CLINICAL AND BIOLOGICAL DISEASE PROGRESSION IN CHILDREN
WITH VERTICALLY-ACQUIRED HIV-1 INFECTION IN EUROPE:
STATISTICAL DETERMINATIONS

A thesis presented for the degree of
Doctor of Philosophy
University of London

Linsay Anne Gray
Institute of Child Health
University College London

2003
Abstract

Clinical status, with HIV (human immunodeficiency virus) RNA (ribonucleic acid) viral load and immunological factors are used by clinicians in decisions regarding the management of HIV-1-infected children. Understanding of these dynamics in children infected from their mothers around birth (vertically-infected) over time thus facilitates clinical supervision.

This thesis aims to elucidate patterns of clinical, virological and immunological progression over age, and how they inter-relate, using data collected since 1986 on nearly 200 vertically-infected children in the European Collaborative Study cohort study, by implementation of appropriate statistical methodology.

Patterns of progression to the Centers for Disease Control categories and to death are clarified by time-to-event methods, and compared across calendar periods, reflecting changing antiretroviral treatment policies. Clinical status is also considered on a visit-by-visit basis to assess fluctuations in illness. HIV RNA viral load over age is elucidated using fractional polynomials and a specific estimation method of Hughes, to account for repeated left-censored measurements. Dynamics of immunological markers over age are investigated with multi-level splines. Methods used for the biological parameters allow stratification by factors such as gender, race and treatment. Early levels of biological markers and clinical indications are related to subsequent progression to serious disease or death by Cox proportional hazards regression.

A novel fluctuating pattern in occurrence of serious HIV-related disease over the first 15 years of life was revealed. Differences in RNA levels and patterns in treatment by gender, and in immunological markers by gender and race over age were identified. Immunological measurements in the first six months of life were found to predict progression before one year of age and early clinical signs were prognostic of progression beyond the first year of life. This work provides an extension of previous knowledge and understanding of clinical and biological disease progression in children vertically-infected with HIV.
CLINICAL AND BIOLOGICAL DISEASE PROGRESSION IN CHILDREN WITH VERTICALLY ACQUIRED HIV-1 INFECTION IN EUROPE:
STATISTICAL DETERMINATIONS

Title page 1
Abstract 3
Contents 4
List of tables 10
List of figures 16
Acknowledgements 21
List of abbreviations 23

Chapter 1 Vertically-acquired paediatric HIV infection 24
1.1 Introduction 24
1.2 Epidemiology of HIV infection in children 25
1.3 Disease progression in children 29
  1.3.1 Clinical disease progression 29
  1.3.2 Viral load patterns over childhood 36
  1.3.3 Immunological patterns over childhood 44
  1.3.4 Determinants of the rate of clinical disease progression during childhood 53
Chapter 2  Data source and methods  

2.1 Aim and Objectives  

2.2 The European Collaborative Study  

2.2.1 Role of the researcher  

2.2.2 Diagnosis of HIV  

2.2.3 Paediatric classification system  

2.2.4 Treatment  

2.2.5 HIV RNA viral load  

2.2.6 Immunological markers  

2.2.7 General definitions  

2.3 Clinical patterns analysis  

2.3.1 Overall disease progression  

2.3.2 Progression to serious disease stratified by ART  

2.3.3 Changing presence of HIV-related symptoms  

2.4 HIV RNA viral load analysis  

2.4.1 Establishing the basic model for the overall pattern of viral load over age  

2.4.2 Accounting for repeated measurements and censoring in HIV RNA viral load  

    Hughes’ modified Monte Carlo expectation-maximisation algorithm  

    Alternative method for accounting for the repeated measurements and censoring of HIV RNA viral load  

2.4.3 Model selection strategy  

2.5 Patterns of immunological markers  

2.5.1 Modelling immunological markers over age  

    Natural cubic splines  

2.5.2 Assessing factors influencing immunological markers  

2.5.3 Comparison of infected and uninfected children
Chapter 3 Results: Clinical disease patterns over childhood in vertically-infected children

3.1 Introduction

3.2 The children and their mothers

3.3 Overall disease progression in infected children
   3.3.1 Clinical status
   3.3.2 Immunological disease progression

3.4 Progression to serious disease by calendar period of birth

3.5 Progression in ART naïve children

3.6 Clinical progression in ART treated children

3.7 Comparison of progression in ART naïve and ART treated children

3.8 Comparison of progression by ZDV vertical transmission prophylaxis

3.9 Risk of serious disease progression by pre-progression ART, PCP prophylaxis and IVIG treatment

3.10 Presence of HIV-related symptoms at six monthly age intervals

3.11 Clinical status five-year look-back

3.12 Immunological status five-year look-back

Key points: Clinical disease patterns over childhood in vertically-infected children
Chapter 4 Results: HIV RNA Viral load over age in infected children

4.1 Data

4.2 HIV RNA viral load over age

4.2.1 Modelling HIV RNA viral load over age using fractional polynomials

4.2.2 Alternatives to fractional polynomials

Conventional polynomial

Change-point regression model

4.2.3 Comparison of methods for modelling HIV RNA viral load over age

4.3 Explanatory variables for HIV RNA viral load levels and patterns over age

4.3.1 Application of Hughes’ modified Monte Carlo EM algorithm

Main effects of factors influencing HIV RNA viral load levels over age

Interactions of factors influencing HIV RNA viral load pattern over age

4.3.2 Comparison with alternative to Hughes’ method

Key points: HIV RNA Viral load over age in infected children

Chapter 5 Results: Patterns of immunological markers in uninfected and infected children

5.1 Data

5.2 Immunological markers in uninfected children

5.2.1 \( \log_{10} \) CD4\(^+\) cell count over age in uninfected children

5.2.2 Factors associated with levels and patterns of \( \log_{10} \) CD4\(^+\) cell counts in uninfected children

5.2.3 CD8\(^+\) cell count over age in uninfected children

5.2.4 Factors associated with levels and patterns of CD8\(^+\) cell counts in uninfected children

5.2.5 Absolute lymphocyte counts over age in uninfected children
5.2.6 Factors associated with levels and patterns of absolute lymphocyte counts in uninfected children

5.2.7 CD4⁺ percentage over age in uninfected children

5.2.8 Factors associated with levels and patterns of CD4⁺ percentage in uninfected children

5.3 Immunological markers in infected children

5.3.1 Log₁₀ CD4⁺ cell count over age in infected children

5.3.2 Factors associated with levels and patterns of CD4⁺ cell counts in infected children

5.3.3 CD8⁺ cell count over age in infected children

5.3.4 Factors associated with levels and patterns of CD8⁺ cell counts in infected children

5.3.5 Absolute lymphocyte counts over age in infected children

5.3.6 Factors associated with levels and patterns of absolute lymphocyte counts in infected children

5.3.7 CD4⁺ percentage over age in infected children

5.3.8 Factors associated with levels and patterns of CD4⁺ percentage in infected children

5.4 Comparison of immunological markers patterns over age in infected and uninfected children

Key points: Patterns of immunological markers in uninfected children

Key points: Patterns of immunological markers in infected children

Key points: Comparison of patterns of immunological markers in uninfected and infected children

Chapter 6 Results: Early-life determinants of clinical disease progression in infected children

6.1 Data

6.2 Association of overall disease progression with early immunological and clinical measurements, child and maternal factors
6.2.1 Univariable Cox regression analysis
6.2.2 Assessing gender- and race-specific effects on prognostic value of clinical and immunological markers
6.2.3 Multivariable Cox regression analysis: Simultaneous predictive value of early clinical and immunological variables
6.3 Rapid progression
6.3.1 Univariable Cox regression analysis
6.3.2 Multivariable Cox regression analysis
6.4 Progression after one year of age
6.4.1 Univariable Cox regression analysis
6.4.2 Multivariable Cox regression analysis
6.5 Threshold for beyond six months CD4⁺ percentage

Key points: Early-life determinants of clinical disease progression in infected children

Chapter 7 Conclusions on the analysis of paediatric HIV data

7.1 Conclusions
7.1.1 Pattern of clinical manifestations over age
7.1.2 HIV RNA viral load
7.1.3 Immunological patterns
7.1.4 Determinants of serious clinical disease progression
7.2 Recommendations
7.3 Limitations of analyses
7.3.1 Data
7.3.2 Statistical methods
7.4 Recommendations for epidemiological and statistical research
Reference List 213
Appendices 238
2.1 European Collaborative Study data collection forms 238
2.2 Collaborators in the paediatric centres of the European Collaborative Study 247
2.3 Publications 248
2.4 Clinical and immunological CDC categories for children with HIV infection 304
2.5 Fractional polynomial models 306
2.6 Change-point models 308
2.7 Hughes’ modified Monte Carlo EM algorithm 309
2.8 Natural cubic splines 311
2.9 Programmes used for selective analyses 312
   2.9.1: STATA programmes for fractional polynomials for HIV RNA viral load 312
   2.9.2: STATA programmes for conventional polynomials for HIV RNA viral load 313
   2.9.3: S-PLUS programme for change-point model for HIV RNA viral load 314
   2.9.4: S-PLUS programme for Hughes' method for HIV RNA viral load 315
   2.9.5: S-PLUS programme for mixed-linear with mid-point censoring for HIV RNA viral load 322
   2.9.6: S-PLUS programmes for splines for immunological markers 325
   2.9.7: STATA programmes for Cox proportional hazards regression for determinants of disease progression 331
4.1 Fitted one-, two- and three-power fractional polynomials 340
4.2 Expressions of models fitted using Hughes’ method to determine important covariate and interaction terms for the final model 341
5.1 Example of establishing choice of knots for spline model for CD4+ cell count of uninfected children 343
5.2 Super smoothers for selected immunological markers by gender and race 344
5.3 Immunological markers by gender and race over age in infected children adjusted for by ART 349
6.1 Considered multivariable models 351
6.2 Multivariable models for determinants analysis adjusted by PCP prophylaxis and IVIG ever before progression 357
List of Tables

Chapter 1

1.1: Features of studies addressing the progression of HIV disease through clinical stages 34
1.2: Methodological aspects of previous studies on HIV-1 RNA viral load in infected children 42
1.3a: Features of previous studies on immunological markers in uninfected children born to women infected with HIV 49
1.3b: Features of previous studies on immunological markers in children vertically-infected with HIV 50
1.3c: Features of previous studies on comparisons of immunological markers in children born to women infected with HIV by infection status 52
1.4: Features of studies addressing the determining factors of serious disease progression 57

Chapter 3

3.1: Mono- and combination-therapy initiation by clinical stage 84
3.2: Progression and Follow-Up of 170 infected children 87
3.3: Clinical and immunological CDC classification of visits 89
3.4: Clinical and immunological CDC classification at treatment initiation 91
3.5: Log-rank test results comparing estimated survival curves for treated and untreated children 94

Chapter 4

4.1: Characteristics of infected children with HIV RNA viral load measurements 106
4.2: Optimal fractional polynomial models for each power 110
4.3: Comparisons of optimal fractional polynomial models 110
4.4: AICs of the three model forms for HIV RNA viral load over age 113
| 4.5: | Likelihoods for models used for the selection of covariates for HIV RNA viral load over age by model comparisons in 5.6 and 5.8 | 115 |
| 4.6: | Significance tests for main effects of covariates for HIV RNA viral load pattern over age based on likelihood values from 4.5 | 116 |
| 4.7: | Univariable results for analyses of RNA viral load pattern over age by gender, treatment, CD4+ cell count and assay | 116 |
| 4.8: | Significance tests for interactions of covariates for HIV RNA viral load pattern over age based on likelihood values from 4.5 | 118 |

Chapter 5

| 5.1: | 1674 uninfected and infected children and their HIV infected mothers | 126 |
| 5.2: | Distribution of laboratory measurements by infection status | 127 |
| 5.3: | Gender, race and log10 CD4+ cell counts over age in uninfected children | 131 |
| 5.4: | Gender, race and CD8+ cell counts over age in uninfected children | 134 |
| 5.5: | Gender, race and log10 absolute lymphocyte counts over age in uninfected children | 138 |
| 5.6: | Gender, race and percentage of CD4+ cell counts over age in uninfected children | 141 |
| 5.7: | Gender, race and log10 CD4+ cell counts over age in infected children | 145 |
| 5.8: | Gender, race and CD8+ cell counts over age in infected children | 149 |
| 5.9: | Gender, race and absolute lymphocyte counts over age in infected children | 152 |
| 5.10: | Gender, race and percentage of CD4+ cell counts over age in infected children | 155 |
| 5.11: | Predicted log10 CD4+ cell count levels for uninfected and infected children | 160 |
| 5.12: | Summary of gender and race effects for lymphocyte patterns in uninfected children | 161 |
| 5.13: | Summary of gender and race effects for lymphocyte patterns in infected children | 162 |
Chapter 6
6.1: Characteristics of 161 infected children born before 1997  165
6.2: Clinical and laboratory data on the first six months of life for 161 infected children born before 1997  169
6.3: Univariable Cox regression results for overall progression to serious disease or death by infant and maternal factors  172
6.4: Multivariable Cox regression results for overall progression to serious disease or death  180
6.5: Univariable Cox regression results for rapid progression  182
6.6: Multivariable Cox regression results for rapid progression  184
6.7: Univariable Cox regression results for long-term progression  186
6.8: Multivariable Cox regression results for long-term progression  188

Appendices
A5.1: Gender, race and CD4+ cell counts over age in infected children adjusted for by ART  349
A5.2: Gender, race and CD8+ cell counts over age in infected children adjusted for by ART  349
A5.3: Gender, race and absolute lymphocyte counts over age in infected children adjusted for by ART  349
A5.4: Gender, race and percentage of CD4+ cell counts over age in infected children adjusted for by ART  350
A6.1.1: Considered multivariable models for predicting overall progression  351
A6.1.2: Considered multivariable models for predicting rapid progression  354
A6.1.3: Considered multivariable models for predicting long-term progression  355
A6.2.1: Timing of progression by PCP prophylaxis and IVIG ever before progression  357
A6.2.2: Univariable results for overall progression to serious disease or death by PCP prophylaxis and IVIG ever before progression  357
A6.2.3: Multivariable results for overall progression to serious disease or death adjusted for PCP and bacterial infection prophylaxes

A6.2.4: Univariable results for rapid progression to serious disease or death by PCP prophylaxis and IVIG ever before progression

A6.2.5: Multivariable results for rapid progression to serious disease or death adjusted for PCP and bacterial infection prophylaxes

A6.2.6: Univariable results for long term progression to serious disease or death by PCP prophylaxis and IVIG ever before progression

A6.2.7: Multivariable results for long-term progression to serious disease or death adjusted for PCP and bacterial infection prophylaxes
List of Figures

Chapter 1
1.1 Treatment practices over time in European paediatric settings 28

Chapter 2
2.1 Location of Paediatric centres of the ECS 62
2.2 ECS children by year of birth and infection status 63
2.3 Two-step procedure for modelling longitudinal HIV-RNA viral load measurements as a function of age, accommodating repeated measures and censoring scheme 71
2.4 Timeline for data for clinical patterns, HIV RNA viral load, immunological patterns and determinants of clinical disease progression analyses 80

Chapter 3
3.1 Eras of birth and treatment administration of 170 infected children 83
3.2 Type of treatment administered initially and at last visit 83
3.3 Estimated progression to CDC clinical categories A or more severe, B or more severe, C or death and death for all children using Kaplan-Meier lifetable analysis 86
3.4 Estimated progression to CDC immunological categories 1 or 2, and 2 using Kaplan-Meier lifetable analysis 88
3.5 Estimated progression to serious disease (CDC stage C) or death for cohort 1, cohort 2 and cohort 3 using Kaplan-Meier lifetable analysis 90
3.6 Estimated progression to CDC clinical categories A or more severe, B or more severe, C or death and death for 55 ART naïve children using Kaplan-Meier lifetable analysis 93
3.7: Estimated progression to CDC clinical categories A or more severe, B or more severe, C or death and death for the 115 children treated with ART using Kaplan-Meier lifetable analysis 93

3.8: Clinical status for infected children assessed cross-sectionally at six-monthly visits 97

3.9: Clinical status for the ART naïve children assessed cross-sectionally at 6-monthly intervals 98

3.10: Clinical status for the ART treated children assessed cross-sectionally at 6-monthly intervals 98

3.11: Clinical status for 45 children who died with AIDS assessed cross-sectionally at 6-monthly intervals 99

3.12: Children alive and in follow-up at five years of age: Clinical status at previous visits by current visit 100

3.13: Children alive and in follow-up at five years of age: CDC immunological status at previous visits by current visit 102

Chapter 4

4.1: Observed HIV-RNA viral load measurements with super smoother 108

4.2: Underlying structure of HIV-RNA viral load by age as shown by the super-smoother along with fractional polynomial, conventional polynomial and change-point model representations 112

4.3: $\log_{10}$ RNA viral load by gender 118

4.4: $\log_{10}$ RNA Viral Load (by Roche assays) by age, gender and treatment 120

4.5: Gender and treatment strata-specific fitted curves of the model from application of Hughes’ modified EM algorithm and fitted curves with confidence bands arising from the application of the LME approach with mid-point censored values. 121
Chapter 5

5.1: Observed, smoothed and fitted values of $\log_{10} CD4^+$ cell count over age in uninfected children 129

5.2: Fitted values of $\log_{10} CD4^+$ cell count over age by gender and race in uninfected children 131

5.3: Observed, smoothed and fitted values of $\log_{10} CD8^+$ cell count over age in uninfected children 133

5.4: Fitted values of $\log_{10} CD8^+$ cell count over age by gender and race in uninfected children 135

5.5: Observed, smoothed and fitted values of $\log_{10}$ absolute lymphocyte count over age in uninfected children 137

5.6: Fitted values of $\log_{10}$ absolute lymphocyte count over age by gender and race in uninfected children 139

5.7: Observed, smoothed and fitted values of CD4$^+$ percentage over age in uninfected children 140

5.8: Fitted values of CD4$^+$ percentage over age by gender and race in uninfected children 142

5.9: Observed, smoothed and fitted values of $\log_{10} CD4^+$ cell count over age in infected children 144

5.10: Fitted values of $\log_{10} CD4^+$ cell count over age by gender and race in infected children 146

5.11: Observed, smoothed and fitted values of $\log_{10} CD8^+$ cell count over age in infected children 148

5.12: Fitted values of $\log_{10} CD8^+$ cell count over age by gender and race in infected children 150

5.13: Observed, smoothed and fitted values of $\log_{10}$ absolute lymphocyte count over age in infected children 151

5.14: Fitted values of $\log_{10}$ absolute lymphocyte count over age by gender and race in infected children 153

5.15: Observed, smoothed and fitted values of CD4$^+$ percentage over age in infected children 154
5.16: Fitted values of CD4\(^+\) percentage over age by gender and race in infected children

5.17: Observed values of log\(_{10}\) CD4\(^+\) cell counts of infected children with 5th percentile for uninfected children

5.18: Observed values of log\(_{10}\) CD4\(^+\) cell counts of uninfected children with 95th percentile for infected children

Chapter 6

6.1: The Kaplan-Meier time-to-serious disease progression or death curve for 161 infected children born before 1997

6.2: The Kaplan-Meier time-to-event curve for serious disease progression or death within the first year of life in 161 infected children born before 1997

6.3: The Kaplan-Meier time-to-event curve for serious disease progression or death beyond the first year of life in 161 infected children born before 1997

6.4: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with a first CD4\(^+\) percentage above 20% and equal to or below 20% in the first six months of life

6.5: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with and those without early persistence of hepatomegaly in the first six months of life

6.6: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with and those without persistence of lymphadenopathy in the first six months of life

6.7: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with and those without persistence of enlarged axillary nodes in the first six months of life
6.8: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with and those without persistence of splenomegaly in the first six months of life 176

6.9: Kaplan-Meier survival plot of progression to CDC clinical category C or death for girls and boys 176

6.10: Kaplan-Meier survival plot of progression to CDC clinical category C or death for black and white children 177

6.11: Kaplan-Meier survival plot of progression to CDC clinical category C or death for full term, moderately premature children and severely premature 177

6.12: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those administered and those not administered neonatal ART prophylaxis for prevention of vertical transmission 178

Appendices
A4.1A: Fitted one-, two- and three-power fractional polynomials 340
A5.1A: Example of establishing choice of knots for spline model for CD4+ cell count of uninfected children 343
A5.1: Super smoothers of log10 CD8+ cell counts over age by gender in uninfected children 344
A5.2: Super smoothers of CD4+ percentage over age by race in uninfected children 345
A5.3: Super smoothers of log10 CD4+ cell counts over age by gender in infected children 346
A5.4: Super smoothers of log10 CD4+ cell counts over age by race in infected children 347
A5.5: Fitted values of log10 CD4+ cell counts over age by vital status at last follow-up in infected children 348
A6.2: Kaplan-Meier survival plot of progression to CDC clinical category C or death by receipt of PCP prophylaxis 358
A6.3: Kaplan-Meier survival plot of progression to CDC clinical category C or death by receipt IVIG as prophylaxis for bacterial infections 359
Acknowledgements

I would like to acknowledge the help of those without whom this work would not have been possible. I begin by expressing my deepest thanks and appreciation to my supervisors, Marie-Louise Newell and Mario Cortina-Borja for their constant support and encouragement.

I would also like to thank all the clinicians and laboratory staff who have contributed to this research through their collaboration in the European Collaborative Study. Particular thanks go to the women and children for their participation in the study.

Many thanks to Catherine Peckham and Tim Cole for their helpful suggestions, and to my other colleagues Claire Thorne, Lucy Pembrey, Simona Fiore and Madeleine Bunders for their help and understanding. Thanks, especially, to Claire Hankin for proof reading, and also to Martin King for his help with some graph formatting.

Finally, I take pleasure in thanking my wonderful family and friends for their reassurance and faith, and Alzy for his extreme patience, and for being my tower of strength.

This work was supported by the European Commission: the ECS is a concerted action of the European Commission. Collaborating Centres were supported by National funding. The Medical Research Council (UK) provides support to the coordinating centre.
In memory of my dearest sister Gillian.
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus (&quot;HIV&quot; refers to HIV-1 subtype)</td>
</tr>
<tr>
<td>AIDS</td>
<td>acquired immune deficiency syndrome</td>
</tr>
<tr>
<td>ART</td>
<td>antiretroviral treatment</td>
</tr>
<tr>
<td>HAART</td>
<td>highly active antiretroviral treatment</td>
</tr>
<tr>
<td>NRTI</td>
<td>nucleoside reverse transcriptase inhibitor</td>
</tr>
<tr>
<td>NNRTI</td>
<td>non-nucleoside reverse transcriptase inhibitor</td>
</tr>
<tr>
<td>PI</td>
<td>protease inhibitors</td>
</tr>
<tr>
<td>PCP</td>
<td><em>Pneumocystis carinii</em> pneumonia</td>
</tr>
<tr>
<td>IVIG</td>
<td>intravenous immunoglobulin</td>
</tr>
<tr>
<td>ECS</td>
<td>European Collaborative Study</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers For Disease Control</td>
</tr>
<tr>
<td>LME</td>
<td>linear mixed effects</td>
</tr>
<tr>
<td>NA</td>
<td>not applicable</td>
</tr>
<tr>
<td>HLA</td>
<td>human leukocyte antigen</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>NASBA</td>
<td>Nucleic Acid Sequence-Based Amplification</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>EM</td>
<td>expectation-maximisation</td>
</tr>
<tr>
<td>LIP</td>
<td>lymphoid interstitial pneumonitis</td>
</tr>
<tr>
<td>RR</td>
<td>Relative risk</td>
</tr>
<tr>
<td>HR</td>
<td>hazard ratios</td>
</tr>
<tr>
<td>df</td>
<td>degrees of freedom</td>
</tr>
<tr>
<td>IQR</td>
<td>inter-quartile range</td>
</tr>
</tbody>
</table>
Chapter 1 Vertically-acquired paediatric HIV infection

1.1 Introduction
At the end of 2002, an estimated 38.6 million adults, half of whom were women, and 3.2 million children under 15 years were living with human immunodeficiency virus (HIV)/acquired immune deficiency syndrome (AIDS) worldwide (1). During that year alone, 610,000 children are thought to have died with AIDS-related illnesses. Every year there are an estimated 800,000 new infections in children, mostly in Sub-Saharan Africa and nearly all are vertically-acquired (2). A small subset, less than 500, of these are children in Western Europe where the annual death toll is less than 100, although by the end of 2001 there were around 5,000 infected individuals under the age of fifteen years estimated to be living with HIV/AIDS (2). The prevalence of HIV in women giving birth varies across Western Europe and within countries (3), with estimates of antenatal prevalence ranging from less than one in 5,000 in Scandinavian countries to one in 700 in Spain (4), although estimates are as high as one in 200 in inner city areas in the UK (3;5). These figures represent increases in the estimates of prevalence of HIV in pregnant women since the late 1990s (3) and with mother-to-child (or vertical) transmission still occurring in a minority of cases, the problem of perinatally acquired paediatric HIV infection remains important.

It has been suggested that affected children may have a unique response to the infection, manifesting clinical deterioration, virological containment and immunological reactions distinct from those seen in individuals acquiring their infection through other routes. Moreover, unlike with other modes of HIV transmission, in infants of HIV infected mothers
the time of infection is known with reasonable accuracy, making the population of vertically-infected children a particularly interesting group for investigating disease progression in HIV infection.

