematical and statistical models for describing changes in lymphocyte populations, methods for describing normal reference ranges and an understanding of the influence of covariates within the population, on lymphocyte development. Below we review foundational literature then modify historical reference range estimation methods to develop our model-based approach. We also extend the double exponential model to incorporate covariates.

2.1 Mathematical and statistical models for lymphocyte cell markers

Both mathematical and statistical models are useful in studying immune system changes. These generally involve simplifying assumptions regarding the system and provide a means to evaluate the underlying biological processes.

2.1.1 Ordinary differential equation models (ODEs)

A variety of studies have employed ODEs to quantify and interpret immunological data from *in vivo* studies [44, 46, 70, 76]. Such models are *mechanistic* in that they require biological assumptions to be made regarding the underlying mechanisms. In cases where they have been used to estimate the rate of T cells division and death, they generally assume that the cells divide and die at constant rates, respectively r and d , over time period, t .

Hazenberg et al. [44] developed an ODE model to describe T cell receptor excision circles (TRECs) by studying two pools of cells; total naive T cells and total TRECs. Total naive CD4+ T-cells at age t , $N(t)$, is described by the following differential equation

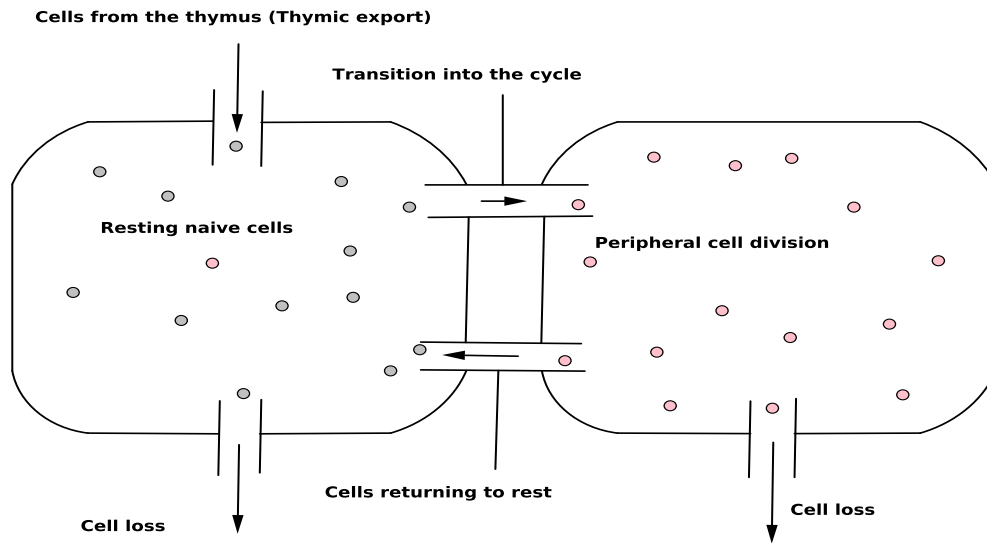
$$\frac{dN(t)}{dt} = \theta(t) + \rho(t)N(t) - \delta(t)N(t),$$

where $\theta(t)$ is the number of CD4+ T-cells exported per day at age t (the rate of thymic export), $\delta(t)$ represents the average rate of naive cell loss at age t (day^{-1}) and $\rho(t)$ denotes per cell rate of addition to the naive population peripheral division at age t (day^{-1}). The authors assume that the rate of loss of TCR excision circle (TREC) content is homogeneous and that the average rate of naive CD4+ T cell loss is equal to the average rate of TREC loss [8]. Under this assumption, the total naive TRECs with age is defined as

$$\frac{dT(t)}{dt} = c(t)\theta(t) - \delta(t)T(t),$$

where $T(t)$ is the total counts of naive TRECs in an individual at age t and $c(t)$ is the average naive TREC content of CD4+ T cells emerging from the thymus while $\delta(t)$ and $\theta(t)$ remains as defined above.

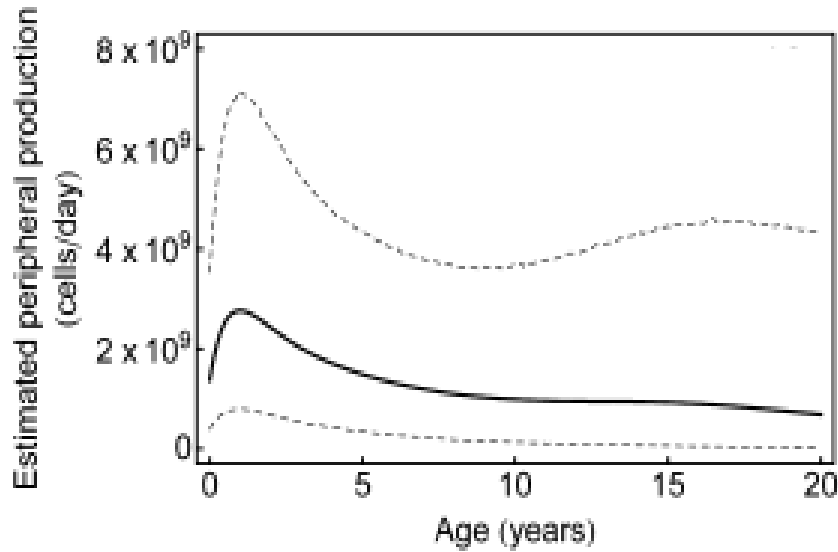
Bains et al. [8] developed a dynamical mathematical model to quantify thymic export by combining naive T cell proliferation and TREC models. The authors propose that naive cell population is divided into compartments of resting and dividing naive cells, as shown in Figure 2.1. Thymic export is expressed as a function of age from birth to 20 years, and the total number of naive T cells is then the sum of all cells within the resting and dividing compartments. After the completion of each cycle, it is also assumed that two daughter cells will move back to the resting pool.



Adapted from Bains et al [7]

Figure 2.1: A simple model of naive CD4+ T cell dynamics. The model divides the naive population into two components. Cell loss is caused by death, differentiation and migration.

This model predicts a decline in the expression of Ki67, the protein specifically associated with cell proliferation, in naive CD4+ T cells and a rise in the rate of thymic export during the first 12 months of life. This is attributed to immune system maturation and increase in body growth in the first year of life (see Figure 2.2).



Adapted from Bains et al [7]

Figure 2.2: Estimated number of naive CD4+ T cells generated by peripheral division per day. Mean (solid line), 2.5% and 97.5% quantiles (dashed lines).

De Boer [28] also used ODEs to estimate the role of thymic output in HIV infection. This model comprises two compartments, including the total number of naive T cells, T , and those derived from the thymus, N . Cells produced in the thymus undergo cell division and eventually die in the periphery leading to the disappearance of TRECs, T . It is argued that cells produced in the thymus are dependent on the age of the person and are also produced or die at a rate which depends on the total population of T cells. Murray et al. [81] used ODE models to describe T cell homeostasis. The authors propose two compartments, those originating from the thymus, N , and those that are proliferating (these are either naive T cells, N_p , memory T cells, M and activate T cells, A). See Equation 2.1.1.

$$\left. \begin{aligned} \dot{N} &= s_0 e^{-\lambda_1 t} s(N_p) - (\lambda_n + \mu_n g(N_p)) N, \\ \dot{N}_p &= \lambda_n N + (ch(N, N_p) - \mu_n) N_p + \lambda_{mn} M, \\ \dot{M} &= \lambda_a A - \mu_m M - \lambda_{mn} M, \end{aligned} \right\} \quad (2.1.1)$$

where $s_0 e^{-\lambda_1 t} s(N_p)$ is the T cells arising from the thymus, $\lambda_n N$ lost into the proliferation compartment and $\mu_n g(N_p) N$ die. The cells in the N_p compartment result from incoming

naive T cells $\lambda_n N$, proliferating cells $ch(N, N_p)N_p$ and those lost by death, $\mu_n N_p$. Memory T cells, M , are reactivated cells $\lambda_a A$ and are lost through re-conversion to naive T cells, $\lambda_{mn} M$ and death μ_M . The model predicts that the contribution of the thymus to the overall CD4+ population at age 25 is approximately 20% per year. Further, this rate decreases to 10% by the age of 40 and to 5% by the age of 55. They conclude that the maintenance of naive T cells within the first 20 years of life is through thymic activity, while after this age, it is through cell proliferation.

2.1.2 Statistical models

Both deterministic and stochastic models can be used to model age-related changes in lymphocyte cell markers. Deterministic models will always produce the same output, given the definition of initial parameter values. Stochastic models enable the introduction of randomness in the parameters. Randomness is produced by specifying a probability distribution for a parameter, which leads to slight variations in the outcome. Early statistical studies of lymphocyte cell markers described changes by grouping particular cell markers into 'age-blocks' and then used various tests to demonstrate differences between groups [24, 33, 50, 93, 95, 116]. Those which applied regression models to describe such changes are of particular interest.

Wade and Ades [110] fitted age-related reference intervals using a maximum likelihood estimation method on CD4 counts of uninfected children born to HIV-1 infected mothers by using an exponential function with a Box-Cox transformation parameter. The Box-Cox transformation ensures that the centiles are asymptotic and conform to the values expected at older ages. Subsequently, they fitted a double exponential function to the data using the LMS (lambda-mu-sigma) based maximum likelihood method. However, the exact double exponential function was not disclosed [111]. The purpose of using exponential models was to allow for the estimation of skewness, spread and to describe the change in the median values of the CD4 counts. The authors used the likelihood ratio test to compare and assess the statistical significance of the estimated model parameters.

A study in Malawi used regression analysis to estimate the linear association between the healthy lymphocyte subsets with age [71]. Lymphocyte counts were log-transformed and estimated cell numbers and percentages per sex category across various age groups, although the study was not designed to explore sex differences in lymphocyte subsets. Shearer et al. [95] used ordinary least squares linear regression on cross-sectional data

of lymphocyte subsets in healthy children below 18 years.

Marie-Quitterie et al. [88] used a semi-mechanistic nonlinear regression model to investigate CD4 reconstitution in HIV-infected children. Their model proposes eventual asymptotic stabilization of CD4 counts:

$$\ln(y_{ij}) = \beta_i - (\beta_i - \lambda_i)e^{c_i t_{ij}},$$

where $\ln(y_i)$ are the CD4 counts for child i at time t_{ij} , λ_i is the count of CD4 at the onset of the treatment, β_i is the long-term $\ln(y)$ estimates, while c_i is the proportional recovery rate from λ_i to β_i to occur. The CD4 counts, y_i , are log-transformed to improve normality. They then used least squares to fit the model to the data of individual subjects followed by nonlinear mixed-effects modeling to account for inter-individual variability. CD4 reconstitution was found to occur more rapidly in younger children. No significant inter-individual recovery rate variability was found among subjects. Similar nonlinear mixed-effects models have been used by Lewis et al. [66] and Claudia et al. [16].

Huenceke et al. [48] used a three parameter exponential model,

$$f(t) = c + b_0(1 - \exp(b_1 t)), \quad (2.1.2)$$

to describe lymphocyte development from childhood to adolescence. This model was fitted using the minimization of the sum of squares of residuals. They also constructed age-related reference ranges around the predicted values using Tukey's method [48]. Profound age-related changes were found at early ages [31, 48]. Although they used an age-continuous exponential function which eliminates the need for age-grouping, this model did not account for the rise, following birth, with subsequent exponential decay after one year of age that is normally empirically measurable in data of this type.

2.2 Factors influencing age-related changes

Prior studies have shown that factors such as gender, age, ethnicity and lifestyle (e.g., alcohol use, smoking, diet and stress levels) may influence lymphocyte cell marker development and quantities.

2.2.1 Age

In paediatric subjects, the immature immune system produces different numbers of peripheral blood cell-surface markers compared to adults [29]. An adult has approximately 3000 lymphocyte *cell/ml* in the peripheral blood, while in children, the number of cell-surface markers rises from birth to a maximum (approximately 4000 lymphocyte *cell/ml*) between six months [50] to one year of age [7], it then follows an exponential decline as the child grows to adulthood. This is thought to be due to a number of inter-related factors. For example, the progressive involution of the thymus in the first 20 years of life, exposure to antigens and the alteration of naive to memory cell ratios associated with immunological 'learning'. Other factors include the rapid change in body size and blood volume associated with growth and the progressive age-related replacement of primary thymic production by peripheral cell division [29].

Various studies estimating reference values for peripheral blood lymphocytes have established the effects of age. For instance, Choi et al. [20] classified healthy individuals aged between 21 to 81 years-old into five groups. These include: 21-30, 31-40, 41-50, 51-60 and 61-80. They found statistically significant differences in the counts of CD3+, CD3+CD8+ and CD3-CD56+ T cells between these age groups. Similar findings have also been reported in other studies [4, 20, 53] in which the aging effects have been attributed to hematopoietic stem cell activities, thymic involution and the functional decline in adaptive immunity.

2.2.2 Sex

Sex has been found to be a cause of differences in both innate and adaptive immune systems. In the innate system, males have less pronounced immune responses compared to females [39]. The influence of sex includes differences in cell counts, cell activities and levels of circulating cytokines. Females have been found to have relatively higher phagocytic activities and more effective antigen presenting cells (APC). Males have been shown to have higher activities of natural killer (NK) cells and increased levels of proin-

flammatory cytokines [13, 14, 39].

In the adaptive immune system, females have higher percentages of circulating T lymphocytes than males. Further, in females, B cells and CD4 Th2 cells mainly drive immunological processes, while in men CD8 and CD4 Th2 cells are more dominant. No differences between the sexes in the CD4/CD8 (Th/Tc) cell ratio have been found [39]. Differences in lymphocyte activities as a result of sex differences have been attributed to hormonal, genetic and environmental factors and microchimerism [13, 39, 40]. Hormonal differences are caused by circulating steroids such as estrogen, progesterone and testosterone, which influence various effector cells of the immune system. For example, estrogen facilitates immune responses while progesterone and androgens reduce it [84].

A study conducted in a healthy Korean population observed significant differences between the sexes [20]. The percentages of CD3+ and CD3+CD4+ T cells were higher in females. On the other hand, the percentages of CD3-CD56+ T cells were lower in females than in males. Rudy et al. [92] established that females had higher counts of CD4+ T cells, CD4+ memory T cells and lower CD16+ cells; higher mean totals of CD3+ and CD4+. Other studies have also shown higher counts (or percentages) of CD4+ T cells in females [11, 78, 103]. Such differences may be due to androgens or estrogen or both. The differences between males and females starts during early development and becomes more pronounced at puberty, mainly due to the regulation and production of sex steroids. The secretion of cytokines, as this influences cell proliferation, chemokines, B and T cells accounts for further differences in T cells counts between males and females [74, 85, 92, 100].

2.2.3 Race

The effects of race on immunological changes, in particularly healthy individuals, has not been thoroughly investigated [60]. However, some studies have suggested that black Africans have relatively lower rates of decline in CD4+ counts, compared to Caucasians, possibly due to genetic differences [5, 35, 60]. Prior studies of normal reference ranges conducted in large Chinese and Indian populations, with large socio-demographic diversity, have shown varying results [20]. For instance, an Indian study categorized the population into Dravidian and Aryan ethnic groups, but found no significant differences in the mean values of the measured immune parameters [107]. Other studies also found no association between ethnicity and the counts of CD4+ T cells [37, 97]. CD8+ and CD19+ cell counts have been found to vary gender and race [92].

2.2.4 Food in the first 6 months following birth

Breast-milk contains antimicrobial and immunomodulatory components which protect infants from pathogenic infections. This is primarily in the form of maternal immunoglobulin A, which is responsible for mucosal respiratory immunity [106]. However, breast-milk also contains lactoferrin, lysozyme and oligosaccharides which are antimicrobial factors [64, 106, 117]. Breast-milk facilitates both short-term and long-term infant immune system development. It has been shown that breast-fed children have enhanced and longer duration immune protection against infections compared to formula-fed children [106].

Breast-feeding is known to influence gene regulation, intestinal cell proliferation and differentiation [6]. The bioactive factors in breast-milk catalyze the differentiation and growth of B lymphocytes and initiate the production of particular antibodies. It also contains pattern-recognition receptors for microorganisms, which are facilitated by the presence of toll-like receptor (TLR) -2 and TLR-4, CD14 co-receptors and soluble CD14 [6, 62]. The composition of the mother's breast-milk may vary depending on her age, diet, ethnicity, weight gain or loss during pregnancy and the infant's birth weight. Studies of HIV transmission have established that exclusively breast-fed infants are at a lower risk of mother-to-child transmission (MTCT) compared to mixed fed infants. However, the risk of transmission of HIV to breast-fed infants increases by approximately 20% for HIV positive versus negative mothers whose risk of transmission is identically zero [83].

2.2.5 History of exposure to illness

Any historical exposure to viruses, bacteria or any other illness may lead to a compromised immune system and a decrease in immune cell counts, such as of CD4+. The rate and level to which cell counts drop depends on the duration of infection and the ability of the immune system to fight it. For example, in newly diagnosed HIV-infected people, CD4+ counts drop due to increased exposure to the virus. After initiation of antiretroviral (ARV) therapy, CD4+ once again increases after the viral load has been suppressed [55]. A proper understanding of the variability of lymphocyte cell markers counts, which occurs in the absence or presence of HIV infection, is needed for both CD4+ T-cells and all other blood lymphocyte subsets.

Studies of individuals with or without chronic illnesses have shown that psychological

depression of any kind is associated with immune suppression. For example, in studies of immunocompromised (HIV-seropositive) gay men, immunological changes in CD4+, CD8+ and CD56+ cell counts and proliferative responses are associated with depression [15, 57]. HIV positive women with symptoms of chronic depression are at twice the risk of mortality compared to non-depressed women. Those most vulnerable have lower CD4+ counts and higher viral loads [55].

Maternal health status has been found to be associated with increased morbidity and mortality in infants. Higher perinatal transmission and disease progression is found in infected infants born to HIV-positive women, particularly those at advanced disease stages [1, 61]. Children born to HIV-seropositive women are at risk of acquiring HIV due to direct or indirect utero or intrapartum viral exposure during birth and breast-feeding [26, 27]. Kuhn et al. [61] established a significant association between infant mortality, low CD4+ T cell counts and maternal hospital admission. Infants born to immunocompromised mothers may have inadequate passive immunity due to low trans-placental transfer of IgG antibodies, which may lead to low counts of lymphocyte cell markers [61]. Ota et al. [86] observed that exposure to HIV may compromise the development of the foetal immune system, which may lead to non-immunogenic vaccination. A study in Gambia [86] showed a reduced rate of formation of bacille Calmette-Guérin scarring among HIV-negative infants born to HIV-infected women.

2.3 Methods for constructing age-related reference ranges

Age-related reference intervals are employed in the monitoring and detection of departures from normality for clinical conditions that may indicate pathology. The reference range is defined as 95% of the predicted population at a particular range. The remaining 5% lie at the lower and upper (2.5% each) limits of the sample. Reference ranges may also be understood as the representation of the intervals between two predetermined centiles (commonly 5th and 95th) of a distribution of a variable, Y , drawn from a 'healthy' population at a given time (age). In reference to the concentration of immunological cell markers, higher or lower values of Y , above the upper limit or below the lower limit represent abnormality, i.e.,

$$\text{Abnormality} = \begin{cases} Y & \text{lies below the lower limit,} \\ Y & \text{lies above the upper limit.} \end{cases}$$

Prior knowledge of the distribution of Y is fundamental in estimating the centile position of the individuals in relation to the reference population. The closeness of the centile location (100% or 0%) measures the extremity of the observed Y . Establishing reference intervals is based on the assumptions of log-normal or normal distributions.

Recent statistical methods have focused on the production of growth-charts, i.e. centiles describing a smooth change over age [96]. According to Chitty and Altman [19], the requirement of simplicity, which was key in the nineties, is becoming less important, owing to improvements in computational technology. Methods for estimating reference intervals include both parametric and non-parametric methods.

2.3.1 The clinical and laboratory standards institute (CLSI) approved guidelines

A working group has published guidelines in view of the challenges associated with empirically determining biomarker reference ranges, namely, the Clinical and Laboratory Standards Institute approved guideline (3rd edition) for defining and establishing reference intervals [21]. Non-parametric and parametric methods were proposed. In the former, the reference intervals are calculated by ranking n observations from smallest to largest. Suppose that the rank of an observation is denoted by r , then the smallest rank is $r = 1$ and the largest rank is $r = n$. The observations corresponding to the 2.5th (corresponding to r_1) and 97.5th (corresponding to r_2) percentiles are $r_1 = 0.025(n + 1)$

and $r_2 = 0.975(n + 1)$ respectively. In the parametric method, suggested by Horn and Pesce [47] and referred to as 'robust', the following is done: the location (centre) and scale (spread) parameters are estimated, represented by median and median absolute deviations (MAD) respectively. Then an iterative process of down-weighting actual observations, depending on their distance from the median, is performed. The median is updated for each iteration until the difference between two consecutive medians is negligible.

The CLSI recommends the use of the non-parametric method owing to its apparent reliability, simplicity and the absence of any assumptions regarding the distribution of the reference values. A key disadvantage is that it requires a minimum sample size of 120 individuals for the 95% reference intervals to be calculated. Small samples may require transformation (such as Box Cox) to eliminate kurtosis and since this involves the weighted sum of observations, it increases susceptibility to the influence of outliers [43].

2.3.2 The LMS (lambda-mu-sigma) method and centile curves

Centile curves are the most commonly used method to estimate age-related reference ranges. They are also used in medicine to compare the individual clinical measurements to the population values. Such curves have been described as the "healthy empirical model" that can be used for both diagnosis and comparison between healthy and unhealthy individuals [25]. Centile curves are drawn from an underlying distribution and obtained by splitting the population into different age groups. The shape of the curves then provides information regarding skewness in the distribution [22].

The LMS method involves the parametrization by Box Cox transformation of the observed variable with the three parameters L , M and S (λ , μ and σ) for skewness, the median and the coefficient of variation respectively, for the variable of interest. Centile curves are then constructed using the parameters [25, 96, 116]. The 100α th centile is denoted by $C_{100\alpha}(t)$ and is defined in (2.3.1).

$$C_{100\alpha}(t) = M(t)[1 + L(t)S(t)Z_\alpha]^{1/L(t)}, \quad (2.3.1)$$

where Z_α is the corresponding normal deviation for the tail area α ; t is age in months; $C_{100\alpha}(t)$ is the Z_α centile. This method accommodates various types of distributions, whether normal, skewed or kurtotic. Although the LMS method produces convincing centile curves, formal statistical inferences (other than location and shape), comparisons

and appropriate formula for obtaining Z-scores remains problematic. The implementation of this method is also computationally intensive [116]. LMS and centile curves are rigid in that the only parameters which can be estimated on the data are λ , μ and σ . Using *semi-mechanistic* or *mechanistic* models, which require the specification of additional parameters to explain physical changes in the variable of interest or of covariates, does not appear to be possible.

Wade and Ades [110] applied the Box-Cox transformation on CD4 lymphocyte counts measured in blood samples of children. They used non-linear (negative exponential power) models to approximate L, M and S parameters for the curves and then compared the models using likelihood ratios. Although the estimated centile curves appeared to fit the data well, the particular fitted functions were not provided. Although the LMS method produces convincing centile curves, formal inferences for model comparisons and appropriate formulae for obtaining Z-scores are not available.

2.3.3 Other approaches

Tsay et al. [104] propose a graphical method for estimating the standard deviation and mean for log-transformed data. A two-parameter log-normal distribution is assumed, and Q-Q plots are then used to obtain the reference limits through interpolation or extrapolation depending on the sample size. This procedure has a subjective element in that it is dependent on the validity of the log-normal assumption.

Lawrence, and Trewin [63], and Baughman et al. [12] propose parameter estimation by maximum likelihood, using data modelled with a mixture of normal distributions. Other methods assume a linear relationship between the centiles and ordered observations. Pearson correlation coefficients between corresponding sample fractions are then used to define the reference intervals [110].

Chapter 3

Appropriate Models for Age-related changes in lymphocyte cell markers

3.1 Introduction

Due to large variations in the measurements of the lymphocyte cell markers at young ages in the South African paediatric data employed in this study, it is not clear on first observation whether the age-related changes are best modelled using a simple negative (single) exponential or a double exponential as previously described (see Section 2.1.2). The formulation of such models is based on the following:

1. In all data sets, a general exponential decline with age is observed and a similar trend has been observed and modelled on comparable European data [48].
2. But we also need to explore the possibility that we need to account for a transient increase in biomarker levels among very young children. Prior empirical studies have demonstrated that cell counts rise in the first year of life and then decline exponentially thereafter [24, 95]. Mechanistic models of T-lymphocyte production also predict a similar pattern [7, 111].