1.2 Epidemiology of HIV infection in children
Globally, the vast majority of paediatric HIV infection is a result of vertical transmission (6;7). This can occur during pregnancy (in utero), around delivery (peripartum), or via breastfeeding (postpartum) (8;9). Timing of in utero transmission varies, with some infants infected earlier in gestation (8), but for most, infection occurs late in pregnancy or during delivery. Postpartum transmission is rare in Europe and the rest of the developed world, as viable alternatives permit the avoidance of breastfeeding (10). In the absence of specific interventions, transmission rates have ranged from 16-30% in Europe and the USA, 25-40% in Africa and 13-48% in South and South East Asia (11-14). Avoidance of breastfeeding contributed to reduction in rates in Western Europe to around 15-20% before the introduction of other effective interventions in the early to mid 1990s, but with additional preventative regimens now in place in developed countries vertical transmission rates have been reduced to less than 1-2% (15-17). The other specific interventions include elective caesarean section (18) and antenatal, perinatal and neonatal prophylactic antiretroviral treatment (ART) (10).

The risk of vertical transmission is associated with vaginal delivery (11;15;18-20), duration of rupture of membranes (21), breastfeeding (22-24) and advanced maternal disease (25;26) (in particular high maternal HIV RNA viral load (15;27-32) and low antenatal CD4⁺ cell count (11;25;33)). There is a well documented association between transmission and prematurity (15;26;33;34), but the mechanisms involved are incompletely understood.
Weaker evidence suggests possible elevated transmission risks associated with sexually transmitted disease co-infection (35).

As vertically-infected children become exposed to the virus at a time when their immune systems are still developing, their response to the infection may differ from that of adults (36). A difference in clinical manifestations between children and adults has been implied since the first report of AIDS in children in 1982 (37) and further documented in populations studies (36). There is suggestion of poorer prognosis in children (38), with findings indicating more rapid disease progression (39-41), possible higher prevalence of invasive bacterial infections (42) and more elevated HIV ribonucleic acid (RNA) viral loads (43-45), although, as viral levels in primary infection are rarely quantified in adults, comparison is difficult. Opportunistic infections may be more aggressive in infants, being exposed for the first time, than in adults whose immune systems are likely to have previously encountered such pathogens (46). Contrary, there has also been suggestion that survival is generally better in children than in adults (36). Clearly, precise disease evolution in adults is difficult to document as infected individuals are not uniformly followed from seroconversion – for instance, those with the most rapid disease progression may not be identified and those with a long asymptomatic period can only be identified in screening programmes. This complicates comparison of immune response between adults and children, and is likely to partly account for the conflicting findings.

Current estimates of AIDS incubation time, and time to death in infected children are longer than estimated in the earlier years of the epidemic (38). This is not only due to the
improvement in available antiretroviral and antibiotic treatment regimens, but is also explained by the comprehensive descriptions of the natural history of vertically-acquired HIV infection, now available, from long duration of follow-up of birth cohorts of large numbers of children. With the development of HIV antibody testing and virus-isolation tests, the previous biases resulting from inclusion of only symptomatic children have now been avoided (47).

Access to ART for the prevention of progression of HIV-related disease is now standard in Western Europe (Figure 1.1). ART first came into clinical paediatric use in Europe as zidovudine monotherapy around 1987 (48), and double therapy of two nucleoside reverse transcriptase inhibitors (NRTIs) around 1994, but highly active ART (HAART) - combination therapy of three and more NRTIs, non-nucleoside reverse transcriptase inhibitors (NNRTIs) and/or protease inhibitors (PIs) - has been administered routinely in developed countries since the mid 1990s. It has been shown in adult (49;50) and child (51-55) clinical trials that regimens of combination ART are more efficacious than monotherapy in preventing the progression of disease and death, and that the use of HAART including PI has led to a reduction in death in HIV-infected children (56;57). Findings from paediatric clinical trials show ART benefits in terms of reduction in HIV RNA viral load and reconstitution of CD4+ cell counts (51-55) and their efficacy in reducing the HIV-related disease progression and mortality rates in perinatally-infected children has been shown in surveillance and cohort studies (48;56). However, decisions on regimens depend on the status of the individual child, and vary across countries as well as between clinicians and over time (58). In addition, prophylactic options such as trimethoprim-sulfamethoxazole for the
prevention of *Pneumocystis carinii* pneumonia (PCP), formerly the most common opportunistic infection and AIDS-defining disease in HIV-infected children (59), and antimycobacterials for tuberculosis have contributed (60) to the decrease in development of serious illness. However, intravenous immunoglobulin (IVIG) used in the past as a prophylaxis for general bacterial infections has been shown to have little lasting benefit in children with HIV (61).

**Figure 1.1 Treatment practices over time in European paediatric settings**

<table>
<thead>
<tr>
<th>Year</th>
<th>Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>No recommendations for treatment</td>
</tr>
<tr>
<td>1981</td>
<td>Mainly monotherapy administration: usually zidovudine (sometimes didanosine) to children with serious HIV-related symptoms.</td>
</tr>
<tr>
<td>1982</td>
<td>Double therapy of NRTIs given to symptomatic children</td>
</tr>
<tr>
<td>1983</td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Initiation of HAART given before the onset of severe HIV-related symptoms, with double therapy popular initially and monotherapy use declining</td>
</tr>
<tr>
<td>2001</td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
</tr>
</tbody>
</table>
1.3 Disease progression in children

The pathogenesis and rate of disease progression resulting from HIV infection can be assessed clinically and biologically. Records of signs and symptoms and of laboratory ascertainment provide details on clinical, immunological and virological parameters. Data from both paediatric prospective studies and clinical trials have previously been used to describe disease progression in children. Here, a brief account is given of what is known to date from those studies, and the methodological issues of their design and statistical approaches are discussed. Outlined are the potential shortcomings of these, and opportunities for expansion and improvement of analyses, setting the scene for the work of this thesis which brings together the various aspects of disease progression in paediatric HIV infection.

1.3.1 Clinical disease progression

As for adults (62), prognosis of children born with vertically-acquired HIV infection in the developed world has improved with advances in the antiretroviral and antibiotic treatment in the last decade (48). The figures of HIV/AIDS Surveillance in Europe in 2003 (63) show that the AIDS cases in vertically-infected children have declined over the previous five years. As reported from the Italian Register for HIV-infected Children and the Italian National AIDS Registry of over 1,000 children, the widespread use of double therapy since 1996 and subsequent introduction of combination therapy of three or more drugs, has seen lessening rates of mortality in infected children (48).

In the early days of the epidemic, knowledge of the progression of disease in children vertically-infected with HIV was based on those coming to the attention of health professionals through identification of symptomatic illness. Reported findings in under 200
hospital patients in the late 1980s implied extremely poor prognosis in affected children, with an expected survival time of three years following diagnosis (64) (Table 1.1). Based on follow-up of symptomatic children, Blanche et al described a bimodal pattern of disease expression in under 100 children with a minority of around one-quarter presenting with immune impairment and HIV-related clinical conditions at very young ages, distinct from the others who had less severe symptoms during their early years (40). Other studies reported similar patterns of early progression where for the rest, rates of disease progression varied throughout childhood (38;65;66). However, since identification of infected children was restricted to only those presenting with symptomatic disease (40;64), estimates were severely biased, likely resulting in artificially high mortality and morbidity rates.

Estimation from registers (48;67) is also potentially biased due to their retrospective component. It is possible that the asymptomatic children were underrepresented, producing a systematic overestimation of true mortality rates (68). Any studies other than those based on birth cohorts suffer from bias arising from the phenomenon of those most acutely affected dying prior to detection of infection.

From more reliable studies of cohorts followed from birth, it was realised that while most children present with some signs or symptoms of infection in the first six months, the majority of children survive their first year free of AIDS, and most children remain stable or initially improve thereafter (69). Even in the absence of widespread ART administration, progression was found to be slower than previously appreciated (39). However, follow-up at
this stage was not long enough for understanding of the impact of HIV-infection into later childhood and beyond.

In the mid-1990s, the Italian Register for HIV Infection in Children reported that around four out of five children expressed some HIV-related morbidity in the first year of life and almost all by age five years (67). The analysis of population- and hospital-based data on over 2,000 children by the American Pediatric Spectrum of Disease Project members around this time showed that durations of both asymptomatic and mild disease were short in relation to time spent with moderate or severe disease (38). Estimated AIDS incubation period was around five years, and from development of AIDS average survival time was around three years with mean survival times overall of less than 10 years. However, since this evidence was based on data from population- and hospital-based surveillance (38), derived estimates are likely to have been affected by bias. The bias due to exclusion of early deaths in the earlier era is apparent in the findings that most deaths in early-life occurred among those born in the later years of the study period (38). Subsequently, the progression rates to AIDS and other associated clinical symptoms are likely to be underestimated. The estimated average of three year survival from presentation of symptoms in a hospital-based cohort study of under 200 children (64) is likely to have been distorted in the same way.

Cohort studies enrolling children born to HIV-infected women from birth avoid such methodological problems. The joint analysis of data from the European Collaborative Study (ECS) and the French Paediatric HIV Infection Study Group in 1997 including almost 400 infected children showed that more than one third had progressed to serious disease (70), but
by six years of age, three-quarters of infected children survived. Death followed within one month of development of the serious illness in around 20% of those progressing. Similar findings on a smaller number of children were reported by investigators of the American Women and Infants Transmission Study Group, who described a one-year mortality rate of just over 5% (71). Researchers analysing data from such birth cohort studies have made use of the revised Centers for Disease Control (CDC) progressive classification system constructed for the surveillance of paediatric HIV infection (72) (see section 2.2.3) (38;70;71). As such, the system reflects lifetime experience, not necessarily actual health status at any given time, and thus does not capture the circumstances of children who improve clinically after an episode of serious disease. Although considered serious, many of the severe category illnesses are acute and treatable, with data now available on older children indicating that many who have previously experienced moderate or severe disease revert to a healthier condition (73). Increasing numbers of infected children are living much longer (74) and those reaching pubertal ages are largely asymptomatic (73), partly because of adequate ART and other supportive treatment and partly reflecting the natural history of vertically-acquired paediatric HIV infection. In using the progressive classification system, the apparent fluctuations in morbidity in these children over age are not being captured.

Since biases in selection of subjects in other types of study lead to distorted estimates of rates of serious disease progression and mortality, analysis of birth cohort study data offers more reliable inference. Only prospective studies which have been ongoing since the beginning of the epidemic offer the length of follow-up required to gain insight into patterns over the whole of childhood. As analysing data based on the monotonic Centers for Disease Control
(CDC) classification system overlooks the occurrence of recovery and sustained health of vertically-infected children it is of interest to take an approach to the assessment of the data which is dynamic and thorough. This is an opportunity to take a novel approach to the analysis of extensive clinical data from the large ECS cohort of children followed from birth into early adulthood. A flexible approach can produce a more complete and composite account of patterns of clinical disease progression throughout childhood in vertically-infected children, avoiding the problems inherent in previous studies.
Table 1.1: Features of studies addressing the progression of HIV disease through clinical stages

<table>
<thead>
<tr>
<th>Paper</th>
<th>Time period</th>
<th>Data source</th>
<th>#</th>
<th>Age range</th>
<th>Methods</th>
<th>Account for ART</th>
<th>Summary of findings</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(71)</td>
<td>1989-1995</td>
<td>American birth cohort</td>
<td>128</td>
<td>Birth - 5 yrs</td>
<td>Kaplan-Meier and Cox estimation of development of progressive CDC system categories</td>
<td>no</td>
<td>One-fifth progressed and 6% died in the first year</td>
<td>Enrollment in era of ART</td>
</tr>
<tr>
<td>(70)</td>
<td>1986-1994</td>
<td>2 European birth cohorts Combined</td>
<td>392</td>
<td>Birth - 6 yrs</td>
<td>Kaplan-Meier estimation of development of progressive CDC system categories.</td>
<td>no</td>
<td>A or B symptoms in first 4 years in most. 20% C in first year. Nearly 5% thereafter. Most commonly encephalopathy or opportunistic infection.</td>
<td>Reflects lifetime experience, not instantaneous health status</td>
</tr>
<tr>
<td>(67)</td>
<td>1980-1994</td>
<td>Population-based Italian Register</td>
<td>200</td>
<td>Birth - 5 yrs</td>
<td>Kaplan-Meier and Cox estimation of onset of each clinical symptom.</td>
<td>no</td>
<td>One-fifth remained asymptomatic by one year. Just over one in twenty by five years.</td>
<td>Register data</td>
</tr>
<tr>
<td>(38)</td>
<td>1982-1993</td>
<td>Population-and hospital-based surveillance</td>
<td>2148</td>
<td>Birth - early teens</td>
<td>Markov model estimating probabilities of progressing to the various CDC stages and mean stage times</td>
<td>no</td>
<td>Short periods of asymptomatic and mild disease, longer in moderate or severe disease status. Half of children progress to C and one quarter die by age 5 years.</td>
<td>Biases in inclusion. More likely to be symptomatic. Model assumptions violated</td>
</tr>
</tbody>
</table>
Table 1.1 *cont: Features of studies addressing the progression of HIV disease through clinical stages*

<table>
<thead>
<tr>
<th>Paper</th>
<th>Time period</th>
<th>Data source</th>
<th>#</th>
<th>Age range</th>
<th>Methods</th>
<th>Account for ART</th>
<th>Summary of findings</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(40)</td>
<td>1983-1988</td>
<td>Symptomatic children French cohort</td>
<td>94</td>
<td>Birth - 3 yrs</td>
<td>descriptive</td>
<td>NA*</td>
<td>Identified rapid and slow progression of clinical and biological symptoms</td>
<td>Restricted to Symptomatic children identified at first clinical event</td>
</tr>
<tr>
<td>(64)</td>
<td>1979-1987</td>
<td>Hospital-based cohort</td>
<td>172</td>
<td>Birth - 6.7 yrs</td>
<td>descriptive</td>
<td>NA</td>
<td>Just over three year survival time from diagnosis, with most becoming symptomatic within the first year</td>
<td>Not followed from birth – selection bias</td>
</tr>
</tbody>
</table>

* NA not applicable
# number of children
1.3.2 Viral load patterns over childhood

There have only been a limited number of studies on levels and patterns of HIV RNA over age in children vertically-infected with HIV, and mainly with few subjects and limited numbers of samples (75-79) (Table 1.2). Their results suggest that levels are low at birth but rise rapidly in the first few months (43) apparently peaking at levels in the region of $10^5$-$10^6$ (43;75;78;80;81) copies/ml at around two months of age, a range 10- to 100-fold greater than reported in adults following primary infection (79). Compared with the rapid viral containment exhibited by adults, where falls of 10- to 100-fold are seen during the month following peak (82), the subsequent decline in levels for children is gradual (75), and appears to decrease over age (43;78;81;83). Conflicting with this, in studies with small numbers of children, levels have been reported to stabilise in the region of $10^4$ copies/ml at around age 5/6 years (76;77).

Possible explanations for high levels of HIV RNA during the first year of life in children infected in utero or intrapartum have included the immaturity of immune system at the time of infection, and the elevated numbers of CD4$^+$ target cells (84) and their probable high replication rate (81). In this latter study, only measurements determined prior to administration of any ART were included (81), providing a largely natural account of the natural history of viral load in vertically-infected children. However, any exclusion of those receiving treatment may result in selection biases, and statistical adjustment made without loss of data provides a more favourable solution. Further, it may only have been necessary to account for measurements determined following HAART. Inaccuracies in estimates arising from such biases may conflict, with both particularly ill children with elevated viral levels preceding treatment and children with suppressed levels resulting from successful therapy.
being under-represented. The subsequent gradual decline near the end of the first year of life would provide evidence of delayed control of viral replication (81). It has been suggested that the decrease over age could be partly explained by increasingly fewer target cells being available for infection as the child gets older (85), although rates of decline in CD4\(^+\) cell counts do not match those in viral load (77).

In some studies on viral dynamics, HIV RNA level estimates have been determined cross-sectionally at specific ages (age-point inference) based on measurements taken over an interval, as opposed to being estimated continuously over age (43;45;75-78;81;83). While cross-sectional studies provide insight into the levels of HIV RNA in vertically-infected children at different ages (77), they do not effectively address the continual changes over age which can only be assessed in long-term cohort studies.

Where data have been collected longitudinally, findings have been based on the application of summary statistical methods such as rates of change of levels (43) or area under the curve (76;78) – empirical means of handling repeated measurements – which may be too crude to fully capture the viral dynamics as local variation is likely to be obscured. In particular, the true peak value and age at occurrence could be missed by such inference. Non-linear regression techniques, such as quadratic models used by some (79) are not adequately sensitive to local variation, and are thus also likely to miss the peak in values of HIV RNA. Further, as the data used in that particular study were obtained from a non-birth cohort (79), they may be non-representative in that enrolment was disposed to symptomatic children. As HIV RNA viral levels in such children are liable to be higher than in infected children in
general, this is likely to have lead to overestimation of levels. Another problematic feature is the use of data that are too sparse to adequately represent the peak in early life (75), and sometimes data are not available until beyond those important early months (44;77).

HIV RNA viral load data on relatively large numbers of children are available from clinical trials (44;80;83); however, there are particular methodological aspects which hinder the interpretation from this type of data. Follow-up tends to be limited to short durations within individual children with few, if any, followed from birth (44;80;83), potentially biasing the overall picture. Further, as inclusion criteria restrict enrolment to healthier children with only mild to moderate symptoms (44;83), those included are not necessarily representative of the population of children with vertically-acquired infection. Indeed, not every participating child has in fact acquired their infection perinatally (83), so the study groups are not necessarily homogeneous in terms of their physiological responses to HIV infection. Analysis has sometimes been confined to graphical presentation (80).

Assessment of correlated, repeated measurements data at different ages from the same children in longitudinal studies can pose methodological problems. Some researchers have applied the use of multiple significance tests of age-specific means (75) in which the fact that the value at one age is likely to influence successive ages is overlooked. In using such methods the possibility of, for instance, more frequent testing in particularly ill children is not accounted for, in which case HIV RNA viral measurements of these children, which are likely to be relatively high, contribute more to the assumed independent observations, possibly biasing the estimates for levels upwards. For this reason assumptions of
independence of measurements, on which multiple testing relies are not valid for longitudinal data and within-child correlation can only be addressed in analysis which accounts for repeated measurements.

Where analysis has involved formal inference such as linear mixed effects (LME) models, arbitrary methods to deal with measurements below the threshold of detection have been used (81;86), which may be inadequate. By the nature of the HIV RNA, low numbers of copies/ml which cannot be exactly quantified are censored at the lower threshold level for detection. (see section 2.2.5). In most studies of viral load in children (43;44;75-80;83;86), censoring has been disregarded and the true, unknown levels have been crudely taken to be the lower detection limit. With the true, undetectable levels of such censored measurements being lower, taking the limit as a surrogate value involves the contribution of systematically higher values to the estimates of the virus RNA levels, resulting in upward bias. Some investigators have set values at some arbitrary point below detection (81), but until now no formal account of the censored nature of quantified HIV RNA has been made in the assessment of data from paediatric studies. Recent studies in adults suggest that to adequately model RNA pattern over time, it is necessary to account for within-individual correlation, due to repeated measures, as well as censoring (11;39;72). The same is potentially true for data in children, and likely to be of increasing importance with HAART administration, for which levels tend to be undetectable.

A further critical aspect of previous studies is that many have involved few samples from small numbers of children (75-79) over relatively short age-spans (75;76;86), which do not
provide the consistent and continuous follow-up of a defined set of children over a long period of follow-up from birth required to elucidate the true dynamics of viral replication and containment.

Whether levels of HIV RNA in infected children vary by factors such as gender and race has not been fully addressed. A meta-analysis of gender and HIV RNA levels in adults indicated that on average women have lower plasma HIV RNA levels than men (87). However, findings may be distorted by differences in durations of HIV-infection (88), although in a few studies accounting for the time of infection or differences in baseline status, HIV RNA concentrations remained lower among women (89;90). In a study in adults addressing gender differences in levels in response to HAART, baseline levels of viral load were similar for men and women, but women achieved virological suppression at a slightly faster rate (91). In the only study on children to have previously looked at gender differences in HIV RNA viral load, investigators did so over the first two years of life in nearly 200 children (86). Results in this conference abstract indicated lower levels in girls, but were not significant, and the possibility of differences being dependent on age (which could be important even in these first two years) was not considered.

Since changes between and within HIV infected individuals in levels and patterns of viral burden can depend on their immunological status (77;83), associations with age should be considered in relation to CD4+ cell counts. There are reported differences in levels between HIV RNA assays (27;92), and these differences need to be regarded in analysis. Although zidovudine monotherapy has little impact on levels of HIV RNA (93), regimens of HAART
suppress viral activity (51-55). A robust analysis of HIV RNA over age must, therefore, take immunological parameters, quantifying assays and ART into account.

In conclusion, previous analysis of paediatric HIV RNA viral load data from birth and non-birth cohorts, cross-sectional studies and clinical trials have given an incomplete impression of levels and patterns over age. More reliable inference from longitudinal HIV RNA viral load measurements which are non-independent due to within-subject correlation and subject to left-censoring is possible. There remains a need for an approach to the modelling of HIV RNA patterns over age using sufficient data over a long range of follow-up from birth which tackles both of these aspects of repeated viral load data commonly overlooked. This has to be performed with due account of ART administered before and after clinical disease progression and quantifying assay used, while investigating the effects of factors such as gender on the patterns over age.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Time period</th>
<th>Data source</th>
<th># children (samples)</th>
<th>Maximum age range of follow-up</th>
<th>Formally accounted for repeat measures</th>
<th>Accounted for censoring</th>
<th>Accounted for ART</th>
<th>Formal Inference of viral load over age?</th>
<th>Summary of relevant findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>(86)</td>
<td>1989-2000</td>
<td>Prospective birth cohort</td>
<td>186 (not stated)</td>
<td>Birth - 2 years</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>Generalized estimating equations</td>
<td>Non-significantly lower RNA in girls throughout</td>
</tr>
<tr>
<td>(79)</td>
<td>1987-1998</td>
<td>Prospective data</td>
<td>32 (156)</td>
<td>Birth - 8 years</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>Quadratic regression</td>
<td>Peak in RNA at 1.5 - 3 months of age, reducing with age up to around 8 years, increasing thereafter</td>
</tr>
<tr>
<td>(81)</td>
<td>1986-1996</td>
<td>Prospective birth cohort</td>
<td>89 (313)</td>
<td>Birth - 3 years</td>
<td>yes</td>
<td>Set to just below detection level</td>
<td>Not administered prior to viral determinations used</td>
<td>Generalized estimating equations to join geometric means, LME</td>
<td>Peak in RNA around two months, decline thereafter</td>
</tr>
<tr>
<td>(78)</td>
<td>1988-1996</td>
<td>Birth cohort</td>
<td>31 (not stated)</td>
<td>Birth - 8 years</td>
<td>no</td>
<td>no</td>
<td>descriptive</td>
<td>no</td>
<td>Peak in RNA between birth and 14 weeks with subsequent decline</td>
</tr>
<tr>
<td>(76)</td>
<td>1987-1995</td>
<td>Sub-group of Prospective birth cohort</td>
<td>11 (69)</td>
<td>Birth - 1 year</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>t-tests, mean curve areas</td>
<td>Spectrum of RNA between individual children</td>
</tr>
<tr>
<td>Paper</td>
<td>Time period</td>
<td>Data source</td>
<td># children (samples)</td>
<td>Maximum age range of follow-up</td>
<td>Formally accounted for repeat measures</td>
<td>Accounted for censoring</td>
<td>Accounted for ART</td>
<td>Formal Inference of viral load over age?</td>
<td>Summary of relevant findings</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------------------</td>
<td>---------------------------------</td>
<td>----------------------------------------</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>---------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>(44)</td>
<td>1992-1995</td>
<td>ART Clinical trial</td>
<td>70 (not stated)</td>
<td>5 months - 12 years</td>
<td>yes</td>
<td>NA</td>
<td>Not administered</td>
<td>LME</td>
<td>Declines in RNA until a nadir at age 6 years</td>
</tr>
<tr>
<td>(43)</td>
<td>1990-1995</td>
<td>Birth cohort</td>
<td>106 (673)</td>
<td>Birth - 5 years</td>
<td>no</td>
<td>no</td>
<td>individual slopes</td>
<td>no</td>
<td>Peak at two months. Wide range of RNA at all ages.</td>
</tr>
<tr>
<td>(80)</td>
<td>1991-1994</td>
<td>ART Clinical trial</td>
<td>566 (not stated)</td>
<td>3 months - 18 years</td>
<td>no</td>
<td>no</td>
<td>Included only treatment naive children</td>
<td>None, presented medians, means and standard deviations</td>
<td>Decreasing decline in RNA from early life, stabilising around 8 years of age</td>
</tr>
<tr>
<td>(77)</td>
<td>1986-1993</td>
<td>Clinic records</td>
<td>48 (not stated)</td>
<td>1-9 years</td>
<td>yes</td>
<td>NA</td>
<td>no</td>
<td>Linear regression, Signed rank tests</td>
<td>Gradual reduction in RNA regardless of immunological status</td>
</tr>
<tr>
<td>(75)</td>
<td>1991-1993</td>
<td>Prospective birth cohort</td>
<td>14 (49)</td>
<td>Birth - 1 year</td>
<td>no</td>
<td>no</td>
<td>Not administered</td>
<td>Means of age groups</td>
<td>Sustained elevated RNA viral levels despite normal CD4^+ counts</td>
</tr>
<tr>
<td>(83)</td>
<td>1988-1991</td>
<td>IVIG Clinical trial</td>
<td>213 (1124)</td>
<td>1 month - 12 years</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
<td>LME, t-tests</td>
<td>Mean RNA decreased continually with age</td>
</tr>
</tbody>
</table>
1.3.3 Immunological patterns over childhood

As all children are born with immune systems which are not fully developed, the immunological response of those vertically-infected with HIV could differ substantially from that of individuals who acquire their infection by other means. Even when vertical transmission has not occurred, the developing immune system in children may be affected by maternal HIV infection. Since it is CD4$^+$ cells which are targeted by the virus, levels of these, along with those of the other immunological components: CD8$^+$ cells, total (absolute) lymphocytes and their inter-relations, have been used to determine the extent of progression of disease in infected individuals.

However, methodological differences between studies to date hinder interpretation of the limited information about the immunological patterns over age in children born to HIV infected mothers. Previous work indicates that in normal healthy children (94), as well as uninfected, perinatally HIV-exposed Western (95;96) and African children (97), there are age-related declines in absolute lymphocyte counts, CD4$^+$ cell counts and relative CD4$^+$ expressed as a percentage of absolute lymphocyte values (Table 1.3a). A single peak in CD4$^+$ cell counts and CD8$^+$ cell count and absolute lymphocyte in early life of uninfected children has been reported (98) but not explored in detail. Levels of immunological markers in infected children also depend on age, with observed decreases in CD4$^+$, CD4$^+$ percentage and lymphocytes over childhood (44;78;96;97) (Table 1.3b). The evolution of CD8$^+$ is unclear with some investigators suggesting levels increase from six months onward (96), but others finding decreasing CD8$^+$ values following initial increases, although findings on increasing relative CD8$^+$ are more consistent (95;96).
Lymphocyte levels at birth of infected and uninfected infants have been reported to be similar in an early study on 61 children (96) (Table 1.3c). However, more rapid declines (95) in absolute lymphocytes and absolute and relative values of CD4+ in vertically-infected children in both the developed (44;78;96) and the developing (97) world, lead to lower levels than those of uninfected children, whereas absolute and relative CD8+ counts tend to be higher (95-97). There are conflicting reports of the age at which differences by infection status in CD4+ cell counts become significant, ranging from three months (96;97), six months to one year (95). Relative CD4+ and CD4+:CD8+ ratio values have been seen to decrease within the first two years (95-97) but for these too there are contrasts in the ages at which relative and absolute levels begin to differ. Reported ages at departures in CD8+ levels of uninfected and infected children range from birth (97) to 12 months (96), and in CD8+ percentage from three to six months (95). In these studies, measurements have been taken at only particular ages (96), or have been grouped in age-wise blocks (95;97), approaches which can distort findings since correlation between measurements within children has not been taken into account. Although absolute lymphocytes have been seen to diverge from birth (97), the clearest differences have been reported in the relative values of CD4+ and CD8+ (96;97), however, a high degree of overlap between levels of all immunological markers in individual children has been noted.

In some of the studies of immune function in HIV-infected children thus far (95;96), the patterns over age have been secondary to the main focus of comparing levels in infected children with exposed but uninfected and unexposed groups of children over short durations of follow-up, with little focus on detail of the long-term dynamics. Many studies have
included relatively few children (78;95;96;99) or have involved samples which are not necessarily representative of the population of vertically-infected children (44;99;100). For instance, in accordance with study protocol, children enrolled in treatment trials (44) tend to be the healthier infected children, who are likely to be immunologically stronger, thus distorting estimates of levels and changes with age. It is probable that studies reporting analyses of data sourced from medical records, with bias arising from inclusion of only symptomatic children, have provided systematic underestimates of CD4⁺ cell count levels and over estimates of CD8⁺ cell count levels. Some have provided interesting graphical and descriptive accounts of patterns over age (78;96), but have not qualified these with sufficient statistical inference to enable robust conclusions. Where formal inference has been applied (97) this has often consisted of, for instance, some simple age-specific point estimates and ranges for median values. Although some studies have allowed for measurements taken on repeated visits within individual children, these have either been based on a relatively short period of follow-up (86) or have been based on a relatively healthy group of children, not all seen from birth. Although findings from previous studies have identified peaks in markers in early life in infected (101) and uninfected children (98), differences by HIV infection status and their exact nature have not fully been explored (102).

Findings relating to the uninfected adult population indicate that women have higher levels of CD4⁺ cell counts than men (103). In a study of the immunological parameters in just over 150 healthy pre-pubescent African American and Hispanic inner-city children, girls were found to have higher absolute lymphocyte levels and significantly higher CD4⁺ percentages (104). In a conference abstract of one of the few studies to address the effect of gender on immunological aspects of infected children, investigators found significantly higher relative
and absolute CD4⁺ cell counts in girls compared with boys during the first two years of life, but these did not lead to differentials in clinical progression rates (86). The dynamics of absolute values of CD4⁺ cell counts over time in untreated, infected American adults have been shown to depend on both gender and race, with women and non-whites (Latino and Afro-Americans) showing more rapid declines over time (90), although Black Americans are not necessarily comparable with individuals born to African women in Europe (105).

In one study on 126 American perinatally exposed uninfected children including fewer than 200 measurements in total, CD4⁺ cell counts did not differ significantly by ethnicity (106), and in an unexposed group, no differences were found between African-American and Latino children (104); gender differences in CD4⁺ percentage were reported. Findings from analysis of data on 750 adults comparing lymphocyte levels in HIV-infected European and West African populations indicate that levels of absolute lymphocyte and CD4⁺ cell counts are higher in West African adults for given relative CD4⁺ values (107).

However, these few studies addressing gender and ethnic differences in lymphocyte levels in uninfected children have presented ambiguous results. The study on unexposed inner-city children was confined to those between the ages of eight and 12 years. Being cross-sectional in nature, clear interpretation of patterns over age in these studies is difficult, and so far conclusive evidence of differences in immunological constitution of HIV-infected children by gender and ethnicity has not yet been presented. Although there is little sustained impact of zidovudine monotherapy on lymphocyte levels (95), delayed immunological reconstitution
occurs in response to potent HAART regimens (51) and this should be allowed for in assessment of the effects of factors on levels and patterns.

There is a need for the modelling of the various immunological markers over age to be carried out using reliable data, prospectively collected from birth, with the application of robust statistical methods which allow for repeated measurements. This will allow the appropriate analysis of how patterns as well as levels differ according to factors such as gender and race, accounting for treatment administration.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Time period</th>
<th>Data source</th>
<th># Children (samples)</th>
<th>Addressed</th>
<th>Maximum age range of follow-up</th>
<th>Methods</th>
<th>Account for RM?</th>
<th>Summary of findings</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(86)</td>
<td>1989-2000</td>
<td>Prospective birth American cohort</td>
<td>1584 (not given)</td>
<td>CD4(^+), CD4(^+)% and CD8(^+)% by gender</td>
<td>Birth - 2 years</td>
<td>Generalized estimating equations</td>
<td>yes</td>
<td>CD4(^+) and CD4(^+)% higher in girls. Genetically programmed non-hormonal mechanisms.</td>
<td>Did not allow the relationship to vary with age</td>
</tr>
<tr>
<td>(97)</td>
<td>1991-2000</td>
<td>Prospective Kenyan perinatally and postnatally infected cohort</td>
<td>398 (1418)</td>
<td>Comparison of lymphocyte subsets in African children</td>
<td>Birth - 10 years</td>
<td>Age-group specific medians and reference ranges</td>
<td>no</td>
<td>High variation in absolute as opposed to relative values. Evidence of peak in AL and CD8(^+) in the first year of life. Age-related declines otherwise.</td>
<td>Simply a presentation of age-specific point estimates and ranges for median</td>
</tr>
<tr>
<td>(95)</td>
<td>Pre-1992</td>
<td>Clinical trials unit cohort</td>
<td>70 (267)</td>
<td>Comparison of lymphocyte subsets over age values over age</td>
<td>Birth - 2 years</td>
<td>3-month block percentiles</td>
<td>no</td>
<td>Age-related declines in AL, CD4, and CD4(^+)% Fluctuating CD8(^+)% and CD4(^+):CD8(^+).</td>
<td>Main interest was in comparisons by infection status</td>
</tr>
<tr>
<td>(96)</td>
<td>Pre-1990</td>
<td>Italian infectious diseases clinic patients</td>
<td>34 (not stated)</td>
<td>Comparison of lymphocyte subsets</td>
<td>Birth - 2 years</td>
<td>Graphical presentation of age-specific means. status</td>
<td>no</td>
<td>Decreasing CD4(^+) and CD4(^+)%. Stable CD8, CD8(^+)% and CD4(^+)/CD8(^+) ratio.</td>
<td>Main interest was in comparisons between infected and seroreverters</td>
</tr>
<tr>
<td>Paper</td>
<td>Time period</td>
<td>Data source</td>
<td>Number of children (samples)</td>
<td>Addressed</td>
<td>Maximum age range of follow-up</td>
<td>Methods</td>
<td>Account for RM?</td>
<td>Account for ART?</td>
<td>Summary of findings</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>-----------</td>
<td>-------------------------------</td>
<td>---------</td>
<td>----------------</td>
<td>----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>(86)</td>
<td>1989-2000</td>
<td>Prospective birth American cohort</td>
<td>186 (not given)</td>
<td>CD4⁺, CD4⁺% and CD8⁺% by gender</td>
<td>Birth - 2 years</td>
<td>Generalized estimating equations</td>
<td>yes</td>
<td>no</td>
<td>CD4⁺ and CD4⁺% higher in infected (and uninfected) girls. Genetically programmed non-hormonal mechanisms.</td>
</tr>
<tr>
<td>(97)</td>
<td>1991-2000</td>
<td>Prospective Kenyan perinatally and post nataelly infected cohort</td>
<td>115 (409)</td>
<td>Reference AL, CD4⁺, CD8⁺, CD4⁺%, CD8⁺% values in African children</td>
<td>Birth - 10 years</td>
<td>Age-group specific medians and reference ranges</td>
<td>no</td>
<td>NA</td>
<td>High variation in absolute as opposed to relative values. Evidence of peak in AL and CD8⁺ in the first year of life. Age-related declines otherwise</td>
</tr>
<tr>
<td>(44)</td>
<td>1992-1995</td>
<td>European ART trial</td>
<td>70 (not given)</td>
<td>CD4⁺ over age in untreated children</td>
<td>5 months - 12 years</td>
<td>LMEs</td>
<td>yes</td>
<td>NA</td>
<td>Steady decline in CD4⁺ and CD4⁺% over age which was not significantly correlated with HIV RNA</td>
</tr>
</tbody>
</table>
### Table 1.3b cont: Features of previous studies on immunological markers in children vertically-infected with HIV

<table>
<thead>
<tr>
<th>Paper</th>
<th>Time period</th>
<th>Data source</th>
<th>Number of children (samples)</th>
<th>Addressed</th>
<th>Maximum age range of follow-up</th>
<th>Methods</th>
<th>Accounted for RM?</th>
<th>Accounted for ART?</th>
<th>Summary of findings</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(99)</td>
<td>Up to 1998</td>
<td>Medical centre follow-up</td>
<td>11 (not stated)</td>
<td>CD4(^+) cell counts in response to therapy</td>
<td>First year - 8 years</td>
<td>Assessed absolute decrease in levels following treatment</td>
<td>no</td>
<td>All treated</td>
<td>Initial increase of 200-1000 cells.</td>
<td>Very small numbers. Not all children vertically-infected.</td>
</tr>
<tr>
<td>(78)</td>
<td>1988 - 1996</td>
<td>Prospective American birth cohort</td>
<td>28 (~340)</td>
<td>CD4(^+) over age</td>
<td>Birth - 8 years</td>
<td>Graphical presentation of subset</td>
<td>no</td>
<td>yes</td>
<td>Initial decline which halted after treatment for some children</td>
<td>Merely descriptive account with no formal inference</td>
</tr>
<tr>
<td>(95)</td>
<td>Pre-1992</td>
<td>Clinical trials unit cohort</td>
<td>46 (175)</td>
<td>Comparison of lymphocyte subsets over age values over age</td>
<td>Birth - 2 years</td>
<td>3-month block percentiles</td>
<td>no</td>
<td>Comparisons of before and after values</td>
<td>Age-related declines in AL, CD4, CD4(^+)% and CD4(^+):CD8(^-). Increases in CD8(^+)% and steady CD8(^-).</td>
<td>Main interest was in comparisons by infection status.</td>
</tr>
<tr>
<td>(96)</td>
<td>Pre-1990</td>
<td>Italian infectious diseases clinic patients</td>
<td>14 (not stated)</td>
<td>Comparison of lymphocyte subsets in African children</td>
<td>Birth - 2 years</td>
<td>Graphical presentation of age-specific means</td>
<td>no</td>
<td>NA</td>
<td>Decreasing CD4(^+) and CD4(^-)% and CD8(^+) and CD8(^-)% and decreasing CD4(^+)/CD8(^-) ratio</td>
<td>Main interest was in comparisons between infected and seroreverters.</td>
</tr>
</tbody>
</table>
Table 1.3c: Features of previous studies on comparisons of immunological markers in children born to women infected with HIV by infection status

<table>
<thead>
<tr>
<th>Paper</th>
<th>Time period</th>
<th>Data source</th>
<th># Children (samples)</th>
<th>Addressed</th>
<th>Maximum age range of follow-up</th>
<th>Methods</th>
<th>Account for RM?</th>
<th>Summary of findings</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(97)</td>
<td>1991-2000</td>
<td>Prospective Kenyan perinatally and postnatally infected cohort</td>
<td>398 (1418) uninfected and 115 (409) infected</td>
<td>Comparision of lymphocyte subsets in African children</td>
<td>Birth - 10 years</td>
<td>Age-group specific medians and reference ranges</td>
<td>no</td>
<td>Differences in lymphocyte subset percentages by 3 months of age</td>
<td>Simply a presentation of age-specific point estimates and ranges for median.</td>
</tr>
<tr>
<td>(95)</td>
<td>Pre-1992</td>
<td>Clinical trials unit cohort</td>
<td>70 (267) and 46 (175)</td>
<td>Comparision of lymphocyte subsets over age</td>
<td>Birth - 2 years</td>
<td>3-month block percentiles</td>
<td>no</td>
<td>Significant differences in CD4⁺ and CD8⁺% from 3 - 6 months; and in CD4⁺% and CD4⁺:CD8⁺ throughout. More rapid declines in CD4⁺ and CD4⁺:CD8⁺ in infected children. Similar CD8⁺ in both.</td>
<td>No formal inference for comparisons of measurements over time</td>
</tr>
<tr>
<td>(96)</td>
<td>Pre-1990</td>
<td>Italian infectious diseases clinic patients</td>
<td>34 (not stated) and 14 (not stated)</td>
<td>Comparision of lymphocyte subsets in African children</td>
<td>Birth - 2 years</td>
<td>Graphical presentation age-specific means. t-tests of levels by infection status</td>
<td>no</td>
<td>No differences at birth. Differences in CD4⁺ at 3 and 24 months. CD8⁺ levels increase from six months. Decreasing CD4⁺, CD4⁺%, steady CD8⁺, CD4⁺:CD8⁺, increasing CD8⁺%</td>
<td>No formal inference for comparisons of measurements over time</td>
</tr>
</tbody>
</table>
1.3.4 Determinants of the rate of clinical disease progression during childhood

Various clinical and biological aspects of HIV-infected individuals have been identified as markers useful in the prognosis of their outcomes of disease. Many of these factors have been found to be applicable to paediatric infection, in particular, but the complex associations have still to be reliably quantified.

Based on follow-up of under 100 symptomatic children, Blanche et al reported correlations of low first CD4+ cell counts with subsequent severe illness (40) (Table 1.4). However, restriction to children with symptoms leads to overestimation of disease severity and thus results cannot be generalised to all children with HIV infection. This selection process could potentially distort estimates of predictive effect of markers in either direction. Associations of low absolute and relative CD4+ cell counts and progression to serious disease have been borne out by findings in a number of studies (45;71;80;83;108;109). However, in studies based on biological data from clinical trials (45;80;83;110), measurements have often been ascertained within a relatively short time-period prior to last follow-up occasion. As levels of biological markers are age-related, it would be necessary to stratify by age when making use of baseline values at enrolment into clinical trials, which has been addressed (80) but not universally (45). Further, studies based on clinical trials data have assessed the predictive ability of baseline data at entry, but measurements have been taken across a range of ages in childhood, very few of which have been quantified very early in life (45;80;83). More reliable work from some birth cohort studies has focused on just a subset of the lymphocyte constituents, and follow-up is over relatively young ages (71;108).
The strong relationship of elevated HIV RNA viral load with risk of serious illness and death has been clearly established (43;78;81;108;110-112) and in line with findings from adult studies (89) indicates that the extent of decline after initial peak during seroconversion may be related to disease progression rate. Immunological and virological determinations have been found to have independent relationships with clinical outcome (45;80;83;108;109). Other reported determining factors positively associated with progression in children include prematurity (113), maternal ART use in pregnancy (112-114), in utero transmission (78) and advanced maternal disease status (71;109;115;116); development of anaemia has been linked to HIV-related deterioration in adults (117). There is also suggestion of elements of progression being attributable to other determinants, such as genetic predisposition in terms of human leukocyte antigen (HLA) haplotype (118) and chemokine receptor genotype (119) and virus phenotype (120;121). No associations of maternal or child CD8+ levels with progression in vertically-infected individuals have been found (109). In a few studies (113;114), children exposed to prophylactic zidovudine were more likely to progress rapidly than unexposed children. This would suggest the effect of zidovudine being preferentially preventative of intrapartum transmission as opposed to in utero transmission, since early vertical transmission is associated with rapid disease progression (113).

As serious disease progression during early life takes hold concurrently with the process of immune system establishment, it is probable that there are separate mechanisms acting on rapid progression to serious disease or death and long-term progression. Thus distinguishing factors are likely to influence progression at the different phases. Although HIV RNA viral load predicts progression at any age (43;78;81;108;110-112), findings indicate distinctions in
other markers for rapid and overall progression (108;109;113). Both relative (109) and absolute CD4$^+$ cell count (108;122) values have been associated with immediate, as opposed to long-term risk, but the prognostic value of total lymphocyte count has not been assessed in this staged way. Early clinical progression to moderate HIV-related illness has been linked with rapid (109;122) and longer-term progression (108;109), although persistence of this effect beyond 18 months was not found in a study on 165 children (108). Reports suggest prematurity (113) influences rapid progression to serious disease but whether this also governed long-term prognosis was not explored. There are methodological drawbacks of the analysis of the study assessing numerous potentially prognostic maternal and infant factors (109). The outcome of clinical progression may have preceded the “prognostic” laboratory determinations in some children. Further, the statistical methods used involved only logistic models, predicting presence or absence of progression as a binary response, not formal survival analysis which is required to account for the time-to-progression and to accommodate censoring of individuals with incomplete follow-up.

There is ongoing debate about the role of gender in determining disease progression in infected individuals (123;124), though most researchers have concluded similar progression rates for males and females; findings in vertically-infected children to date indicate no gender differentials (125;126). Neither have differences in rates of progression by race been found in children (125;126). A separate issue is whether the predictive value of laboratory levels or clinical markers differs for boys and girls (86) or among ethnic groups (107) and this has not previously been addressed in any depth.
There has thus been a lack of comprehensive information on predictors of disease progression from birth into late childhood. Rigorous methodological assessment of the prognostic value of various laboratory and clinical factors from early life using data truly representative of disease evolution in vertically-acquired infection has yet to be performed. A thorough investigation will tackle the determinants of both rapid and long-term progression, as well as progression overall, in such a way.
<table>
<thead>
<tr>
<th>Paper</th>
<th>Time period</th>
<th>Data source</th>
<th># children</th>
<th>Age range</th>
<th>Factors addressed</th>
<th>Methods</th>
<th>Account ART</th>
<th>Findings</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(113)</td>
<td>1990-1997</td>
<td>Retrospective hospital records data</td>
<td>51</td>
<td>First 2 years</td>
<td>zidovudine prophylaxis and prematurity</td>
<td>Kaplan-Meier and Cox estimation of development of progressive CDC system categories</td>
<td>no</td>
<td>Higher risk of progression in premature infants and those born to treated mothers</td>
<td>Data source Small numbers</td>
</tr>
<tr>
<td>(45)</td>
<td>1988-1996</td>
<td>American IVIG trial</td>
<td>218</td>
<td>1 month - 8 years</td>
<td>Associations of p24 antibody, RNA and CD4+ with mortality</td>
<td>$\chi^2$, Kaplan-Meier, Cox regression</td>
<td>yes</td>
<td>p24 antibody, RNA and CD4+ independently associated with mortality</td>
<td>Baseline values at enrolment into trial. Categories not age-related.</td>
</tr>
<tr>
<td>(122)</td>
<td>1986-1995</td>
<td>French birth cohort</td>
<td>267</td>
<td>Birth - 12 months</td>
<td>Associations of clinical and laboratory measures at birth with serious disease progression within first year of life</td>
<td>Cox regression</td>
<td></td>
<td>Hepatomegaly, splenomegaly, and/or lymphoadenopathy low CD4+ cell count and virus positivity associated with rapidly progressive disease</td>
<td>As clinical signs were categorised together it was not possible to differentiate individual prognostic values</td>
</tr>
<tr>
<td>Paper</td>
<td>Time period</td>
<td>Data source</td>
<td># child-ren</td>
<td>Age range</td>
<td>Factors addressed</td>
<td>Methods</td>
<td>Account ART</td>
<td>Findings</td>
<td>Limitations</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------------</td>
<td>---------</td>
<td>--------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>(109)</td>
<td>1989 - 1996</td>
<td>American birth cohort</td>
<td>122</td>
<td>Birth - 18 months</td>
<td>Maternal and infant HIV RNA and immunological parameters, and infant clinical status for progression at 6 and 18 months of age</td>
<td>Logistic regression</td>
<td>no</td>
<td>6 month progression: low maternal CD4(^+) and low CD4(^+)% lymphadenopathy, hepatomegaly and splenomegaly. 18 month progression: moderate disease by 6 months and high RNA in infant</td>
<td>Statistical analysis accounted only for occurrence of serious disease progression, not timing. Progression may have taken place before measurements in some children</td>
</tr>
<tr>
<td>(71)</td>
<td>1989 - 1995</td>
<td>American birth cohort</td>
<td>128</td>
<td>Birth - 5 years</td>
<td>Relationship of early clinical signs and immunosuppression</td>
<td>Kaplan-Meier and Cox estimation</td>
<td>no</td>
<td>Immuno-suppression and early clinical signs associated with progression to serious disease</td>
<td>ART not adjusted for, despite enrolment in era of ART</td>
</tr>
<tr>
<td>(80)</td>
<td>1991 - 1994</td>
<td>ART Clinical trial</td>
<td>566</td>
<td>3 months - 18 years</td>
<td>Prognostic value of baseline values of HIV RNA and CD4(^+)</td>
<td>Cox stratified by age</td>
<td>NA</td>
<td>Decline in progression-free survival with increasing baseline RNA and low CD4(^+) count</td>
<td>Children enrolled at different ages. Grouping by age unlikely to capture all of the age-dependence of either marker. Bias toward healthier children</td>
</tr>
<tr>
<td>Paper</td>
<td>Time period</td>
<td>Data source</td>
<td># child-REN</td>
<td>Age range</td>
<td>Factors addressed</td>
<td>Methods</td>
<td>Account ART</td>
<td>Findings</td>
<td>Limitations</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-----------</td>
<td>------------------</td>
<td>---------</td>
<td>-------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>(108)</td>
<td>1989-1993</td>
<td>American birth cohort</td>
<td>165</td>
<td>Birth - 5 years</td>
<td>HIV RNA, CD4(^+) and clinical class for predicting disease progression</td>
<td>Kaplan-Meier and Cox</td>
<td>no</td>
<td>RNA independent of immunological and clinical status in its association with progression. CD4(^+) levels relate to current risk but not long-term risk.</td>
<td>Did not adjust for ART. Did not confine attention to first C, so effects could be confused</td>
</tr>
<tr>
<td>(83)</td>
<td>1988-1991</td>
<td>American IVIG trial</td>
<td>254</td>
<td>1 month - 12 years</td>
<td>Association of HIV RNA, CD4(^+) % and mortality</td>
<td>Baseline HIV RNA, CD4(^+) % and change in RNA. (\chi^2). Kaplan-Meier, Cox</td>
<td>yes</td>
<td>Independent association of RNA and CD4(^+)% with death</td>
<td>Potential bias from inclusion of only children with mild to moderate symptoms. Not all vertically-infected children, not systematically seen from birth.</td>
</tr>
<tr>
<td>(40)</td>
<td>1983-1988</td>
<td>Symptomatic children French cohort</td>
<td>94</td>
<td>Early-life - 7 years</td>
<td>Initial CD4(^+) related to life-threatening events</td>
<td>Compared proportions</td>
<td>NA</td>
<td>Low first CD4(^+) associated with encephalopathy and opportunistic infection</td>
<td>Restricted to symptomatic children identified at first clinical event</td>
</tr>
</tbody>
</table>
Chapter 2 Data source and methods

Understanding of the natural evolution of HIV in vertically-infected children from birth will aid in the elucidation of the pathogenesis of paediatric HIV infection, informing the specific prognosis and therapeutic interventions for children. Families, designers of clinical trials, health-care planners and clinicians in charge of the treatment management of HIV infected children would benefit from a more thorough knowledge of the long-term consequences of the disease: Information can be used by clinicians when advising families about the timing and nature of expected clinical outcomes; aspects of the evolution of disease inform the design and implementation of treatment trials, and reliable inference is a necessary base for decisions on resource provision for paediatric HIV care.

Since the development of HIV-infection is reflected in the clinical, virological and immunological processes, it is important to have a thorough understanding of these aspects of progression of disease in children as well as how they inter-relate. A thorough account of patterns of HIV-related signs and symptoms is required to provide a reliable picture of clinical disease progression. To clarify the viral dynamics in paediatric infection it is important to elucidate the pattern of HIV RNA viral load over age. Of immunological interest are the patterns of CD4\(^+\), CD8\(^+\) and absolute lymphocyte counts as well as relative CD4\(^+\) percentage. Identifying which characteristics are associated with risk of disease progression enables focus on effective intervention for those at increased risk to improve an otherwise poor long-term prognosis.
2.1 Aim and Objectives

The aim of this work is thus to elucidate patterns of clinical progression, and virological and immunological dynamics over age in vertically HIV infected children. Further, investigations are intended to explore the early-life clinical and biological circumstances related to disease progression.