3.2 Study data

3.2.1 Participants

Three hundred and eighty one (381) children aged from 2 weeks to 12.5 years were recruited from a 'Child Wellness Clinic' (CWC) at a community health clinic in an informal settlement of Cape Town, South Africa [65]. The CWC was established primarily as a research clinic, which also aimed to benefit the participants and the wider community in terms of health promotion, education and screening. Attendance at the CWC was voluntary, and the criteria for recruitment was that the child was well at the time with no chronic medical condition or prescription medications, registered at the health clinic, and attended with its biological mother and hand-held medical record. Maternal HIV-exposure was not excluded. Informed consent was obtained in English or via translator in Afrikaans or Xhosa. The session included clinical history and examination by a paediatrician, anthropometry plotting, assessment of vaccination status (and catch-up as needed), provision of nutritional supplements and a food voucher. Each participant had phlebotomy of 2-3mls of blood used for rapid HIV-antibody analysis (Alere Determine[®]), full blood count (FBC) and basic immunological profile. HIV-infected children were not included but were separately referred for treatment by the clinic. The University of Stellenbosch granted ethical approval, (M12/01/005) and permission for the study was given by Cape Town Department of Health.

3.2.2 Laboratory testing

Blood samples (in EDTA) were taken between 9am and 1pm, with 500 μ l of the original samples couriered at room temperature by air to Johannesburg and processed the following morning at the National Health Laboratory Service, Charlotte Maxeke Johannesburg Academic Hospital (SANAS M0106B). Immunophenotypic analysis was performed at the Johannesburg flowcytometry laboratory according to standard operating procedures. Directly labelled antibodies CD3 APC, CD3 FITC, CD16 PE, CD19 FITC, CD45 PerCP, CD45RO PE, CD45RA FITC, HLA Dr APC (Becton Dickinson Immunocytometry Systems (BDIS), San Jose, CA), CD4 FITC, CD8 PE and CD56 PE (Beckman Coulter, Inc. Miami, Florida) were added in pre-titrated manufacturer optimised concentrations to tubes with 50 μ l of well-mixed whole blood. Stained samples were vortexed once and incubated for 30 minutes. Red blood cells were then lysed using FACS Lysing Solution (BDIS, San Jose, CA). All samples were run on a Becton Dickinson FACSCalibur[™] and acquired, and analysed, using CellQuest[™] Pro software. Prior to analysis, basic daily flow cytometer set-up included assessment of

Chapter 3. Appropriate Models for Age-related changes in lymphocyte cell markers

Calibrite™ 3 and Calibrite™ APC beads (BDIS, San Jose, CA) to monitor laser, optics, fluidic alignment, linearity and instrument performance for the FACSCalibur™, according to the manufacturers' standards. Listmode data were stored for retrospective analysis. External CD4 Quality Assessment for CD4 testing was performed through the U.K. NEQAS Immune Monitoring scheme and the NHLS CD4 African Regional External Quality Assessment Scheme [65]. Lymphocyte subsets were expressed as a proportion of total lymphocytes determined using bright CD45 expression and side scatter. Specific lymphoid subsets assessed included: total CD3, CD3+/CD4+, CD3+/CD8+, CD3-/CD56+, CD16+/56+, CD3+/HLA Dr+, CD3+/CD4+/HLA Dr+, CD3+/CD8+/HLA Dr+, CD3+/CD4+/45RA+, CD3+/CD4+/45RO+, CD3+/CD8+/45RA+, CD3+/CD8+/45RO+ and total CD19. Absolute cell counts were obtained using a dual platform method; total lymphocyte counts on all samples were obtained on a Beckman Coulter LH750 haematological analyzer. All laboratory work and data analysis were performed by technician blinded to the patients. In this study we describe seven lymphocyte cell markers. These include: CD3+, CD19+, CD8+, CD4+, ratio of CD4+ naive/memory, CD16+CD56+ and CD3–CD56+.

Table 3.1: Explanation of cell-markers investigated in this study.

Cell-surface markers	Cellular Expression	Functions
CD3+	Thymocytes, T-cells Small amounts in Purkinje cells	Associated with the T cell receptor (TCR). Required for cell surface expression and signal transduction by TCR.
CD4+	Thymocytes, Helper T-cells Inflammatory T-cells About 2/3 of the peripheral T-cells	Sends signal to other immune cells (CD8 and killer cells). Co-receptor for MHC class II molecules. Binds lck on cytoplasmic face of membrane. Receptor for HIV-1 and HIV-2 gp120. CD4+ counts are used to determine when to start HIV therapy.
CD8+	Thymocyte subsets, cytotoxic T-cells About 1/3 of the peripheral T-cells	Co-receptor for MHC class I molecules. Binds lck on cytoplasmic face of membrane.
CD19+	B cells	B cell co-receptor in conjunction with CD21 and CD81. Required for binding of Src-family, kinases and recruitment of PI-3 kinase . Diagnosis of cancers arising from B-cell lymphomas.
CD16+	Natural killer cells, macrophages	Component of low affinity Fc receptors, FcγRIIIa (CD16a) and FcγRIIIb (CD16b). These Fcs mediates phagocytosis and antibody-dependent cell-mediated cytotoxicity (ADCC).
CD56+	Natural killer cells	An isoform of Neural Cell Adhesion Molecule (NCAM) which signals to induce neurite outgrowth via the Fibroblast growth factor receptor (FGFR).

Adapted from [2, 115]

The data generated from the 381 SA children was entered into Microsoft Excel (2010). The immunophenotypes counts was then tabulated by age. Data analysis was performed using R statistical software (version 3.2.3).

3.3 Methods

3.3.1 Formulation of models

The models are *semi-mechanistic*, and assume that cell markers either decrease exponentially from birth, or alternately rise from birth to a maximum at 12 months and then decline to an asymptote with advancing age. Specifically:

3.3.1.1 The single exponential model

This is a simple three parameter model, mathematically equivalent to that applied by Huenceke et al. [48]:

$$f_1(t) = \beta_0 + \beta_1 \exp(-\beta_2 t), \quad (3.3.1)$$

where the sum of β_0 and β_1 is the estimated cell count at birth while β_2 is the rate of change in cell markers over time. From the graph of this function, showed in Figure 3.1 in green, it assumes a simple exponential decline from birth.

3.3.1.2 The double exponential model

This model is formulated from the single exponential model by adding a fourth parameter β_3 and then multiplying (3.3.1) by $\frac{1}{1+\exp(-\beta_3 t)}$. The resulting function is defined as

$$f_2(t) = \frac{\beta_0 + \beta_1 \exp(-\beta_2 t)}{1 + \exp(-\beta_3 t)}. \quad (3.3.2)$$

The average of β_0 and β_1 estimates the cell count at birth ($t = 0$), β_2 estimates the rate of change of cell markers over time while (β_3) determines the shape of the curve. For large values of t , the function reduces to the single exponential model described above and thus β_0 estimates the cell count at large t 's, i.e. older ages. This function, when plotted, rises and then descends to an asymptote as t increases, as shown in Figure 3.1 in blue.

Chapter 3. Appropriate Models for Age-related changes in lymphocyte cell markers

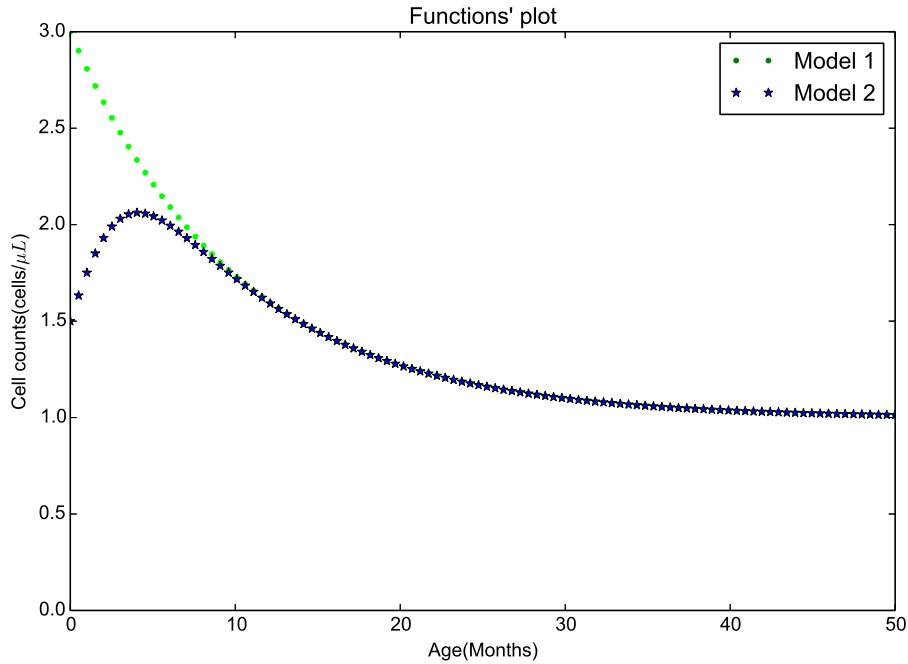


Figure 3.1: A comparison of the formulated models. Model 1 (the single exponential) showed an exponential decay immediately from birth, Model 2 (the double exponential) demonstrates an exponential growth from birth followed by an exponential decline.

3.3.2 Parameter estimation

In the two models, we assume that the parameters are associated with the population, rather than individuals, as the data is cross-sectional in nature.

Definition 3.3.1. Given predictor values t_1, t_2, \dots, t_n and the observed values Y_i , where $i = 1, \dots, n$, the unknown mean function, $\mu(t) = E(y_i|t_i)$, can be approximated by a parametric functions f_j , $j = 1, 2$. These functions are formally called regression functions. Define

$$Y_i = f_j(t_i, \boldsymbol{\beta}) + \epsilon_i,$$

where the unknown, $\boldsymbol{\beta}$, is a vector of parameters in f_j and t is the predictor variable.

The random error, ϵ_i , by construction, is the difference between the observed Y and the mean functions $f_j(t, \boldsymbol{\beta})$. We assume that $E(\epsilon_i) = 0$ and $\text{var}(\epsilon_i) = \sigma_i^2$. The goal is to estimate the unknown parameter $\boldsymbol{\beta}$ by minimizing the distance between the observed Y and $f_j(t, \boldsymbol{\beta})$. The nonlinear least squares method aims at obtaining the estimates of $\boldsymbol{\beta}$ that minimizes the residual sum of squares (SSR) defined as

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n (Y_i - f_j(t_i, \boldsymbol{\beta}))^2. \quad (3.3.3)$$

Let $\hat{\boldsymbol{\beta}}$ be the solution of (3.3.3), then $\hat{\boldsymbol{\beta}}$ is called the least square parameter estimate of $\boldsymbol{\beta}$. Suppose we assume homogeneity in the variance of ϵ_i and let this value to be $\text{var}(\epsilon_i) = \sigma^2$, then the estimated variance is given by

$$\hat{\sigma}^2 = \frac{S(\hat{\boldsymbol{\beta}})}{n}. \quad (3.3.4)$$

Considering (3.3.3), we obtain the solution set for the parameters $\hat{\beta}_j$ from the partial derivative with respect to β_k , $k = 0, 1, 2, \dots$, if f_j , $j = 1$ and β_k , $k = 0, 1, 2, 3$, if f_j , $j = 2$ equated to zero, i.e.,

$$\begin{aligned} \frac{\partial S(\beta_k)}{\partial \beta_k} &= -2 \sum_{i=1}^n (Y_i - f_j(t_i, \boldsymbol{\beta})) \frac{\partial f_j(t_i, \beta_k)}{\partial \beta_k} \\ \implies -2 \sum_{i=1}^n (Y_i - f_j(t_i, \boldsymbol{\beta})) \frac{\partial f_j(t_i, \beta_k)}{\partial \beta_k} &= 0 \end{aligned} \quad (3.3.5)$$

This is referred to as *nonlinear least squares (nls)*. Due to the nonlinear nature of $f_j(t_i, \boldsymbol{\beta})$, the solution to (3.3.5) cannot easily be explicitly obtained. However, iterative numerical procedures such as the Gauss-Newton algorithm may be used for this (see discussion by Huet [49] and Bates & Pinheiro [89]). Here, Y_i , $i = 1, 2, \dots, n$, are the counts of the lymphocyte cell markers corresponding to particular ages represented by t_i of subject i .

3.3.2.1 Modelling heteroscedasticity

One of the underlying assumptions in modelling is that of equal variances of data points about the deterministic part of the model (fitted values). In other words, the deviations of the error terms in the regression model are constant across all the explanatory variables. However, this assumption normally fails, and did so in the present case. Heteroscedasticity is present when the magnitude of the error term is variable within the predictor variable. Mathematically, heteroscedasticity is defined as

$$\text{var}(\epsilon_i) = \sigma_i^2.$$

Weights in generalized nonlinear least squares were used to address this problem. Weights were expressed as the power-of-the-mean (fitted) response (*varPower()*) in *gnls* [89]. Plotting the data revealed that the variance structure around the expected mean function declined with advancing age (see Figure 3.2). Weights are defined as

Chapter 3. Appropriate Models for Age-related changes in lymphocyte cell markers

$$\sigma_i^2 = \sigma^2 g(t_i, \boldsymbol{\beta}, \tau) = \sigma^2 f_j(t_i, \boldsymbol{\beta})^\tau.$$

To estimate the above parameters, the generalized nonlinear least squares *gnls* library in the *NLME* package in the R environment was used, specifically because of the heteroscedastic variance structure and the non-linear mean function.

3.3.3 Model comparison and selection

In this study, we used the likelihood ratio test, Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

3.3.3.1 Likelihood ratio test

The likelihood ratio test is used to compare nested models. Let β_1 denote the set of estimated parameters from the less restricted model and β_2 denote the set of estimated parameters for the restricted model. Then the likelihood ratio is defined as

$$\begin{aligned} LR &= 2 \log \left[\frac{L(\beta_1)}{L(\beta_2)} \right] \\ &= 2[\log L(\beta_1) - \log L(\beta_2)]. \end{aligned}$$

$LR \sim \mathcal{X}_p$, where p is the degrees of freedom.

3.3.3.2 Information criteria

Suppose we have nested models M_1, \dots, M_q which we wish to compare. We rank the models $M_1 \prec M_2 \prec \dots \prec M_q$ in terms of their complexity, so that $M_i \prec M_j$, $i, j = 1, 2, \dots, q$. The most common methods which are based on information criterion are the AIC and BIC which combines both the model complexity and the given number of parameters.

a) Akaike's Information Criterion (AIC)

This method was proposed by Akaike in 1973 following the works by Kullback-Leibler information for model selection [73]. The AIC is defined as

$$AIC = -2(\log \text{likelihood}) + 2p,$$

where p is the number of the estimated parameters included in the model. Lower AIC values indicate better model. To compare the models, delta AIC and Akaike

weights were obtained. The delta AIC is defined as

$$\text{Akaike } \Delta_i = AIC_i - \min AIC$$

while the Akaike weights are defined as

$$\text{Akaike } w_i = \frac{\exp(-0.5\Delta_i)}{\sum_{r=1}^M \exp(-0.5\Delta_r)}.$$

Akaike weights represents the ratio delta AIC for each model relative to the collection of candidate models and provides the strength of evidence for each model. They indicate the probability that the model is the best among the whole set of candidate models. The interpretation is simple, i.e. the greater the weight, the better the model [73].

b) Bayesian information criterion (BIC)

BIC, similar to the AIC, assesses the overall fit of a model and can be used to compare both nested and non-nested models. It identifies the model which is most to have generated the data. The BIC is defined as

$$BIC = -2(\log\text{likelihood}) + p\log(n)$$

The criteria for model comparison using BIC is the same as for the AIC,

$$\text{Bayesian } \Delta_i = BIC_i - \min BIC.$$

Δ_i provides the evidence against model i .

Chapter 3. Appropriate Models for Age-related changes in lymphocyte cell markers 64

3.4 Results

Figure 3.2 provides a summary of empirical changes in three lymphocyte cell markers (CD19+, CD8+ and CD3+) by age-grouping. The ‘age-blocks’ are the same as those applied by Comans-Bitter et al. [24] and Shearer et al. [95].

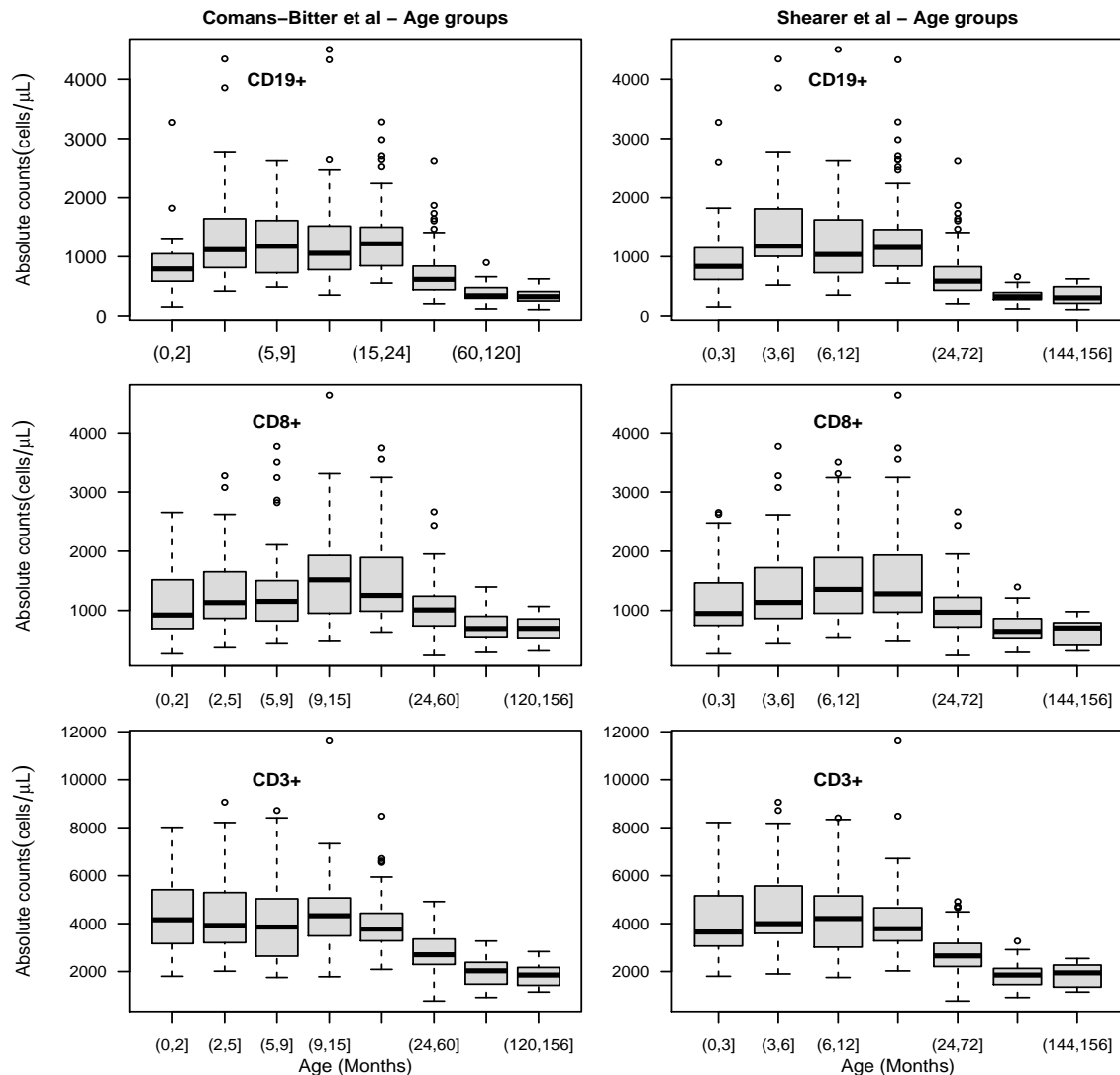


Figure 3.2: The current data represented using the Comans-Bitter et al. [24] and Shearer et al. [95] ‘age-blocks’. For CD19+ and CD8+ the markers clearly increase in the first 12 months and then decline thereafter. In the case of CD3+ the trend is also present but less pronounced.

3.4.1 Parameter estimates

Model parameters were estimated using weighted *gnls* for the CD3+, CD19+, CD8+, CD4+, the ratio of CD4+ naive/memory, CD16+CD56+ and CD3-CD56+ groups:

Table 3.2: Comparison of parameter estimates for the two models.

Lymphocytes	Parameters	Double Exponential Model			Single Exponential Model		
		Estimates	95% Confidence Interval	p-value	Estimates	95% Confidence Interval	p-value
CD3+	β_0	1838.114	(1674.603, 2001.625)	0.000	1665.714	(1419.235, 1912.192)	0.000
	β_1	6390.389	(5240.055, 7540.723)	0.000	3218.810	(2878.038, 3559.582)	0.000
	β_2	0.047	(0.040, 0.055)	0.000	0.026	(0.019, 0.033)	0.000
	β_3	0.108	(0.045, 0.171)	0.001			
CD19+	β_0	308.142	(259.398, 356.885)	0.000	279.957	(216.938, 342.975)	0.000
	β_1	1776.413	(1201.939, 2350.888)	0.000	1285.270	(1131.061, 1439.479)	0.000
	β_2	0.038	(0.028, 0.048)	0.000	0.028	(0.021, 0.035)	0.000
	β_3	0.207	(-0.029, 0.442)	0.085			
CD8+	β_0	728.810	(643.107, 814.512)	0.000	588.653	(437.002, 740.305)	0.000
	β_1	1384.083	(809.443, 1958.723)	0.000	1004.887	(840.705, 1169.069)	0.000
	β_2	0.038	(0.023, 0.054)	0.000	0.021	(0.011, 0.031)	0.000
	β_3	0.223	(-0.019, 0.464)	0.071			
CD4+	β_0	1127.695	(1008.358, 1247.032)	0.000	1070.594	(935.140, 1206.048)	0.000
	β_1	4988.450	(4260.474, 5716.426)	0.000	2174.424	(1940.508, 2408.341)	0.000
	β_2	0.050	(0.035, 0.064)	0.000	0.032	(0.024, 0.040)	0.000
	β_3	0.061	(0.006, 0.116)	0.030			
Ratio CD4+ naive/memory	β_0	1.025	(0.737, 1.314)	0.000	0.925	(0.597, 1.253)	0.000
	β_1	2.788	(2.332, 3.244)	0.000	2.606	(2.259, 2.953)	0.000
	β_2	0.026	(0.015, 0.037)	0.000	0.021	(0.013, 0.030)	0.000
	β_3	0.874	(0.200, 1.549)	0.011			
CD16+CD56+	β_0	265.371	(217.848, 312.894)	0.000	259.848	(206.680, 313.016)	0.000
	β_1	1012.223	(651.045, 1373.402)	0.000	432.207	(316.128, 548.286)	0.000
	β_2	0.057	(0.023, 0.092)	0.001	0.038	(0.017, 0.060)	0.000
	β_3	0.074	(-0.065, 0.213)	0.295			
CD3-CD56+	β_0	335.200	(-1948.868, 2619.268)	0.773	260.010	(209.763, 310.256)	0.000
	β_1	1031.178	(-1055.107, 3117.462)	0.332	429.836	(312.918, 546.753)	0.000
	β_2	0.036	(-0.012, 0.084)	0.138	0.040	(0.018, 0.062)	0.000
	β_3	0.010	(-0.230, 0.250)	0.936			

All the parameter estimates for all markers using the single exponential model were statistically significant (at alpha = 0.05). In the double exponential model, the estimate of the additional β_3 parameter was not statistically significant for CD19+, CD8+, CD16+CD56+ and CD3-CD56+, thus, there is no reason to prefer the double exponential model. Although, the double exponential model is not as likely as the single exponential, but contains more inherent information.

Chapter 3. Appropriate Models for Age-related changes in lymphocyte cell markers

Little 'structural' differences are visible in the overall fits of the two models, as shown in Figures 3.3.