The objectives are, through the use of appropriate statistical methods:

to clarify clinical patterns of disease progression;

to elucidate HIV RNA viral load over age;

to investigate immunological markers over age;

and to address the associations of biological and clinical parameters with progression of disease.
2.2 The European Collaborative Study

Data were available from the prospective ECS, on children born from 1984. Children born to HIV infected mothers are followed according to the standard ECS protocol (Appendix 2.1), with detailed clinical and laboratory information collected regularly by clinicians in paediatric referral centres in major hospitals in Padua, Berlin, Edinburgh, Madrid, Valencia, Amsterdam, Stockholm, Brussels, Genoa and Barcelona (Figure 2.1 and Appendix 2.2). Children are seen at birth, around three and six weeks, 3, 4.5 and six months and then at 3-monthly intervals until 24 months. Subsequently, infected children are seen at least twice a year, uninfected children once a year, and clinical and laboratory information and current treatment recorded on standard forms. Data are also recorded when children make visits, because of symptoms, outwith the protocol schedule. Parental consent is obtained before enrolment in the ECS, and the study is approved by the local ethics committees.

Figure 2.1 Location of Paediatric centres of the ECS
Results presented in this thesis are based on ECS data available at the time of each analysis. The last of the four analyses used data collected before September 2002. By this time, 190 (10.2%) infected and 1666 (89.8%) uninfected children had been enrolled in ECS paediatric centres (Figure 2.2).

**Figure 2.2: ECS children by year of birth and infection status**

![Graph showing ECS children by year of birth and infection status](image)

2.2.1 Role of the researcher

As this work was carried out as part of the ongoing study, which I joined in 1999 as the ECS statistician, it is appropriate to clarify my role in the research presented in this thesis. The data collection was the responsibility of local clinicians, and data verification, entry and database management was carried out by epidemiologists at the ECS coordinating centre at the Institute of Child Health, London. I was given the responsibility for the analysis of the data relating to
infected children in particular, with the aim of clarifying the patterns of disease progression in vertically-infected children in Europe.

The direction and nature of the analyses were agreed in consultation with my supervisor, Professor Marie-Louise Newell, and co-supervisor Mario Cortina-Borja. The strategic and technical development of all analyses presented here were entirely my responsibility. The work has resulted in five manuscripts for publication in peer-reviewed journals (Appendix 2.3). I was involved in the preparation of all five manuscripts, as part of the ECS team.

2.2.2 Diagnosis of HIV
Until the development of polymerase chain reaction (PCR) and virus culture tests, detection of HIV-infection in newborn children of infected mothers was based on anti-HIV immunoglobulin-G antibodies. As maternal antibodies cross the placenta and remaining in the newborn system for up to 18 months of life or more, antibody testing is not specific in these infants, and diagnosis by PCR and virus culture is the informative test (72). A child is regarded as HIV-infected if it has at least two positive tests or meets the criteria for AIDS. Children are considered uninfected with HIV if they test negative for HIV-antibodies at least twice at 6 to 18 months of age, there is no other laboratory indication of infection and they do not meet the AIDS diagnosis criteria (127).

2.2.3 Paediatric classification system
The CDC developed the Public Health Service classification system for HIV infection for the monitoring of health service needs. The system for children less than 13 years of age was revised in 1994 (72) in response to knowledge gained on the development of the clinical symptoms and
immunological status of HIV in children. The refinement of the classification provides an easy to use system comprised of mutually exclusive categories which reflects the stage of the disease for HIV-infected children. In clinical practice, the classification is progressive in that once a child is classified they are not reclassified into less severe categories even if there is a subsequent clinical or immunological improvement.

The CDC classification system for paediatric HIV infection has proved a useful framework for analysis of clinical information on infected children. There are four clinical categories based on signs and symptoms and three immunological categories which classify children according to the severity of their immunosuppression attributable to HIV infection (72). The two category structures represent the progressive staging of the disease (Appendix 2.4 A and B). The four categories for clinical manifestations in infected children are: N (asymptomatic), A (mildly symptomatic), B (moderately severe symptoms, including lymphoid interstitial pneumonitis (LIP), and C (severe symptoms) according to the CDC (72) (Appendix 2.4 A). Children who died with HIV-related disease were classified in a separate category. Normal (category 1), and moderate (2), and severe (3) immune suppression was assessed by CD4+ cell counts or percentages appropriate for age at the time of assessment (Appendix 2.4 B).

2.2.4 Treatment
Practice of ART administration for the prevention of HIV-related disease progression in the ECS has varied between centres and among children, and has not always been consistent with policy (58). Policies on initiation and change of therapy have varied from routine administration upon
confirmation of infection, to decisions being determined by the specific clinical, immunological and virological circumstances of the individual child.

As of the last analysis, 67 (35.3%) remained ART-naïve, 28 (14.7%) had been administered monotherapy only, 14 (7.4%) double therapy and 81 (42.6%) had been treated with a combination of three or more ART drugs, at some point during follow-up. The proportion of children treated is different for all analyses presented, and breakdown is given in the corresponding results chapters. Different categorisations of ART have been used in different analyses, depending on the context and data constraints, and are outlined in the relevant methods sections. Generally, combinations of two or more therapies have been categorised together, owing to the restricted numbers of observations. These treatment categorisations always excluded ART given neonatally as prophylaxis for prevention of vertical transmission.

In some analyses, children were also categorised according to receipt, or not, of PCP prophylaxis and IVIG treatment.

2.2.5 HIV RNA viral load
HIV RNA viral load is quantified by the number of copies per millilitre (copies/ml) of blood plasma. In the ECS, laboratory tests including HIV RNA viral load, are carried out locally according to standard procedures. HIV RNA viral load is measured by quantitative reverse transcriptase PCR assay (AmpliCor Monitor, Roche Diagnostic Systems, Branchburg, New Jersey, USA) or Nucleic Acid Sequence-Based Amplification (NASBA) (bioMérieux, Durham, USA). Values are conventionally log_{10} transformed to resolve non-Normality and heteroscedasticity.
An aspect of assay systems used in the quantification of RNA viral load is that they measure values above particular cut-off values for detection, below which the assays used are not sufficiently sensitive; in this way measurements are potentially left-censored. This problem is more serious for data ascertained using older generations of assay systems which are less accurate due to higher detection cut-off points in the region of 2000-4000 copies/ml. Assays in use in laboratories in the developed world today have lower limits in the region of just 50-100 copies/ml.

2.2.6 Immunological markers
In the ECS, standard, locally carried out laboratory tests for CD4⁺, CD8⁺ cell count and absolute lymphocyte measurements are based on flow cytometry (FACScan) with Becton-Dickinson antibodies. Venous blood specimens are anti-coagulated using ethylene diamine tetra acid or heparin and processed and an automated haemocytometer is used to obtain total, or absolute, lymphocyte counts and the subtypes of white blood cell. Counts are expressed in units of 10⁶ cells per litre (x 10⁶ cells/l). As for HIV RNA viral load, analyses of CD4⁺, CD8⁺ cell counts and absolute lymphocyte measurements were log₁₀ transformed throughout as it was successful in Normalising residuals and resolving heteroscedasticity, and was thus appropriate for the application of linear models. CD4⁺ cell counts measurements were also expressed as a percentage of absolute lymphocyte count.

2.2.7 General definitions
Black children in the ECS cohort are almost all born to mothers from sub-Saharan Africa. As most of the enrolled children are born to Caucasian or black African women, investigation of differences by race was restricted to comparison of data on children born to white mothers to
data on children born to black mothers. With the exception of the clinical patterns investigation, stratification by gender was carried out in the analyses. Children were classified as severely premature if born at or before 34 weeks of gestation, as moderately premature between 34 and 37 weeks and as full-term beyond 37 weeks. Children were categorised according to receipt, or not, of zidovudine (or other ART) neonatally as prophylaxis for the prevention of vertical transmission. Maternal CD4⁺ cell count measurement were categorised above or below $500 \times 10^6$ cells/l; the dichotomy maximised the power of the assessment of the effect.

2.3 Clinical patterns analysis
The methods used for the analyses of clinical patterns of disease progression presented in Chapter 3 are described here.

2.3.1 Overall disease progression
Standard Kaplan-Meier time-to-event methods (128) were used in the analysis of clinical and immunological symptoms. Kaplan-Meier product-limit estimates and confidence intervals (CIs) were obtained for age at progression to each of the CDC categories A (or worse), B (or worse), C (or death), and to death. Children were censored at the age when last seen. Kaplan-Meier analysis was also performed for progression to immunological categories 2 and 3.

2.3.2 Progression to serious disease stratified by ART
Patterns of disease are described for children grouped according to therapeutic policy recommendations. Children were considered in three separate cohorts which were defined a priori according to treatment policy at the time of birth: Cohort 1 (1985 – 1988): no recommendation for treatment; Cohort 2 (1989 – 1994): treatment policy restricted to mono-therapy for symptomatic children and Cohort 3 (1995 – 1999): initiation of combination
therapies recommended at an early stage (48). Progression to serious disease was assessed by Kaplan-Meier estimation for each of these three cohorts separately, and compared using the log-rank test (128).

The patterns of disease are also described separately for treated and untreated children: disregarding vertical transmission prophylaxis, children were identified as ever or never having received any ART, irrespective of timing of initiation; progression to CDC categories A, B, C, and death was examined separately for the two groups. Comparisons between progression in treated and untreated children were made using the log rank test (128). Similar analyses were conducted according to receipt of prophylactic ART in the first weeks of life (129).

The relative risk (RR) of serious disease progression by ART, PCP prophylaxis and IVIG treatment administered prior to development was assessed by time-dependent Cox proportional hazards regression (130). A similar analysis was performed to assess the effect of ART on the risk of progression to severe immune deficiency (category 3).

2.3.3 Changing presence of HIV-related symptoms
Allocation of CDC categories at each visit based on information regarding current clinical symptoms and signs (72) enables dynamic reclassification with changing clinical status, providing the possibility of children being moved to a less serious category if their clinical or immunological status improves. Cross-sectional analysis was performed using the CDC classification allocated on the basis of symptoms or signs present at each visit. At six monthly intervals from birth, the numbers of children in each of the CDC classes at that time were ascertained. The intermittency of HIV related symptoms was also assessed through a look-back
exercise for children in follow-up at five years of age, providing an example of the clinical and immunological history of children up to that age.

2.4 HIV RNA viral load analysis
Chapter 4 presents the results of the analyses addressing HIV RNA viral load over age in considerable methodological detail. The approach to modelling the HIV RNA viral load data is described briefly here, with components expanded in the following sections. As well as the non-linear pattern of HIV RNA viral load in vertically-infected children over age, further challenges in description are posed by the repeated measures and left-censored nature of longitudinal data (see section 2.2.5). Thus the modelling of viral patterns over age requires the use of sophisticated modelling techniques. Fractional polynomials, conventional polynomials and a change-point procedure (131) were compared in the modelling of the basic pattern of viral load over age. Mixed-effects models account for the within-child correlation between repeated measurements; Hughes' method (132) is a particular form of such a model, developed to accommodate censoring. The alternative is the standard LME model with midway censoring. Since the overall pattern of RNA viral load over age is not likely to be affected substantially by the censoring pattern or the estimation procedures, a two-step approach was followed as shown in Figure 2.3. Initially treating the data as independent uncensored observations, the basic model for the overall pattern of viral load over age was established. This chosen model's parameters were estimated within the LME with censoring framework allowing both repeated measures and censoring.
Figure 2.3: Two-step procedure for modelling longitudinal HIV-RNA viral load measurements as a function of age, accommodating repeated measures and censoring scheme.

<table>
<thead>
<tr>
<th>Identifying the basic model for the overall pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Fractional polynomial</td>
</tr>
<tr>
<td>- (Conventional polynomial)</td>
</tr>
<tr>
<td>- (Change-point)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accounting for repeated measurements and censoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Hughes' method</td>
</tr>
<tr>
<td>- (Linear mixed-effects with mid-point censoring cut-off value)</td>
</tr>
</tbody>
</table>

For these analyses, monotherapy was categorised with no therapy since it is unlikely to be associated with a substantial reduction in viral load, or increase in CD4+ cell count or delayed clinical progression (44;93). Children were thus categorised as either treatment naïve / receiving monotherapy or treated with a combination of two or more antiretroviral drugs. As assay system and type of ART could vary over age, both factors were introduced in a time-dependent manner (130). Any measurements taken on a child’s day of birth were set to day one to overcome numerical indeterminations (131).

2.4.1 Establishing the basic model for the overall pattern of viral load over age
Fractional polynomials encompass a group of curves, intermediate between conventional polynomials and curves non-linear in their parameters, with power terms usually restricted to a predetermined set of integer and non-integer values (131). The advantages of fractional polynomial models lie in their ability to reveal the “true” curve shape and in the possibilities for
modelling asymptotes, using a small number of parameters. The formal details on fractional polynomials are given in Appendix 2.5.

To ensure use of optimal methods for modelling HIV RNA viral load over age, alternatives to fractional polynomials were considered. These were conventional polynomial and change-point models, as shown in Figure 2.3. Within the framework of conventional polynomials, models involving terms up to quartic powers were considered to be as elaborate as necessary without loss of parsimony. Change-point analysis involves the fitting of a segmented regression model with an intercept and slope for the initial increase to peak in RNA viral load and another slope for the decline. The formal details on change-point models are given in Appendix 2.6.

In comparing methods for establishing the linear predictor of responses over age, super smoothers provide visual guides for model selection (133). The most suitable of the three approaches was identified by graphical comparison to the super smoother and by evaluation of the Akaike information criterion (AIC) (134). The model with the lowest AIC is considered the best model.

2.4.2 Accounting for repeated measurements and censoring in HIV RNA viral load
The repeated-measures nature of the censored HIV RNA viral load data requires models which can account for within-child correlation and left censoring of measurements. Mixed effects models comprise fixed effects which represent the population components and random effects for variations at the individual level.
**Hughes' modified Monte Carlo expectation-maximisation algorithm**
The methods described by Hughes (132) offer a modification of the usual expectation-maximisation (EM) estimation procedure for fitting mixed effects models with Normal errors (135,136), by accommodation of such censored observations arising from lower (as well as upper) detection limits. Hughes makes use of a general solution that works with any censoring scheme with arbitrarily complex design matrices for both the fixed and random effects, using the Gibbs sampler, a Monte Carlo procedure, to maximise the likelihood function (132). The procedure for calculating the likelihood function described by Jacqmin-Gadda et al (137), for methods which include Hughes' as a particular case, was used. The methodology of Hughes' is outlined in Appendix 2.7.

**Alternative method for accounting for the repeated measurements and censoring of HIV RNA viral load**
The performance of another, less complex estimation procedure for accounting for repeated measures and censoring, was compared to that of Hughes' method. This alternative approach involved an LME model in which censoring is dealt with informally by replacing the assay's cut-off by half its value.

Comparison of Hughes' and LME with mid-way censored values was based on graphical comparison of fitted prediction curves to those created from Hughes' method. Bootstrap resampling was used to construct 95% confidence bands (138).
2.4.3 Model selection strategy
To allow for other factors, the LME model was expanded from that involving only the age terms, through introduction of stratification by gender, treatment and CD4$^+$ cell count separately, accounting for assay type to assess the significance of these factors. The inclusion of a main effect of an explanatory factor in a model imposes a parallel effect and testing its significance informs whether there are differences which are constant over age. Interaction terms allow testing of differences which depend on age, with significance indicating that the effect of a factor differs by age, signifying differential patterns over age. The factors were assessed for significance of interaction with the age terms and, where appropriate, with one another. Significance of individual terms was assessed by comparing nested models using likelihood ratio tests. There were sufficient data to expand the model as far as three-way interaction. Wald tests inform the significance of specific terms within individual models.

2.5 Patterns of immunological markers
In the analysis of immunological markers over age in uninfected as well as infected children presented in Chapter 5, measurements were truncated at 12 years of age for purposes of model stability.

2.5.1 Modelling immunological markers over age
Natural cubic splines were used to model the patterns of CD4$^+$ cell counts, CD8$^+$ cell counts, absolute lymphocytes and CD4$^+$ percentage over age. Identifying the appropriate model for each marker involved comparison of the super smoother (133) of the underlying location of the data over age with the candidate spline model. Since they are adequately flexible to represent local variation, natural cubic spline models (139) which best reflected the structure of the mean
function were chosen. The repeated measurement nature of the data was allowed for in LME (136) with a general form for the random effects covariance of the intercepts.

Natural cubic splines

A cubic polynomial is a linear expression which has terms up to power three. A cubic spline is comprised of piecewise cubic polynomials joined together at points known as knots (140). For natural cubic splines, intervals need not be equally spaced. This flexibility allows the placement of knots to best represent the structure of the data. As these models are non-parametric, the magnitudes of their coefficients cannot be interpreted in a way that is meaningful in the conventional quantitative sense. The formal definition for natural cubic splines is given in Appendix 2.8.

2.5.2 Assessing factors influencing immunological markers

The effects of gender and maternal race on the levels and patterns of markers over age were investigated. Interactions between the two were also examined for the possibility of, differences by gender being dependent on race (or, equivalently, differences by race being dependent on gender). Differences in CD4\(^+\) cell count of uninfected children by prematurity and maternal CD4\(^+\) cell count at delivery were also assessed. To account for any confounding effect of ART administration in infected children, analyses were also performed with adjustment by combination of two or more ARTs.

Significance of the main effects and interactions of gender, maternal race, gestational age and maternal CD4\(^+\) cell count at delivery was assessed using likelihood ratio tests. Additionally, original and fitted values were visually examined to aid the assessment of differences in patterns.
2.5.3 Comparison of infected and uninfected children

To establish the extent of overlap by infection status, immunological data on uninfected children were compared to those of infected children. As the methodology of splines does not lend itself to the construction of inference based age-related standards, age-related percentiles represent an appropriate alternative. The \((100p)^{th}\) percentile of a set of data values is the number such that \((100p)\%\) of the ordered values are less than it. Age-specific percentiles, or running percentiles, were constructed from the data from uninfected children used in the comparison with data for infected children. \(p\)-Values comparing individual predictions for combined data on uninfected and infected children at particular ages were derived from two sample \(t\)-tests of interpolated predicted values.

2.6 Early-life determinants of clinical disease progression analysis

The aim of the analysis presented in Chapter 6 was to identify early life predictors of clinical progression. The earliest values for markers available during the first six months of life are considered. Severe disease progression is defined by CDC category C (72), or death; progression occurring prior to the first marker value was disregarded. Overall disease progression was considered and also divided into rapid progression, defined as occurring within the first year of life, and long-term progression beyond age one year. ART information was categorised into no, mono, and double or combination (three or more) therapy, with the exception of the analysis for rapid progression, in which categorisation was no or combination (two or more) therapy (since there were no rapid progressors treated with monotherapy). Categories were updated at each visit, allowing children to be re-assigned if the number of drugs they were prescribed increased, thus ART treatment was included as a time-dependent variable (130). Because of widespread use
of neonatal ART and the initiation of ART early in life (for the prevention of vertical transmission and disease progression respectively) in recent years, disease progression and HIV-related death in the ECS cohort has been almost entirely restricted to those born before 1997 (141). Therefore, for this analysis, only children born before 1997 were considered but follow-up information until September 2002 was included.

The predictive values of first available log_{10} CD4⁺ cell count, absolute lymphocyte count and CD4⁺ percentage, persistence of (present at two or more visits during the first six months of life) hepatomegaly, splenomegaly and lymphadenopathy (and specifically axillary node enlargement as that was found to be associated with disease progression in an earlier stage of the study) were estimated, accounting for gender, prematurity, race, neonatal prophylactic ART to reduce vertical transmission risk, and subsequent ART administered for clinical indications. CD4⁺ percentage values were dichotomised. From exploratory analysis of the distribution of CD4⁺ percentage, considering values above or below 20% (midway between the CDC thresholds for moderate (25%) and severe (15%) immunosuppression (72)) gave the best categorisation in the first six months, with lower values being appropriate for later in life.

Gender- and race-specific effects of these early life indicators were assessed. Further, measurements at age two were assessed for effects on disease progression specific to boys and to girls.

As assays for quantifying HIV-RNA viral load have become routinely available only since 1997 and there is now much less vertical transmission of HIV to children being born to infected
women, there is a lack of available viral load measurements early in life of infected children and of their mothers. Where HIV RNA viral load data are available, there are few available measurements early in life for those who have progressed. Considering the entire cohort, of the 48 infected children with available viral load data in the first six months of life, only six (13%) progressed to serious disease, thus limiting the possibilities for extensive analyses as they are too few observations to allow reliable estimations.

2.6.1 Univariable Analysis
Separate univariable analyses were performed for overall, rapid and long-term progression to serious disease or death. For each, the overall survivor function was estimated with the Kaplan-Meier method. Cox proportional hazards models (142) were fitted to estimate risk of progression, expressed as hazard ratios (HRs) allowing for covariates.

Candidate terms for the multivariable model included ART treatment and factors reaching significance at the 15% level in univariable models. Setting the significance level above the more conventional 5% as the limit protects against the possibility of overlooking important factors which only become significant whilst adjusting for other factors.

2.6.2 Multivariable Analysis
Multivariable models describe the independent effects of included factors. On establishment of the candidate variables from the univariable analysis, all plausible multivariable models were constructed. Only combinations of explanatory factors which are not highly correlated are feasible due to collinearity. For instance, since they are correlated, simultaneous investigation of more than two lymphocyte determinations was not possible.
Comparisons of potential final models were made on the basis of the significance of individual terms and of the models overall. Robust estimates of standard errors minimize bias even in the presence of outlying observations and were calculated throughout (143).

The relative predictive capabilities of the various terms in the final model were determined by changes in the deviance, excluding each in turn; a large difference would indicate substantial contribution to the model.

2.6.3 Classification tree analysis of threshold for CD4⁺ percentage
In a separate analysis, classification trees (144) were used to detect the most predictive threshold for longitudinal CD4⁺ percentage measurements at any age after the first six months of life prior to disease progression.

2.7 Software
With the exception of the application of Hughes’ methodology, all analyses were carried out in a Windows environment. Analyses of clinical patterns were conducted using SAS (SAS Version 6.12, Cary, NC, USA, SAS Institute, Inc. 1996) and STATA (STATA Version 6.0, College Station, Texas, USA) statistical software packages. For HIV RNA viral load analyses, STATA (Version 6.0) (programmes given in Appendices 2.9.1 and 2.9.2) and S-PLUS 6.0 (Insightful, Seattle, USA), in a Unix environment (programmes given in Appendices 2.9.3, 2.9.4 and 2.9.5), were used, utilising a modified version of the software provided by Hughes in 1999 (132). The software described in (137) was used to compute the multivariate normal distribution integrals in the likelihood function. Immunological markers analyses were carried out using S-PLUS 2000 (programme given in Appendix 2.9.6). In the investigation of determinants of serious clinical
disease progression, all survival analyses were performed using STATA (Version 7.0) (programme given in Appendix 2.9.7) and classification trees were fitted using S-PLUS 2000.

2.8 Timeline for analyses data
Data on children known to be HIV infected according to the CDC definition of paediatric HIV infection, recorded before the beginning of 2000 were available for the analysis of clinical disease patterns presented in Chapter 3 (Figure 2.4). For the HIV RNA viral load analysis in Chapter 4, data recorded by April 2001 were included. Data available by February 2002 on both confirmed uninfected and infected children contributed to the assessment of immunological patterns in Chapter 5 and finally, the results presented in Chapter 6 on the investigation of determinants of clinical disease progression were based on data on infected children born before 1997, collected by the end of August 2002.

Figure 2.4: Timeline for data for clinical patterns (1), HIV RNA viral load (2), immunological patterns (3) and determinants of clinical disease progression (4) analyses.
Chapter 3 Results: Clinical disease patterns over childhood in vertically-infected children

3.1 Introduction
This analysis is undertaken to achieve a reliable picture of the pattern of clinical disease in vertically-acquired infection using the abundant data of the ECS spanning a long duration of follow up, taking a dynamic approach. Any instance of progression to the CDC categories A, B and C as well as death is considered. Clinical manifestations in children who were either never treated or, if treated, before and after they received ART are described. Also the effect of temporal changes in ART administration over time is investigated. For all other analyses in this section, children are classified on a visit-by-visit basis, regardless of previous status.

3.2 The children and their mothers
Up to the end of 1999, information on 170 infected children born to 170 mothers enrolled in the ECS was available for this analysis. For 93 mothers, the likely mode of HIV acquisition was injecting drug use, for 21 women an injecting drug using partner and 36 women acquired their infection through heterosexual contact. Of the remaining 20 women, four were infected via a blood transfusion and for 16, the route of HIV transmission was unknown. The women are generally in good health with three-quarters classified with only mild symptoms and no recorded severe immunological impairment.

Data were available from 3361 visits by the 170 infected children. The median number of visits per child was 15, ranging from one up to 68. The vast majority (143/170, 84.1%) made at least five visits.
The median duration of follow-up of these children was 5.8 years, with a maximum of 15 years. Fifteen children are no longer followed up by ECS clinicians: four moved and are now under the care of other clinicians, eight were living in socially disadvantaged circumstances and their mothers did not return for further visits and for three the reasons for loss of contact were unclear. The age at which these 15 children were last seen ranged from five months to nine years, and the number of recorded visits from three to 26.

Of the 170 children, 48 (28.2%) were born between 1985 and 1988, before the era of widely administered treatment (cohort 1) (Figure 3.1). The majority, 101 (59.4%), were born during 1989 to 1994, the time when ART was recommended only for symptomatic children using monotherapy or combination therapy with two NRTIs (cohort 2). There were 21 (12.4%) born after 1994, when treatment policy was to initiate combination therapy at an early stage (cohort 3).

The distinction by birth cohorts on the basis of treatment policy only partially reflected actual treatment received. Of the 170 infected children, 115 (67.6%) received ART at some point during follow up. Of the 48 children in cohort 1, over half (26/48 54.1%) actually received ART at some point during their follow-up. However, as expected, most (38/48, 79.2%) of these children were started on treatment only after they had progressed to serious disease. By comparison, 76 of the 101 cohort 2 children (75.2%) were ever administered ART, 75.0% (57/76) of these before onset of serious disease. Finally, 61.9% (13/21) of the cohort 3 received ART, which for 53.8% (7/13) preceded serious disease progression.
At start of ART initiation, 81 \((\frac{33+48}{170}, 70.4\%)\) of the 115 treated children received monotherapy (usually zidovudine), 61.7% \((\frac{15+35}{81})\) of whom were asymptomatic (Figure 3.2 and Table 3.1). The rest, 34 (29.6%) were initiated on combination therapy, of whom only one (2.9%) had a regimen that included PI. Reflecting changing clinical practice over time, treatment was modified in many children and at the last visit included in this analysis just 33 (28.7%) of the treated children were still receiving monotherapy, 62 (53.9%) combination without PI and 20 (17.4%) combination therapy with PI.
Table 3.1: Mono- and combination-therapy initiation by clinical stage

<table>
<thead>
<tr>
<th>CDC clinical stage at ART</th>
<th>Monotherapy only (%)</th>
<th>Initiated on mono-therapy then changed to combination therapy (%)</th>
<th>Initiated on combination-therapy (%)</th>
<th>All (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CDC clinical stage when changed to combination-therapy (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>All (%)</td>
<td></td>
<td>33 (100)</td>
<td>28</td>
<td>7</td>
</tr>
<tr>
<td>N</td>
<td>15 (45)</td>
<td>21 (75)</td>
<td>7 (100)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>A</td>
<td>1 (3)</td>
<td>2 (7)</td>
<td>0 (0)</td>
<td>3 (25)</td>
</tr>
<tr>
<td>B</td>
<td>12 (36)</td>
<td>5 (18)</td>
<td>0 (0)</td>
<td>3 (25)</td>
</tr>
<tr>
<td>C</td>
<td>5 (15)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

3.3 Overall disease progression in infected children

3.3.1 Clinical status

By last follow-up visit, 45 children (26.4%) had died of AIDS, a further two of non-HIV-related causes (one died of a cardiac arrest resulting from congenital heart tumour and the other died at home in suspicious circumstances) and 66 (38.8%) had progressed to serious disease or death (Figure 3.3). An additional four children in category B with symptomatic lymphoid interstitial pneumonitis (LIP) brought the total with AIDS to 70 (41%). Of the 135 (79.4%) children who were classified with at least moderate disease (class B), 19 (14.1%) had had symptomatic LIP, all born before 1993. Ten (52.6%) of these 19 died, three (30.0%) within 10 days of first diagnosis of LIP, suggesting that they may have been misdiagnosed and could have been suffering from PCP. Other children were classified with moderate disease due mainly to bronchitis. Mild disease was most commonly persistent.
lymphadenopathy, hepatomegaly, splenomegaly and upper respiratory tract infection; 153 (90%) of children had at least A defining symptoms, leaving a subset of 17 (10%) who remained asymptomatic throughout follow-up.

Progression to each of the CDC categories occurred most frequently during the first year of life, decelerating over one to five years of age and further from five to ten years of age (Figure 3.3). Cumulatively, more than three-quarters (77.4%) of infected children were estimated to have ever developed category A or more severe symptoms by one year of age (Table 3.2). This extends to more than nine out of ten children (91.5%) by five years and nearly all infected children (95.5%) by ten years of age. Progression to category B or more extreme, was estimated to have occurred in 59.7% children by one year of age, in 80.2% by five years and 85.3% by ten years of age. Over 15% (17.7%) of infected children will have progressed to category C or death by one year of age rising annually thereafter by 5% to just over 35% (36.9%) by five years and then by 2% to just over 45% (46.6%) by 10 years. Almost one tenth (8.5%) of all infected children will have died by one year of age. A further 17.1% were estimated to have died over the following four years of life, with an estimated cumulative probability of dying by five years of age of one quarter (25.6%). By ten years of age an estimated one third (35.0%) of infected children will have died. At ten years of age over half (53.4%) of the children will have survived and remain free of serious disease development.
Figure 3.3: Estimated progression to CDC clinical categories A or more severe (___), B or more severe (___), C or death (___) and death (___) for all children using Kaplan-Meier lifetable analysis.
Table 3.2: Progression and Follow-Up of 170 infected children

<table>
<thead>
<tr>
<th>CDC Category</th>
<th>Number progressing</th>
<th>Mean age at progression in years (95% CI)</th>
<th>percentage (95% CI) progressed by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 year</td>
</tr>
<tr>
<td>A or more severe</td>
<td>153</td>
<td>1.3 (0.9 to 1.7)</td>
<td>77.4 (71.0 to 83.7)</td>
</tr>
<tr>
<td>B or more severe</td>
<td>135</td>
<td>2.8 (2.1 to 3.5)</td>
<td>59.7 (52.2 to 67.2)</td>
</tr>
<tr>
<td>C or death</td>
<td>66</td>
<td>5.8 (5.2 to 6.3)</td>
<td>17.7 (11.9 to 23.6)</td>
</tr>
<tr>
<td>Death</td>
<td>47*</td>
<td>7.1 (6.6 to 7.6)</td>
<td>8.5 (4.3 to 12.8)</td>
</tr>
</tbody>
</table>

* 45 HIV deaths + two deaths from non-HIV related causes
3.3.2 Immunological disease progression

Figure 3.4 shows the Kaplan-Meier graph for immune status as defined by the CDC system. Overall, immune impairment was common and rapid, with an estimated half of all children (50.7%) progressing to moderate or severe immune deficiency by one year of age. By five years of age less than one in ten children has remained immunologically normal (92.5% progressed to at least moderate immune deficiency) and by ten years virtually all children have shown evidence of immunosuppression (99.2%). After one year of follow up, just over one fifth (21.2%) of infected children will have severe immune deficiency (category 3), rising to over half (53.4%) by five years and about three-quarters (76.6%) by 10 years of age. Immunological status poorly reflected clinical condition with three-fifths (248/426, 58.2%) of immunological category 3 visits classified as clinical category N (Table 3.3).

Figure 3.4: Estimated progression to CDC immunological categories 1 or 2 (___) and 2 (___) using Kaplan-Meier lifetable analysis
Table 3.3: Clinical and immunological CDC classification of visits

<table>
<thead>
<tr>
<th>CDC classification</th>
<th>Clinical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (n = 1024)</td>
</tr>
<tr>
<td></td>
<td>A (37.9%)</td>
</tr>
<tr>
<td></td>
<td>B (8.4%)</td>
</tr>
<tr>
<td></td>
<td>C (4.4%)</td>
</tr>
<tr>
<td></td>
<td>Total (37.7%)</td>
</tr>
<tr>
<td>Immunological</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>759</td>
</tr>
<tr>
<td></td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1024 (51.1%)</td>
</tr>
<tr>
<td>2</td>
<td>410</td>
</tr>
<tr>
<td></td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>555 (27.7%)</td>
</tr>
<tr>
<td>3</td>
<td>248</td>
</tr>
<tr>
<td></td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>426 (21.3%)</td>
</tr>
<tr>
<td>Total</td>
<td>1417 (70.7%)</td>
</tr>
<tr>
<td></td>
<td>328 (16.4%)</td>
</tr>
<tr>
<td></td>
<td>211 (10.5%)</td>
</tr>
<tr>
<td></td>
<td>49 (2.4%)</td>
</tr>
<tr>
<td></td>
<td>2005 (100%)</td>
</tr>
</tbody>
</table>

3.4 Progression to serious disease by calendar period of birth

Figure 3.5 displays the progression to serious disease or death by era of birth. Patterns of serious disease progression differed significantly according to era of birth ($\chi^2 = 9.10$, $p = 0.011$). By one year of age nearly one third (29.8%) of the children in the earliest cohort will have progressed to category C or death compared to 14.2% for those in cohort 2 and only 5.6% for those born in the most recent era. By four years of age, the gap has widened further, with the estimated proportion for cohort 3 remaining at just above one in twenty, while 34.5% of the middle cohort and 45.5% of the earliest cohort are estimated to have progressed.
3.5 Progression in ART naïve children

Children not receiving ART at any point in life provide the basis for the account of the natural history of vertically acquired HIV infection. Of the 170 infected children, 55 (32.4%) never received ART. Only eight (14.5%) of them were born at a time when early initiation of combination therapy was the recommended policy, 22 (40.0%) were born in the pre-treatment era, and 25 (45.5%) when monotherapy was recommended for symptomatic children (Figure 3.1). Ten untreated children (18.2%) remained asymptomatic when last seen between the ages of 1.5 months and 7.6 years, with a median of 4.5 years. Figure 3.6 presents the Kaplan-Meier estimates for progression to the three CDC symptomatic groups and death for untreated children. Forty-five (81.2%) were diagnosed with category A or more severe HIV-related symptoms or signs. This occurred in the first year of life in all but two children.
Progression was rapid and more than one-quarter (27.9%) of untreated children were estimated to have progressed to serious disease or death by one year of age. Overall, 20 (36.4%) of the 55 treatment-naive children died.

3.6 Clinical progression in ART treated children

For this analysis, children were classed as treated, regardless of whether progression took place before or after initiation of treatment. Of the 115 treated children, immunological information was available for 82 at start of treatment: 25 had no evidence of immune suppression, 26 had moderate immune suppression and 31 were severely immune deficient (Table 3.4).

The median age by clinical stage at initiation of treatment was 2.5 years for the five children with C symptoms, 4.0 years for the 14 in category B, 7.0 years for the eight with mild symptoms (A) and 6.0 years for the 55 asymptomatic children. The healthier the child at treatment initiation, the longer they lived: 90% (49/55) of the children who were asymptomatic at treatment initiation were still alive five years later compared to only 25% (14/55) of the children with clinical category C symptoms at treatment.

Table 3.4: Clinical and immunological CDC classification at treatment initiation

<table>
<thead>
<tr>
<th>CDC classification</th>
<th>Clinical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (22.0)</td>
</tr>
<tr>
<td>Immunological</td>
<td>25 (30.5)</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>Sub-total</td>
<td>55 (67.1)</td>
</tr>
<tr>
<td>unknown</td>
<td>19 (57.6)</td>
</tr>
<tr>
<td>Total</td>
<td>74 (64.3)</td>
</tr>
</tbody>
</table>
In total, 27 (23.5%) treated children died during follow-up, including two who died of non-HIV-related causes. Patterns of progression in ART treated children were similar to those in children overall (Figure 3.7). The five- and ten-year survival estimates of progression to at least moderate symptoms (80.4% and 87.8%, respectively) and severe symptoms or death (33.1% and 42.3% respectively) were similar to those in all children. However, a lower proportion of treated children will have progressed to mild disease (57.8%) and none died within the first year.

Of the 88 ever treated children alive at the last visit, most were asymptomatic (76.5%), only seven (7.9%) had category A symptoms and the remaining 12 (13.6%) had clinical category B symptoms. Seven (6.1%) of the treated children remained asymptomatic throughout their follow-up of between three months and 10.5 years (median 3.6 years).
Figure 3.6: Estimated progression to CDC clinical categories A or more severe (___), B or more severe (___), C or death (___) and death (___) for 55 ART naïve children using Kaplan-Meier lifetable analysis.

Figure 3.7: Estimated progression to CDC clinical categories A or more severe (___), B or more severe (___), C or death (___) and death (___) for the 115 children treated with ART using Kaplan-Meier lifetable analysis.
3.7 Comparison of progression in ART naïve and ART treated children

Overall, progression to mild, moderate and severe disease and to death happened more quickly for untreated children. The estimated survival curves for each category were different for treated and untreated children as shown in Table 3.4. For onset of at least A-defining illness, the one year progression rate of nearly 70% (69.7%) in untreated children (Figure 3.6) compares with less than 60% (58.2%) in treated children (Figure 3.7). Similarly, the difference in one-year rates for progression to clinical category B or more severe is around 10% (63.1% and 53.8%, respectively). By ten years of age the trend has reversed, with a greater proportion (91.2%) of treated children having progressed to at least category B compared with untreated (73.5%), reflecting the history of preferential treatment administration to symptomatic children of that age. More obvious differences are seen in progression to serious disease and death. Estimated progression to category C and to death was initially more rapid in the untreated children (30.2% (untreated) versus 12.3% (treated) for C and 26.4% (untreated) versus 0.01% (treated) for death). This almost certainly reflects treatment opportunity: these children were enrolled in the early years of the study and died before having had the chance to be treated. By 10 years of follow-up, the estimated cumulative progression to C is more similar in the two groups (61.1% versus 42.8%) although the overall patterns were significantly different \( p = 0.036 \) (Table 3.5). The overall difference in the age-related death rates was most significant \( p<0.0001 \).

Table 3.5: Log-rank test results comparing estimated survival curves for treated and untreated children

<table>
<thead>
<tr>
<th>CDC Category</th>
<th>( \chi^2 )</th>
<th>df*</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A or more severe</td>
<td>5.00</td>
<td>1</td>
<td>0.0253</td>
</tr>
<tr>
<td>B or more severe</td>
<td>12.76</td>
<td>1</td>
<td>0.0004</td>
</tr>
<tr>
<td>C or more severe</td>
<td>4.41</td>
<td>1</td>
<td>0.036</td>
</tr>
<tr>
<td>Death</td>
<td>18.60</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* df, degrees of freedom
3.8 Comparison of progression by ART vertical transmission prophylaxis
Of the 170 infected children, 13 (7.6%) had and 157 (91.4%) had not received ART as prophylaxis for vertical transmission. Nearly one in five (19.2%) of the non-exposed children have progressed to C or death in the first year of life compared to 0% of the children exposed to prophylaxis. Disease progression patterns were marginally distinct, but not statistically significant (log rank test $\chi^2 = 2.84$, $p$-value=0.092). However, these results should be interpreted in the light of the more widespread use of ART therapy for infected children after 1995. Of the 13 children who received prophylaxis with ART, 11 (85%) were also subsequently ART treated; 91% (10/11) with combination therapy.

3.9 Risk of serious disease progression by pre-progression ART, PCP prophylaxis and IVIG treatment
The immediate risk of progression to CDC C category or death in children after initiation of ART, PCP prophylaxis and IVIG were investigated separately by time-dependent Cox regression analysis. In this analysis, any child who had progressed to C before receiving treatment was regarded as untreated. At any time children receiving ART (RR = 0.58, $p$-value = 0.037) or PCP (RR = 0.63, $p$-value = 0.071) prophylaxis were less likely to progress to C than those not receiving prophylaxis. Most children who started ART before progression to CDC C disease were started on combination therapy. Of the 115 children receiving ART, 76 (66.1%) also received PCP prophylaxis, and just three other children received PCP prophylaxis but not ART; it is thus not possible to perform reliable further analysis to distinguish independent effects of these treatments. There was no evidence for a difference in progression to C between children receiving or not receiving IVIG (RR =1.02, $p = 0.923$).
Risk of progressing to severe immune deficiency (immunological category 3) was 64% higher when receiving ART (RR = 1.64, 95% CI: 0.96 to 2.82) with borderline significance ($p = 0.070$).

### 3.10 Presence of HIV-related symptoms at six monthly age intervals

For this analysis, clinical status was classified according to the CDC system on a cross-sectional basis every six months, regardless of previous disease manifestations. Figure 3.8 shows the breakdown of the classifications of the children for every 6-month visit. In total, the 160 children (160/170, 94.1%) who were still alive at the visit closest to six months were included in this analysis.

At each 6-month snapshot, most infected children were without serious symptoms or signs. For each of the 6-month visits before three years of life, over half were asymptomatic (443/752 child-visits, 58.9%), from three to twelve years of age, for just over a quarter (257/946, 27.2%) children had symptoms, and at almost all visits (27/29, 93.1%), teenage children were without symptoms.
Figure 3.8: Clinical status for infected* children assessed cross-sectionally at six-monthly visits

* the 10 children who died before the first 6 month visit are not included. There were 4 children not seen before 6 months.

Figures 3.9 and 3.10 show the equivalent breakdown separately for ART treated and untreated children, respectively. The proportions at each visit are very similar for both children treated with ART and those untreated.
Figure 3.9: Clinical status for the ART naïve children* assessed cross-sectionally at 6-monthly intervals

* the nine children not treated with ART who died before the first 6 month visit are not included.

Figure 3.10: Clinical status for the ART treated children* assessed cross-sectionally at 6-monthly intervals

* the one child treated with ART who died before the first 6 month visit is not included.
Infected children who died contribute most of their information in the first four years of life, and after that age the picture largely relates to the experience of survivors. However, restricting the analysis to 45 children who died with AIDS showed a similar pattern to that of the overall cohort, as shown in Figure 3.11. Over half the visits reflected an absence of clinical symptoms (129/231, 55.8%), but where there were symptoms these were likely to be serious, such as encephalopathy and PCP.

Figure 3.11: Clinical status for 45 children who died with AIDS assessed cross-sectionally at 6-monthly intervals

3.11 Clinical status five-year look-back

The 76 (76/170, 44.7%) children who were still alive and in follow-up at or beyond age five years were included in the 5-year look-back analysis (Figure 3.12). Represented are the clinical classifications of the 1283 visits made by these children. These constitute 944 visits made during the five years of life by the 58 children who had no signs or symptoms of HIV at five years of age; 247 visits relating to the 11 who had mild disease symptoms; 76 visits by
the six who displayed moderate disease manifestations and 16 of the one child with a C-defining illness at age five. Of the previous visits of the children with no, or only minor symptoms at five years, 14 had been categorised as category C visits. On the other hand, the child who had category C symptoms at age five had been largely symptom free up to that age. The one previous C visit was, in fact, the 6-month visit previous to the 5-year visit when the child had oesophageal candida. Conventional cumulative progression of disease thus does not predict clinical status at subsequent visits.

Figure 3.12: Children alive and in follow-up at five years of age: Clinical status at previous visits by current visit
Of these 76 children, 12 were ART-naïve, and 64 had been treated with ART. The 12 ART naïve children were all asymptomatic at the last visit closest to age five years, and none of them had ever had a C-defining illness. Of the treated children, 46 (72%) were asymptomatic at age five years and had been asymptomatic most (521) of the previous 740 visits too. One treated child was diagnosed with a C-defining illness at the 5-year visit.

With the exception of the one child in category C at the five-year visit, children in clinical category N were most likely to be those on combination therapy: 19% (11/58) in category N compared to 9% (1/11) in category A and none in category B.

3.12 Immunological status five-year look-back

Equivalent to that for clinical status, the immunological look back of five-year olds is given in Figure 3.13. This includes the 60 (60/67, 89.6%) children who were still alive and in follow-up at or beyond age five years with available immunological status information. Displayed are the immunological classifications of 165 visits made during the five years of life by the 13 children without evidence of immunosuppression at five years of age; 396 visits relating to the 27 who had evidence of moderate immunosuppression and the 353 of the children severely immunosuppressed at age five. This picture shows that although none of the 13 children who had no evidence of immunosuppression at age five years had been previously severely immunosuppressed, 15 (15/396, 3.8%) visits of moderately suppressed five year olds had been classified as immunological category 3. Nearly one third (105/353, 29.7%) of the previous visits for the 31 children who were severely immunosuppressed at age five years, were classified as category 3, but nearly 40% (38.8%) had been category 1.
Figure 3.13: Children alive and in follow-up at five years of age: CDC immunological status at previous visits by current visit.
Key points: Clinical disease patterns over childhood in vertically-infected children

- An estimated 15% of children vertically-infected with HIV will progress to serious disease or die before age one.
- Up to half will have progressed to serious clinical disease or have died by ten years of age.
- Compared with development of disease in the first year of life, progression is less rapid between the ages of 1 and 5 years and slow between 5 and 10 years.
- Onset of category A-defining symptoms or more severe is common in early life.
- Less than one tenth of infected children remain asymptomatic over the first five years of life.
- However, both ever ART treated and ART naïve children are well most of the time.
- A subset of vertically infected children can do well, even without early initiation of treatment.
- No evidence was found of the existence of a subgroup of infected children who are always severely symptomatic.
- Progression to serious disease differed among children born at a time when highly effective combination treatment was becoming widespread and initiated at an early stage, those born in the early years of the study when few therapy options were available, and when therapy was confined to those with serious disease.
Key points cont: Clinical disease patterns over childhood in vertically-infected children

- Children exposed to ART prophylaxis to reduce the risk of vertical transmission in their first weeks of life were less likely to progress to serious disease, although of this could be a reflection of the effects of clinical experience and advanced therapy.

- Infants who survive an initial episode of serious opportunistic disease were subsequently likely to become asymptomatic or to have only mild symptoms for prolonged periods, even without aggressive ART.

- Fluctuations in clinical symptoms were mainly associated with intermittent serious infections, while children diagnosed with encephalopathy remained seriously ill or died.

- Findings indicate a generally better than appreciated clinical picture for children with vertically acquired HIV infection.
Chapter 4 Results: HIV RNA Viral load over age in infected children

The aim of this analysis is to estimate RNA viral load patterns over age in vertically infected children accounting for between and within individual variation and assay cut-off detection level. Interest was also in possible gender and treatment-based differences allowing for CD4\(^+\) cell counts and assay type. The various methodological approaches are compared.

4.1 Data

Of the 178 infected children enrolled by 1\(^{st}\) April 2001, 19 were no longer in follow-up in any ECS centre and 41 had died prior to the introduction of routine viral load assays. A total of 894 viral load measurements were thus available on 118 infected children (Table 4.1). Of these, 163 (18.2\%) were below the assay specific cut-off value. The median RNA viral load was 14,690 copies/ml plasma. Of the 90.2\% of measurements with known assay type, the majority (639/806, 79.3\%) were ascertained by Roche and the remaining (20.7\%) by NASBA. Most (88, 74.6\%) children had received anti-retroviral therapy in a combination of at least two drugs at some point, with a median age at treatment initiation of 31.5 months. Twenty-one (11 girls and 10 boys) of the 118 children had been diagnosed with serious HIV related disease (CDC class C or death), of whom six had died.
Table 4.1: Characteristics of infected children with HIV RNA viral load measurements.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Children $n = 118$</th>
<th>Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of measurements</td>
<td>Median (range)</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>7 (1 to 23)</td>
<td>894</td>
</tr>
<tr>
<td>Viral load values (copies/ml)</td>
<td>-</td>
<td>14,690</td>
</tr>
<tr>
<td>Median (range)</td>
<td>( &lt;40 to 8,100,000 )</td>
<td></td>
</tr>
<tr>
<td>Viral load measurement $n$ (%)</td>
<td>Actual</td>
<td>70 (59.3)</td>
</tr>
<tr>
<td></td>
<td>Below detection level*</td>
<td>48 (40.7)§</td>
</tr>
<tr>
<td>Sex $n$ (%)</td>
<td>Female</td>
<td>60 (50.8)</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>58 (49.1)</td>
</tr>
<tr>
<td>Treatment Status $n$ (%)</td>
<td>No or monotherapy</td>
<td>30 (25.4)◊</td>
</tr>
<tr>
<td></td>
<td>Combination therapy</td>
<td>88 (74.6)◊</td>
</tr>
<tr>
<td>CD4$^+$ cell count ($x 10^5$ cells/l)</td>
<td>Median (range)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>670 (0.1 to 482)</td>
</tr>
<tr>
<td>Assay $n$ (%)</td>
<td>NASBA</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Roche</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>-</td>
</tr>
<tr>
<td>Age at treatment initiation (months) Median (range)</td>
<td>31.5 (0.76 to 151.1)</td>
<td></td>
</tr>
<tr>
<td>Age at last visit (months) Median (range)</td>
<td>93.1 (0.2 to 179.0)</td>
<td></td>
</tr>
</tbody>
</table>

* threshold values: 4000, 1000, 800, 500, 400, 200, 150, 100, 85, 80, 50, 40
§ at least one censored measurement per child
◊ treatment status at last measurement
Table 4.1 cont: Characteristics of infected children with HIV RNA viral load measurements.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Children n = 118</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC immunological stage at last measurement</td>
<td>n (%)</td>
</tr>
<tr>
<td>1</td>
<td>12 (10.2)</td>
</tr>
<tr>
<td>2</td>
<td>50 (42.4)</td>
</tr>
<tr>
<td>3</td>
<td>50 (42.4)</td>
</tr>
<tr>
<td>Died</td>
<td>6 (5.1)</td>
</tr>
<tr>
<td>CDC clinical stage at last measurement</td>
<td>n (%)</td>
</tr>
<tr>
<td>N</td>
<td>15 (12.7)</td>
</tr>
<tr>
<td>A</td>
<td>12 (10.2)</td>
</tr>
<tr>
<td>B</td>
<td>70 (59.3)</td>
</tr>
<tr>
<td>C</td>
<td>15 (12.7)</td>
</tr>
<tr>
<td>Died</td>
<td>6 (5.1)</td>
</tr>
</tbody>
</table>

4.2 HIV RNA viral load over age

The techniques of non-parametric running smoothers were applied to reveal the structure of the raw log_{10} HIV RNA viral load data over age. The running smoother representing is shown in Figure 4.1, with log_{10} HIV RNA viral load peaking at around one and a half months of age, and gradual declining thereafter.
The methods of fractional polynomials, conventional polynomials and change-point regression were applied and compared in their ability to represent the HIV RNA viral load data structure over age.