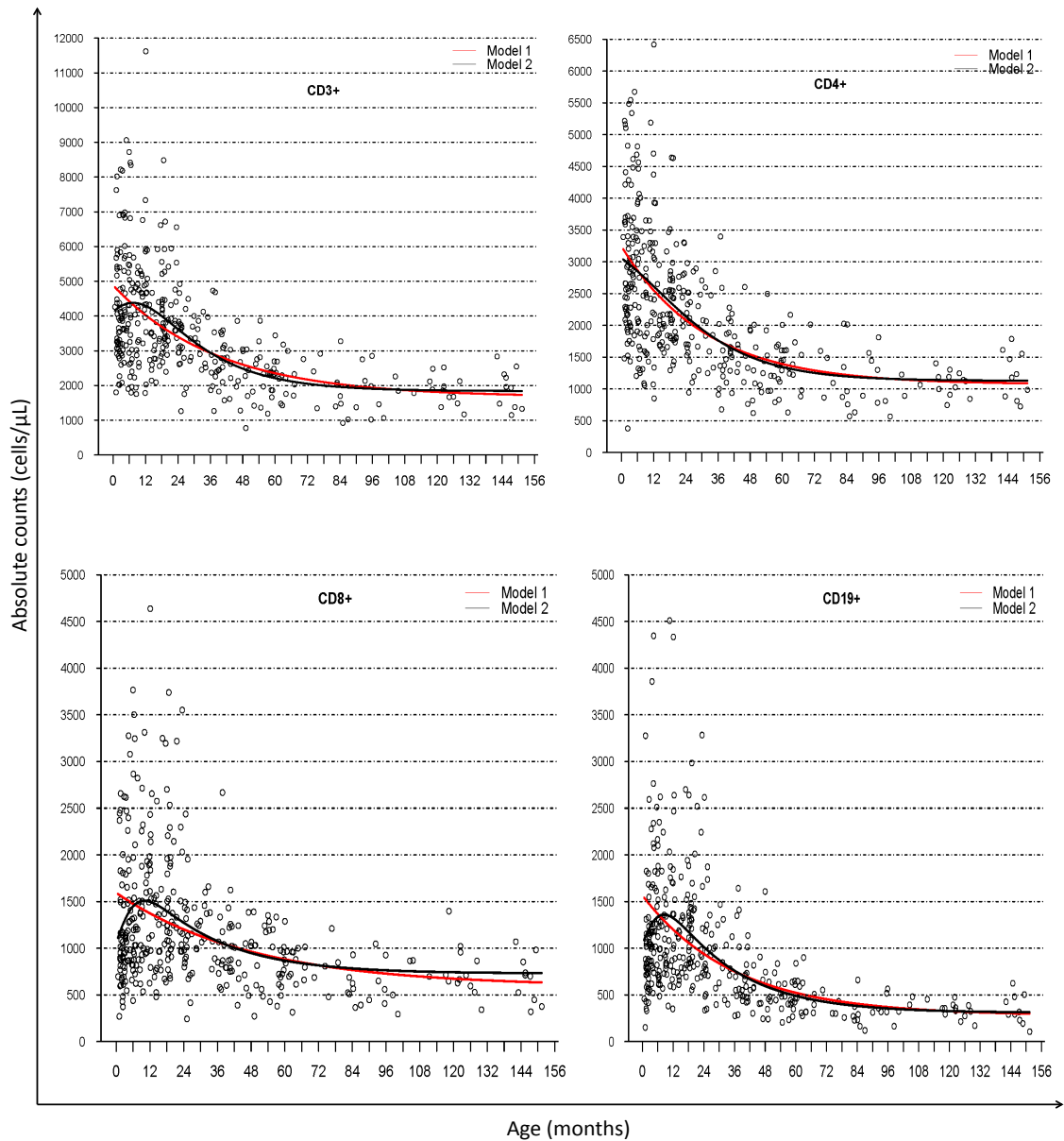


Figure 3.3: Plots showing changes in absolute counts of CD3+, CD4+, CD8+ and CD19+ over time for healthy South African children. For CD3+, CD8+ and CD19+, the double exponential model predicts a rise in the counts of cell makers from birth to about 12 months followed by an exponential decline to older ages. This pattern is less apparent for CD4+.

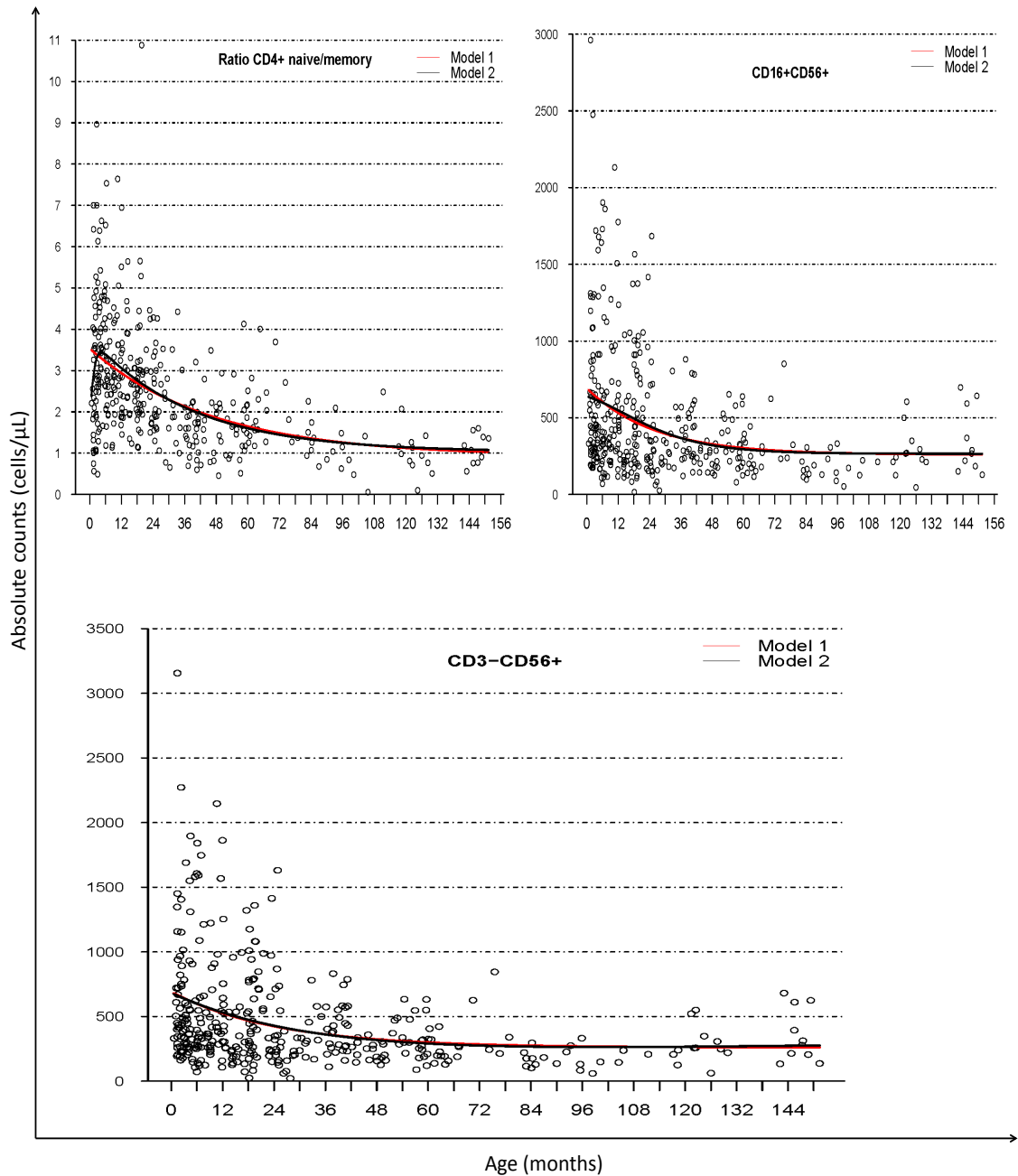


Figure 3.4: Plots showing changes in absolute counts of ratio of CD4+ naive to memory T cells, CD16+CD56+ and CD3-CD56+ over time for healthy South African children. For the ratio of CD4+ naive to memory T cells, the double exponential model predicts an initial rise in the counts from birth to about 12 months followed by an exponential decline. For CD16+CD56+ and CD3-CD56+, the prediction for the double exponential model is similar to that of the single exponential model and thus either of the models is preferred.

Chapter 3. Appropriate Models for Age-related changes in lymphocyte cell markers

3.4.2 Models-fit comparison and assessment

For CD3+, CD19+, CD8+, CD16+CD56+ and CD3 CD56+ the double exponential model had higher AIC-weights of 0.998, 1.000, 1.000, 0.790 and 0.820 respectively. The single exponential model had higher AIC-weights in ratio of CD4+ naive/memory (0.650) and for CD4+ (0.540).

Table 3.3: Log-likelihoods and AIC for the two fitted models

Lymphocytes	Model	Loglik	δ Loglik	AIC	δ AIC	AIC weight
CD3 +	2	-3209	7.5	6430.2	0.0	0.998
	1	-3216.5	0.0	6443.2	13.0	0.002
CD19 +	2	-2861.9	21.6	5736	0.0	1.000
	1	-2883.5	0.0	5777.1	41.1	< 0.001
CD8 +	2	-2926.5	18.6	5865.2	0.0	1.000
	1	-2945	0.0	5900.2	35.1	< 0.001
CD4 +	1	-3046	0.0	6102.2	0.0	0.540
	2	-3045.2	0.9	6102.5	0.3	0.460
Ratio CD4+ naïve/memory	1	-590	0.0	1190.1	0.0	0.650
	2	-589.6	0.4	1191.3	1.2	0.350
CD16+CD56+	2	-2736.4	2.3	5485.1	0.0	0.790
	1	-2738.8	0.0	5487.7	2.6	0.210
CD3–CD56+	2	-2728.1	2.6	5468.4	0.0	0.820
	1	-2730.7	0.0	5471.5	3.1	0.180

*Model 1 refers to the single exponential model while Model 2 refers to the double exponential model. Higher AIC-weight implies better model.

Overall, based on the information content and the AIC the double exponential may be considered the better of the candidate models. In assessing the quality of generalized nonlinear least squares (*gnls*) fits we consequently focus our analysis on the double exponential model from this point forward. As shown in Figure 3.5, plots of residuals versus fitted values are used to assess the adequacy of the heteroscedastic fit. For most of the markers, the standardized residuals are symmetrically distributed around the zero line, with fairly uniform variance. This indicates that the power variance model we employed adequately models the heteroscedasticity for the majority of the markers.

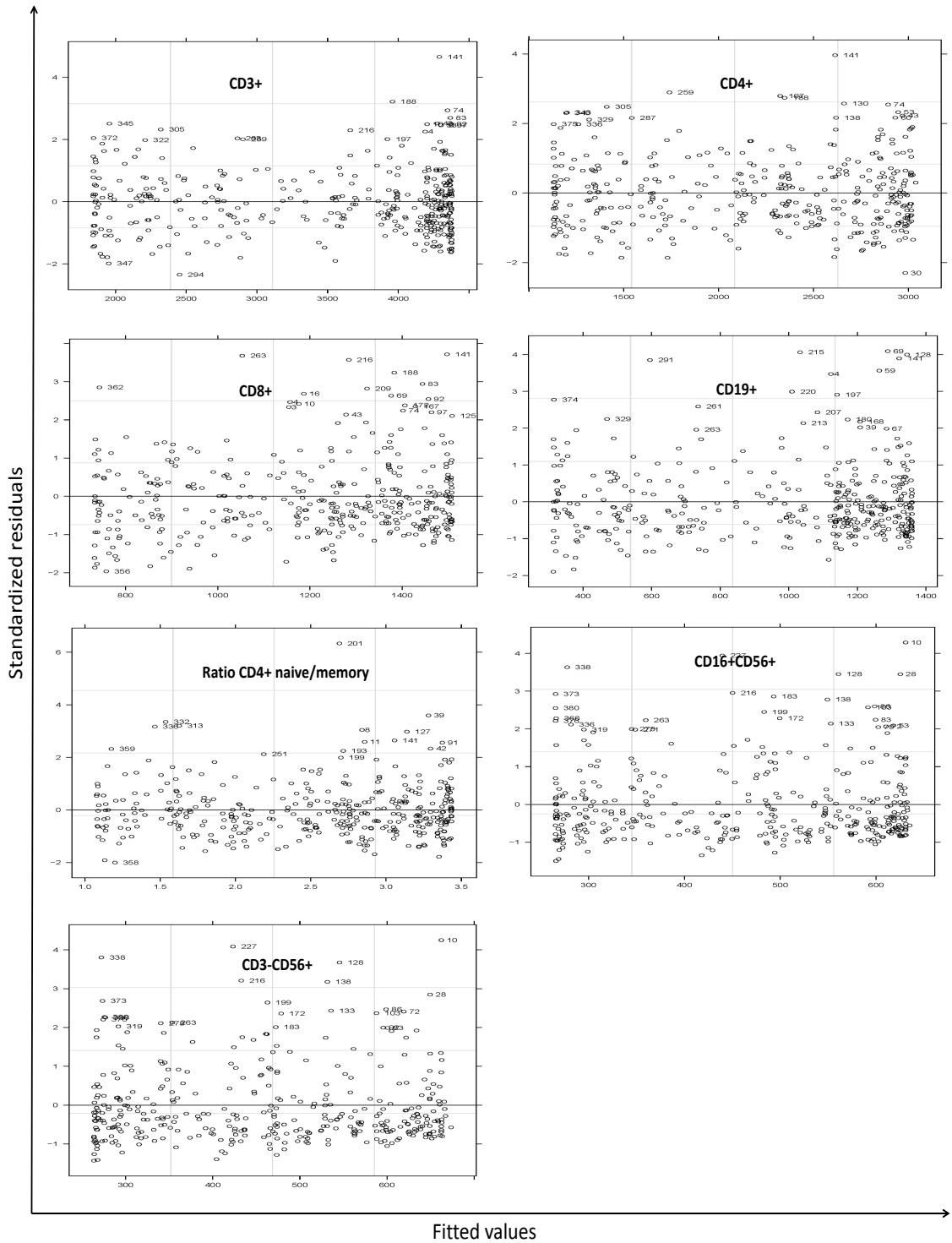


Figure 3.5: Plots of standardized residuals against the fitted values to assess heteroscedasticity using *gnls* for the double exponential.

Chapter 3. Appropriate Models for Age-related changes in lymphocyte cell markers

The correlation of the error terms are assessed using empirical autocorrelation function plots shown in Figure 3.6. No significant autocorrelations between the residuals are visible. For some cell markers, in some lags, the autocorrelations were slightly outside the 5% significance (lower and upper critical) levels, but this does not seem sufficient to conclude that the errors are non-random.

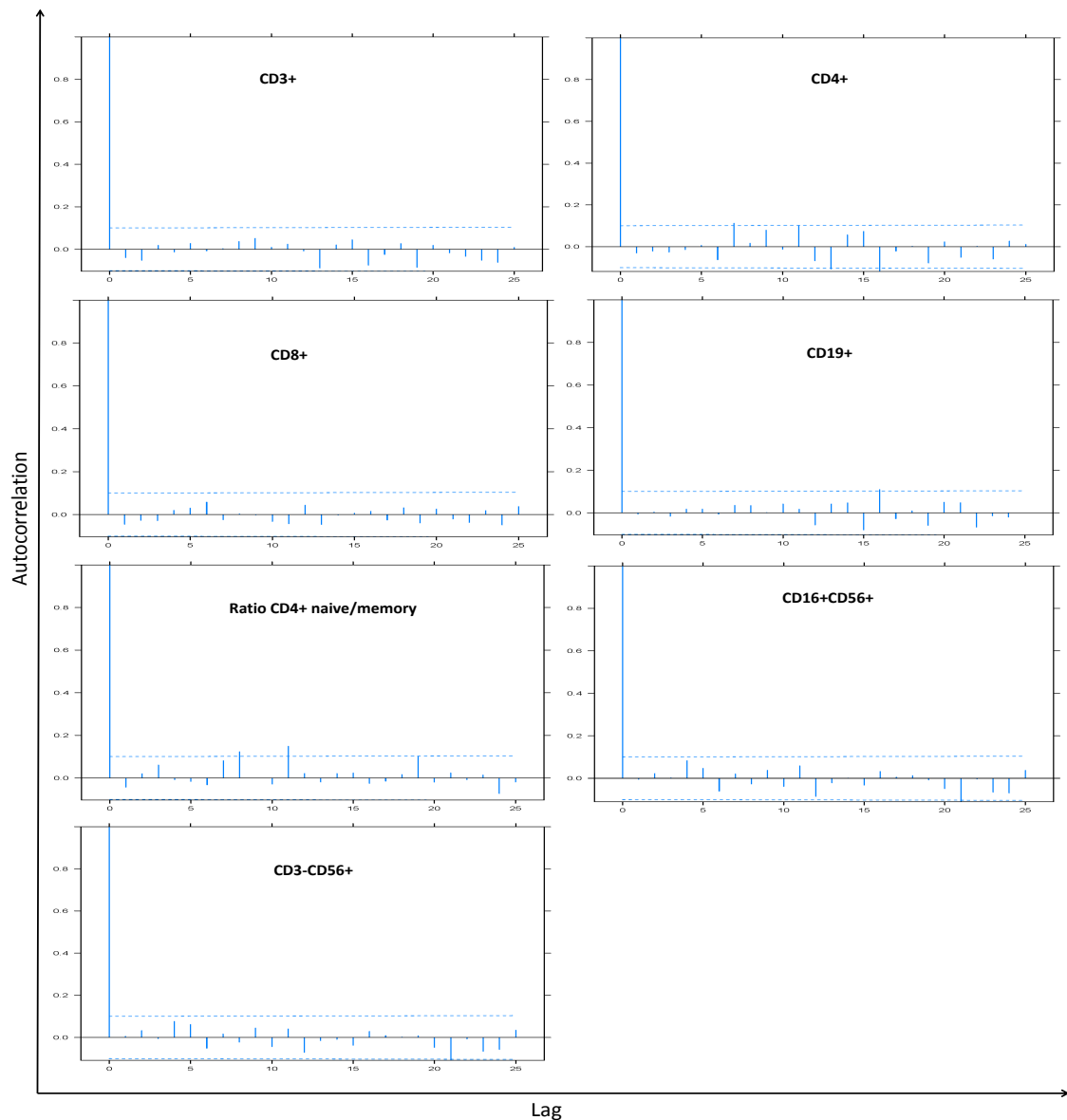


Figure 3.6: Plot of empirical autocorrelation function for standardized residuals to assess the correlation of the errors for the double exponential model using *gnls*.

3.5 Summary

Our analysis demonstrated that age-related changes in count of peripheral blood lymphocyte cell-markers (immune biomarkers) with advances in age were present for all markers investigated. In all cases, in the first year of life, the population count of lymphocyte cell marker was larger compared to that at greater age. This finding is in agreement with prior studies [8, 48, 50, 69, 93, 105]. For some markers the nature of the change in the first year was not clear. For this reason, two models were fitted to the data to determine which of them provided the mathematically better fit. Both models demonstrated relatively similar fits and also provided continuous predictions over time (age). The single exponential model predicts that the absolute count in the mean population of the blood lymphocyte cell markers only declines from a maximum at birth to an asymptote in adulthood. This trend is the same as that proposed by a prior European study [48]. On the other hand, our double exponential model predicts a rise in absolute count to a maximum at approximately 12 months of age followed by an exponential decline thereafter. The latter findings agree with prior mechanistic models and empirical study measurements in the young age reference ranges [8, 24, 33, 95, 111]. Our double exponential model also incorporates to some extent the single exponential model in that it reduces to its mathematical form under particular conditions. The mathematical and statistical comparison between the two models also singled out the double exponential as the best in predicting the change in cell markers with age, as seen in higher AIC-weights for most of the cell markers in the case of double exponential model.

Chapter 4

The influence of covariate factors on Age-related changes in lymphocyte cell markers

4.1 Introduction

An important function of circulating blood lymphocytes is the regulation of immune homeostasis. Invasive pathogens may affect the immune balance which might lead to immune-related disorders. Quantifying lymphocyte cell markers is important in evaluating disease progression and optimizing treatment interventions [20]. Regional, i.e. African and international, variations in cell marker quantities have been attributed to differences in age, sex, ethnicity, altitude, diet, endemic infections, autoimmune conditions and the testing methodologies employed [4, 13, 20, 39, 60].

In HIV, various factors have been found to be associated with low CD4 counts, such as age, sex, ethnicity and geographic region. Patients undergoing ARV therapy with low CD4 counts have been shown to be at high risk of acquiring opportunistic infections. Other factors such as treatment adherence, immunosuppressive drugs, socio-economic and psychosocial factors are correlated to CD4 T cell counts in both HIV positive and negative patients [77]. In this chapter, we investigate the influence of covariate factors for healthy South African children.

4.2 Methods

In this section we extend the double exponential to incorporate covariate factors and describe the parameter estimation methods.

4.2.1 Generalized nonlinear least squares

Mixed effects models are useful in determining the relationships between variables of interest and covariates which may be grouped according to at least one classification factor (covariate levels). Longitudinal or repeated measurement data is required for mixed effects models, which model the variance-covariance structure of within-group dependencies. In the present case, we only have cross-sectional data with covariates as the only grouping factors. We consequently apply an extended generalized nonlinear least squares regression, a simplification of nonlinear mixed effects, to model within-group variance-covariance structure without random effects [89], i.e. the current models are at a *population* level only.

Considering the i th observation in the j th covariate group, the extended nonlinear least squares is defined as

$$\begin{aligned} y_{ij} &= f(\theta_{ij}, b_{ij}) + \epsilon_{ij}, & i = 1, \dots, n_j; \quad j = 1, \dots, m. \\ \theta_{ij} &= \mathbf{A}_{ij}\boldsymbol{\beta}, \end{aligned} \quad (4.2.1)$$

where b_{ij} are the covariates, θ_{ij} are the parameters, m is the number of covariate groups and ϵ_{ij} are the normally distributed error terms. The generalized nonlinear least square (*gnls*) which does not involve two-stage modeling, described in Equation 4.2.1, by the covariate grouping parameters θ_{ij} , is defined as

$$\begin{aligned} y_i &= f_i(\theta_i, b_i) + \epsilon_i \\ \theta_i &= \mathbf{A}_i\boldsymbol{\beta}, \quad \epsilon_i \sim N(0, \sigma^2\Delta_i). \end{aligned}$$

Inference and estimation of this model is through generalized nonlinear least squares. The transformation suggested by Thisted [102] and Pinheiro & Bates [89] on this model leads to the ‘classic’ nonlinear regression model defined as

$$\begin{aligned} y_i^* &= f_i^*(\theta_i, b_i) + \epsilon_i^* \\ \theta_i &= \mathbf{A}_i\boldsymbol{\beta}, \quad \epsilon_i^* \sim N(0, \sigma^2I). \end{aligned} \quad (4.2.2)$$

Chapter 4. The influence of covariate factors on Age-related changes in lymphocyte cell markers 44

Parameters in 4.2.2 are estimated using both maximum likelihood and least squares. By letting N be the total observations, the log-likelihood function of a *gnls* model is defined as

$$l(\theta, \sigma^2 | \mathbf{y}) = -\frac{1}{2} \left[N \log(2\pi\sigma^2) + \sum_{i=1}^M \left(\frac{(y_i^* - f_i^*(\theta_i, b_i))^2}{\sigma^2} + \log(\Delta_i) \right) \right]. \quad (4.2.3)$$

Define $f_i^*(\boldsymbol{\beta}) = f_i^*(\theta_i, b_i)$, where $\boldsymbol{\beta}$ is a fixed parameter. Also let λ be the profile log-likelihood estimator, then the maximum likelihood estimator of σ^2 is given by

$$\hat{\sigma}^2(\boldsymbol{\beta}, \lambda) = \sum_{i=1}^M \left(\frac{(y_i^* - f_i^*(\boldsymbol{\beta}))^2}{N} \right). \quad (4.2.4)$$

We replace $f_i^*(\theta_i, b_i)$ with $f_i^*(\boldsymbol{\beta})$ and σ^2 with $\hat{\sigma}^2(\boldsymbol{\beta}, \lambda)$ in Equation 4.2.3 to obtain the profile log-likelihood defined as

$$l(\boldsymbol{\beta}, \lambda | \mathbf{y}) = -\frac{1}{2} \left[\left(N \log \left(\frac{2\pi}{N} \right) + N \right) + \log \left(\sum_{i=1}^M (y_i^* - f_i^*(\boldsymbol{\beta}))^2 \right) + \sum_{i=1}^M \log(\Delta_i) \right]. \quad (4.2.5)$$

Pinheiro and Bates [89] describe the Gauss-Seidel algorithm for obtaining maximum likelihood estimates for $\boldsymbol{\beta}$ and λ from Equation 4.2.5. The computational method uses restricted maximum likelihood estimation (REML), thus, the estimate of σ^2 in Equation 4.2.4 becomes

$$\tilde{\sigma} = \sum_{i=1}^M \left(\hat{\Delta}_i^{-\frac{-T}{2}} \frac{(y_i - f_i(\hat{\boldsymbol{\beta}}))^2}{(N - p)} \right),$$

where p is the number of parameters.

4.2.2 Extending the double exponential model

We incorporate the covariates into our model by extending the double exponential model 3.3.2 described in 3.3.1.2 and then describe model fitting process.

4.2.2.1 Adding covariates

To investigate the influence of covariates on the overall age-related changes of the lymphocytes, the parameters β_0 , β_1 , β_2 and β_3 are adjusted as a sum of the intercept term and the covariate effect sizes. We are interested in observing the differences at birth and changes with advancing age. The average of β_0 and β_1 gives the expected cell counts at birth. We then assume that effects on either β_0 or β_1 implies an effect on the average of the two. β_2 estimates the rate of cell decline over time. Further, we assume that the shape parameter, β_3 , is the same across the covariates. Therefore, we only adjust β_0 and β_2 .

We define Ω as the combined effect sizes of the covariates. It follows that Ω can be expressed as a linear combination of the covariates. Let G represent sex (female = 0, male = 1), R represent race (African black = 0, Mixed race/Coloured = 1), E represent history of exposure to illness (No history of exposure = 0, history of exposure = 1), and F represents the type of feeding within the first six months following birth (exclusive breast feeding = 0, bottle feeding = 1, mixed feeding = 2). Then the terms incorporating covariates are defined as

$$\begin{aligned}\Omega_0 &= \gamma_0 G_{g \in \{0,1\}} + \alpha_0 R_{r \in \{0,1\}} + \tau_0 E_{e \in \{0,1\}} + \phi_0 F_{f \in \{0,1,2\}} \\ \Omega_2 &= \gamma_2 G_{g \in \{0,1\}} + \alpha_2 R_{r \in \{0,1\}} + \tau_2 E_{e \in \{0,1\}} + \phi_2 F_{f \in \{0,1,2\}},\end{aligned}$$

where γ_i , α_i , τ_i and ϕ_i , $i = 0, 2$, are the effect sizes as a result of incorporating sex, race, history of exposure and feeding, respectively, into the model. Therefore, we define the extended double exponential as

$$f^*(t, \beta, \Omega) = \frac{(\beta_0^* + \Omega_0) + \beta_1 \exp(-(\beta_2^* + \Omega_2)t)}{1 + \exp(-\beta_3 t)}, \quad (4.2.6)$$

where $f^*(t, \beta, \Omega)$ is the expected cell count at age t , the average of $(\beta_0^* + \Omega_0)$ and β_1 estimates the cell counts at birth ($t = 0$), $(\beta_2^* + \Omega_2)$ estimates the rate of change of cell markers over time while β_0^* and β_2^* are the estimated intercept values for β_0 and β_2 as a result of the adjustment.

Chapter 4. The influence of covariate factors on Age-related changes in lymphocyte cell markers 46

4.2.2.2 Parameter estimation using *gnls*

The extended model describing lymphocyte counts cell markers y_{ij} in the j th covariate at age i , where j represents the 4 covariates described above is then redefined as

$$y_{ij} = \frac{(\beta_0^* + \Omega_0) + \beta_1 \exp(-(\beta_2^* + \Omega_2)age_{ij})}{1 + \exp(-\beta_3 age_{ij})} + e_{ij}, \quad (4.2.7)$$

The parameters β_0^* , β_1 , β_2^* , β_3 and covariate adjustment parameters γ_i , α_i , τ_i and ϕ_i , $i = 0, 2$ are estimated using using generalised nonlinear least squares (gnls) described above and implemented in library *gnls* in the *nlme* package in R.

4.3 Results

4.3.1 Descriptives

The population characteristics and covariates for the healthy South African children are presented in Table 4.1.

Table 4.1: Characteristics of the study population.

Characteristics	N	Percentage
Children	381	100%
Sex		
Male	174	46%
Female	207	54%
Race		
Black	85	22%
Coloured	296	78%
Food in the first 6 months following birth		
Bottle	46	12%
Mixed	103	27%
Breast	231	61%
History of Exposure to Illness		
Exposed	162	43%
Not Exposed	219	57%

*1 respondent had no information on type of food during 6 months. History of exposure to illness included maternal HIV-status, antenatal/neonatal/past medical event requiring admission or recent infection within last month.

The mean age of the children was 32.08 (sd = 34.9) months, minimum and maximum age was 0.62 and 151.33 months respectively. 54% of the subjects were female, with a mean age of 31.19 months, a minimum age of 0.62 and a maximum of 147.41 months.

4.3.2 Model prediction comparisons and extended model estimates

4.3.2.1 CD3+ T cells

Figure 4.1 shows a comparison of independently predicted counts of CD3+ for each of the investigated covariates.

1. **Sex and history of exposure to illness:** Female subjects had higher predicted cell count than males at almost all ages. The same was observed for maternal exposure to illness. Those who had previous exposure had increased predicted cell counts at birth compared to those who had not.
2. **Race:** There was no major difference in the trends of predicted CD3+ T cells counts between African blacks and Mixed Race/Coloured children.
3. **Type of food:** Children who were fed on bottle feeds had higher and lower predicted cell counts at birth and older ages respectively, compared to the other two groups. Breast-fed children had slightly higher counts at birth and older ages compared to those with mixed-feeding.

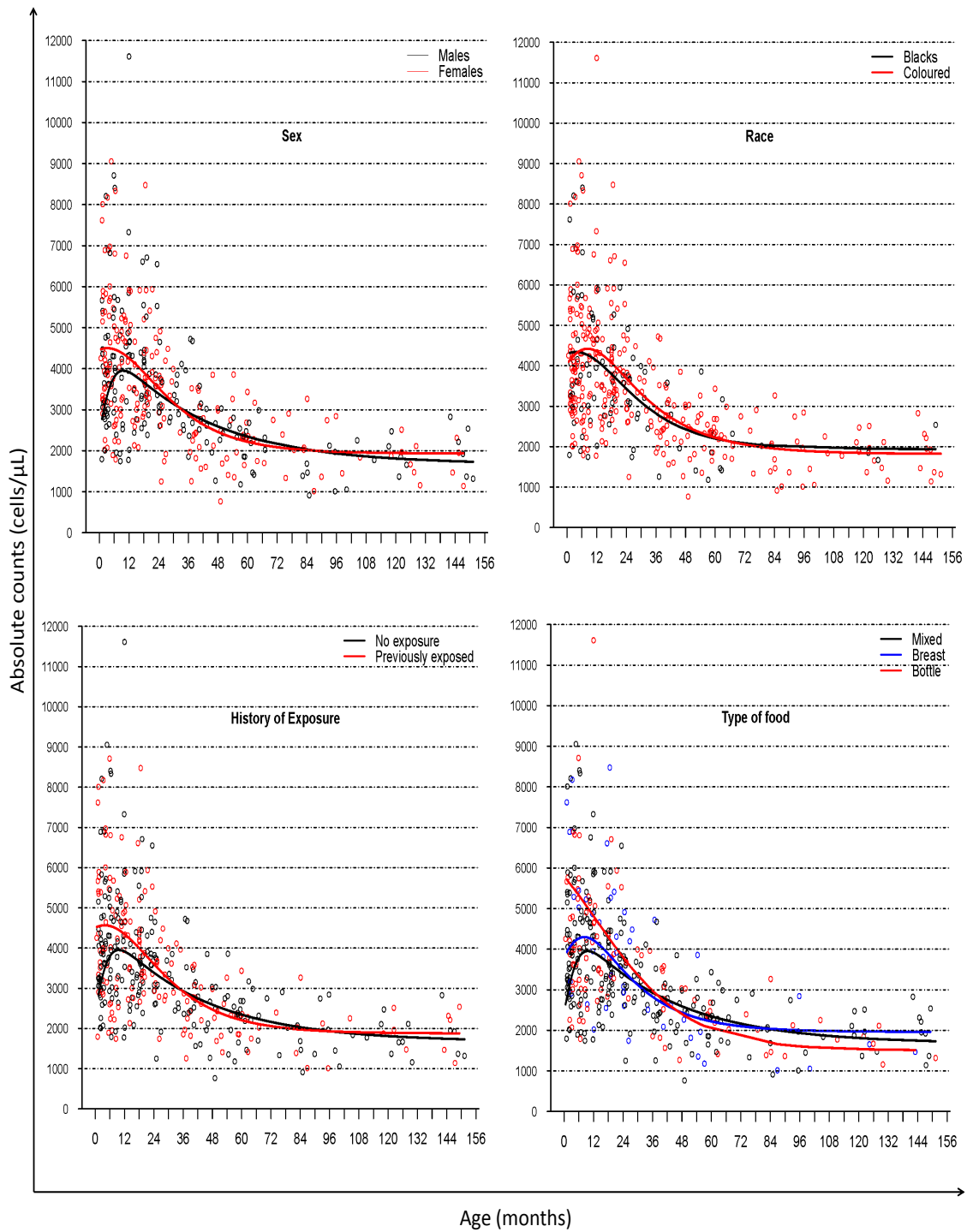


Figure 4.1: A plot showing age-related changes in CD3+ T cells across the different covariates.

Chapter 4. The influence of covariate factors on Age-related changes in lymphocyte cell markers 50

When using independent predictions for the covariates some differences were observable in the figures. However, in the extended model, the covariates had no statistically significant influence on age-related changes in the counts of CD3+ cell markers (Table 4.2).

Table 4.2: Parameter estimates for the extended model with CD3+.

	Estimate	95% Confidence Interval	p-value
β_1	6373.047	[5178.343, 7567.752]	0.000
β_3	0.114	[0.045, 0.183]	0.001
β_0			
β_0 .(Intercept)	2146.510	[1617.890, 2675.130]	0.000
γ_0 .(Sex - Male)	-197.740	[-508.784, 113.303]	0.212
α_0 .(Race - Coloured)	-150.512	[-606.080, 305.057]	0.516
τ_0 .(Exposure - Exposed)	18.339	[-299.076, 335.754]	0.910
ϕ_0 .(Feeding - Bottle)	-302.435	[-788.300, 183.430]	0.222
ϕ_0 .(Feeding - Mixed)	-220.855	[-598.793, 157.084]	0.251
β_2			
β_2 .(Intercept)	0.063	[0.041, 0.085]	0.000
γ_2 .(Sex - Male)	-0.006	[-0.018, 0.006]	0.295
α_2 .(Race - Coloured)	-0.009	[-0.027, 0.009]	0.331
τ_2 .(Exposure - Exposed)	0.000	[-0.012, 0.012]	0.962
ϕ_2 .(Feeding - Bottle)	-0.011	[-0.027, 0.006]	0.207
ϕ_2 .(Feeding - Mixed)	-0.011	[-0.025, 0.003]	0.112

*The reference groups are: females for sex; blacks for race; non-exposed for history of exposure to illness and breast-feeding for type of feeding during 6 months following birth.

4.3.2.2 CD4+ T cells

Figure 4.2 shows a comparison of independently predicted counts of CD4+ for each of the investigated covariates.

1. **Sex:** At birth, female subjects demonstrate higher cell counts compared to males. All male subjects have an initial rise in cell counts from birth to a maximum at 12 months followed by exponential decline.
2. **History of exposure to illness:** During early life, subjects that had no maternal exposure to illness had increased cell counts compared to those who were previously exposed. All previously exposed subjects had an initial rise in cell counts from birth to a maximum at 12 months followed by an exponential decline.
3. **Race:** Black children had slightly increased predicted counts at birth and older ages versus Mixed race/Coloured subjects. For black children there was an exponential decline immediately from birth while for coloured children, cell counts initially rose to approximately 12 months and then declined thereafter.
4. **Type of food:** Children who were bottle-fed had higher and lower predictions of cell counts at birth and older ages respectively, compared to the other two groups. Breast-fed children had slightly higher counts at birth and older ages compared to those fed on mixed meals. Breast-fed children had an initial rise in cell counts followed by an exponential decline.

Chapter 4. The influence of covariate factors on Age-related changes in lymphocyte cell markers

52

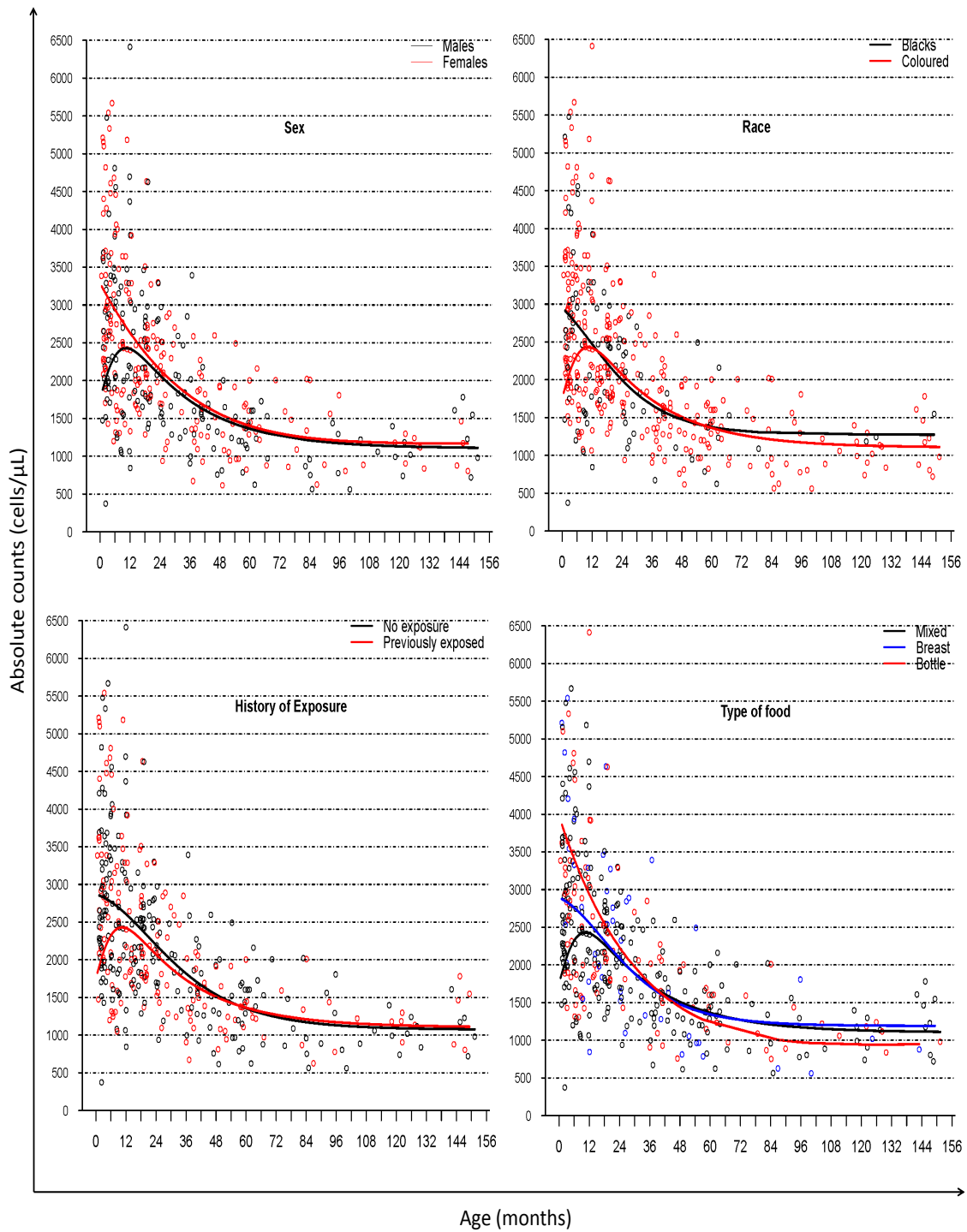


Figure 4.2: A plot showing age-related changes in CD4+ T cells across the different covariates.

The extended model estimates did not show any significantly different effect sizes versus independently estimated predictions for the covariates for CD4+ (Table 4.3)

Table 4.3: Parameter estimates for the extended model for Age-related changes in CD4+.

	Estimate	95% Confidence Interval	p-value
β_1	4857.763	[4088.510, 5627.015]	0.000
β_3	0.079	[0.021, 0.137]	0.008
β_0			
β_0 .(Intercept)	1384.107	[1056.467, 1711.747]	0.000
γ_0 .(Sex - Male)	-187.005	[-381.902, 7.893]	0.060
α_0 .(Race - Coloured)	-147.640	[-429.455, 134.174]	0.304
τ_0 .(Exposure - Exposed)	59.952	[-137.843, 257.747]	0.552
ϕ_0 .(Feeding - Bottle)	-218.728	[-516.265, 78.809]	0.149
ϕ_0 .(Feeding - Mixed)	-76.213	[-307.224, 154.798]	0.517
β_2			
β_2 .(Intercept)	0.069	[0.043, 0.095]	0.000
γ_2 .(Sex - Male)	-0.005	[-0.017, 0.007]	0.407
α_2 .(Race - Coloured)	-0.013	[-0.031, 0.006]	0.182
τ_2 .(Exposure - Exposed)	0.005	[-0.007, 0.018]	0.394
ϕ_2 .(Feeding - Bottle)	-0.011	[-0.027, 0.005]	0.166
ϕ_2 .(Feeding - Mixed)	-0.009	[-0.023, 0.006]	0.230

*The reference groups are: females for sex; blacks for race; non-exposed for history of exposure to illness and breast-feeding for type of feeding during 6 months following birth.

4.3.2.3 CD8+ T cells

Figure 4.3 shows a comparison of independently predicted counts of CD8+ T for each of the investigated covariate.

1. **Sex:** There was no observable difference for male and female subjects.
2. **History of exposure to illness:** Those subjects who had a history of maternal exposure to illness had higher cell count predictions at birth compared to those who had no history of exposure to illness. At older ages, the two fits were not substantially different.
3. **Race:** Black children show higher predicted cell counts at birth, compared to Mixed race/Coloureds for the first 2.5 years after which predicted cell counts remain approximately the same for both groups.
4. **Type of food:** There was little difference between the food types. However, bottle-fed children had slightly increased cell counts at birth compared to those on breast-milk or mixed meals.

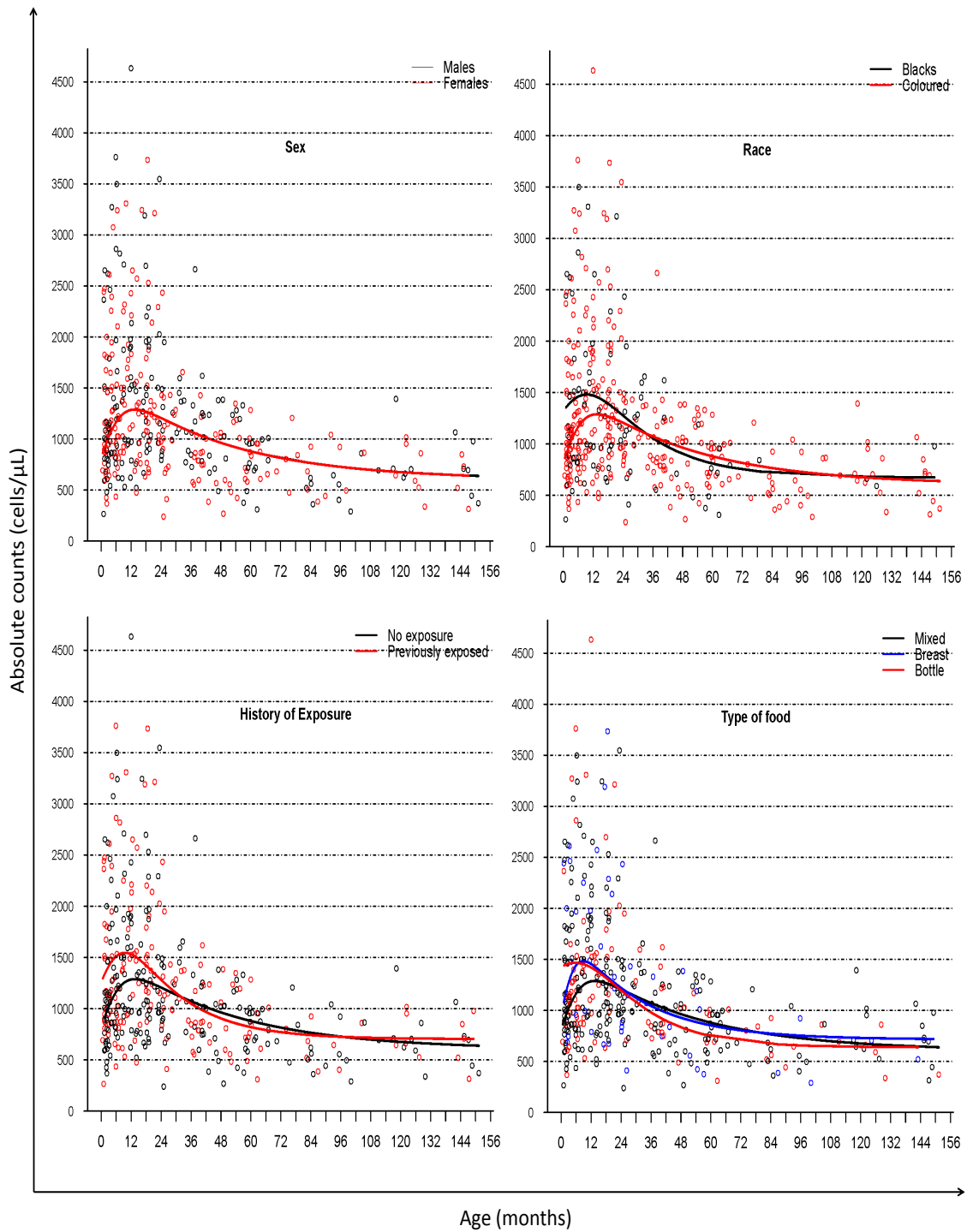


Figure 4.3: A plot showing age-related changes in CD8+ across different covariates.

Table 4.4: Parameter estimates for the extended model for CD8+.

	Estimate	95% Confidence Interval	p-value
β_1	1467.138	[1006.002, 1928.274]	0.000
β_3	0.232	[0.009, 0.455]	0.041
β_0			
β_0 .(Intercept)	612.498	[344.004, 880.992]	0.000
γ_0 .(Sex - Male)	105.622	[-46.678, 257.923]	0.173
α_0 .(Race - Coloured)	49.498	[-176.563, 275.558]	0.667
τ_0 .(Exposure - Exposed)	50.235	[-100.307, 200.778]	0.512
ϕ_0 .(Feeding - Bottle)	-64.146	[-357.480, 229.188]	0.667
ϕ_0 .(Feeding - Mixed)	-295.367	[-578.467, -12.268]	0.041
β_2			
β_2 .(Intercept)	0.039	[0.017, 0.061]	0.001
γ_2 .(Sex - Male)	0.004	[-0.008, 0.015]	0.529
α_2 .(Race - Coloured)	0.004	[-0.009, 0.017]	0.581
τ_2 .(Exposure - Exposed)	0.001	[-0.011, 0.012]	0.912
ϕ_2 .(Feeding - Bottle)	-0.011	[-0.039, 0.017]	0.430
ϕ_2 .(Feeding - Mixed)	-0.025	[-0.045, -0.006]	0.011

*The reference groups are: females for sex; blacks for race; non-exposed for history of exposure to illness and breast-feeding for type of feed during 6 months.

A statistically significant difference was observed in adjustment parameters for feeding type. Children on mixed meals had low CD8+ cells at birth compared to those on bottle or breast-milk ($p = 0.041$). The rate of cell decline was also slower in children on mixed meals ($p = 0.011$) compared to the others. However, Age-related changes in CD8+ cells was not statistically different for breast-fed versus bottle-fed children. There was no statistical differences in sex, race and exposure to illness for CD8+ T cells.

4.3.2.4 CD19+ B cells

Figure 4.4 shows a comparison of independently predicted counts of CD19+ B cells for each of the investigated covariate.

1. **Sex:** The predicted counts for male and female subjects did not significantly differ.
2. **History of exposure to illness:** Subjects without a history of maternal exposure to illness had slightly higher counts at birth but at older ages those who were exposed to illness had slightly higher counts.
3. **Race:** Black children had slightly higher predicted counts compared to Mixed race/Coloureds in the first year of life, after which the trend was similar in the two groups.
4. **Type of food:** Bottle-fed children had higher predictions at birth compared to the other two groups. Breast-fed children had slightly higher counts at birth compared to those on mixed meals in early life. At older ages predictions were similar for the three groups.