4.2.1 Modelling HIV RNA viral load over age using fractional polynomials
The complexity of fractional polynomial models depends on the number of powers, or explanatory terms, in the model. Thus, a one-power model for HIV RNA viral load over age has one term for age, a two-power model has two terms for age, and so on. There are various models of equivalent power and the first step in fractional polynomial involves selection of the superior
one, identified as that with lowest deviance. This is performed for models of different powers. The overall optimal model is then identified by significance tests comparing the superior models for different powers.

In this application, from among eight different one-power fractional polynomial models for log_{10} RNA viral load and age, the optimal was identified as the one with a singular age term (deviance = 2797.9), equivalent to the simple linear regression model (Table 4.2). The best two-power fractional polynomial model from among 44 different ones was identified as the one with inverse square root of age and natural log age terms (deviance = 2779.3). From among 164 different three-power fractional polynomial models for log_{10} RNA viral load and age, the optimal was identified as the one with inverse of age, natural log age by inverse of age and natural log age terms (deviance = 2776.4). The fitted curves corresponding to the optimal one-, two- and three-power models are given in Appendix 4.1.

The identified two-power fractional polynomial was a significantly better fit than the best one-power model ($\chi^2 = 18.6$, df = 2, $p<0.0001$) (Table 4.3). The best fit three-power model was not statistically significantly better fitting than the optimal two-power one ($p = 0.235$), establishing the two-power model as the chosen one. The optimal fractional polynomial model for log_{10} RNA viral load and age was therefore the one with inverse square root of age and log age terms. The predicted pattern of HIV RNA viral load over age for this model is given in Figure 4.2. This shows the peak at one and a half months of age and a gradual decline thereafter, characteristic of the running smoother.
Table 4.2: Optimal fractional polynomial models for each power

<table>
<thead>
<tr>
<th>Model</th>
<th>Power</th>
<th>Expression</th>
<th>df</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP1</td>
<td>1</td>
<td>$\log_{10}(\text{Viral Load}_y) = 3.96 - 0.005(Age_y) + \varepsilon_y$</td>
<td>2</td>
<td>2797.9</td>
</tr>
<tr>
<td>FP2</td>
<td>2</td>
<td>$\log_{10}(\text{Viral Load}_y) = 3.90 + \frac{0.089}{\sqrt{Age_y}} - 0.338 \times \ln(Age_y) + \varepsilon_y$</td>
<td>4</td>
<td>2779.3</td>
</tr>
<tr>
<td>FP3</td>
<td>3</td>
<td>$\log_{10}(\text{Viral Load}_y) = 3.89 - \frac{0.014}{Age_y} - \frac{0.002 \times \ln(Age_y)}{Age_y} - 0.309 \times \ln(Age_y) + \varepsilon_y$</td>
<td>6</td>
<td>2776.4</td>
</tr>
</tbody>
</table>

Table 4.3: Comparisons of optimal fractional polynomial models

<table>
<thead>
<tr>
<th>Model comparison</th>
<th>df</th>
<th>$\chi^2$ (Deviance difference)</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP2 to FP1</td>
<td>4 - 2 = 2</td>
<td>2797.9 - 2779.3 = 18.6</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>FP3 to FP2</td>
<td>6 - 4 = 2</td>
<td>2779.3 - 2776.4 = 2.9</td>
<td>0.235</td>
</tr>
</tbody>
</table>

4.2.2 Alternatives to fractional polynomials

*Conventional polynomial*

As for fractional polynomials, convention polynomials of different powers can be constructed, with increasing sophistication as the number of powers increases. With this methodology, there is just one possible model within each power. In the application of convention polynomials to the HIV RNA viral load over age, representation of the data structure was not improved by use of models with beyond power four, that is beyond the quartic polynomial. Since none of the quartic, cubic or quadratic terms were significant ($p = 0.452$, $p = 0.564$ and $p = 0.319$, respectively), the quartic polynomial model had little structure. The trajectory of the predicted values of the quartic polynomial model is given in Figure 4.2. When compared to the running smoother, the predicted
curve did not visually reflect the data. The main drawback of conventional polynomials for modelling paediatric HIV RNA viral load over age was their inability to represent the important rapid peak at young ages.

**Change-point regression model**

Change-point regression models consist of at least two joined segments, simultaneously modelled to characterise different ranges of the explanatory variable. In this application, two joined linear regressions were used in an attempt to represent the data at ages before and after peak in HIV RNA viral load. From maximisation of the running smoother, the peak was estimated to occur at \( \tau = 2.6 \) months. The predicted conjoined regression lines for the resulting change-point model are shown in Figure 4.2. It is clear that the change-point fails to achieve the extent of the peak displayed by the running smoother.
Figure 4.2: Underlying structure of HIV-RNA viral load by age as shown by the super-smoother (SS) along with fractional polynomial (FP), conventional polynomial (quartic) and change-point model (PW) representations.
4.2.3 Comparison of methods for modelling HIV RNA viral load over age
Comparing the resultant models of the three approaches: the quartic polynomial fails to detect
the early peak and the peak of the change-point model does not reflect the magnitude of the
peak of the running smoother. The peak of the fractional polynomial comes closest to that of
the running smoother. Formal comparison of the three possibilities was performed by
evaluation of the respective AICs (Table 4.4). Since it has the lowest value AIC, the
fractional polynomial is the most suitable (AIC = 2517.24 compared to 2536.94 for the
quartic polynomial and 2528.42 for the change-point model).

Table 4.4: AICs of the three model forms for HIV RNA viral load over age

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-power fractional polynomial</td>
<td>2517.24</td>
</tr>
<tr>
<td>Quartic polynomial</td>
<td>2536.94</td>
</tr>
<tr>
<td>Change-point model</td>
<td>2528.42</td>
</tr>
</tbody>
</table>

4.3 Explanatory variables for HIV RNA viral load levels and patterns over age
Interest was in which factors determine the level and pattern of HIV RNA viral load over age.
Gender, ART administration and CD4^+ cell count were considered, accounting for assay type.
The first step in establishing the final multivariable model was to assess the main effects of
factors, which investigated any constant differences over time. For instance, the investigation
of a main effect of treatment addresses any overall difference there may be in HIV RNA
levels by treatment status over age. Differences may not, however, be constant over age. The
second step was to explore interactions of interest, revealing any differences which depend on
the levels of other factors. Introduction of a treatment by age interaction allows for
differences by treatment which depend on age. The final stage of modelling involved the fitting of three-way interaction, allowing for other relevant factors.

For 88 of the 894 HIV RNA viral load measurements determining assay type was unknown and for a further one CD4$^+$ cell count had not been measured at the time of the HIV RNA determination. Thus 805 observations with complete information on assay type, age, gender, CD4$^+$ cell count and ART were available in the subsequent modelling. These measurements had equivalent mean (14,800 copies/ml), range (40-8,100,000 copies/ml) and proportion of censored values (18.2%) to the original 894 measurements. Of the total number of these 805 observations nearly 80% (638/805) were assessed by a Roche assay and 20% (167/805) by NASBA assay.

4.3.1 Application of Hughes’ modified Monte Carlo EM algorithm
Table 4.5 presents the likelihoods for all the models used in the selection of covariates for HIV RNA viral load over age. Values form the basis of the likelihood ratio tests for assessing the significance of the main effects and interaction terms of interest, reported in tables 4.6 and 4.8. The full expressions for these models are given in Appendix 4.2.
Table 4.5: Likelihoods for models used for the selection of covariates for HIV RNA viral load over age by model comparisons in Table 5.6 and Table 5.8.

<table>
<thead>
<tr>
<th>Model Number</th>
<th>Model</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Age</td>
<td>-1333.71</td>
</tr>
<tr>
<td>2</td>
<td>Age + Gender</td>
<td>-1333.70</td>
</tr>
<tr>
<td>3</td>
<td>Age + Treatment</td>
<td>-1268.62</td>
</tr>
<tr>
<td>4</td>
<td>Age + log_{10} CD4</td>
<td>-1288.83</td>
</tr>
<tr>
<td>5</td>
<td>Age + Assay</td>
<td>-1328.00</td>
</tr>
<tr>
<td>6</td>
<td>Age + Gender + Treatment</td>
<td>-1268.60</td>
</tr>
<tr>
<td>7</td>
<td>Age + Gender + Age×Gender</td>
<td>-1323.87</td>
</tr>
<tr>
<td>8</td>
<td>Age + Treatment + Age×Treatment</td>
<td>-1268.32</td>
</tr>
<tr>
<td>9</td>
<td>Age + log_{10} CD4^ + Age×log_{10} CD4^</td>
<td>-1288.71</td>
</tr>
<tr>
<td>10</td>
<td>Age + Gender + Treatment + Gender×Treatment</td>
<td>-1261.85</td>
</tr>
<tr>
<td>11</td>
<td>Age + Gender + Treatment + log_{10} CD4^ + Assay + Age×Gender + Age×Treatment + Gender×Treatment</td>
<td>-1211.50</td>
</tr>
<tr>
<td>12</td>
<td>Age + Gender + Treatment + log_{10} CD4^ + Assay + Age×Gender + Age×Treatment + Gender×Treatment + Age×Gender×Treatment</td>
<td>-1208.34</td>
</tr>
</tbody>
</table>

**Main effects of factors influencing HIV RNA viral load levels over age**

The significance of the main effects of gender, treatment, CD4^+ cell count and assay was assessed, in turn by likelihood ratio tests comparing each of the main effects models (Models 2, 3, 4, and 5, respectively) to the basic model (Model 1). Table 4.6 presents the results of the likelihood ratio tests for these covariate main effects. From the tests of the main effects, log_{10} RNA viral load did not differ by gender in univariable analysis (p = 0.903), but there were significant differences by treatment, CD4^+ cell count and assay (p<0.001 for all three). Combination therapy was associated with a 1.28 lower log_{10} viral load compared with the viral load in the no/monotherapy category (Wald test: p<0.001) (Table 4.7). A one log_{10} CD4^+
cell count increase resulted in a significant decrease of 0.895 in $\log_{10}$ RNA viral load ($p<0.001$) and the Roche assay was associated with a 0.404 lower $\log_{10}$ viral load than the NASBA assay ($p = 0.0128$).

Table 4.6: Significance tests for main effects of covariates for HIV RNA viral load pattern over age based on likelihood values from Table 4.5.

<table>
<thead>
<tr>
<th>Model Comparison</th>
<th>Term</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &amp; 2</td>
<td>Gender Main Effect</td>
<td>0.015</td>
<td>1</td>
<td>0.903</td>
</tr>
<tr>
<td>1 &amp; 3</td>
<td>Treatment Main Effect</td>
<td>130.2</td>
<td>1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>1 &amp; 4</td>
<td>$\log_{10}$ CD4$^+$ Main Effect</td>
<td>89.76</td>
<td>1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>1 &amp; 5</td>
<td>Assay Main Effect</td>
<td>11.41</td>
<td>1</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Table 4.7: Univariable results for analyses of RNA viral load pattern over age by gender, treatment, CD4$^+$ cell count and assay

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>Univariable Coefficient</th>
<th>95% CI</th>
<th>Wald</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$1/\sqrt{\text{Age}}$ increase</td>
<td>-1.16</td>
<td>-1.50 to -0.822</td>
<td>6.73</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>$\log_{10}(\text{Age})$ increase</td>
<td>-0.496</td>
<td>-0.625 to -0.366</td>
<td>7.49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>Gender</td>
<td>Females Reference</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Males</td>
<td>-0.006</td>
<td>-0.342 to 0.355</td>
<td>0.036</td>
<td>0.971</td>
</tr>
<tr>
<td>3</td>
<td>ART</td>
<td>No/Monotherapy Reference</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Combination therapy</td>
<td>-1.28</td>
<td>-1.56 to -0.994</td>
<td>8.86</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>$1 \log_{10}$ CD4$^+$ cell count increase</td>
<td>-0.895</td>
<td>-1.14 to -0.651</td>
<td>7.19</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5</td>
<td>Assay</td>
<td>NASBA Reference</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Roche</td>
<td>-0.404</td>
<td>-0.722 to -0.086</td>
<td>2.49</td>
<td>0.0128</td>
</tr>
</tbody>
</table>
Interactions of factors influencing HIV RNA viral load pattern over age

Significance of two-way interactions for age by gender, age by treatment, age by CD4\(^+\) cell count and gender by treatment was assessed by individual likelihood ratio tests comparing each of the two-way interaction models (Models 7, 8, 9 and 10, respectively) to the corresponding main effects models (Models 2, 3, 4, and 6, respectively). The results of the significance of the two-way interactions of each of the factors gender, treatment and CD4\(^+\) cell count with age are shown in table 4.8. The sex by age interaction was significant ($\chi^2 = 19.7, p < 0.001$) and this is represented in Figure 4.3. Thus the pattern of viral load over age in girls differed from that of boys; this difference was masked in the previous model which accounted for only the main effect of sex. Viral load in girls peaks earlier (1.1 months) than in boys (2.1 months) and at a higher value and declines more sharply in girls than boys, with a cross-over at about four years of age. After this age, viral load in girls is consistently lower than in boys. At the peak the estimated HIV RNA viral load is more than one log\(_{10}\) above the peak level for boys, while after the cross-over girls have a predicted viral load 0.25 to 0.5 log\(_{10}\) below that for boys. Interactions of treatment ($\chi^2 = 0.599, p = 0.741$) and CD4\(^+\) cell count ($\chi^2 = 0.235, p = 0.628$) with age were not significant.

Allowing for an interaction between gender and treatment, there was evidence of a difference in the effect of treatment by sex ($\chi^2 = 5.18, p = 0.023$), suggesting that differences in RNA viral load by treatment were different for boys and girls.
Table 4.8: Significance tests for interactions of covariates for HIV RNA viral load pattern over age based on likelihood values from Table 4.5.

<table>
<thead>
<tr>
<th>Model Comparison</th>
<th>Effect</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 &amp; 7</td>
<td>Age x Gender Interaction</td>
<td>19.7</td>
<td>2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3 &amp; 8</td>
<td>Age x Treatment Interaction</td>
<td>0.599</td>
<td>2</td>
<td>0.741</td>
</tr>
<tr>
<td>4 &amp; 9</td>
<td>Age x log_{10} CD4${^+}$ Interaction</td>
<td>0.235</td>
<td>2</td>
<td>0.628</td>
</tr>
<tr>
<td>6 &amp; 10</td>
<td>Gender x Treatment Interaction</td>
<td>13.5</td>
<td>1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>11 &amp; 12</td>
<td>Age x Gender x Treatment Interaction</td>
<td>6.31</td>
<td>2</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Figure 4.3. Log_{10} RNA viral load by gender
Finally, modelling was performed to investigate the significance of three-way interaction of age by sex by treatment, accounting for assay type and CD4⁺ cell count. For any model, the presence of a three-way interaction term leads to the automatic inclusion of the corresponding two-way interactions and main effects terms, regardless of their significance in simpler models. Hence, a full model (Model 12), incorporating main effects of gender, treatment, assay type and log₁₀ CD4⁺ cell count and two-way interactions of age by sex, age by treatment (despite lack of significance in the two-way model comparison (Models 3 and 8)) and gender by treatment, and three-way interaction of age by sex by treatment was fitted. The likelihood ratio test comparing the full model (Model 12) with the corresponding comparison model (Model 11) yields a significant (p = 0.043) three-way age by sex by treatment interaction (Table 4.8). Hence, HIV RNA viral load pattern over age is presented separately for the four strata defined by gender and treatment. For the purposes of presentation, levels are fitted for Roche assay, at the median CD4⁺ cell count (Figure 4.4). The picture for NASBA observations is similar, with identical patterns, but slightly higher levels than for Roche measurements. Again, patterns for HIV RNA viral load are identical for any value of CD4⁺ cell count, but decrease with increasing CD4⁺ cell count. This model predicts that, for example, at 12 months of age the difference in RNA viral load is about 2 log₁₀ between untreated and treated girls, and 1.5 log₁₀ for treated and untreated boys. At 15 years of age, the predicted differences would be one log₁₀ for girls and 2 log₁₀ for boys.
4.3.2 Comparison with alternative to Hughes’ method

As an alternative methodology to that developed by Hughes, the less complex LME approach with censored values taken as the mid-point between zero and the detection threshold was explored. The results of this alternative are compared to those obtained when using Hughes’ approach. For this comparison, results of the application of Hughes’ method are considered as the standard (because this was the most thorough in accounting for censored values).

To determine how closely the models represented the gender- and treatment-specific profiles over age, fitted curves were compared graphically. Figure 4.5 presents the gender and treatment strata-specific fitted curves from the applications of both Hughes’ modified EM
algorithm the LME approach with mid-point censored values. Confidence bands are those constructed from the application of the LME approach with mid-point censored values using bootstrap methods.

Curves of the model arising from the LME approach with mid-point censoring lie close to those created from the application of Hughes’ method. For each stratum, the 95% confidence bands around the fitted curves of the model from the LME approach contain the fitted curves from Hughes’ method. However, the agreement was not universally consistent, with less agreement between the fitted curves at ages and within strata, where values of viral load were more likely to be censored. Censoring is more frequent where, observed and thus estimated values are low in contrast to sporadic censoring when values are typically higher. For instance, it is clear from Figure 4.5 that there is less agreement between results of the two procedures for measurements under treatment in both girls and boys, when viral levels are relatively low. Further, in general, differences are larger at older years, when measurements are lower overall.
Figure 4.5: Gender and treatment strata-specific fitted curves of the model from application of Hughes’ modified EM algorithm (solid line) and fitted curves with confidence bands (dotted lines) arising from the application of the LME approach with mid-point censored values.
Key points: HIV RNA Viral load over age in infected children

- There is a marked peak in HIV RNA viral load at about one and a half months of age, followed by a rapid decline in the first five years of life, slowing thereafter.
- Fractional polynomial models adequately reflect the actual shape of the viral load pattern over age in vertically infected children.
- It is essential to account for the within-subject correlation arising from the repeated measures, and for the left-censoring induced by assay detection thresholds when modelling the pattern of longitudinal paediatric HIV RNA viral load data.
- However, simple means for dealing with censoring of setting the value at midpoint between zero and cut-off, are sufficient to identify the model which best describes the age-related dynamics of RNA viral load.
- RNA viral load levels were estimated to peak a little earlier and at substantially higher levels in girls than in boys with a difference of more than one \( \log_{10} \).
- After about 5 years of age the levels of RNA in girls are up to half a \( \log_{10} \) below that of boys.
- For both boys and girls viral load observations under treatment were consistently and substantially lower.
- However, whereas the difference between RNA viral load measurements obtained under treatment and those not, slightly increased with age for boys, it narrowed substantially for girls.
Key points *cont*: HIV RNA Viral load over age in infected children

- Without treatment, measurements of RNA viral load are initially higher for girls than boys, but levels cross over about 3 years of age, and thereafter the measurements relating to untreated girls are consistently lower than in boys.

- With treatment, viral load measurements for girls started off being higher than for boys, but after about 9 months of age they became lower than for boys with very little difference at later ages.
Chapter 5 Results: Patterns of immunological markers in uninfected and infected children

Uniquely, in this chapter, measurements on enrolled uninfected as well as infected children are included in the analyses. Interest was in establishing immunological patterns in uninfected children for comparison with those in infected ones. This analysis aimed to elucidate the patterns of $CD4^+$ cell, $CD8^+$ cell, absolute lymphocyte counts and $CD4^+$ percentage over childhood in uninfected and infected children. There was further interest in whether age-related patterns of immunological markers differed by gender and race.

5.1 Data

By February 2002, 1674 children, 1488 uninfected and 186 infected, enrolled had at least one $CD4^+$ cell, $CD8^+$ cell or absolute lymphocyte count measurements between birth and 12 years of age (Table 5.1). Overall, maternal $CD4^+$ cell counts were below 500 cells in 294 of 477 mothers with available data. Only 8.3% of mothers had been diagnosed with AIDS. Two-thirds of infected children received combination ART to delay disease progression, with a median age of 3.2 years at initiation. The ethnic distributions were similar for girls and boys for both uninfected (26.4% and 28.1% black, respectively) and infected (24.1% and 29.1% black, respectively) children.
Table 5.1: Characteristics of 1674 uninfected and infected children and their HIV infected mothers

<table>
<thead>
<tr>
<th></th>
<th>Uninfected children</th>
<th>Infected children</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 1488$ (%)$^a$</td>
<td>$n = 186$ (%)$^a$</td>
</tr>
<tr>
<td><strong>CHILD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median age last seen (months) (IQR$^b$)</td>
<td>21.6 (12.1 to 51.7)</td>
<td>70.5 (26.7 to 119.7)</td>
</tr>
<tr>
<td>(cum %) C or death</td>
<td>-</td>
<td>65</td>
</tr>
<tr>
<td><strong>Vital Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>1483 (99.7)</td>
<td>139 (74.7)</td>
</tr>
<tr>
<td>Dead</td>
<td>5 (0.3)</td>
<td>47 (25.3)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>767 (51.8)</td>
<td>90 (48.4)</td>
</tr>
<tr>
<td>Female</td>
<td>713 (48.2)</td>
<td>96 (51.6)</td>
</tr>
<tr>
<td>Unknown</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1002 (69.6)</td>
<td>123 (68.3)</td>
</tr>
<tr>
<td>Black</td>
<td>364 (25.3)</td>
<td>43 (23.9)</td>
</tr>
<tr>
<td>Other</td>
<td>74 (5.1)</td>
<td>14 (7.8)</td>
</tr>
<tr>
<td>Unknown</td>
<td>48</td>
<td>6</td>
</tr>
<tr>
<td><strong>Prematurity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;=34 weeks</td>
<td>110 (7.6)</td>
<td>11 (6.3)</td>
</tr>
<tr>
<td>&gt;34,&lt;=37</td>
<td>347 (24.1)</td>
<td>36 (20.5)</td>
</tr>
<tr>
<td>&gt;37 weeks</td>
<td>985 (68.3)</td>
<td>129 (73.3)</td>
</tr>
<tr>
<td>unknown</td>
<td>46</td>
<td>10</td>
</tr>
<tr>
<td><strong>ART initiation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/Mono</td>
<td>-</td>
<td>61 (32.8)</td>
</tr>
<tr>
<td>Combination</td>
<td>-</td>
<td>125 (67.2)</td>
</tr>
<tr>
<td><strong>Median age at ART initiation (years) (IQR$^b$)</strong></td>
<td>-</td>
<td>3.2 (0.8 to 4.6)</td>
</tr>
<tr>
<td><strong>MATERNAL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ART prophylaxis</td>
<td>No</td>
<td>1246 (83.7)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>242 (16.3)</td>
</tr>
<tr>
<td>Maternal CD4 at delivery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;=500</td>
<td>168 (38.4)</td>
<td>15 (38.5)</td>
</tr>
<tr>
<td>&lt;500</td>
<td>270 (61.6)</td>
<td>24 (61.5)</td>
</tr>
<tr>
<td>unknown</td>
<td>1050</td>
<td>147</td>
</tr>
</tbody>
</table>

$^a$ percentages are of total for each category excluding “unknown”

$^b$ IQR – inter-quartile range
A total of 13203 records with at least one CD4⁺ cell, CD8⁺ cell or absolute lymphocyte count measurements (Table 5.2) were available: 9789 lymphocyte counts related to the uninfected and 3414 to the infected children. Generally, log₁₀ CD4⁺ cell counts, log₁₀ AL counts and CD4⁺ percentages were lower and log₁₀ CD8⁺ cell counts were higher for infected children.

Table 5.2: Distribution of laboratory measurements by infection status

<table>
<thead>
<tr>
<th>Laboratory measurement</th>
<th>1488 uninfected children</th>
<th>186 infected children</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 9789 )</td>
<td>( n = 3414 )</td>
</tr>
<tr>
<td>CD4⁺ cell counts (x 10^6 cells/l) [log₁₀] mean (IQR)</td>
<td>2213 (1589 to 3228)</td>
<td>693 (460 to 1570)</td>
</tr>
<tr>
<td></td>
<td>3.345 (3.201 to 3.509)</td>
<td>2.841 (2.663 to 3.196)</td>
</tr>
<tr>
<td>CD8⁺ cell counts (x 10^6 cells/l) [log₁₀] mean (IQR)</td>
<td>1094 (780 to 1531)</td>
<td>1212 (849 to 1950)</td>
</tr>
<tr>
<td></td>
<td>3.039 (2.892 to 3.185)</td>
<td>3.087 (2.929 to 3.290)</td>
</tr>
<tr>
<td>absolute lymphocyte cell counts (x 10^6 cells/l) [log₁₀] mean (IQR)</td>
<td>5284 (3999 to 7096)</td>
<td>3273 (2188 to 5284)</td>
</tr>
<tr>
<td></td>
<td>3.723 (3.602 to 3.851)</td>
<td>3.515 (3.340 to 3.723)</td>
</tr>
<tr>
<td>CD4⁺ percentage mean (IQR)</td>
<td>43.3 (36.6 to 49.3)</td>
<td>26.2 (16.9 to 34.9)</td>
</tr>
</tbody>
</table>
5.2 Immunological markers in uninfected children

5.2.1 $\log_{10}$ CD4$^+$ cell count over age in uninfected children

Figure 5.1 shows the observed data and running smoother plot revealing the underlying patterns for the $\log_{10}$ CD4$^+$ cell count of uninfected children. $\log_{10}$ CD4$^+$ cell counts peaked at three weeks of age, then dipped before peaking again at six months declining gradually thereafter.

The natural cubic spline linear mixed effects model closest in fit to the $\log_{10}$ CD4$^+$ cell count data has knots positioned at one and three weeks, three and 10 months and 10 years (Appendix 5.1). The fitted pattern over age is given by the predicted values from the model, shown in Figure 5.1. For instance, (on the original scale) the predicted CD4$^+$ cell counts peaks at 3238 per mm$^3$ at three weeks, then again at 3009 per mm$^3$ at six months, dropping to 2597 per mm$^3$ at one year of age and to 1096 per mm$^3$ by five years of age. CD4$^+$ cell counts approach normal adult values some time after age six years.
5.2.2 Factors associated with levels and patterns of $\log_{10} CD4^+$ cell counts in uninfected children

The results for the effects of gender and race on $\log_{10} CD4^+$ cell counts in uninfected children are given in Table 5.3. The pattern of $\log_{10} CD4^+$ cell counts over age differs significantly by gender in both level and pattern. There is a systematic main effect difference ($p = 0.0003$) and that difference is of different magnitude at different ages, as given by the gender by age interaction ($p=0.0001$). The fitted $\log_{10} CD4^+$ cell counts values by gender are plotted in Figure 5.2. It can be seen that the level of $CD4^+$ cell counts is marginally higher for girls than for boys during the first year of life and again from around age three years onward, but not between one and three years. For example, at eight years of age the fitted value for girls is
997 \times 10^6 \text{ cells/l} \text{ compared to } 827 \times 10^6 \text{ cells/l} \text{ for boys, but at two years of age values are } 1906 \times 10^6 \text{ cells/l} \text{ and } 1856 \times 10^6 \text{ cells/l}, \text{ respectively.}

The levels and patterns of \log_{10} \text{ CD4}^+ \text{ cell counts differ by race. The main effect of the levels differs overall (} p<0.0001) \text{ and this difference is dependent on age with borderline significance (} p = 0.0586). Figure 5.2 shows the fitted \log_{10} \text{ CD4}^+ \text{ cell count levels by race over age. Other than at older ages, levels for white children are systematically higher than those of black children, and this difference appears marginally larger after age four years. For instance, at age four years, the value for white children is } 1283 \times 10^6 \text{ cells/l} \text{ compared to } 1002 \times 10^6 \text{ cells/l} \text{ for black children. Although sparse data on black children after eight years of age yielded estimates of fitted values from the model which were unreliable at older ages, there is apparent crossing over of levels for black and white children.}

The non-significant gender by race interaction term (\chi^2_6 = 2.62; \ p = 0.8541) \text{ means neither the differences in levels or patterns by gender over age depend on the race of the child, nor do the differences by race depend on the gender of the child.}

Estimated levels of \log_{10} \text{ CD4}^+ \text{ cell counts in the first year of life were similar for premature and full-term uninfected infants (} p = 0.615). In a limited subset of the data, where maternal \text{ CD4}^+ \text{ cell count had been systematically collected, maternal \text{ CD4}^+ \text{ cell count at delivery was estimated to have a significant impact on the level of the child's } \log_{10} \text{ CD4}^+ \text{ cell count. Children of mothers with } \text{ CD4}^+ \text{ cell counts below } 500 \times 10^6 \text{ cells/l} \text{ had consistently lower values (} p<0.001), \text{ with the difference at peak being about } 450 \times 10^6 \text{ cells/l.}
Table 5.3: Gender, race and log_{10} CD4^+ cell counts over age in uninfected children

<table>
<thead>
<tr>
<th>Factor</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>12.85</td>
<td>1</td>
<td>0.0003</td>
<td>32.52</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Race</td>
<td>103.0</td>
<td>1</td>
<td>&lt;0.0001</td>
<td>12.16</td>
<td>1</td>
<td>0.0586</td>
</tr>
</tbody>
</table>

Figure 5.2: Fitted values of log_{10} CD4^+ cell count over age by gender and race in uninfected children
5.2.3 CD8\(^+\) cell count over age in uninfected children

The running smoother plot in Figure 5.3 shows the underlying patterns for the observed data on CD8\(^+\) cell counts of uninfected children. Similar to that of log\(_{10}\) CD4\(^+\) cell counts, the pattern for log\(_{10}\) CD8\(^+\) cell counts peaked at around the first few weeks of age, then dipped before peaking again and declining gradually thereafter, although with a more protracted second peak over the latter part of the first, at the end of the second year of life.

The natural cubic spline mixed effect model which best represented CD8\(^+\) cell counts of uninfected children had knots placed at one and three weeks, three and 10 months, and two, five and ten years of age (Figure 5.3). The predicted values peak at 1343 per mm\(^3\) at two weeks and then at 11 months at 1219 per mm\(^3\); the subsequent decrease in values is slow, reaching an estimated 739 per mm\(^3\) at five years.
5.2.4 Factors associated with levels and patterns of CD8$^+$ cell counts in uninfected children

Table 5.4 presents the results for the effects of gender and race on CD8$^+$ cell counts. Although the overall main effect of difference in CD8$^+$ cell counts by gender is not significant ($p = 0.4985$), there are statistically significant differences in patterns which depend on age ($p = 0.0067$). Figure 5.4 shows the fitted CD8$^+$ cell counts values over age by gender. The smoothed values suggest the CD8$^+$ levels differ beyond age seven years, which may be more reliable than the fluctuating pattern displayed by the fitted values (Appendix 5.2, Figure A5.1).
Both the main effect ($p < 0.0001$) and the age interaction ($p = 0.0002$) of race are highly statistically significant, indicating levels and patterns of CD8$^+$ cell counts are different for white and black children. This is evident from the fitted values (Figure 5.4) which indicate levels are higher for white children at almost all ages, and this difference becomes smaller at older ages. For example, at one year of age the fitted value for white children is $351 \times 10^6$ cells/l higher than that of black children, whereas at age nine years there is virtually no difference.

From the lack of significance of the interaction of gender and race for CD8$^+$ cell counts ($\chi^2 = 3.77; p = 0.877$), the differences in levels and patterns by race do not vary by the gender of the child.

<p>| Table 5.4: Gender, race and CD8$^+$ cell counts over age in uninfected children |
|-----------------------------------------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Factor</th>
<th>Main effect</th>
<th>Interaction with age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>$\chi^2$</td>
<td>df</td>
</tr>
<tr>
<td>Gender</td>
<td>0.46</td>
<td>1</td>
</tr>
<tr>
<td>Race</td>
<td>105.0</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 5.4: Fitted values of $\log_{10}$ CD8$^+$ cell count over age by gender and race in uninfected children
5.2.5 Absolute lymphocyte counts over age in uninfected children

The running smoother plot of absolute lymphocyte counts of uninfected children shows the age-specific location of the observed data (Figure 5.5). The double peaks of $\log_{10} CD4^+$ and $\log_{10} CD8^+$ cell counts are also present in the pattern for the $\log_{10}$ absolute lymphocytes over time, but now the peaks are more shallow and smoothed out. In contrast to the patterns for $\log_{10} CD4^+$ and $\log_{10} CD8^+$ cell counts, the second peak is higher than the first.

Figure 5.5 displays the fitted values of the natural cubic spline mixed effects model for $\log_{10}$ absolute lymphocyte values which has knots at one and three weeks, one and three months and one and ten years. The predicted absolute lymphocyte values peak at 6470 per mm$^3$ at three weeks, again at 6702 per mm$^3$ at seven months; there was a subsequent gradual decline to 6149 per mm$^3$ at one year falling to 3016 per mm$^3$ by five years.
5.2.6 Factors associated with levels and patterns of absolute lymphocyte counts in uninfected children

The results of the main effects and interactions by age for gender and race for absolute lymphocyte counts are shown in Table 5.5. The overall levels differ by gender ($p = 0.0405$) with the levels for girls above those for boys (Figure 5.6). The difference in levels depends on age ($p = 0.0111$), with slight differences in the first year of life and larger ones beyond age four years, with no differences during the second, third or fourth years of life. For instance, at age eight years, the level for girls of $2617 \times 10^6$ cells/l compares with $2368 \times 10^6$ cells/l for boys.
The main effect of race for absolute lymphocyte counts is significant \((p<0.0001)\) but the age-interaction is not \((p = 0.5056)\), indicating that the difference does not significantly vary over age. As for CD4\(^+\) and CD8\(^+\) cell counts, levels are higher for white than for black children (Figure 5.6). For example there is a difference of \(593 \times 10^6\) cells/l at age five years.

Patterns of \(\log_{10}\) absolute lymphocyte counts over age by gender and race were strikingly similar to those for \(\log_{10}\) CD4\(^+\) cell counts. Allowing for interactions between gender and race for \(\log_{10}\) absolute lymphocyte counts did not statistically significantly improve the model \((\chi^2_7 = 2.14; p = 0.952)\).

**Table 5.5: Gender, race and \(\log_{10}\) absolute lymphocyte counts over age in uninfected children**

<table>
<thead>
<tr>
<th>Factor</th>
<th>(\chi^2)</th>
<th>df</th>
<th>(p)-value</th>
<th>(\chi^2)</th>
<th>df</th>
<th>(p)-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>4.20</td>
<td>1</td>
<td>0.0405</td>
<td>18.20</td>
<td>1</td>
<td>0.0111</td>
</tr>
<tr>
<td>Race</td>
<td>94.15</td>
<td>1</td>
<td>&lt;0.0001</td>
<td>6.29</td>
<td>1</td>
<td>0.5056</td>
</tr>
</tbody>
</table>
Figure 5.6: Fitted values of $\log_{10}$ absolute lymphocyte count over age by gender and race in uninfected children
5.2.7 CD4$^+$ percentage over age in uninfected children

The smoothed values of CD4$^+$ cell counts as a percentage of absolute lymphocyte counts indicate that levels fall rapidly during the first year or so, stabilising beyond age two years (Figure 5.7).

Predicted values of CD4$^+$ percentage, obtained from the fitting of the cubic spline model with knots at one week, three and six months and two and ten years, fell rapidly from around 55% at birth, with a distinct change at 10 weeks of age, levelling off to just below 40% by two years of age, asymptoting thereafter (Figure 5.7).

Figure 5.7: Observed, smoothed and fitted values of CD4$^+$ percentage over age in uninfected children
5.2.8 Factors associated with levels and patterns of CD4⁺ percentage in uninfected children

Levels and patterns of CD4⁺ percentages vary by gender as indicated by the statistically significant main effect \( (p<0.0001) \) and age-interaction \( (p = 0.0139) \) terms (Table 5.6). Differences occur during the first year and beyond the second year of life. Where there are differences, levels for girls are again higher than those of boys (Figure 5.8). The most notable difference in the fitted values is around age six years, with levels for girls about 38.1%, 2.8% higher than the value 35.3% for boys.

Differences by race are significant \( (p<0.0001) \), but do not significantly depend on age \( (p = 0.328) \). Although in conflict with the fitted patterns (Figure 5.8), smoothed values (Appendix 5.2, Figure A5.2), which appear more stable, concurrent with findings for other markers, levels are higher for white children.

The non-significant interaction of gender with race \( (\chi^2 = 6.25; \ p = 0.396) \) means that patterns of one were not modified by the effects of the other.

These findings indicate that the gender and race differences seen in CD4⁺ cell counts were not due to a general increase in absolute lymphocytes.

Table 5.6: Gender, race and percentage of CD4⁺ cell counts over age in uninfected children

<table>
<thead>
<tr>
<th>Factor</th>
<th>Main effect</th>
<th>Interaction with age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \chi^2 )</td>
<td>df</td>
</tr>
<tr>
<td>Gender</td>
<td>10.75</td>
<td>1</td>
</tr>
<tr>
<td>Race</td>
<td>14.62</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 5.8: Fitted values of CD4\(^+\) percentage over age by gender and race in uninfected children.
5.3 Immunological markers in infected children

5.3.1 $\log_{10}$ CD$^+$ cell count over age in infected children

Figure 5.9 shows the raw data and the running smoother for $\log_{10}$ CD$^+$ cell counts for infected children. Observed $\log_{10}$ CD$^+$ cell counts peaked once briefly then fell quickly before recovering slightly at around the middle of the age range.

The appropriate mixed effect natural cubic spline model for $\log_{10}$ CD$^+$ cell counts had knots at three weeks, three months and six years (Figure 5.9). This gives predicted values of CD$^+$ cell counts which peak at seven weeks at 2334 per mm$^3$, falling to 1391 per mm$^3$, approximately half the estimated value for uninfected children, at one year, and reaching a low-point of 278 per mm$^3$, about one third of that for uninfected children at around six years.
5.3.2 Factors associated with levels and patterns of CD4\(^+\) cell counts in infected children

Table 5.7 presents the results for the effects of gender and race for log\(_{10}\) CD4\(^+\) cell counts in infected children. Although the main effect of gender did not reach statistical significance \(p = 0.1192\), the age-interaction term did \(p < 0.0001\), indicating there are significant differences which depend on the age of the child at measurement. The fitted log\(_{10}\) CD4\(^+\) cell count values shown in Figure 5.10 can be compared with the gender-specific super-smoothers of the raw data (Appendix 5.2, Figure A5.3), which may be more reliable. In contrast to the uninfected children, levels are higher for boys than for girls other than very early in life and possibly at ages beyond six years, when levels are unlikely to differ by gender (smoothed values, (Appendix 5.2, Figure A5.3)). Values are most different around age three years, where the fitted level for girls of \(486 \times 10^6\) cells/l compares with a value of \(741 \times 10^6\) cells/l.
for boys. Conclusions are not altered by the account of ART, with the adjusted $p$-values for the main effect and interaction being similar to those of the unadjusted (Appendix 5.3, Table A5.1).

Likewise for gender, significant race differences are seen when an interaction with age is introduced ($p = 0.0006$). Similar to the models for uninfected children, CD4$^+$ cell counts for white children are generally higher than those for black children, differing beyond the first year of life. For instance, at three years of age, the fitted value for white children ($633 \times 10^6$ cells/l) is $160 \times 10^6$ cells/l lower than that for black children. Conflicting patterns in smoothed (Appendix 5.2, Figure A5.4) and fitted values beyond age six make inference uncertain, although there is evidence the levels for black children become higher at older ages (fitted values). Adjustment by ART does not alter the conclusions (Appendix 5.3, Table A5.1). From the testing of the interaction of gender with race, patterns by gender did not vary according to race ($\chi^2_4 = 6.28; p = 0.179$).

Table 5.7: Gender, race and log$_{10}$ CD4$^+$ cell counts over age in infected children

<table>
<thead>
<tr>
<th>Factor</th>
<th>Main effect</th>
<th>Interaction with age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>df</td>
</tr>
<tr>
<td>Gender</td>
<td>2.43</td>
<td>1</td>
</tr>
<tr>
<td>Race</td>
<td>0.13</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 5.10: Fitted values of $\log_{10} CD4^+$ cell count over age by gender and race in infected children
5.3.3 CD8\(^+\) cell count over age in infected children
The smoothed values of the raw CD8\(^+\) cell count data shown in Figure 5.11 are seen to rise sharply to peak before six months of age, decreasing very gradually before reconstituting beyond eight years.

The knots for the cubic spline for CD8\(^+\) cell counts in infected children are positioned at three weeks, three months and two and ten years. Similar to predictions for uninfected children, the peak for the predicted level of CD8\(^+\) cell counts occurs at four months but has a higher value (1723 per mm\(^3\), falling to 1639 per mm\(^3\) at one year, reaching a nadir of 719 per mm\(^3\) just after eight years. The levels were higher than the levels of CD4\(^+\) cell counts from eight months of age onward.
5.3.4 Factors associated with levels and patterns of CD8^+ cell counts in infected children

The results of the analysis on CD8^+ cell counts by gender and race of infected children are displayed in Table 5.8. The differences in levels by gender (p = 0.0034) vary significantly by age (p < 0.0001). The gender- and race-specific fitted curves for CD8^+ cell counts in infected children are plotted in Figure 5.12. As for CD4^+ cell counts of infected children, CD8^+ cell counts are generally higher in boys, with evident distinction from the first year of life to around 10 years of age. For instance, at age four the level for girls is 807 x 10^6 cells/l compared to 1244 x 10^6 cells/l for boys. Difference persisted after adjustment by ART administration (p = 0.0040 and p < 0.0001 for main effect and interaction terms, respectively (Appendix 5.3, Table A5.2)).
From the results of the likelihood ratio tests, there is evidence of varying race differentials for CD8\(^+\) cell counts in infected children (p = 0.0368). Levels for white children are higher beyond the early months of life until around age five years (Figure 5.12). Thereafter, one cannot easily distinguish between levels by race. Where race differences exist, they are not as pronounced as differences by gender, with the largest difference being 326 x 10\(^6\) cells/l at age eight months. Accounting for ART did not affect conclusions on race differences (main effect and interaction significances are p = 0.1984 and p = 0.0132, respectively (Appendix 5.3, Table A5.2)).

Introduction of a gender by race interaction term did not significantly improve the model (\(\chi^2 = 4.14; p = 0.530\)).

Table 5.8: Gender, race and CD8\(^+\) cell counts over age in infected children

<table>
<thead>
<tr>
<th>Factor</th>
<th>Main effect</th>
<th>Interaction with age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\chi^2)</td>
<td>df</td>
</tr>
<tr>
<td>Gender</td>
<td>8.56</td>
<td>1</td>
</tr>
<tr>
<td>Race</td>
<td>2.24</td>
<td>1</td>
</tr>
</tbody>
</table>


Figure 5.12: Fitted values of $\log_{10}$ CD8$^+$ cell count over age by gender and race in infected children.
5.3.5 Absolute lymphocyte counts over age in infected children
The trajectory of absolute lymphocyte counts over age in infected children is similar to that of CD8$^+$ cell counts, with around half a log_{10} upward shift (Figure 5.13).

The cubic spline for absolute lymphocytes has knots at three weeks, three months and ten years. The peak occurs at two months at 6738 per mm$^3$, declining to 5098 per mm$^3$ at one year and falling to the lowest value of 1718 per mm$^3$ at 8.6 years.

Figure 5.13: Observed, smoothed and fitted values of log_{10} absolute lymphocyte count over age in infected children
5.3.6 Factors associated with levels and patterns of absolute lymphocyte counts in infected children

The results for explanatory factors of absolute lymphocyte counts over age in infected children are presented in Table 5.9. The clear differences by gender \( (p = 0.0066) \) are significantly dependent on age \( (p<0.0001) \). As for the lymphocyte specific cell counts, levels of absolute lymphocyte counts in boys are higher than those in girls, as shown in Figure 5.14. However, although they vary over age, gender differences for this immunological marker are more consistent than for other markers. An absolute difference of \( 585 \times 10^6 \) cells/l at age five years is typical of differences over ages beyond one year.

There is a marginal significant difference in levels over age by race \( (p = 0.0769) \). Figure 5.14 illustrates levels in white children are marginally higher before age five years and lower thereafter. The largest difference is just \( 210 \times 10^6 \) cells/l at age eight years. On adjustment by ART administration differences become more significant \( (p = 0.0367) \), with no revision of conclusions (Appendix 5.3, Table A5.3).

The patterns by gender were not altered by race \( (\chi^2_s = 6.23; p = 0.286) \).

Table 5.9: Gender, race and absolute lymphocyte counts over age in infected children

<table>
<thead>
<tr>
<th>Factor</th>
<th>Main effect</th>
<th>Interaction with age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \chi^2 )</td>
<td>df</td>
</tr>
<tr>
<td>Gender</td>
<td>7.37</td>
<td>1</td>
</tr>
<tr>
<td>Race</td>
<td>0.40</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 5.14: Fitted values of $\log_{10}$ absolute lymphocyte count over age by gender and race in infected children
5.3.7 CD4\(^+\) percentage over age in infected children

The running smoother of the raw data for CD4\(^+\) as a percentage of absolute lymphocytes indicates that levels plummet rapidly very early in life then display a slower decline which steadies around school age (Figure 5.15).

The spline representing CD4\(^+\) as a percentage of total lymphocyte, with knots at one week, three and six months, and two, six and eight years, had a sharp fall from around 55% at birth to around 25% by two years falling more gradually thereafter before levelling off at 20% after four years of age, substantially lower than the levels for uninfected children.

**Figure 5.15: Observed, smoothed and fitted values of CD4\(^+\) percentage over age in infected children**

![Figure 5.15: Observed, smoothed and fitted values of CD4\(^+\) percentage over age in infected children](image-url)
5.3.8 Factors associated with levels and patterns of CD4\(^+\) percentage in infected children

Results for CD4\(^+\) percentage assessed by stratification of gender and race are given in Table 5.10. Although the non-significance of the main effect of gender \( (p = 0.9925) \) would indicate no difference in CD4\(^+\) percentage, the highly significant interaction term \( (p = 0.0002) \) suggests age-dependent differences. However, inspection of the gender-specific fitted values (Figure 5.16) shows that the differences are small in magnitude and fluctuate with no consistent pattern, indicating that there are no important differences in CD4\(^+\) percentage by gender.

Similarly, based on the likelihood ratio test results, there appear to be age-dependent race differences \( (p = 0.0001) \), but the inconsistent pattern is revealed by the fitted values in Figure 5.16. Once again, gender and race results are not altered by adjustment by ART administration \( (p = 0.0003 \) and \( p = 0.0001 \), respectfully) (Appendix 5.3, Table A5.1).

It was not necessary to allow for interaction between gender and race \( (\chi^2 = 5.73; p = 0.334) \).

Table 5.10: Gender, race and percentage of CD4\(^+\) cell counts over age in infected children

<table>
<thead>
<tr>
<th>Factor</th>
<th>Main effect</th>
<th>Interaction with age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \chi^2 )</td>
<td>df</td>
</tr>
<tr>
<td>Gender</td>
<td>&lt;0.0001</td>
<td>1</td>
</tr>
<tr>
<td>Race</td>
<td>0.0030</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 5.16: Fitted values of CD4$^+$ percentage over age by gender and race in infected children
5.4 Comparison of immunological markers patterns over age in infected and uninfected children

Although the $\log_{10} CD4^+$ cell count measurements were consistently lower for infected children, there was considerable overlap with measurements for uninfected children. For instance, of the 2781 measurements on infected children, 1511 (54.3%) were above the running 5th percentile for uninfected children (Figure 5.17). Overlap was especially pronounced at young ages (73.9% under one year) but decreased thereafter (56.8% at one to five years and 38.8% over five years). On the other hand, 73.5% of measurements for uninfected children were below the 95th percentile for infected children (Figure 5.18). Nearly 10 percent (8.3%) of $\log_{10} CD4^+$ cell count measurements in uninfected children fell into the CDC moderately (7.5%) or severely (0.8%) immunosuppressed categories, and 45.6% (26.4% and 19.2% respectively) of those in infected children.
Figure 5.17: Observed values of $\log_{10}$ CD4$^+$ cell counts of infected children with 5th percentile for uninfected children
An alternative way of quantifying differences in levels for uninfected and infected children is through the fitting of a spline-based model for data on all children with terms for interactions with infection status. This model included knots at one and three weeks, 3, six and 10 months, 3, six and 10 years and indicated substantial and significant differences in estimated log_{10} CD4^+ cell counts between infected and uninfected children from birth onwards (Table 5.11). The largest absolute difference of 1296 x 10^6 cells/l was at 10 months, relating to the measurements of 993 uninfected and 145 infected children. This difference subsequently decreased, to 509 x 10^6 cells/l at age 10 years, although the number of children tested at that age was small, especially among uninfected children. Infected children who died consistently
had significantly lower log₁₀ CD4⁺ cell counts than infected children who survived
(p<0.0001) (Appendix 5.2, Figure A5.5), and those tested at later ages were thus a selected
population of survivors. Nearly one third (14/47) of children who died did so during the first
year of life. Similar analyses revealed significantly lower values in infected children of CD4⁺
percent (p<0.001) and from three months onward of CD8⁺ cell counts (p<0.001) and
absolute lymphocyte counts (p<0.001) throughout the 12 years of follow-up.

| Table 5.11: Predicted log₁₀ CD4⁺ cell count levels for uninfected and infected children |
|----------------------------------------|-----------------|-----------------|-------------|---|
| Age                                   | Uninfected children (absolute) n | Infected children (absolute) n | Absolute difference | p-value* |
| Intercept, age=0                       | 3.253 (1791) 90 | 3.191 (1552) 15 | 239 | 0.0152 |
| 1 week                                | 3.453 (2836) 251 | 3.345 (2215) 30 | 621 | 0.0001 |
| 3 weeks                               | 3.501 (3168) 351 | 3.406 (2545) 48 | 623 | 0.0003 |
| 3 months                              | 3.447 (2798) 801 | 3.331 (2142) 105 | 656 | <0.0001 |
| 6 months                              | 3.479 (3010) 1022 | 3.265 (1840) 138 | 1170 | <0.0001 |
| 10 months                             | 3.444 (2782) 993 | 3.172 (1486) 145 | 1296 | <0.0001 |
| 3 years                               | 3.160 (1444) 460 | 2.788 (614) 112 | 830 | <0.0001 |
| 6 years                               | 3.007 (1016) 117 | 2.401 (252) 79 | 764 | <0.0001 |
| 10 years                              | 2.906 (805) 18 | 2.472 (296) 35 | 509 | <0.0001 |

* from t-tests
Key points: Patterns of immunological markers in uninfected children

- Levels of CD4⁺ cell counts peaked twice early and did not approach adult values until after age 6 years.

- Estimated patterns for CD8⁺ and absolute lymphocytes were similar to those of CD4⁺ cell counts.

- The early double peaks in predicted CD4⁺, CD8⁺ cell counts and absolute lymphocytes for uninfected children highlights the complex and dynamic nature of a developing immune system.

- CD4⁺ cell counts, absolute lymphocytes and CD4⁺ percentage values were higher for girls.

- Levels of all four markers were higher in white children.

- Gender and race differences were not explained by a general increase in absolute lymphocytes.