Chapter 4. The influence of covariate factors on Age-related changes in lymphocyte cell markers

58

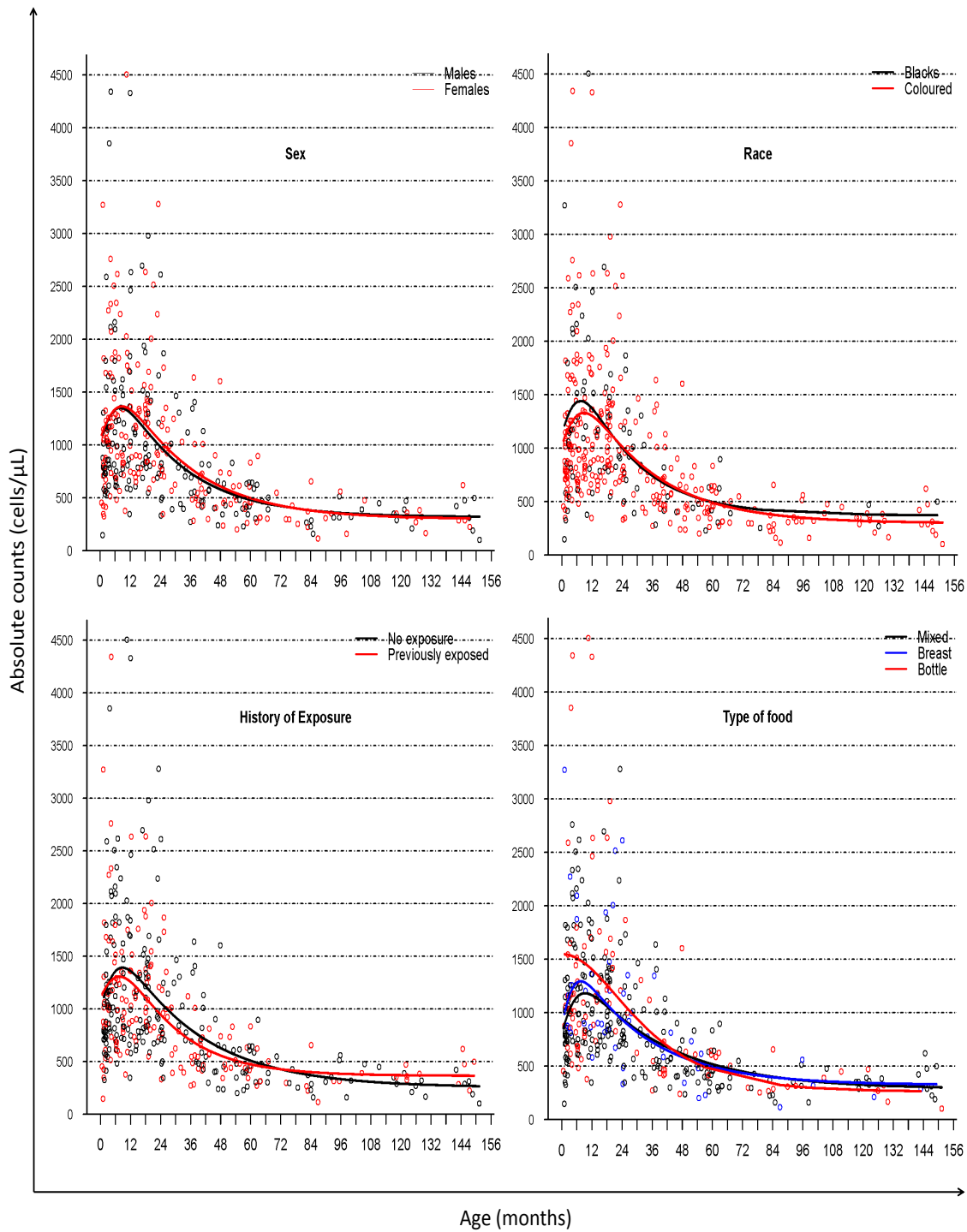


Figure 4.4: A plot showing age-related changes in CD19+ across different covariates.

Table 4.5: Parameter estimates for the extended model for CD19+ B lymphocytes.

	Estimate	95% Confidence Interval	p-value
β_1	1832.428	[1301.979, 2362.877]	0.000
β_3	0.198	[-0.010, 0.406]	0.061
β_0			
β_0 .(Intercept)	359.313	[190.086, 528.540]	0.000
γ_0 .(Sex - Male)	16.675	[-72.716, 106.066]	0.714
α_0 .(Race - Coloured)	-93.485	[-240.543, 53.572]	0.212
τ_0 .(Exposure - Exposed)	114.128	[22.949, 205.306]	0.014
ϕ_0 .(Feeding - Bottle)	-151.953	[-296.598, -7.309]	0.040
ϕ_0 .(Feeding - Mixed)	-94.961	[-200.481, 10.559]	0.078
β_2			
β_2 .(Intercept)	0.044	[0.025, 0.063]	0.000
γ_2 .(Sex - Male)	0.003	[-0.006, 0.011]	0.516
α_2 .(Race - Coloured)	-0.008	[-0.024, 0.007]	0.285
τ_2 .(Exposure - Exposed)	0.014	[0.003, 0.025]	0.010
ϕ_2 .(Feeding - Bottle)	-0.015	[-0.026, -0.004]	0.011
ϕ_2 .(Feeding - Mixed)	-0.012	[-0.022, -0.001]	0.027

*The reference groups are: females for sex; blacks for race; non-exposed for history of exposure to illness and breast-feeding for type of feeding during 6 months following birth.

Type of feeding and history of exposure to illness parameter estimates were statistically significantly different for CD19+. Bottle-fed children had lower CD19+ at birth ($p=0.040$) compared to those on mixed meals or breast-milk. There was no significant difference at birth between those on mixed meals and breast-milk. Cell decline was lower in children on bottle meals ($p=0.011$) and mixed meals ($p=0.027$) compared to those fed on breast-milk. Children who had a history of prior exposure to illness had higher counts at birth ($p=0.014$) compared to those who had no history of exposure. The rate of cell decline was faster for exposed children ($p=0.010$) compared to those without a history of exposure. Sex and race had no apparent influence on the parameters.

4.3.2.5 Ratio of CD4+ naive/memory T cells

Figure 4.5 compares independently predicted counts of ratio of CD4+ naive/memory T cells for the investigated covariates.

1. **Sex:** Predicted counts for females were slightly higher than those for males at birth but slightly lower at older ages. Both males and females had a double exponential trend.
2. **History of exposure to illness:** Subjects who had no history of exposure to illness had slightly lower predicted cell counts compared to those with previous exposure in the first 6 months from birth, but this rapidly increased above the previously exposed group. At older ages, the predicted counts for the previously exposed subjects were higher.
3. **Race:** Black children had higher predicted counts compared to Mixed race/Coloureds in the first year of life, after which their predicted cell counts descended to lower values than those of the mixed race children.
4. **Type of food:** Children fed on mixed meals and breast-milk had similar trend of predicted counts. Bottle-fed children had lower predicted counts than those of the above two feeding groups at all ages.

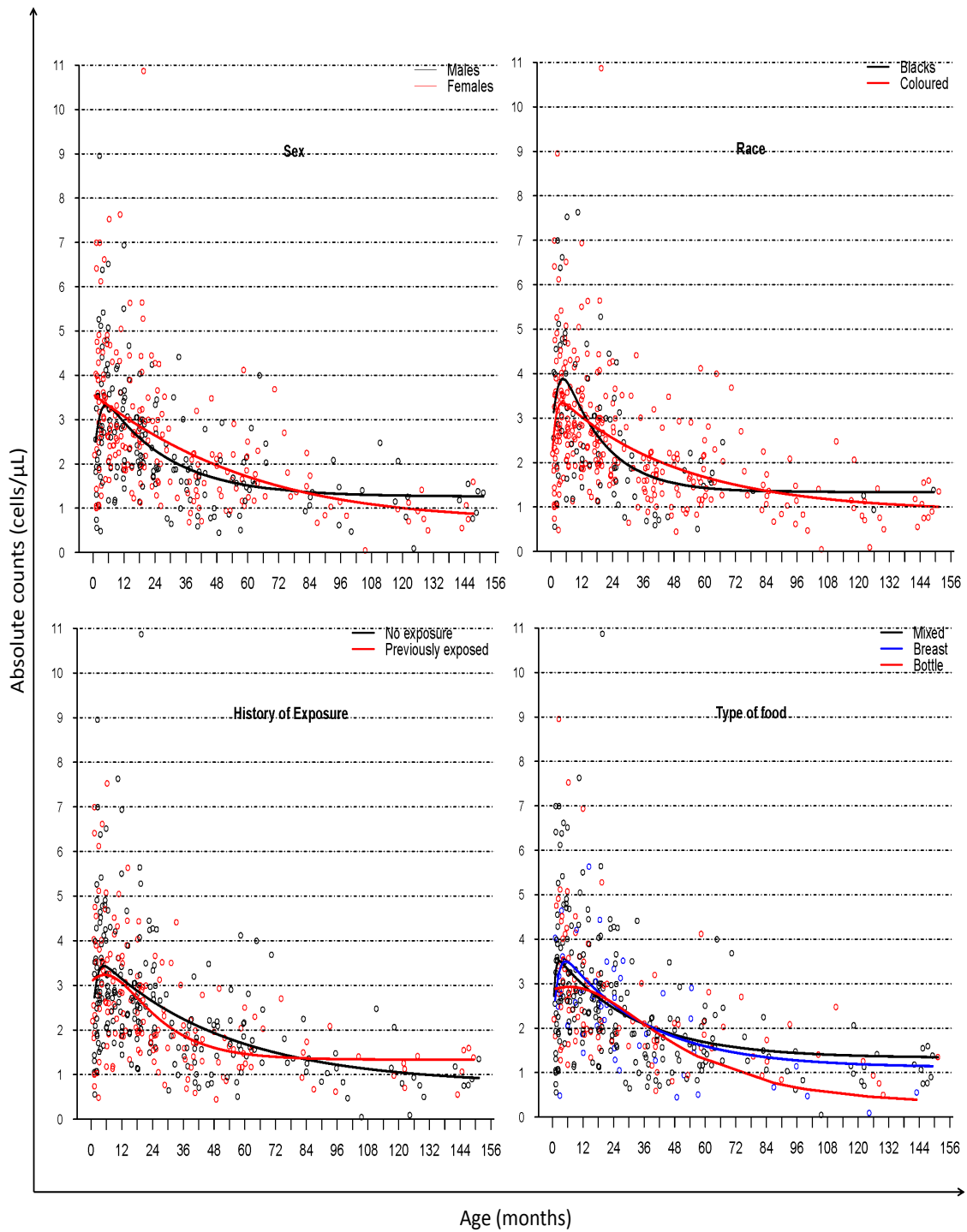


Figure 4.5: plot showing age-related changes in Ratio CD4+ naive/memory T cells across the different covariates.

Chapter 4. The influence of covariate factors on Age-related changes in lymphocyte cell markers 62

Table 4.6: Parameter estimates for the extended model for Age-related changes in the ratio of CD4+ naive/memory T cells.

	Estimate	95% Confidence Interval	p-value
β_1	3.951	[3.131, 4.771]	0.000
β_3	0.287	[0.081, 0.494]	0.006
β_0			
β_0 .(Intercept)	1.240	[0.747, 1.732]	0.000
γ_0 .(Sex - Male)	0.028	[-0.270, 0.326]	0.852
α_0 .(Race - Coloured)	-0.047	[-0.432, 0.339]	0.811
τ_0 .(Exposure - Exposed)	0.013	[-0.264, 0.291]	0.924
ϕ_0 .(Feeding - Bottle)	-1.413	[-2.006, -0.820]	0.000
ϕ_0 .(Feeding - Mixed)	0.385	[-0.012, 0.782]	0.057
β_2			
β_2 .(Intercept)	0.044	[0.025, 0.064]	0.000
γ_2 .(Sex - Male)	0.011	[0.001, 0.021]	0.046
α_2 .(Race - Coloured)	-0.006	[-0.019, 0.006]	0.302
τ_2 .(Exposure - Exposed)	0.003	[-0.004, 0.010]	0.362
ϕ_2 .(Feeding - Bottle)	-0.029	[-0.044, -0.013]	0.000
ϕ_2 .(Feeding - Mixed)	0.029	[-0.006, 0.064]	0.102

*The reference groups are: females for sex; blacks for race; non-exposed for history of exposure to illness and breast-feeding for type of feed during 6 months.

The dynamics in the ratio of CD4+ naive/memory T cells was not statistically different in terms race ($p=0.811$) and history of prior exposure to illness ($p=0.924$). Bottle-fed children had lower cell counts at birth ($p=0.000$) and a lower rate of cell decline with advancing age ($p=0.000$) as compared to the other groups. Breast-fed and mixed meal fed children were not statistically different. Although the number of cell counts at birth for female children was not statistically different from that of males ($p=0.852$), the rate of cell decline with advancing age was lower in females compared to males ($p=0.046$).

4.3.2.6 CD16+CD56+ T cells

Figure 4.6 compares independently predicted counts of ratio of CD16+CD56+ for each of the investigated covariates.

1. **Sex:** At birth, female children had slightly higher predicted cell counts than males. Thereafter, females and males had similar trends. Male children had a double exponential trend while females had a general exponential decline after birth.
2. **History of exposure to illness:** Children who had a history of exposure to maternal illness had slightly higher counts than those who were without a history of exposure to illness at all ages.
3. **Race:** Mixed race/Coloured children had higher predicted cell counts than blacks in early life. However, predicted cell counts for black children rise from birth to 1 year, followed by an exponential decline, while mixed race children decline exponentially from birth.
4. **Type of food:** There was no observable difference in the trend of predicted count for children fed on bottle versus mixed food. Predicted cell counts were higher for breast-fed children versus those on bottle and mixed feeds at all ages.

Chapter 4. The influence of covariate factors on Age-related changes in lymphocyte cell markers

64

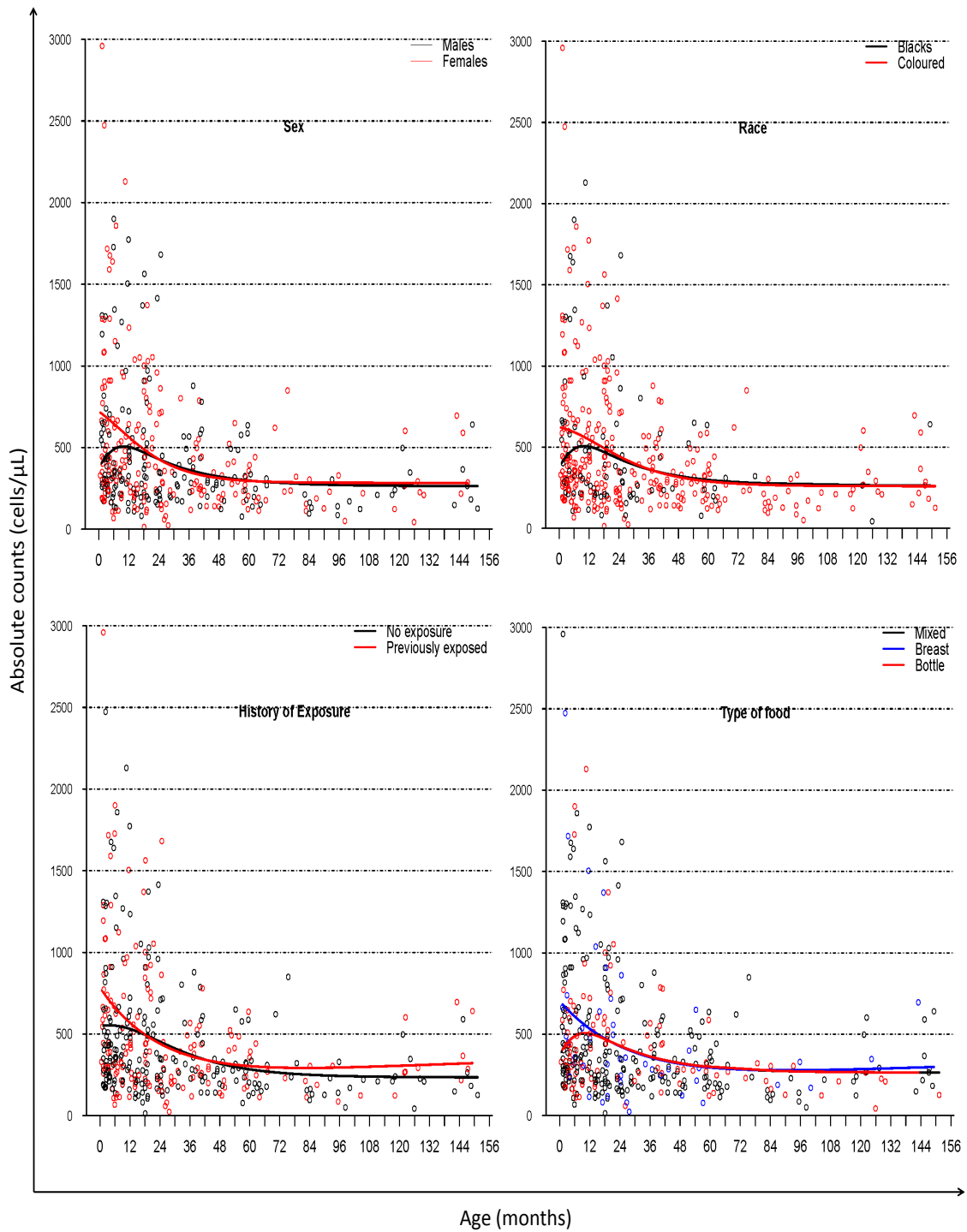


Figure 4.6: A plot showing age-related changes in CD16+CD56+ across the covariates.

Table 4.7: Parameter estimates for the extended model for Age-related changes in CD16+CD56+ T cells.

	Estimate	95% Confidence Interval	p-value
β_1	546.951	[307.274, 786.628]	0.000
β_3	1.561	[-1.861, 4.983]	0.370
β_0			
β_0 .(Intercept)	374.918	[273.012, 476.824]	0.000
γ_0 .(Sex - Male)	-48.268	[-121.463, 24.926]	0.196
α_0 .(Race - Coloured)	-132.074	[-230.296, -33.852]	0.009
τ_0 .(Exposure - Exposed)	79.296	[3.949, 154.644]	0.039
ϕ_0 .(Feeding - Bottle)	29.459	[-90.077, 148.995]	0.628
ϕ_0 .(Feeding - Mixed)	136.156	[51.010, 221.302]	0.002
β_2			
β_2 .(Intercept)	0.202	[-0.037, 0.440]	0.097
τ_2 .(Sex - Male)	-0.021	[-0.056, 0.014]	0.237
ϕ_2 .(Race - Coloured)	-0.155	[-0.386, 0.076]	0.187
ϕ_2 .(Exposure - Exposed)	0.030	[-0.019, 0.079]	0.231
γ_2 .(Feeding - Bottle)	0.012	[-0.045, 0.069]	0.671
α_2 .(Feeding - Mixed)	1.128	[-1.500, 3.757]	0.399

*The reference groups are: females for sex; blacks for race; non-exposed for history of exposure to illness and breast-feeding for type of feed during 6 months.

Females and males were not statistically different ($p = 0.196$). Children on mixed meals had higher cell counts at birth ($p=0.002$) compared to the breast-fed and bottle-fed, which were not statistically different. Significant differences in cell counts at birth were visible in race and exposure to illness. Mixed race/coloured children had low cell counts at birth ($p = 0.009$) compared to black children. Children who were exposed to maternal illness had higher cell counts at birth ($p = 0.039$) compared to those not exposed. However, no statistically significant differences were visible in the rate of decline in cell counts with advancing age, with regard to race and history of exposure to illness.

4.3.2.7 CD3–CD56+ T cells

Figure 4.6 compares the ratio of CD3-CD56+ T cells for the investigated covariates.

1. **Sex:** Predicted cell counts for females and males were not different after the first year from birth. Within the first year from birth, females had higher predicted cell counts than males. In addition, the predicted cell count for females exhibits a general exponential decline trend immediately from birth, while that of males shows a double exponential trend.
2. **History of exposure to illness:** There is no observable differences in the trend of predicted cell counts for children who had an history of exposure to maternal illness and those who had no history of exposure to illness both at birth and older ages.
3. **Race:** Coloured children had higher predicted cell counts than black ones at early stages of life. However, predicted cell counts for black children seem to rise from birth up to about 1 year, followed by an exponential decline to older ages while that of coloured children starts to decline exponentially from birth.
4. **Type of food:** There is no observable differences in the trend of predicted cell counts for the children fed on bottle and mixed feeds. However, at birth and older ages, the predicted cell counts are higher for children fed on breast milk than those fed on either bottle and mixed feeds.

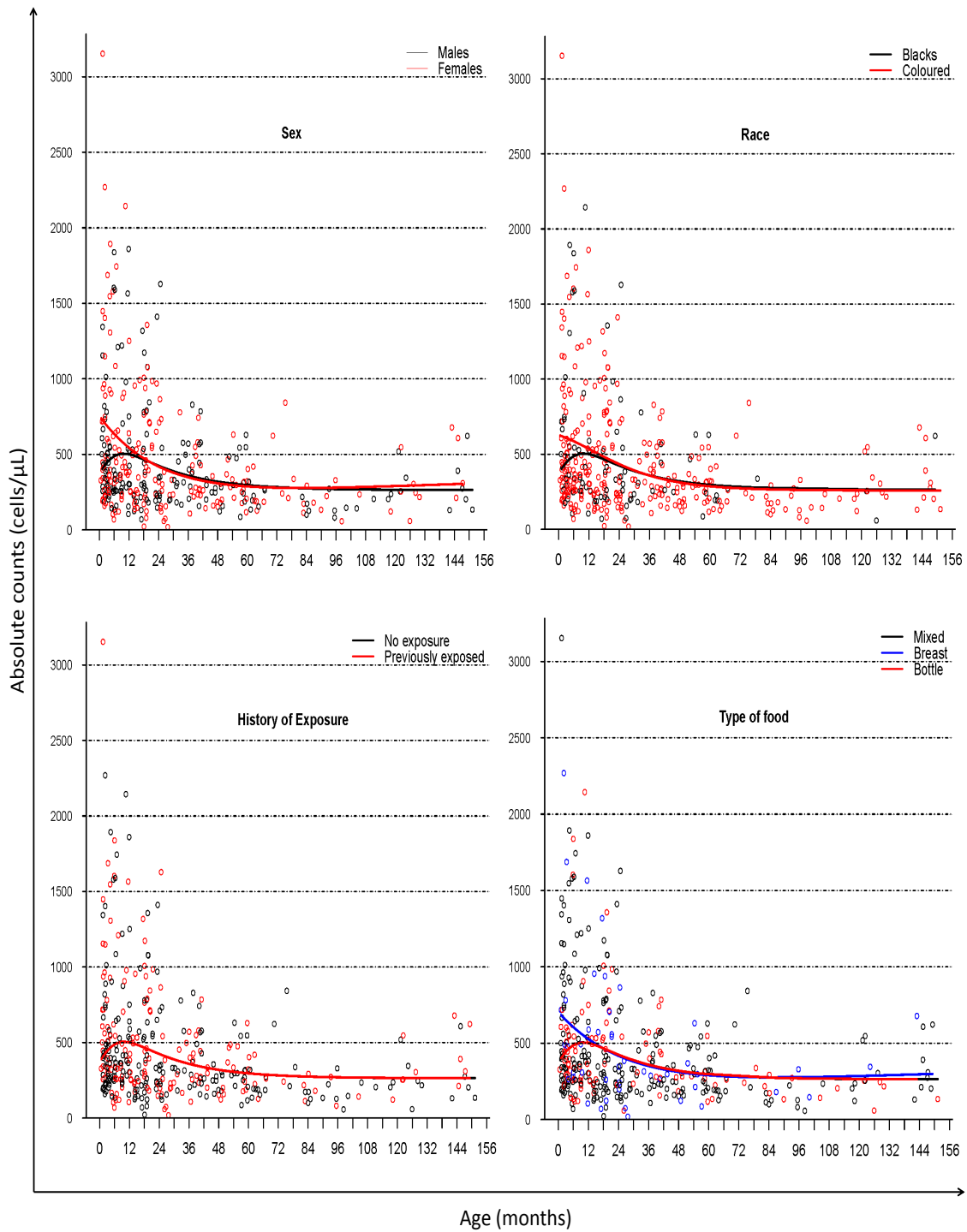


Figure 4.7: A plot showing age-related changes in CD3-CD56+ across the covariates.

Chapter 4. The influence of covariate factors on Age-related changes in lymphocyte cell markers 68

Table 4.8: Parameter estimates for the extended model for Age-related changes in CD3–CD56+ T cells.

	Estimate	95% Confidence Interval	p-value
β_1	1073.496	[821.155, 1325.837]	0.000
β_3	0.009	[-0.013, 0.030]	0.439
β_0			
β_0 .(Intercept)	446.406	[105.253, 787.559]	0.010
γ_0 .(Sex-Male)	-31.054	[-136.882, 74.775]	0.564
α_0 .(Race-Coloured)	-95.874	[-290.628, 98.881]	0.334
τ_0 .(Exposure-Exposed)	61.881	[-54.097, 177.858]	0.295
ϕ_0 .(Feeding-Bottle)	55.999	[-165.194, 277.192]	0.619
ϕ_0 .(Feeding-Mixed)	-402.551	[-776.150, -28.952]	0.035
β_2			
β_2 .(Intercept)	0.045	[0.023, 0.068]	0.000
γ_2 .(Sex-Male)	-0.001	[-0.006, 0.004]	0.731
α_2 .(Race-Coloured)	-0.005	[-0.019, 0.010]	0.509
τ_2 .(Exposure-Exposed)	0.001	[-0.005, 0.008]	0.683
ϕ_2 .(Feeding-Bottle)	0.013	[-0.038, 0.064]	0.610
ϕ_2 .(Feeding-Mixed)	-0.030	[-0.054, -0.007]	0.012

*The reference groups are: females for sex; blacks for race; non-exposed for history of exposure to illness and breast-feeding for type of feed during 6 months.

CD3-CD56+ T cells dynamics were not statistically different with regard to sex, race and history of exposure to illness. Children fed on mixed meal had lower counts ($p = 0.035$) and lower rates of cell decline ($p = 0.012$) with advancing age compared to those fed on the other two types of food.

4.4 Summary

4.4.0.1 Influence of sex

Separate empirical fits for males and females demonstrated measurable differences in CD3+, CD4+, ratio of CD4+ naive/memory, CD16+CD56+, CD3-CD56+ T cells. There have been similar findings in prior studies [20, 39, 92]. However, in the extended analysis, statistically significant sex differences were only observed in the ratio of CD4+ naive/memory T cells. A study by Rudy et al. [92] observed sex differences in CD4+ and CD3-CD56+ T cells.

4.4.0.2 Influence of race

Our extended covariate regression analysis did not demonstrate any statistically significant racial differences for the majority of the investigated lymphocytes subsets, CD3+, CD4+, CD8+, CD19+, ratio of CD4+ naive/memory, and CD3-CD56+ T cells. Similar findings have been found in other regions [37, 92, 97, 107]. Our study only demonstrated statistically significant racial differences in CD16+CD56+ T cells. Other studies have shown evidence of ethnic immunological variations, mostly between blacks and white HIV positive individuals [5, 35, 60].

4.4.0.3 Influence of history of exposure to illness

This study demonstrated an influence of exposure to illness in the CD19+ and CD16+CD56+ lymphocyte subsets. A prior study on HIV-positive gay men showed similar influences on cell counts and proliferative responses in CD4+, CD8+ and CD56+ [55].

4.4.0.4 Food in the first 6 months following birth

Our results on separate empirical fits for the breast-fed, bottle-fed and mixed meal-fed consistently indicated differences in age-related changes for lymphocyte subsets. Further, our extended regression also demonstrated differences for CD8+, CD19+, ratio of CD4+ naive/memory, CD16+CD56+ and CD3-CD56+ T cells. Prior studies have reported similar results [6, 62].

Chapter 5

An Age-continuous Model-based Biomarker Reference Range estimation method

5.1 Introduction

Prior studies of age-related reference ranges in industrialized countries have variously proposed the use of centile curves, or growth-charts, obtained by applying Box-Cox transformations to reference data [22, 23], or by empirically grouping particular age ranges into non-continuous ‘age-blocks’ [24, 26, 50, 67, 69, 71, 93, 105]. However, studies of thymic maturation have shown that immunological biomarkers change continuously from birth towards adulthood [7, 29, 50]. For this reason, ‘age-block’ methods have the disadvantage of a loss of precision in indicating whether patients that lie close to the limits of particular ranges are normal or not. Ideally, age-related changes should be described as a continuous process. Furthermore, although centile curves are continuous they provide no opportunity for *mechanistic* biological inferences regarding observed changes over time. In the present study we review prior methods for calculating age-related reference ranges (see Section 2.3) and then present a model-based, age-continuous statistical method without the above disadvantages.

5.2 Methods

Normal reference ranges are those sets of values falling within the inner 95% for a laboratory test measured on healthy subjects of a population. Using a Gaussian normality assumption, reference ranges are values measured in the 'healthy' group with 2 standard deviations on either side of the mean. This accounts for 95% of the population.

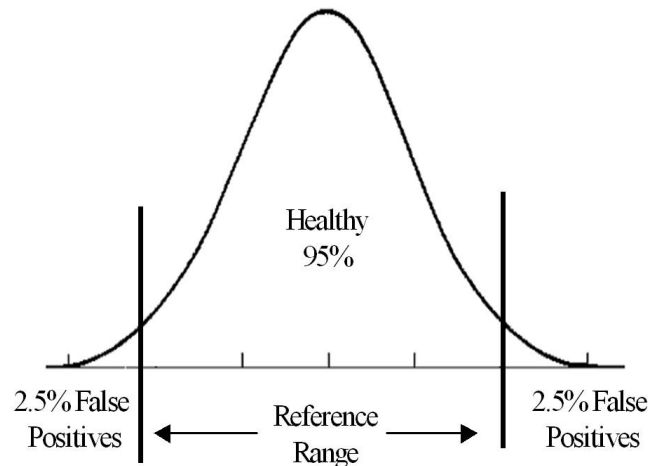


Figure 5.1: An illustration of how deviation from the mean accounts for different percentages of the sample. For the normal distribution, the values less than one standard deviation away from the mean account for 68.27% of the set; while two standard deviations from the mean account for 95.45%; and three standard deviations account for 99.73%. The values within two standard deviations are the normal reference range.

Definition 5.2.1. Model-based Reference Range

An $\alpha\%$ reference range is the set of values of an observed variable Y , comprising $\alpha\%$ of the data distributed about a given measure of central tendency, either the mean or median function/s of the data in terms of a probability scale. An $\alpha\%$ model based reference range can, therefore, be defined as the set of values bounded by upper and lower fitted curves corresponding to the upper and lower data limits for the observed Y .

Assumptions:

1. Reference ranges are a particular Z_α -shift of the standard deviations of residuals, on a continuous time scale, away from the fitted central function. A Z_α is the Z-score corresponding to $\alpha\%$ significance level. This level is adjustable to any desired value.

2. The positive and negative, Z_α -shifted residuals are considered independent and are modelled separately.
3. Since the continuous reference ranges curves are fitted independently on either side of the central fitted function, each intrinsically incorporates the distributional characteristics of the data above or below the fit of the central model-function.
4. Both upper and lower Z_α -shifted curves are modelled with their own parameter values, rather than by the parameters of the data as a whole.
5. Although the fitted Z_α -shifted curves are independent, they employ the same class of function as the central curve, with the same number of parameters, although these differ numerically. For example, if the central function is considered a 4-parameter double exponential, then so too is the particular Z_α shifted curve, above or below it.

Proposition 5.2.2. *The upper and lower data thresholds at, for example, a 95% reference range of a random variable $Y(t)$, are denoted by $U(t)$ and $L(t)$ respectively. Let $f(t, \boldsymbol{\beta})$ be the fitted values of $Y(t)$ and z the Z-score corresponding to $\alpha\%$, then*

$$U(t) = f(t, \boldsymbol{\beta}) + z(sd^+(t)) \quad (5.2.1)$$

$$(5.2.2)$$

and

$$L(t) = f(t, \boldsymbol{\beta}) - z(sd^-(t)), \quad (5.2.3)$$

$$(5.2.4)$$

where t is the age in months, $sd^+(t)$ and $sd^-(t)$ are vectors, or 'running' standard deviation for the n samples of positive and negative residuals respectively.

Let ψ denote the probability that X lies within the interval $[\mathcal{U}, \mathcal{L}]$, then

$$\begin{aligned} Pr(\mathcal{L} < X < \mathcal{U}) &= \psi \\ Pr\left(\frac{\mathcal{L} - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{\mathcal{U} - \mu}{\sigma}\right) &= \psi. \end{aligned} \quad (5.2.5)$$

For a normally distributed X , the Z-score is given by $Z = \frac{X - \mu}{\sigma}$, thus equation (5.2.5) becomes

$$Pr\left(\frac{\mathcal{L} - \mu}{\sigma} < Z < \frac{\mathcal{U} - \mu}{\sigma}\right) = \psi.$$

Thus

$$\frac{\mathcal{L} - \mu}{\sigma} = -z \quad \text{and} \quad \frac{\mathcal{U} - \mu}{\sigma} = z$$

or

$$\mathcal{U} = \mu + z\sigma \quad (5.2.6)$$

and

$$\mathcal{L} = \mu - z\sigma. \quad (5.2.7)$$

Which illustrates 5.2.1 and 5.2.3.

The main idea is that $U(t)$ and $L(t)$ are two datasets that can be modelled independently to obtain the parameter estimates for the upper and lower fitted values. The model-based reference ranges are then the sets of values falling within the upper and lower fitted values.

The calculation of $U(t)$ and $L(t)$ in 5.2.1 and 5.2.1, respectively, involves a ‘running’ partitioning of residuals with an arbitrary size k , where k refers to nearest neighbours. Any value of k is possible, but we empirically determined that values of 4 or 5 were optimal.

Suppose we have r_1, r_2, \dots, r_n residuals defined as

$$r_i(t) = y_i(t) - f(t_i, \beta), \quad i = 1, 2, 3, \dots, n,$$

where $y_i(t)$ are the observed values and $f(t_i, \beta)$ are the fitted values. Positive residuals are $r_i^+(t)$ if $r_i(t) > 0$ while the negative residuals are $r_i^-(t)$ if $r_i(t) < 0$. The k -nearest neighbours ‘running’ standard deviations of the positive and negative residuals are defined as in 5.2.8 and 5.2.9, respectively,

$$sd^+(t) = \left(\sqrt{\frac{1}{k} \sum_{i=1}^k (r_i^+(t))^2}, \sqrt{\frac{1}{k} \sum_{i=2}^{k+1} (r_i^+(t))^2}, \sqrt{\frac{1}{k} \sum_{i=3}^{k+2} (r_i^+(t))^2}, \dots, \sqrt{\frac{1}{k} \sum_{i=n-k}^n (r_i^+(t))^2} \right) \quad (5.2.8)$$

and

$$sd^-(t) = \left(\sqrt{\frac{1}{k} \sum_{i=1}^k (r_i^-(t))^2}, \sqrt{\frac{1}{k} \sum_{i=2}^{k+1} (r_i^-(t))^2}, \sqrt{\frac{1}{k} \sum_{i=3}^{k+2} (r_i^-(t))^2}, \dots, \sqrt{\frac{1}{k} \sum_{i=n-k}^n (r_i^-(t))^2} \right). \quad (5.2.9)$$

At 95% reference ranges $Z_{\alpha=95\%} \simeq 2$ and thus Equations 5.2.1 and 5.2.1 becomes

$$U(t) = f(t, \beta) + 2(sd^+(t)) \quad (5.2.10)$$

and

$$L(t) = f(t, \boldsymbol{\beta}) - 2(sd^-(t)), \quad (5.2.11)$$

The terms $2(sd^+)$ and $2(sd^-)$ are the Z_α -shifts of the standard deviation of residuals, away from the fitted values, $f(t, \boldsymbol{\beta})$. The positive and negative residuals are not symmetrical about the fitted central function. $U(t)$ and $L(t)$ datasets are modelled independently to obtain the best fits at the two reference range limits and implicitly automatically adjusts for skewness, kurtosis and any other distributional characteristics in the population as a whole.

Summary of procedure:

1. Sort the data by ascending age, i.e. from younger to older.
2. Choose the value of k . The choice of an optimal value for k is user defined and ideally requires an initial inspection of the data. We found that for small samples ($n \leq 100$) the central value along with its 3 or 4 nearest neighbours ($k = 4$ or $k = 5$) and for larger samples, $4 < k \leq 10$. Ideally, the 'window' (the value of k) should be as small as possible.
3. Calculate the fitted values, $f(t_i, \boldsymbol{\beta})$, of the central regression (function) model.
4. Calculate the residuals, $r_i(t) = y_i(t) - f(t_i, \boldsymbol{\beta})$.
5. Determine positive residuals, denoted by $r_i^+(t)$, if $r_i(t) > 0$ and negative residuals, denoted by $r_i^-(t)$, if $r_i(t) < 0$.
6. Independently, calculate the k-nearest neighbours 'running' standard deviations, $sd^+(t)$ and $sd^-(t)$.
7. Calculate the Z-score corresponding to the desired $\alpha\%$ reference ranges.
8. Calculate $U(t)$ and $L(t)$ datasets using Equations 5.2.10 and 5.2.11.
9. Fit the appropriate model to the result in (8). This generates both the parameter values at the respective upper or lower data thresholds and gives the desired age-related reference ranges.

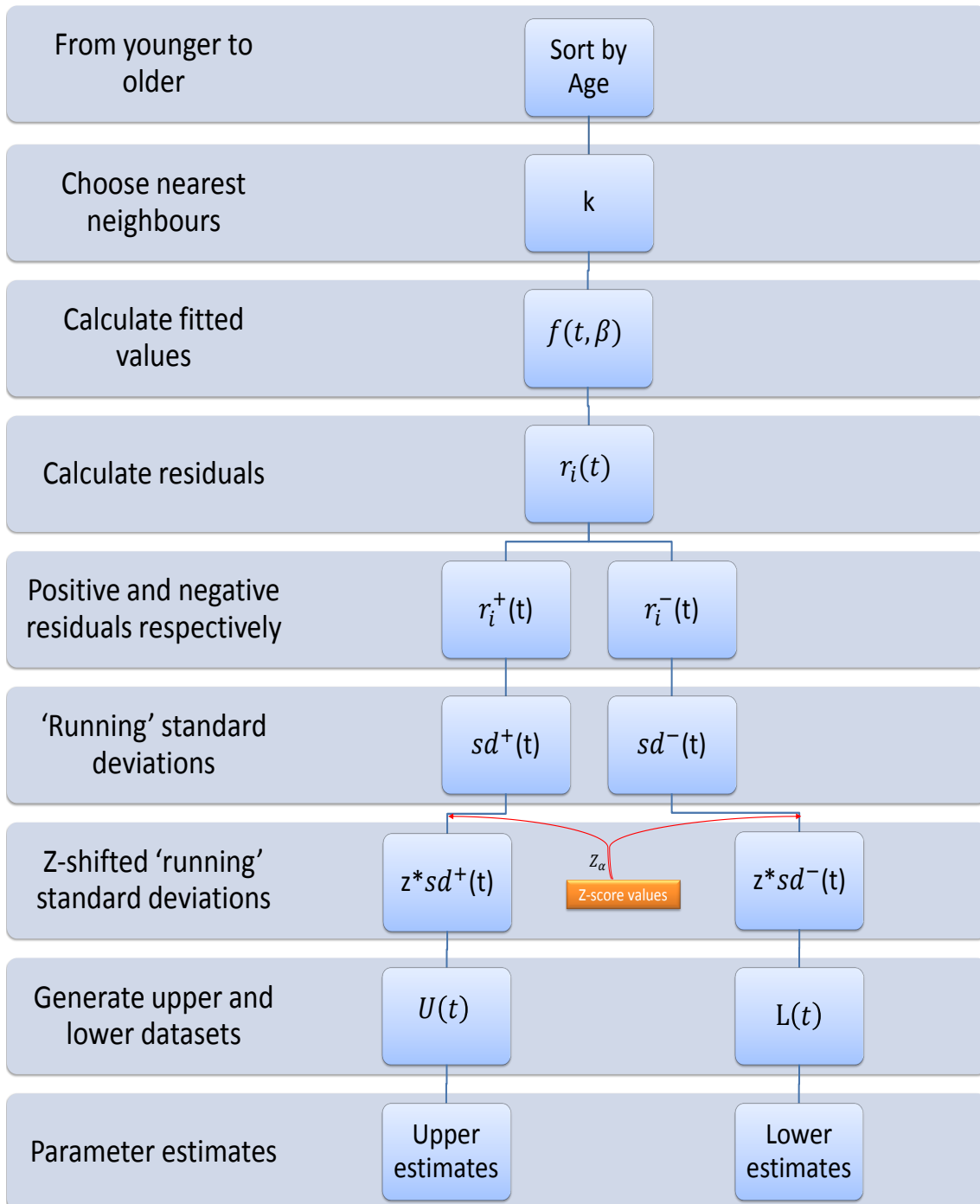


Figure 5.2: Flowchart of the model-based estimation procedure.

Chapter 5. An Age-continuous Model-based Biomarker Reference Range estimation method

76

In the present case, we fit the double exponential model (equation 3.3.2) on the seven lymphocyte subsets and parameter values obtained using the weighted generalized nonlinear least squares estimation method as is available in package nlme in the R statistical environment. Thereafter, the model is re-used to estimate the new set of parameters for $U(t)$ and $L(t)$, thus providing the model-fits at the upper and lower 95% thresholds respectively. We also compare our results to those obtained in centile curves for some of the lymphocyte cell markers.

5.3 Results

Table 5.1 summarizes the parameter estimates and significance for the central (mean), lower and upper fits for a 95% model-based reference ranges. An age-continuous reference ranges prediction can be obtained by inputting the parameter estimates back into the model. The statistical significance level for the parameters is set at $p < 0.05$.

Chapter 5. An Age-continuous Model-based Biomarker Reference Range estimation method

78

Table 5.1: Parameter estimates for the central (mean), lower and upper fits for a 95% model-based reference ranges.

Cell Markers	Parameters	Central Estimates			Lower Estimates			Upper Estimates		
		Estimate	95% CI	p-value	Estimate	95% CI	p-value	Estimate	95% CI	p-value
CD3+	β_0	1838.114	[1674.603, 2001.625]	0.000	1154.236	[1084.804, 1223.668]	0.000	2660.391	[2607.060, 2713.723]	0.000
	β_1	6390.389	[5240.055, 7540.723]	0.000	3413.560	[3002.210, 3824.910]	0.000	10266.219	[9295.176, 11237.262]	0.000
	β_2	0.047	[0.040, 0.055]	0.000	0.041	[0.036, 0.045]	0.000	0.056	[0.053, 0.059]	0.000
	β_3	0.108	[0.045, 0.171]	0.001	0.107	[0.068, 0.147]	0.000	0.126	[0.090, 0.162]	0.000
CD4+	β_0	1127.695	[1008.358, 1247.032]	0.000	726.011	[686.827, 765.195]	0.000	1602.427	[1547.103, 1657.751]	0.000
	β_1	4988.450	[4260.474, 5716.426]	0.000	2446.855	[2135.886, 2757.825]	0.000	8210.455	[7752.771, 8668.140]	0.000
	β_2	0.050	[0.035, 0.064]	0.000	0.050	[0.043, 0.057]	0.000	0.046	[0.031, 0.060]	0.000
	β_3	0.061	[0.006, 0.116]	0.030	0.084	[0.045, 0.124]	0.000	0.038	[0.001, 0.075]	0.045
CD8+	β_0	728.810	[643.107, 814.512]	0.000	347.664	[301.996, 393.331]	0.000	989.083	[949.876, 1028.290]	0.000
	β_1	1384.083	[809.443, 1958.723]	0.000	615.484	[490.708, 740.259]	0.000	3071.153	[2627.006, 3515.301]	0.000
	β_2	0.038	[0.023, 0.054]	0.000	0.026	[0.019, 0.033]	0.000	0.043	[0.038, 0.048]	0.000
	β_3	0.223	[-0.019, 0.464]	0.071	0.156	[0.076, 0.236]	0.000	0.226	[0.127, 0.326]	0.000
CD19+	β_0	308.142	[259.398, 356.885]	0.000	182.289	[165.183, 199.395]	0.000	493.475	[469.959, 516.992]	0.000
	β_1	1776.413	[1201.939, 2350.888]	0.000	668.427	[531.866, 804.989]	0.000	3496.154	[2944.952, 4047.356]	0.000
	β_2	0.038	[0.028, 0.048]	0.000	0.033	[0.027, 0.040]	0.000	0.043	[0.039, 0.048]	0.000
	β_3	0.207	[-0.029, 0.442]	0.085	0.218	[0.070, 0.366]	0.004	0.196	[0.084, 0.309]	0.001
Ratio CD4+ naive/memory	β_0	1.025	[0.737, 1.314]	0.000	0.220	[0.017, 0.423]	0.034	1.250	[1.001, 1.498]	0.000
	β_1	2.788	[2.332, 3.244]	0.000	1.640	[1.462, 1.818]	0.000	4.636	[4.369, 4.903]	0.000
	β_2	0.026	[0.015, 0.037]	0.000	0.017	[0.011, 0.024]	0.000	0.019	[0.016, 0.023]	0.000
	β_3	0.874	[0.200, 1.549]	0.011	0.462	[0.208, 0.715]	0.000	2.604	[-0.031, 5.239]	0.053
CD16+CD56+	β_0	265.371	[217.848, 312.894]	0.000	100.000	[82.428, 117.572]	0.000	500.921	[466.686, 535.157]	0.000
	β_1	1012.223	[651.045, 1373.402]	0.000	100.000	[50.934, 149.066]	0.000	2490.847	[2146.922, 2834.772]	0.000
	β_2	0.057	[0.023, 0.092]	0.001	0.044	[-0.001, 0.089]	0.054	0.070	[0.059, 0.081]	0.000
	β_3	0.074	[-0.065, 0.213]	0.295	0.100	[-0.026, 0.226]	0.121	0.102	[0.038, 0.165]	0.002
CD3-CD56+	β_0	335.200	[-1948.868, 2619.268]	0.773	106.696	[77.015, 136.376]	0.000	490.553	[457.229, 523.878]	0.000
	β_1	1031.178	[-1055.107, 3117.462]	0.332	112.495	[59.817, 165.174]	0.000	2518.308	[2157.418, 2879.198]	0.000
	β_2	0.036	[-0.012, 0.084]	0.138	0.033	[-0.021, 0.086]	0.229	0.071	[0.060, 0.083]	0.000
	β_3	0.010	[-0.230, 0.250]	0.936	0.058	[-0.033, 0.149]	0.212	0.099	[0.032, 0.166]	0.004

Figures 5.3 and 5.4 show the age-continuous fits for model-based 95% reference ranges. Each plot is comprised of central, lower and upper fitted curves. The age-continuous reference ranges are the points falling within the two red lines. In all the cell markers described below, the lower and upper limits of the reference ranges have the same mechanistic properties assumed by the central fit, as also described in the model formulation.

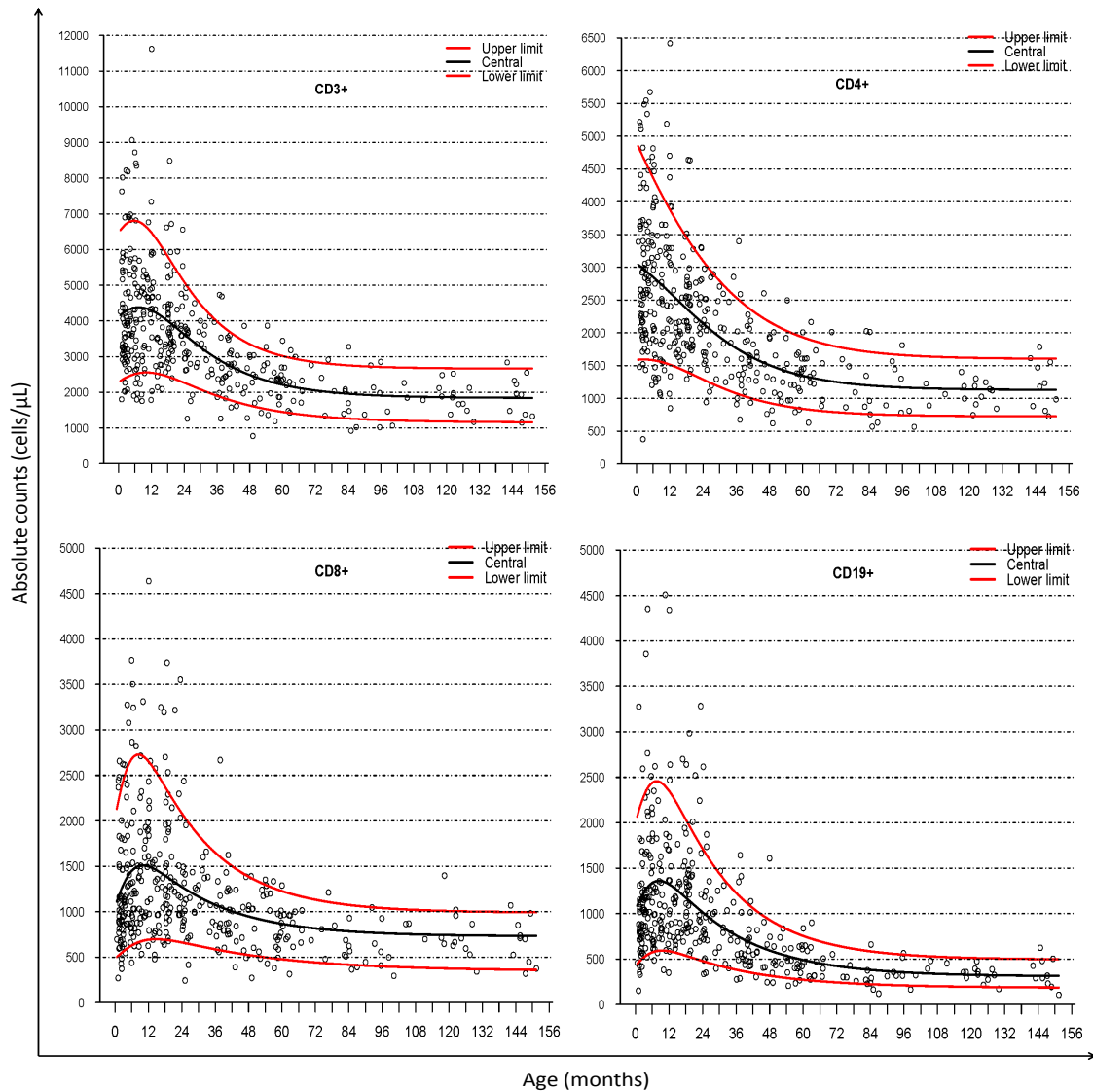


Figure 5.3: Plots showing model-based, age-continuous reference ranges for absolute counts of CD3+, CD4+, CD8+ and CD19+.

Chapter 5. An Age-continuous Model-based Biomarker Reference Range estimation method

80

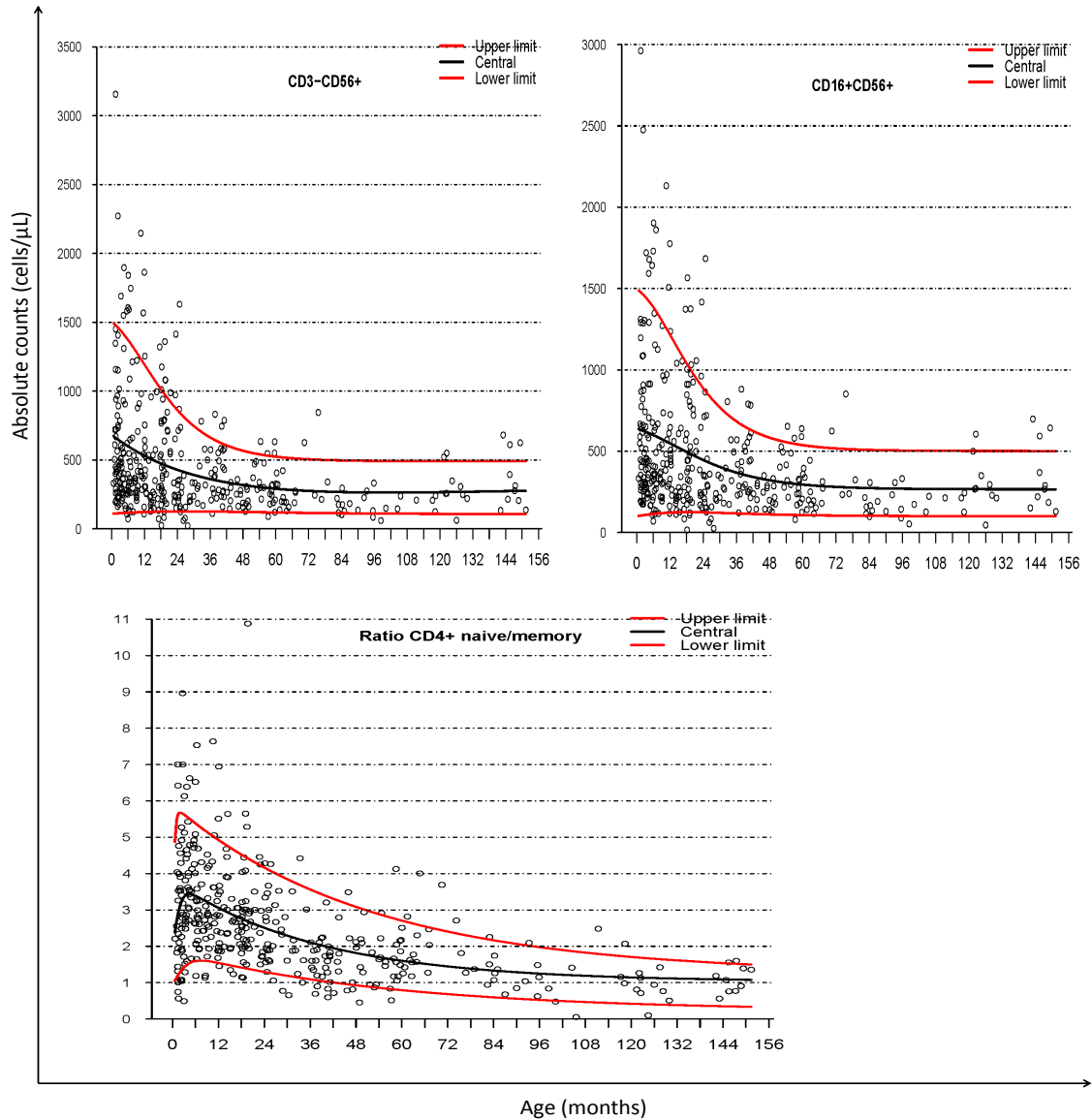


Figure 5.4: Plots showing model-based, age-continuous reference ranges for absolute counts of CD16+CD56+, CD3-CD56+ and ratio of CD4+ naive to memory T cells for healthy South African children aged 12.5 years and below. The fits assume similar mechanistic properties of the described model.

Figure 5.5 compares the fits obtained using centile curves with the model-based approach. The two methods provide somewhat similar central predictions, with the exception of the double exponential pattern that is noticeable at earlier ages using the model-based method. The methods predict a similar exponential decline at later ages.

Overall, centile curves give wider reference ranges compared to the model-based approach, i.e. given that the chosen Z_{α} -shift from the central function was chosen as the 95% range the latter method is less sensitive to residual 'noise' than the former. This was found to be uniformly true for all the biomarkers the method was applied to.

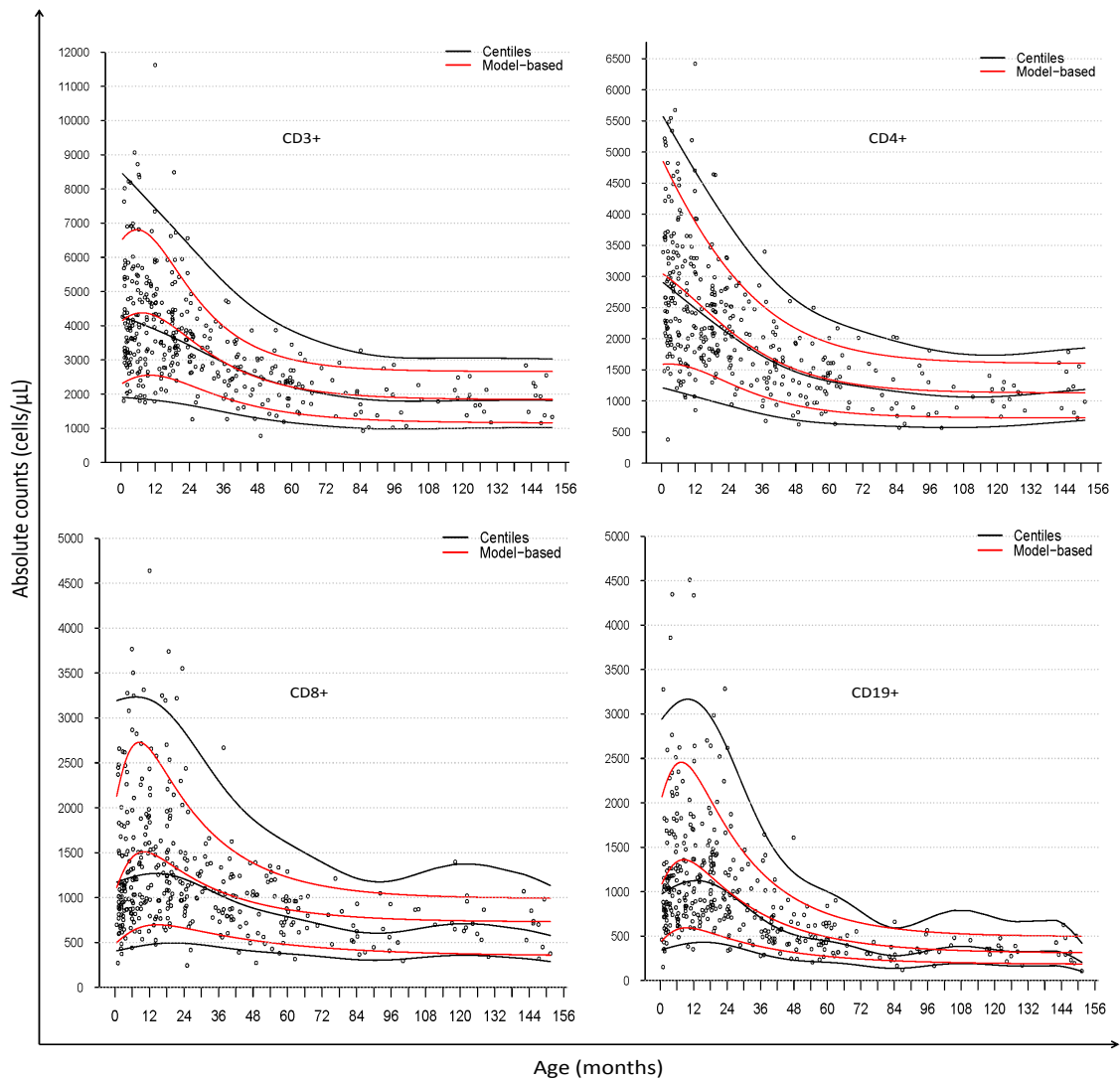


Figure 5.5: A comparison between centile curves (black lines) and model-based fits for absolute counts of CD3+, CD4+, CD8+ and CD19+. At early ages, centile curves (black lines) estimate wider, and simpler single-exponential-like, reference ranges than the double exponential model (red lines).

5.4 Summary

In this chapter we presented an alternative method for estimating age-related reference ranges based on *semi-mechanistic* models. Not all lymphocyte cell-markers follow a simple exponential decline over time, and a reference range estimation method which incorporate alternate models is consequently necessary. The model-based method allows for parameter estimation at the 95% limits which may be compared with other reference populations. Reference range estimation is also possible over age-continuous intervals. Centile curves remain appropriate in cases where the family of the underlying distribution of the smoothing spline is known. However, in the above examples this approach predicted only a simple exponential decline with age and gave generally wider reference ranges at early ages. This was the same as that proposed by a prior European study [48]. The reference ranges of our model-based method follows the same trend as that of the fitted central function, which is in agreement with prior empirical investigations and studies [7, 8, 24, 50, 71, 95, 111].

Chapter 6

Discussion and Conclusion

In this thesis we have reviewed the biological background of the study, mathematically described age-related changes in immunological biomarkers and described a method to construct age-continuous model-based reference ranges for healthy South African children. We compared the quality of fits for double and single exponential models, extended the double exponential model to incorporate the covariates and later described and implemented a model-based method for estimating reference ranges.

6.1 Discussion

In Chapter 3, our analysis demonstrated age-related changes in lymphocyte cell markers and established that the double exponential model was the best candidate model for describing such changes. As indicated above the paediatric levels are indeed higher than in adults. This has been attributed to the lack of functional maturity of the immune system, leading to it undergoing sequential development stimulated by both genetic factors and induced activation following exposure to new antigens [51]. The proliferation and maturation processes continue until adequate levels of immune surveillance have been reached [24]. Findings in highly ‘immune-exposed’ resource limited African settings have demonstrated that the levels of some markers may decrease in an exponential manner from a maximum at birth, [50, 69], in the same way as that proposed by a prior European study [48]. However, empirical measurements from our own data demonstrate that the counts of some of the immunophenotypes initially increase in the first year and then decrease thereafter which is in agreement with several prior studies [24, 33, 95].

The observed patterns in the SSA data may be due to poor diets with vitamin and iron deficiencies and exposure to more varied and more pathogenic antigens at younger ages. Iron deficiency may lead to impaired hematopoiesis, chronic blood loss may occur due to hookworm infestation and in SSA exposure to tuberculosis is more common. Such exposures would tend to drive up immunophenotype counts at younger ages. A limitation of the current approach was the assumption of an exponential decline in cell markers with age in the absence of a purely *mechanistic* model of the underlying biological processes involved. Our models were *semi-mechanistic*; however, their form agrees with that proposed in prior *mechanistic* studies [24, 33, 95].

In Chapter 4 we presented an extension of the double exponential model to investigate the effect of covariates, ie. sex, race, type of food within the first 6 months following birth and history of maternal exposure to illness. A large historical study has identified the need for covariate and region specific investigations of lymphocyte subsets. Each population should ideally have its own reference range, which might be updated as socio-demographic influences change [20]. The differences between males and females that we demonstrate might be due to differences in sex hormones [92], genetic and environmental factors and harboring of small numbers of cells that originated in a genetically different individual (microchimerism); and become more pronounced with the onset of puberty [13, 39, 40]. We observed very little effect of race in our study. However, the observed racial differences in previous studies might be due to the genetic variations across the racial groups [5, 35, 60]. We observed that changes in the majority of the cell markers were associated with the type of food given to the infants in the first 6 months following birth. Malnourishment may compromise immune system development. Breast milk provides passive immunotherapy through maternal immunoglobulin, and consequent gene expression, all of which prevents common infant infections [62, 106, 117]. Prior studies have demonstrated that bioactive factors in breast milk catalyze the differentiation and growth of B lymphocytes and initiate the production of particular antibodies [6, 62]. Our study did demonstrate that exposure to maternal illness influenced some of the lymphocyte cell markers investigated. However, the effect of exposure to such risk factors remains unclear [1, 26, 27, 61]. The interpretation of such findings is also complicated by the fact that the paediatric subjects in this study were specifically selected only if they were judged to be healthy.

In Chapter 5 we presented the implementation procedure and results of age-continuous, model-based reference ranges estimation method. Prior empirical studies have used dif-

fering age-block ranges [24, 95], which adds to the difficulty of comparing results across cohorts. The two studies which the South African National Health Laboratories were using as the de facto reference ranges were conducted in the EU and the US [24, 95]. For resource-limited, SSA circumstances, this is obviously not ideal, as the the environmental and immunological milieu of these populations are likely to differ. Prior empirical studies have also transformed data using, for the most part, logarithmic or Box-Cox transformations [30, 96, 116]. Back transformation is then required to obtain parameter estimates. Furthermore, centile curves are continuous but use smoothing splines, which do not allow for *mechanistic* interpretations of data patterns. Our approach eliminates the need for age-blocks, forward or back transformation, enables the estimation of age continuous reference ranges, and incorporates *mechanistic* model-based assumptions. Such models, when applied to paediatric immunological markers can estimate the rate of cell death, predict cell population numbers at birth, or at any given age at the thresholds of the 95% reference range. In addition, this method is relatively flexible and can be extended to incorporate covariates affecting the distribution of the reference population, e.g. sex, race, maternal smoking, maternal HIV status and child feeding habits.

A potential limitation of our method is the assumption that particular reference ranges are a Z_α -shifted 'running' standard deviations of residuals from the central fitted-function. However, all approaches to reference range estimation employ assumptions of some sort, including the fact that what is 'normal' incorporates a particular percentile range of the population under study. As the Z_α -shift is adjustable in our method, any particular range can be chosen based on an evaluation of the data in question.

6.2 Conclusion

Using 'age snap-shot' cross-sectional data we established that the continuous double exponential model was the ideal for predicting change in blood lymphocyte cell markers for the cohort of healthy South African children. This study has methodological value in both resource-limited and industrialized settings, in that it may lead to the development of an easy-to-use laboratory tool. This might take the form of a spreadsheet, into which a technician could enter the patient's age and particular cell-count and immediately determine whether the sample is within normal ranges or not. We have proposed an alternative method for constructing reference ranges, which has the flexibility to incorporate prior *mechanistic* assumptions, is age-continuous, and which also demonstrates agreement with empirical measurements and prior *mechanistic* studies. The method has

been implemented as R code and will be made available as an R software package.

Future studies might involve additional mathematical verification, such as evaluation of model covariates in data for patients that are unhealthy, i.e. not healthy subjects, sensitivity analysis for further evaluation of distributional assumptions and additional comparisons to prior reference range determination methods.

List of references

- [1] Elaine J Abrams, Jeffrey Wiener, Rosalind Carter, Louise Kuhn, Paul Palumbo, Stephen Nesheim, Francis Lee, Peter Vink, Marc Bulterys, et al. Maternal health factors and early pediatric antiretroviral therapy influence the rate of perinatal HIV-1 disease progression in children. *Aids*, 17(6):867–877, 2003.
- [2] Jeffrey K. Actor. Cells and Organs of the Immune System. <https://med.uth.edu/pathology/documents/2014/12/cells-and-organs-of-the-immune-system.pdf>, 2010. Date retrieved: 24 April 2015 13:28 UTC.
- [3] Rafi Ahmed, Michael J Bevan, Steven L Reiner, and Douglas T Fearon. The precursors of memory: models and controversies. *Nature Reviews Immunology*, 9(9):662–668, 2009.
- [4] Adhra Al-Mawali, Avinash Daniel Pinto, Raiya Al Busaidi, and Ibrahim Al-Zakwani. Lymphocyte subsets: Reference ranges in an age-and gender-balanced population of Omani healthy adults. *Cytometry Part A*, 83(8):739–744, 2013.
- [5] Kathryn Anastos, Stephen J Gange, Bryan Lau, Barbara Weiser, Roger Detels, Janis V Giorgi, Joseph B Margolick, Mardge Cohen, John Phair, Sandra Melnick, et al. Association of race and gender with HIV-1 RNA levels and immunologic progression. *Journal of acquired immune deficiency syndromes (1999)*, 24(3):218–226, 2000.
- [6] Nicholas J Andreas, Beate Kampmann, and Kirsty Mehring Le-Doare. Human breast milk: A review on its composition and bioactivity. *Early human development*, 91(11):629–635, 2015.
- [7] Iren Bains, Rustom Antia, Robin Callard, and Andrew J Yates. Quantifying the development of the peripheral naive CD4+ T-cell pool in humans. *Blood*, 113(22):5480–5487, 2009.

- [8] Iren Bains, Rodolphe Thiébaud, Andrew J Yates, and Robin Callard. Quantifying thymic export: combining models of naive T cell proliferation and TCR excision circle dynamics gives an explicit measure of thymic output. *The Journal of Immunology*, 183(7):4329–4336, 2009.
- [9] Irenjeet Kaur Bains. *Mathematical modelling of T cell homeostasis*. PhD thesis, UCL (University College London), 2010.
- [10] Charles RM Bangham. The immune control and cell-to-cell spread of human T-lymphotropic virus type 1. *Journal of General Virology*, 84(12):3177–3189, 2003.
- [11] Jacqueline A Bartlett, Steven J Schleifer, Melissa K Demetrikopoulos, Beverly R Delaney, Samuel C Shiflett, and Steven E Keller. Immune function in healthy adolescents. *Clinical and diagnostic laboratory immunology*, 5(1):105–113, 1998.
- [12] Andrew L Baughman, Kristine M Bisgard, Freyja Lynn, and Bruce D Meade. Mixture model analysis for establishing a diagnostic cut-off point for pertussis antibody levels. *Statistics in medicine*, 25(17):2994–3010, 2006.
- [13] Annechien Bouman, Martin Schipper, Maas Jan Heineman, and Marijke M Faas. Gender difference in the non-specific and specific immune response in humans. *American Journal of Reproductive Immunology*, 52(1):19–26, 2004.
- [14] Faas MM Bouman A, Heineman MJ. Sex hormones and the immune response in humans. pages 11:411–423, 2005.
- [15] Jeffrey H Burack, Donald C Barrett, Ron D Stall, Margaret A Chesney, Maria L Ekstrand, and Thomas J Coates. Depressive symptoms and CD4 lymphocyte decline among HIV-infected men. *Jama*, 270(21):2568–2573, 1993.
- [16] Giulia Casorati, Claudia de Lalla, and Paolo Dellabona. Invariant natural killer T cells reconstitution and the control of leukemia relapse in pediatric haploidentical hematopoietic stem cell transplantation. *OncoImmunology*, 1(3):355–357, 2012.
- [17] John T Chang, Vikram R Palanivel, Ichiko Kinjyo, Felix Schambach, Andrew M Intlekofer, Arnob Banerjee, Sarah A Longworth, Kristine E Vinup, Paul Mrass, Jane Oliaro, et al. Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *science*, 315(5819):1687–1691, 2007.
- [18] Zhi Chen, Arian Laurence, and John J O’Shea. Signal transduction pathways and transcriptional regulation in the control of Th17 differentiation. In *Seminars in immunology*, volume 19, pages 400–408. Elsevier, 2007.

- [19] Lyn S Chitty and Douglas G Altman. Charts of fetal size: limb bones. *BJOG: An International Journal of Obstetrics & Gynaecology*, 109(8):919–929, 2002.
- [20] Joungbum Choi, Su Jin Lee, Yun A Lee, Hyung Gun Maeng, Jong Kyun Lee, and Yong Won Kang. Reference Values for Peripheral Blood Lymphocyte Subsets in a Healthy Korean Population. *Immune network*, 14(6):289–295, 2014.
- [21] CLSI. *Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline*. Clinical and Laboratory Standards Institute, third edition edition, 2010.
- [22] Timothy J Cole. Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 385–418, 1988.
- [23] Timothy J Cole and Pamela J Green. Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in medicine*, 11(10):1305–1319, 1992.
- [24] W Marieke Comans-Bitter, Ronald de Groot, René van den Beemd, Herman J Neijens, Wim CJ Hop, Kees Groeneveld, Herbert Hooijkaas, and Jacques JM van Dongen. Immunophenotyping of blood lymphocytes in childhood. Reference values for lymphocyte subpopulations. *The Journal of pediatrics*, 130(3):388–393, 1997.
- [25] Wolney L Conde and Carlos A Monteiro. Body mass index cutoff points for evaluation of nutritional status in Brazilian children and adolescents. *Jornal de Pediatria*, 82(4):266–272, 2006.
- [26] Edward M Connor, Rhoda S Sperling, Richard Gelber, Pavel Kiselev, Gwendolyn Scott, Mary Jo O’Sullivan, Russell VanDyke, Mohammed Bey, William Shearer, Robert L Jacobson, et al. Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. *New England Journal of Medicine*, 331(18):1173–1180, 1994.
- [27] François Dabis, Philippe Msellati, Nicolas Meda, Christiane Welffens-Ekra, Bruno You, Olivier Manigart, Valérie Leroy, Arlette Simonon, Michel Cartoux, Patrice Combe, et al. 6-month efficacy, tolerance, and acceptability of a short regimen of oral zidovudine to reduce vertical transmission of HIV in breastfed children in Côte d’Ivoire and Burkina Faso: a double-blind placebo-controlled multicentre trial. *The Lancet*, 353(9155):786–792, 1999.
- [28] R J De Boer. Estimating the role of thymic output in HIV infection. *Current Opinion in HIV and AIDS*, 1(1):16–21, 2006.

- [29] R J De Boer and A S Perelson. Quantifying T Lymphocyte Turnover. *Journal of Theoretical Biology*, 327:45–87, 2013. doi:10.1016/j.jtbi.2012.12.025.
- [30] Mercedes de Onis, Cutberto Garza, Adelheid W Onyango, and Elaine Borghi. Comparison of the WHO child growth standards and the CDC 2000 growth charts. *The Journal of nutrition*, 137(1):144–148, 2007.
- [31] Esther De Vries, Sandra de Bruin-Versteeg, W Marieke Comans-Bitter, Ronald De Groot, Wim CJ Hop, Geert JM Boerma, Fred K Lotgering, and Jacques JM Van Dongen. Longitudinal survey of lymphocyte subpopulations in the first year of life. *Pediatric research*, 47(4):528–537, 2000.
- [32] Ineke den Braber, Tendai Mugwagwa, Nienke Vrisekoop, Liset Westera, Ramona Mögling, Anne Bregje de Boer, Neeltje Willems, Elise HR Schrijver, Gerrit Spierenburg, Koos Gaiser, et al. Maintenance of peripheral naive T cells is sustained by thymus output in mice but not humans. *Immunity*, 36(2):288–297, 2012.
- [33] Thomas Denny, Ram Yogev, Rebecca Gelman, Christopher Skuza, James Oleske, Ellen Chadwick, Su-Chun Cheng, and Edward Connor. Lymphocyte subsets in healthy children during the first 5 years of life. *Jama*, 267(11):1484–1488, 1992.
- [34] Daniel C Douek, Richard D McFarland, Philip H Keiser, Earl A Gage, Janice M Massey, Barton F Haynes, Michael A Polis, Ashley T Haase, Mark B Feinberg, John L Sullivan, et al. Changes in thymic function with age and during the treatment of HIV infection. *Nature*, 396(6712):690–695, 1998.
- [35] Philippa J Easterbrook, Homayoon Farzadegan, Donald R Hoover, John Palenicek, Joan S Chmiel, Richard A Kaslow, and Alfred J Saah. Racial differences in rate of CD4 decline in HIV-1-infected homosexual men. *AIDS*, 10(10):1147–1155, 1996.
- [36] Feza M Erkeller-Yuksel, V Deneys, B Yuksel, I Hannel, F Hulstaert, C Hamilton, H Mackinnon, L Turner Stokes, V al Munhyeshuli, F Vanlangendonck, et al. Age-related changes in human blood lymphocyte subpopulations. *The Journal of pediatrics*, 120(2):216–222, 1992.
- [37] Alexander J Frater, David T Dunn, Alison J Beardall, Koya Ariyoshi, John R Clarke, Myra O McClure, and Jonathan N Weber. Comparative response of African HIV-1-infected individuals to highly active antiretroviral therapy. *Aids*, 16(8):1139–1146, 2002.

- [38] Jens Geginat, Antonio Lanzavecchia, and Federica Sallusto. Proliferation and differentiation potential of human CD8⁺ memory T-cell subsets in response to antigen or homeostatic cytokines. *Blood*, 101(11):4260–4266, 2003.
- [39] Ghina Ghazeeri, Lina Abdullah, and Ossama Abbas. Immunological differences in women compared with men: overview and contributing factors. *American Journal of Reproductive Immunology*, 66(3):163–169, 2011.
- [40] EJ Giltay, JCM Fonk, BME Von Blomberg, HA Drexhage, Casper Schalkwijk, and LJG Gooren. In vivo effects of sex steroids on lymphocyte responsiveness and immunoglobulin levels in humans. *The Journal of Clinical Endocrinology & Metabolism*, 85(4):1648–1657, 2000.
- [41] Ananda W Goldrath, Lisa Y Bogatzki, and Michael J Bevan. Naive T cells transiently acquire a memory-like phenotype during homeostasis-driven proliferation. *The Journal of experimental medicine*, 192(4):557–564, 2000.
- [42] AL Gruver, LL Hudson, and GD Sempowski. Immunosenescence of ageing. *The Journal of pathology*, 211(2):144–156, 2007.
- [43] Eugene K Harris and James C Boyd. *Statistical bases of reference values in laboratory medicine*. CRC Press, 1995.
- [44] Mette D Hazenberg, Sigrid A Otto, James WT Cohen Stuart, Martie CM Verschuren, Jan CC Borleffs, Charles AB Boucher, Roel A Coutinho, Joep MA Lange, Tobias F Rinke de Wit, Aster Tsegaye, et al. Increased cell division but not thymic dysfunction rapidly affects the T-cell receptor excision circle content of the naive T cell population in HIV-1 infection. *Nature medicine*, 6(9):1036–1042, 2000.
- [45] Mette D Hazenberg, Sigrid A Otto, Annemarie MC van Rossum, Henriëtte J Scherpbier, Ronald de Groot, Taco W Kuijpers, Joep MA Lange, Dörte Hamann, Rob J de Boer, José AM Borghans, et al. Establishment of the CD4⁺ T-cell pool in healthy children and untreated children infected with HIV-1. *Blood*, 104(12):3513–3519, 2004.
- [46] Marc Hellerstein, MB Hanley, D Cesar, S Siler, C Papageorgopoulos, Eric Wieder, D Schmidt, R Hoh, R Neese, D Macallan, et al. Directly measured kinetics of circulating T lymphocytes in normal and HIV-1-infected humans. *Nature medicine*, 5(1):83–89, 1999.

- [47] Paul S Horn and Amadeo J Pesce. *Reference intervals: a user's guide*. American Association for Clinical Chemistry, Incorporated, 2005.
- [48] Sabine Huenecke, Michael Behl, Carla Fadler, Stefanie Y Zimmermann, Konrad Bochennek, Lars Tramsen, Ruth Esser, Dieter Klarmann, Martina Kamper, Alexandra Sattler, et al. Age-matched lymphocyte subpopulation reference values in childhood and adolescence: application of exponential regression analysis. *European journal of haematology*, 80(6):532–539, 2008.
- [49] Sylvie Huet. *Statistical tools for nonlinear regression: a practical guide with S-PLUS and R examples*. Springer Science & Business Media, 2004.
- [50] Emmanuel Oni Idigbe, Rosemary A Audu, Edna O Iroha, Adebola O Akinsulie, Edamisan Olusoji Temiye, Veronica C Ezeaka, Ifedayo MO Adetifa, Adesola Z Musa, Joseph Onyewuche, and Sylvester U Ikundu. T-lymphocyte subsets in apparently healthy Nigerian children. *International journal of pediatrics*, 2010, 2010.
- [51] Aydan İkinçioğulları, Tanıl Kendirli, Figen Doğu, Yonca Eğin, İsmail Reisli, Şükrü Cin, and Emel Babacan. Peripheral blood lymphocyte subsets in healthy Turkish children. *The Turkish journal of pediatrics*, 46(2):125–130, 2004.
- [52] Stephen C Jameson. Maintaining the norm: T-cell homeostasis. *Nature Reviews Immunology*, 2(8):547–556, 2002.
- [53] Yang Jiao, ZhiFeng Qiu, Jing Xie, DongJing Li, and TaiSheng Li. Reference ranges and age-related changes of peripheral blood lymphocyte subsets in Chinese healthy adults. *Science in China Series C: Life Sciences*, 52(7):643–650, 2009.
- [54] Rong Jin, Wei Wang, Jin-Yan Yao, Yu-Bin Zhou, Xiao-Ping Qian, Jun Zhang, Yu Zhang, and Wei-Feng Chen. Characterization of the in vivo dynamics of medullary CD4+ CD8- thymocyte development. *The Journal of Immunology*, 180(4):2256–2263, 2008.
- [55] Ickovics JR, Hamburger ME, Vlahov D, and et al. Mortality, cd4 cell count decline, and depressive symptoms among hiv-seropositive women: Longitudinal analysis from the hiv epidemiology research study. *JAMA*, 285(11):1466–1474, 2001.
- [56] Susan M Kaech, E John Wherry, and Rafi Ahmed. Effector and memory T-cell differentiation: implications for vaccine development. *Nature Reviews Immunology*, 2(4):251–262, 2002.

- [57] Margaret E Kemeny, Herbert Weiner, Shelley E Taylor, Steven Schneider, Barbara Visscher, and John L Fahey. Repeated bereavement, depressed mood, and immune parameters in HIV seropositive and seronegative gay men. *Health Psychology*, 13(1):14, 1994.
- [58] Ryan D Kilpatrick, Tammy Rickabaugh, Lance E Hultin, Patricia Hultin, Mary Ann Hausner, Roger Detels, John Phair, and Beth D Jamieson. Homeostasis of the naive CD4+ T cell compartment during aging. *The Journal of Immunology*, 180(3):1499–1507, 2008.
- [59] Ludger Klein, Maria Hinterberger, Gerald Wirnsberger, and Bruno Kyewski. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nature Reviews Immunology*, 9(12):833–844, 2009.
- [60] MA Kolber, MO Saenz, O Gómez-Marín, and LJ Tamariz. Race and ethnicity impact on the maximum proliferative response in peripheral blood lymphocytes from HIV-seropositive individuals. *HIV medicine*, 8(6):401–405, 2007.
- [61] Louise Kuhn, Prisca Kasonde, Moses Sinkala, Chipepo Kankasa, Katherine Semrau, Nancy Scott, Wei-Yann Tsai, Sten H Vermund, Grace M Aldrovandi, and Donald M Thea. Does severity of HIV disease in HIV-infected mothers affect mortality and morbidity among their uninfected infants? *Clinical Infectious Diseases*, 41(11):1654–1661, 2005.
- [62] Mario O Labéta, Karine Vidal, Julia E Rey Nores, Mauricio Arias, Natalio Vita, B Paul Morgan, Jean Claude Guillemot, Denis Loyaux, Pascual Ferrara, Daniel Schmid, et al. Innate recognition of bacteria in human milk is mediated by a milk-derived highly expressed pattern recognition receptor, soluble CD14. *The Journal of experimental medicine*, 191(10):1807–1812, 2000.
- [63] CJ Lawrence and VF Trewin. The construction of biochemical reference ranges and the identification of possible adverse drug reactions in the elderly. *Statistics in medicine*, 10(6):831–837, 1991.
- [64] Robert M Lawrence and Camille A Pane. Human breast milk: current concepts of immunology and infectious diseases. *Current problems in pediatric and adolescent health care*, 37(1):7–36, 2007.
- [65] Denise Lawrie, Helen Payne, Martin Nieuwoudt, and Deborah Kim Glencross. Observed full blood count and lymphocyte subset values in a cohort of clinically

- healthy South African children from a semi-informal settlement in Cape Town. *SAMJ: South African Medical Journal*, 105(7):589–595, 2015.
- [66] Joanna Lewis, A Sarah Walker, Hannah Castro, Anita De Rossi, Diana M Gibb, Carlo Giaquinto, Nigel Klein, and Robin Callard. Age and CD4 count at initiation of antiretroviral therapy in HIV-infected children: effects on long-term T-cell reconstitution. *Journal of Infectious Diseases*, page jir787, 2011.
- [67] Ida Maria Lisse, Peter Aaby, Hilton Whittle, Henrik Jensen, Mads Engelmann, and Lone Bech Christensen. T-lymphocyte subsets in West African children: impact of age, sex, and season. *The Journal of pediatrics*, 130(1):77–85, 1997.
- [68] Rishi Vishal Luckheeram, Rui Zhou, Asha Devi Verma, and Bing Xia. CD4+T Cells: Differentiation and Function. *Clinical and Developmental Immunology*, 2012(925135):12, 2012. doi:10.1155/2012/925135.
- [69] Eric S Lugada, Jonathan Mermin, Frank Kaharuza, Elling Ulvestad, Willy Were, Nina Langeland, Birgitta Asjo, Sam Malamba, and Robert Downing. Population-based hematologic and immunologic reference values for a healthy Ugandan population. *Clinical and Diagnostic Laboratory Immunology*, 11(1):29–34, 2004.
- [70] Derek C Macallan, Becca Asquith, Andrew J Irvine, Diana L Wallace, Andrew Worth, Hala Ghattas, Yan Zhang, George E Griffin, David F Tough, and Peter C Beverley. Measurement and modeling of human T cell kinetics. *European journal of immunology*, 33(8):2316–2326, 2003.
- [71] Wilson L Mandala, Jenny M MacLennan, Esther N Gondwe, Steven A Ward, Malcolm E Molyneux, and Calman A MacLennan. Lymphocyte subsets in healthy Malawians: implications for immunologic assessment of HIV infection in Africa. *Journal of Allergy and Clinical Immunology*, 125(1):203–208, 2010.
- [72] David Masopust, Vaiva Vezys, Amanda L Marzo, and Leo Lefrançois. Preferential localization of effector memory cells in nonlymphoid tissue. *Science*, 291(5512):2413–2417, 2001.
- [73] MJ Mazerolle. Making sense out of Akaike’s Information Criterion (AIC): its use and interpretation in model selection and inference from ecological data. *World Wide Web*. <<http://www.theses.ulaval.ca/2004/21842/apa.html>>, 2004.
- [74] Robert W McMurray, Robert W Hoffman, Wanda Nelson, and Sara E Walker. Cytokine mRNA expression in the B/W mouse model of systemic lupus

- erythematosus-analyses of strain, gender, and age effects. *Clinical immunology and immunopathology*, 84(3):260–268, 1997.
- [75] Jacques FAP Miller. Immunological function of the thymus. *The Lancet*, 278(7205):748–749, 1961.
- [76] Hiroshi Mohri, Alan S Perelson, Keith Tung, Ruy M Ribeiro, Bharat Ramratnam, Martin Markowitz, Rhonda Kost, Leor Weinberger, Denise Cesar, Marc K Hellenstein, et al. Increased turnover of T lymphocytes in HIV-1 infection and its reduction by antiretroviral therapy. *The Journal of experimental medicine*, 194(9):1277–1288, 2001.
- [77] Ulisses Ramos Montarroyos, Demócrito Barros Miranda-Filho, Cibele Comini César, Wayner Vieira Souza, Heloisa Ramos Lacerda, Maria de Fátima Pessoa Militão Albuquerque, Mariana Freitas Aguiar, and Ricardo Arraes de Alencar Ximenes. Factors related to changes in CD4+ T-cell counts over time in patients living with HIV/AIDS: a multilevel analysis. *PloS one*, 9(2):e84276, 2014.
- [78] D Muller, M Chen, A Vikingsson, D Hildeman, and K Pederson. Oestrogen influences CD4+ T-lymphocyte activity in vivo and in vitro in beta 2-microglobulin-deficient mice. *Immunology*, 86(2):162, 1995.
- [79] Kaja Murali-Krishna and Rafi Ahmed. Cutting edge: naive T cells masquerading as memory cells. *The Journal of Immunology*, 165(4):1733–1737, 2000.
- [80] Kaja Murali-Krishna, John D Altman, M Suresh, David JD Sourdive, Allan J Zajac, Joseph D Miller, Jill Slansky, and Rafi Ahmed. Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity*, 8(2):177–187, 1998.
- [81] John M Murray, Gilbert R Kaufmann, Philip D Hodgkin, Sharon R Lewin, Anthony D Kelleher, Miles P Davenport, and John J Zaunders. Naive T cells are maintained by thymic output in early ages but by proliferation without phenotypic change after age twenty. *Immunology and cell biology*, 81(6):487–495, 2003.
- [82] Keith Naylor, Guangjin Li, Abbe N Vallejo, Won-Woo Lee, Kerstin Koetz, Ewa Bryl, Jacek Witkowski, James Fulbright, Cornelia M Weyand, and Jörg J Goronzy. The influence of age on T cell generation and TCR diversity. *The Journal of Immunology*, 174(11):7446–7452, 2005.

- [83] Stella Ngwende, Notion T Gombe, Stanley Midzi, Mufuta Tshimanga, Gerald Shambira, and Addmore Chadambuka. Factors associated with HIV infection among children born to mothers on the prevention of mother to child transmission programme at Chitungwiza Hospital, Zimbabwe, 2008. *BMC public health*, 13(1):1, 2013.
- [84] Sabine Oertelt-Prigione. The influence of sex and gender on the immune response. *Autoimmunity reviews*, 11(6):A479–A485, 2012.
- [85] Nancy J Olsen, Susan M Viselli, Jin Fan, and William J Kovacs. Androgens Accelerate Thymocyte Apoptosis 1. *Endocrinology*, 139(2):748–752, 1998.
- [86] Martin OC Ota, O Diarmuid, Arnaud Marchant, Lawrence Yamuah, Elizabeth Harding, Shabbar Jaffar, Keith PWJ McAdam, Tumani Corrah, Hilton Whittle, et al. HIV-negative infants born to HIV-1 but not HIV-2-positive mothers fail to develop a Bacillus Calmette-Guerin scar. *AIDS*, 13(8):996, 1999.
- [87] Reinhard Pabst. The spleen in lymphocyte migration. *Immunology today*, 9(2):43–45, 1988.
- [88] Marie-Quitterie Picat, Joanna Lewis, Victor Musiime, Andrew Prendergast, Kusum Nathoo, Addy Kekitiinwa, Patricia Nahirya Ntege, Diana M Gibb, Rodolphe Thiebaut, A Sarah Walker, et al. Predicting patterns of long-term CD4 reconstitution in HIV-infected children starting antiretroviral therapy in sub-Saharan Africa: a cohort-based modelling study. *PLoS Med*, 10(10):e1001542, 2013.
- [89] Jose Pinheiro and Douglas Bates. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media, 2006.
- [90] Joseph R Podojil and Stephen D Miller. Molecular mechanisms of t-cell receptor and costimulatory molecule ligation/blockade in autoimmune disease therapy. *Immunological reviews*, 229(1):337–355, 2009.
- [91] Benedita Rocha, Nicole Dautigny, and Pablo Pereira. Peripheral T lymphocytes: expansion potential and homeostatic regulation of pool sizes and CD4/CD8 ratios in vivo. *European journal of immunology*, 19(5):905–911, 1989.
- [92] Bret J Rudy, Craig M Wilson, Stephen Durako, Anna-Barbara Moscicki, Larry Muenz, and Steven D Douglas. Peripheral blood lymphocyte subsets in adolescents: a longitudinal analysis from the REACH project. *Clinical and diagnostic laboratory immunology*, 9(5):959–965, 2002.

- [93] Bertrand Sagnia, Francis Ateba Ndongo, Suzie Ndiang Moyo Tetang, Judith Ndongo Torimiro, Cristiana Cairo, Irenée Domkam, Geraldine Agbor, Emmanuel Mve, Olive Tocke, Emilien Fouda, et al. Reference values of lymphocyte subsets in healthy, HIV-negative children in Cameroon. *Clinical and Vaccine Immunology*, 18(5):790–795, 2011.
- [94] Liisa K Selin, Meei Y Lin, Kristy A Kraemer, Drew M Pardoll, Jonathan P Schneck, Steven M Varga, Paul A Santolucito, Amelia K Pinto, and Raymond M Welsh. Attrition of T cell memory: selective loss of LCMV epitope-specific memory CD8 T cells following infections with heterologous viruses. *Immunity*, 11(6):733–742, 1999.
- [95] William T Shearer, Howard M Rosenblatt, Rebecca S Gelman, Rebecca Oyomopito, Susan Plaeger, E Richard Stiehm, Diane W Wara, Steven D Douglas, Katherine Luzuriaga, Elizabeth J McFarland, et al. Lymphocyte subsets in healthy children from birth through 18 years of age: the Pediatric AIDS Clinical Trials Group P1009 study. *Journal of Allergy and Clinical Immunology*, 112(5):973–980, 2003.
- [96] RJ Silverwood and TJ Cole. Statistical methods for constructing gestational age-related reference intervals and centile charts for fetal size. *Ultrasound in obstetrics & gynecology*, 29(1):6–13, 2007.
- [97] Colette J Smith, Caroline A Sabin, Mike S Youle, Sabine Kinloch-de Loes, Fiona C Lampe, Sara Madge, Ian Cropley, Margaret A Johnson, and Andrew N Phillips. Factors influencing increases in CD4 cell counts of HIV-positive persons receiving long-term highly active antiretroviral therapy. *Journal of Infectious Diseases*, 190(10):1860–1868, 2004.
- [98] ME Smith and WL Ford. The recirculating lymphocyte pool of the rat: a systematic description of the migratory behaviour of recirculating lymphocytes. *Immunology*, 49(1):83, 1983.
- [99] GG Steinmann, B Klaus, and H-K Müller-hermelink. The involution of the ageing human thymic epithelium is independent of puberty. *Scandinavian journal of immunology*, 22(5):563–575, 1985.
- [100] WH Stimson. Oestrogen and Human T Lymphocytes: Presence of Specific Receptors in the T-Suppressor/Cytotoxic Subset. *Scandinavian journal of immunology*, 28(3):345–350, 1988.

- [101] Charles D Surh and Jonathan Sprent. Homeostasis of naive and memory T cells. *Immunity*, 29(6):848–862, 2008.
- [102] Ronald A Thisted. *Elements of statistical computing: numerical computation*, volume 1. CRC Press, 1988.
- [103] David J Tollerud, Suzanne T Ildstad, Linda Morris Brown, Jeffrey W Clark, William A Blattner, Dean L Mann, Carolyn Y Neuland, Luba Pankiw-Trost, and Robert N Hoover. T-cell subsets in healthy teenagers: transition to the adult phenotype. *Clinical immunology and immunopathology*, 56(1):88–96, 1990.
- [104] JY Tsay, I-Wen Chen, HR Maxon, and L Heminger. A statistical method for determining normal ranges from laboratory data including values below the minimum detectable value. *Clinical chemistry*, 25(12):2011–2014, 1979.
- [105] Aster Tsegaye, Tsehaynesh Messele, Tesfaye Tilahun, Ermias Hailu, Tefera Sahlu, Ronan Doorly, Arnaud L Fontanet, and Tobias F Rinke de Wit. Immunohematological reference ranges for adult Ethiopians. *Clinical and diagnostic laboratory immunology*, 6(3):410–414, 1999.
- [106] Mathilde Turfkruyer and Valerie Verhasselt. Breast milk and its impact on maturation of the neonatal immune system. *Current opinion in infectious diseases*, 28(3):199–206, 2015.
- [107] SS Uppal, Shashi Verma, and PS Dhot. Normal values of CD4 and CD8 lymphocyte subsets in healthy Indian adults and the effects of sex, age, ethnicity, and smoking. *Cytometry Part B: Clinical Cytometry*, 52(1):32–36, 2003.
- [108] Ulrich H von Andrian and Thorsten R Mempel. Homing and cellular traffic in lymph nodes. *Nature Reviews Immunology*, 3(11):867–878, 2003.
- [109] Harald Von Boehmer, Iannis Aifantis, Fotini Gounari, Orly Azogui, Lorelee Haughn, Irina Apostolou, Elmar Jaeckel, Fabio Grassi, and Ludger Klein. Thymic selection revisited: how essential is it? *Immunological reviews*, 191(1):62–78, 2003.
- [110] AM Wade and AE Ades. Age-related reference ranges: Significance tests for models and confidence intervals for centiles. *Statistics in medicine*, 13(22):2359–2367, 1994.
- [111] AM Wade and AE Ades. Incorporating correlations between measurements into the estimation of age-related reference ranges. *Statistics in medicine*, 17(17):1989–2002, 1998.

- [112] Michael A Weinreich and Kristin A Hogquist. Thymic emigration: when and how T cells leave home. *The Journal of Immunology*, 181(4):2265–2270, 2008.
- [113] Wolfgang Weninger, Maura A Crowley, N Manjunath, and Ulrich H von Andrian. Migratory properties of naive, effector, and memory CD8+ T cells. *The Journal of experimental medicine*, 194(7):953–966, 2001.
- [114] E John Wherry, Volker Teichgräber, Todd C Becker, David Masopust, Susan M Kaech, Rustom Antia, Ulrich H Von Andrian, and Rafi Ahmed. Lineage relationship and protective immunity of memory CD8 T cell subsets. *Nature immunology*, 4(3):225–234, 2003.
- [115] Wikipedia contributors. Cluster of differentiation. Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Cluster_of_differentiation&oldid=651807740, April 2015. Date retrieved: 25 April 2015 12:28 UTC.
- [116] Eileen M Wright and Patrick Royston. A comparison of statistical methods for age-related reference intervals. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(1):47–69, 1997.
- [117] Masaru Yoshida, Kanna Kobayashi, Timothy T Kuo, Lynn Bry, Jonathan N Glickman, Steven M Claypool, Arthur Kaser, Takashi Nagaishi, Darren E Higgins, Emiko Mizoguchi, et al. Neonatal Fc receptor for IgG regulates mucosal immune responses to luminal bacteria. *The Journal of clinical investigation*, 116(8):2142–2151, 2006.
- [118] Liang Zhou, Mark MW Chong, and Dan R Littman. Plasticity of CD4+ T cell lineage differentiation. *Immunity*, 30(5):646–655, 2009.