- There was a significant association of maternal CD4⁺ cell count with child CD4⁺ cell count

Table 5.12: Summary of gender and race effects for lymphocyte patterns in uninfected children

<table>
<thead>
<tr>
<th>Factor</th>
<th>Level or Pattern</th>
<th>log₁₀ CD4⁺ cell count</th>
<th>log₁₀ CD8⁺ cell count</th>
<th>log₁₀ absolute lymphocyte</th>
<th>CD4⁺ percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>level</td>
<td>girls higher</td>
<td>no</td>
<td>girls higher</td>
<td>girls higher</td>
</tr>
<tr>
<td></td>
<td>pattern</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Race</td>
<td>level</td>
<td>white higher</td>
<td>white higher</td>
<td>white higher</td>
<td>white higher</td>
</tr>
<tr>
<td></td>
<td>pattern</td>
<td>borderline</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>
Key points: Patterns of immunological markers in infected children

- CD4\(^+\) values peaked in children at seven weeks of age, below the peak seen in uninfected children.
- CD8\(^+\) cell counts were higher than CD4\(^+\) cell counts after eight months of age.
- Gender differences in CD4\(^+\), CD8\(^+\) cell counts and absolute lymphocytes levels in infected children were in an opposite direction to those seen in uninfected children, with higher levels in boys.
- Gender differences for CD4\(^+\) percentage were not consistent for infected children.
- Levels for all four markers were generally higher for white than black children, though there was evidence of a switch in later childhood.

Table 5.13: Summary of gender and race effects for lymphocyte patterns in infected children

<table>
<thead>
<tr>
<th>Factor</th>
<th>Level or Pattern</th>
<th>log(_{10}) CD4(^+) cell count</th>
<th>log(_{10}) CD8(^+) cell count</th>
<th>log(_{10}) absolute lymphocyte</th>
<th>CD4(^+) percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>level</td>
<td>boys higher</td>
<td>boys higher</td>
<td>boys higher</td>
<td>Not consistent</td>
</tr>
<tr>
<td></td>
<td>pattern</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Race</td>
<td>level</td>
<td>white higher</td>
<td>white higher until 3 years</td>
<td>borderline</td>
<td>white higher</td>
</tr>
<tr>
<td></td>
<td>pattern</td>
<td>yes</td>
<td>yes</td>
<td>borderline</td>
<td>yes</td>
</tr>
</tbody>
</table>
Key points: Comparison of patterns of immunological markers in uninfected and infected children

- Levels were substantially lower in infected children and patterns were distinctly different.

- There was considerable overlap between measurements for uninfected and infected children with more than half the measurements for infected children exceeding the 5th percentile for uninfected children.

- However, estimated CD4\(^+\) cell counts differed by as much as almost 1300 \(\times 10^6\) cells/l at 10 months of age.

- Significant differences existed by race, which, when compared with gender differences, were more consistent by infection status with generally higher levels for white than black children, though there was evidence of a switch in later childhood.

- Race differences were independent of gender.
Chapter 6 Results: Early-life determinants of clinical disease progression in infected children

This analysis aims to identify early life predictors of overall, rapid and long-term clinical progression. Differences in effect of laboratory markers on risks of progression to serious disease by gender and race were also addressed.

6.1 Data
Since severe illness and HIV-related death in the ECS cohort has been almost entirely restricted to those born up to the mid-1990s - due to widespread prevention of vertical transmission and serious disease progression - only data available on children born before 1997 were included in this analysis. Laboratory measurements and clinical follow-up information, up to the end of August 2002, were available for the 161 infected children born before 1997. Of these, 65 (40.4%) progressed to serious disease or died during follow-up (Table 6.1). The Kaplan-Meier estimate representing progression to stage C-defining illness or death for this subset of children is displayed in Figure 6.1. This shows that by one year of age, an estimated 17.7% (95% CI: 12.6% to 24.6%) of children have progressed to serious disease or have died; by age five, this will have increased to 36.3% (95% CI: 29.2% to 44.4%) and to 44.1% (95% CI: 36.3% to 52.7%) by 10 years. These 161 children are a subset of the 170 children included in the clinical disease patterns analysis presented in Chapter 3, and survival estimates are very similar.
Table 6.1: Characteristics of 161 infected children born before 1997

<table>
<thead>
<tr>
<th>Factor</th>
<th>Rapid Progressors</th>
<th>Long-term Progressors</th>
<th>Progressors</th>
<th>Non-Progressors</th>
<th>Total</th>
<th>$n = 161$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 28$ (43.1%)</td>
<td>$n = 37$ (56.9%)</td>
<td>$n = 65$ (40.4%)</td>
<td>$n = 96$ (59.6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>12 (42.9%)</td>
<td>20 (54.1%)</td>
<td>32 (49.2%)</td>
<td>53 (55.2%)</td>
<td>85 (52.8%)</td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>16 (57.1%)</td>
<td>17 (45.9%)</td>
<td>33 (50.8%)</td>
<td>43 (44.8%)</td>
<td>76 (47.2%)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>23 (88.5%)</td>
<td>29 (78.4%)</td>
<td>52 (82.3%)</td>
<td>61 (65.6%)</td>
<td>113 (72.4%)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>3 (11.5%)</td>
<td>7 (18.9%)</td>
<td>10 (15.9%)</td>
<td>22 (23.7%)</td>
<td>32 (20.5%)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>2 (7.1%)</td>
<td>1 (2.7%)</td>
<td>3 (4.6%)</td>
<td>13 (20.0%)</td>
<td>16 (9.9%)</td>
<td></td>
</tr>
<tr>
<td>Prematurity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;=34 weeks</td>
<td>1 (3.6%)</td>
<td>5 (13.5%)</td>
<td>6 (9.2%)</td>
<td>5 (5.2%)</td>
<td>11 (6.8%)</td>
<td></td>
</tr>
<tr>
<td>&gt;34,&lt;=37 weeks</td>
<td>7 (25.0%)</td>
<td>7 (18.9%)</td>
<td>14 (21.5%)</td>
<td>17 (17.7%)</td>
<td>31 (19.3%)</td>
<td></td>
</tr>
<tr>
<td>&gt;37 weeks</td>
<td>20 (71.4%)</td>
<td>24 (64.9%)</td>
<td>44 (67.7%)</td>
<td>72 (75%)</td>
<td>116 (72.0%)</td>
<td></td>
</tr>
<tr>
<td>unknown</td>
<td>0 (0.0%)</td>
<td>1 (2.7%)</td>
<td>1 (1.5%)</td>
<td>2 (2.1%)</td>
<td>3 (1.9%)</td>
<td></td>
</tr>
<tr>
<td>ART ever before</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>progression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>27 (96.4%)</td>
<td>19 (51.4%)</td>
<td>46 (70.8%)</td>
<td>24 (25.0%)</td>
<td>70 (43.5%)</td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>0 (0)</td>
<td>16 (43.2%)</td>
<td>16 (24.6%)</td>
<td>7 (7.3%)</td>
<td>23 (14.3%)</td>
<td></td>
</tr>
<tr>
<td>Combination</td>
<td>1 (3.6%)</td>
<td>2 (5.4%)</td>
<td>3 (4.6%)</td>
<td>65 (67.7%)</td>
<td>68 (42.2%)</td>
<td></td>
</tr>
<tr>
<td>ART neonatal prophylaxis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>28 (100)</td>
<td>36 (97.3%)</td>
<td>64 (98.5%)</td>
<td>86 (89.6%)</td>
<td>150 (93.2%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0 (0)</td>
<td>1 (2.7%)</td>
<td>1 (1.5%)</td>
<td>10 (10.4%)</td>
<td>11 (6.8%)</td>
<td></td>
</tr>
<tr>
<td>Maternal CD4+ cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>count&lt;sup&gt;$\ast$&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;500 x 10&lt;sup&gt;6&lt;/sup&gt; cells/l</td>
<td>2 (40.0%)</td>
<td>14 (58.3%)</td>
<td>5 (50.0%)</td>
<td>11 (57.9%)</td>
<td>16 (55.2%)</td>
<td></td>
</tr>
<tr>
<td>≥500 x 10&lt;sup&gt;6&lt;/sup&gt; cells/l</td>
<td>3 (60.0%)</td>
<td>10 (41.7%)</td>
<td>5 (50.0%)</td>
<td>8 (42.1%)</td>
<td>13 (44.8%)</td>
<td></td>
</tr>
<tr>
<td>(not recorded)</td>
<td>23</td>
<td>13</td>
<td>55</td>
<td>77</td>
<td>132</td>
<td></td>
</tr>
</tbody>
</table>

* rapid progression defined as progression to serious disease or death before one year of age

◊ long-term progression defined as progression to serious disease or death beyond one year of age

§ percentages exclude unrecorded measurements
Figure 6.1: The Kaplan-Meier time-to-serious disease progression or death curve for 161 infected children born before 1997.

Of the 161 infected children included in this determinants analysis, 28 (17.4%) progressed to serious disease or died within the first year of life (Table 6.1). These rapidly progressing children represent 43% of the 65 who progressed to serious disease or who died at some point during follow-up. The corresponding survival curve, estimated by Kaplan-Meier is shown in Figure 6.2.
By one year of age, five (3.8%) of the 133 infected children who had not previously progressed were lost to follow-up. This left 128 (79.5%) of the original 161 infected children who had not progressed to serious disease or death and remained in follow-up for the analysis of progression beyond one year of age. The remaining 37 (57%) of the 65 children who ever progressed, developed serious disease or died after the first year of life (Table 6.1). These long-term progressors constitute 23.0% of the total 161 included children. The Kaplan-Meier survival curve estimate for progression beyond the first year of life is presented in Figure 6.3.
Overall, a similar number of girls (33/76, 43%) and boys (32/85, 38%) progressed to serious disease or death (Table 6.1). Although there was some evidence of differences in proportion by race with 52/113 (46%) white, 10/32 (31%) black and only 3/16 (23%) of the other (mainly Asian) children progressing (p = 0.065), this reflects later enrolment of children of non-white ethnicity when ART use was increasingly widespread (145).

Table 6.2 provides the early-life laboratory and clinical data of the children included in this analysis. First laboratory measurements, taken at a mean age of 1.5 months (range: birth to 5.8 months), were used provided they were taken prior to the outcome of serious disease progression. The mean first CD4\(^+\) percentage for rapid progressors was 35.0% (IQR: 24.6 to
43.5), substantially lower than the 43.1% (36.5% to 54.5%) for children progressing to serious
disease or death beyond one year of age.

Table 6.2: Laboratory data on the first six months of life for 161 infected children born
before 1997

<table>
<thead>
<tr>
<th>Factor</th>
<th>Rapid Progressors</th>
<th>Long-term Progressors</th>
<th>Progressors</th>
<th>Non-Progressors</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 28 )</td>
<td>( n = 37 )</td>
<td>( n = 65 )</td>
<td>( n = 96 )</td>
<td>( n = 161 )</td>
</tr>
<tr>
<td>Median age (months) at last visit (IQR)</td>
<td>12.9 (4.9 to 41.2)</td>
<td>61.5 (43.4 to 95.2)</td>
<td>44.8 (17.2 to 85.0)</td>
<td>113.8 (76.0 to 140.4)</td>
<td>85.0 (37.5 to 85.0)</td>
</tr>
<tr>
<td>Median number of visits (IQR)</td>
<td>8.5 (3.5 to 24.0)</td>
<td>23.0 (17.0 to 3.4)</td>
<td>19.0 (9.0 to 30.0)</td>
<td>28.0 (16.5 to 47.0)</td>
<td>24.0 (14.0 to 37.0)</td>
</tr>
<tr>
<td>first available ( \log_{10} ) CD4(^+ ) cell count (x 10(^6 ) cells/l) mean (IQR)</td>
<td>3.25 (3.01 to 3.46)</td>
<td>3.39 (3.26 to 3.53)</td>
<td>3.33 (3.18 to 3.53)</td>
<td>3.38 (3.28 to 3.50)</td>
<td>3.36 (3.25 to 3.52)</td>
</tr>
<tr>
<td>first available ( \log_{10} ) CD8(^+ ) cell count (x 10(^6 ) cells/l) mean (IQR)</td>
<td>3.13 (3.09 to 3.30)</td>
<td>3.17 (3.08 to 3.29)</td>
<td>3.15 (3.08 to 3.29)</td>
<td>3.15 (2.97 to 3.32)</td>
<td>3.15 (2.99 to 3.30)</td>
</tr>
<tr>
<td>first available log(_{10}) absolute lymphocyte count (x 10(^6 ) cells/l) mean (IQR)</td>
<td>3.69 (3.63 to 3.85)</td>
<td>3.76 (3.69 to 3.89)</td>
<td>3.73 (3.67 to 3.88)</td>
<td>3.79 (3.69 to 3.90)</td>
<td>3.77 (3.68 to 3.89)</td>
</tr>
<tr>
<td>first available CD4(^+ ) percentage mean (IQR)</td>
<td>36.2 (25.8 to 45.0)</td>
<td>44.6 (35.9 to 54.6)</td>
<td>41.1 (28.9 to 51.9)</td>
<td>41.5 (32.8 to 50.0)</td>
<td>41.3 (30.2 to 50.4)</td>
</tr>
</tbody>
</table>
Table 6.2 cont: Clinical data on the first six months of life for 161 infected children born before 1997

<table>
<thead>
<tr>
<th>Factor</th>
<th>Rapid Progressors</th>
<th>Long-term Progressors</th>
<th>Progressors</th>
<th>Non-Progressors</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 28 ) (43.1%)</td>
<td>( n = 37 ) (56.9%)</td>
<td>( n = 65 ) (40.4%)</td>
<td>( n = 96 ) (59.6%)</td>
<td>( n = (161) )</td>
</tr>
<tr>
<td>Persistent Hepatomegaly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>23 (82.1%)</td>
<td>26 (70.3%)</td>
<td>49 (75.4%)</td>
<td>83 (86.5%)</td>
<td>132 (82.0%)</td>
</tr>
<tr>
<td>Yes</td>
<td>5 (17.9%)</td>
<td>11 (29.7%)</td>
<td>16 (24.6%)</td>
<td>13 (13.5%)</td>
<td>29 (18.0%)</td>
</tr>
<tr>
<td>Persistent Splenomegaly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>25 (89.3%)</td>
<td>33 (89.2%)</td>
<td>58 (89.2%)</td>
<td>85 (88.5%)</td>
<td>143 (88.8%)</td>
</tr>
<tr>
<td>Yes</td>
<td>3 (10.7%)</td>
<td>4 (10.8%)</td>
<td>7 (10.8%)</td>
<td>11 (11.5%)</td>
<td>18 (11.2%)</td>
</tr>
<tr>
<td>Persistent Lymphadenopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>24 (85.7%)</td>
<td>29 (78.4%)</td>
<td>53 (81.5%)</td>
<td>85 (88.5%)</td>
<td>138 (85.7%)</td>
</tr>
<tr>
<td>Yes</td>
<td>4 (14.3%)</td>
<td>8 (21.6%)</td>
<td>12 (18.5%)</td>
<td>11 (11.5%)</td>
<td>23 (14.3%)</td>
</tr>
<tr>
<td>Persistence of enlarged axillary nodes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>23 (82.1%)</td>
<td>30 (81.1%)</td>
<td>53 (81.5%)</td>
<td>79 (82.3%)</td>
<td>132 (82.0%)</td>
</tr>
<tr>
<td>Yes</td>
<td>5 (17.9%)</td>
<td>7 (18.9%)</td>
<td>12 (18.5%)</td>
<td>17 (17.7%)</td>
<td>29 (18.0%)</td>
</tr>
</tbody>
</table>

6.2 Association of overall disease progression with early immunological and clinical measurements, child and maternal factors

6.2.1 Univariable Cox regression analysis
Table 6.3 presents the univariable results of the Cox proportional hazards model fitting for overall disease progression. For each of the factors in turn, results are given in the form of the
risk ratios, confidence intervals and p-values. In univariable analysis, a one log increase in the first available log\(_{10}\) absolute lymphocyte count (p = 0.049) (Table 6.3) and the first CD4\(^+\) percentage category (HR=2.15, p = 0.107) (Figure 6.4) were predictive of subsequent progression to serious disease or death at any point during follow-up, but log increases in first available log\(_{10}\) CD4\(^+\) (p = 0.253) and CD8\(^+\) (p = 0.953) cell count were not (Table 6.3). The risk of progression to serious disease or death was associated with early persistent hepatomegaly (HR=1.83, p = 0.028) (Figure 6.5) and lymphadenopathy (HR=1.76, p = 0.052) (Figure 6.6), but not axillary nodes enlargement in particular (p = 0.460) (Figure 6.7) or splenomegaly (p = 0.896) (Figure 6.8). Overall, progression did not significantly differ by gender (p = 0.506) (Figure 6.9), race (p = 0.240) (Figure 6.10), prematurity (p = 0.337 and p = 0.353) (Figure 6.11) or neonatal ART prophylaxis (Figure 6.12) (p = 0.100).

As maternal CD4\(^+\) cell counts near the time of delivery were only recorded routinely in the participating centres since around 1995, there was a limited subset (29/161 (18.0%)) of mothers with available measurements at delivery. Although multivariable analysis involving maternal CD4\(^+\) cell count was thus not possible, in univariable analysis a one log increase in maternal CD4\(^+\) count was compatible with a 75% reduction in the risk of progressing to serious disease for the child although this was not statistically significant (p = 0.171).
Table 6.3: Univariable Cox regression results for overall progression to serious disease or death by infant and maternal factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>1.18 (0.73 to 1.91)</td>
<td>0.506</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.67 (0.34 to 1.31)</td>
<td>0.240</td>
</tr>
<tr>
<td>Prematurity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-term</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Moderately premature</td>
<td>1.32 (0.71 to 2.43)</td>
<td>0.377</td>
</tr>
<tr>
<td>Severely premature</td>
<td>1.44 (0.67 to 3.09)</td>
<td>0.353</td>
</tr>
<tr>
<td>ART administered neonatally</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.19 (0.03 to 1.37)</td>
<td>0.100</td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>first available log_{10} CD4⁺ cell count (x 10^6 cells/l) (per log increase)</td>
<td>0.44 (0.11 to 1.79)</td>
<td>0.253</td>
</tr>
<tr>
<td>first available log_{10} CD8⁺ cell count (x 10^6 cells/l) (per log increase)</td>
<td>0.96 (0.29 to 3.18)</td>
<td>0.953</td>
</tr>
<tr>
<td>first available log_{10} absolute lymphocyte count (x 10^6 cells/l) (per log increase)</td>
<td>0.31 (0.10 to 0.99)</td>
<td>0.049</td>
</tr>
<tr>
<td>first available CD4 percentage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20%</td>
<td>2.15 (0.85 to 5.48)</td>
<td>0.107</td>
</tr>
<tr>
<td>&gt;= 20%</td>
<td>Reference</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.3 cont: Univariable Cox regression results for overall progression to serious disease or death by infant and maternal factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)*</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatomegaly (2 or more episodes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.83 (1.07 to 3.12)</td>
<td>0.028</td>
</tr>
<tr>
<td>Splenomegaly (2 or more episodes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.05 (0.48 to 2.31)</td>
<td>0.896</td>
</tr>
<tr>
<td>Lymphadenopathy (2 or more episodes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.76 (0.99 to 3.13)</td>
<td>0.052</td>
</tr>
<tr>
<td>Axillary nodes (2 or more episodes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.27 (0.67 to 2.39)</td>
<td>0.460</td>
</tr>
<tr>
<td>ART ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.75 (0.93 to 3.32)</td>
<td>0.084</td>
</tr>
<tr>
<td>Combination</td>
<td>0.61 (0.16 to 2.29)</td>
<td>0.466</td>
</tr>
<tr>
<td>Maternal CD4&lt;sup&gt;+&lt;/sup&gt; cell count at delivery (x 10&lt;sup&gt;6&lt;/sup&gt; cells/l) (per log increase)</td>
<td>0.274 (0.043 to 1.746)</td>
<td>0.171</td>
</tr>
</tbody>
</table>
Figure 6.4: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with a first CD4* percentage above 20% (brown) and equal to or below 20% (green) in the first six months of life.

Figure 6.5: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with (brown) and those without (green) early persistence of hepatomegaly in the first six months of life.
Figure 6.6: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with (brown) and those without (green) persistence of lymphadenopathy in the first six months of life.

Proportion progressing to serious disease or death

Figure 6.7: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with (brown) and those without (green) persistence of enlarged axillary nodes in the first six months of life.

Proportion progressing to serious disease or death
Figure 6.8: Kaplan-Meier survival plot of progression to CDC clinical category C or death for those with (brown) and those without (green) persistence of splenomegaly in the first six months of life.

Figure 6.9: Kaplan-Meier survival plot of progression to CDC clinical category C or death for girls (brown) and boys (green).
Figure 6.10: Kaplan-Meier survival plot of progression to CDC clinical category C or death for black (brown) and white (green) children.

Figure 6.11: Kaplan-Meier survival plot of progression to CDC clinical category C or death for full term (orange), moderately premature (brown) children and severely premature (green).
6.2.2 Assessing gender- and race-specific effects on prognostic value of clinical and immunological markers

In analyses of models including terms representing interactions of gender and race with each of the clinical and immunological measurements, the effects on disease progression of the early immunological markers and clinical indicators were not modified by gender or race ($p > 0.3$, all tests for gender, $p > 0.5$, all tests for race).

In a subset of 22 children who had not progressed to serious disease or death by age two years and who had available laboratory measurements at that age, the predictive values of markers did not depend on gender or race (all $p > 0.2$).
6.2.3 Multivariable Cox regression analysis: Simultaneous predictive value of early clinical and immunological variables

The significant univariable factors, first available log₁₀ absolute lymphocyte count, first CD₄⁺ percentage category, early persistent hepatomegaly and lymphadenopathy, were proposed as candidates for multivariable analysis. Although not significant univariably, ART administration at any time prior to serious disease progression was allowed for in all multivariable analyses to account for any confounding effects on associations between markers and disease progression. The most appropriate multivariable model for overall disease progression by early clinical and immunological factors, allowing for ART administration, was identified as the one with CD₄⁺ percentage category, absolute lymphocyte count and lymphadenopathy (Appendix 6.1, Table A6.1.1).

Children with a first CD₄⁺ percentage below 20% had nearly a three-fold increased risk of progression to serious disease or death compared with those with a value above 20% \( (p = 0.041) \) (Table 6.4). A one log₁₀ (ten-fold) increase in the first absolute lymphocyte count independently reduced the risk of progression by 77% \( (p = 0.014) \) and lymphadenopathy at two or more visits within the first six months of life increased the risk by 88% \( (p = 0.045) \). Additionally, hepatomegaly \( (HR = 1.885, p = 0.035; \text{Model } \chi^2 = 17.32; p = 0.0039) \) and lymphadenopathy had similar and interchangeable, but not independent predictive effects. Based on model comparisons, a CD₄⁺ percentage category was substantially more predictive of disease progression than absolute lymphocyte count or lymphadenopathy (deviance differences = 81.4, 5.2 and 3.2 respectively). Allowing for PCP prophylaxis and for the use of IVIG before disease progression had occurred did not significantly alter the estimates (Appendix 6.2, Table A6.2.3).
Table 6.4: Multivariable Cox regression results for overall progression to serious disease or death

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>first available log_{10} absolute lymphocyte count (x 10^6 cells/l) (per log increase)</td>
<td>0.233 (0.074 to 0.741)</td>
<td>0.014</td>
</tr>
<tr>
<td>first available CD4(^+) percentage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20%</td>
<td>2.79 (1.04 to 7.46)</td>
<td>0.041</td>
</tr>
<tr>
<td>&gt;= 20%</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Lymphadenopathy (2 or more episodes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.883 (1.015 to 3.493)</td>
<td>0.045</td>
</tr>
<tr>
<td>ART ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.495 (0.732 to 3.054)</td>
<td>0.270</td>
</tr>
<tr>
<td>Combination</td>
<td>0.504 (0.121 to 2.091)</td>
<td>0.345</td>
</tr>
</tbody>
</table>

* Model \( \chi^2 = 17.51; \) 5 df; \( p = 0.0036 \)

Illustrative example

For instance, keeping other variables constant (no early persistent lymphadenopathy and an absolute lymphocyte count of 6000 cells (3.78 in log units), say) and accounting for treatment, compared to a child with an early CD4\(^+\) percentage above 20\% (with a risk of progression by age 5 years of 25\%), one with a CD4\(^+\) percentage value below 20\% would be 2.79 times more likely to progress (with a risk of progression by age 5 years of 70\% (2.79 x 25\%).
6.3 Rapid progression

The majority (19/28,68%) of those progressing in the first year of life had opportunistic infections, mostly PCP (12/19,63%); there were four cases of encephalopathy; two with serious bacterial infections and the remaining three children died at home most likely of an undiagnosed opportunistic infection.

6.3.1 Univariable Cox regression analysis

Table 6.5 presents the univariable Cox proportional hazards results for rapid progression. Neither gender \( (p = 0.250) \), nor race \( (p = 0.172) \) of the child were predictive of rapid disease progression (Table 6.5). Of the first available laboratory measurements, \( \log_{10} \) CD8\(^+\) cell count (per log increase) \( (p = 0.689) \), was not associated with serious disease progression during the first year of life. On the other hand, the predictive value of a one log increase in the first CD4\(^+\) cell count was significant and reduced risk of rapid progression by 89% \( (HR = 0.11, p = 0.021) \) while a first CD4\(^+\) percentage of below 20% was associated with a fourfold increase in risk \( (HR = 4.16, p = 0.002) \). The reduction in risk of rapid progression for a one log increase in the first absolute lymphocyte measurement was 81% \( (HR = 0.19, p = 0.003) \). None of the recurrent early clinical signs were associated with rapid disease progression: hepatomegaly \( (p = 0.482) \), splenomegaly \( (p = 0.788) \), lymphadenopathy \( (p = 0.562) \), and enlarged axillary nodes \( (p = 0.529) \). Prematurity was not associated with rapid disease progression \( (p = 0.532 \) and 0.518).
Table 6.5: Univariable Cox regression results for rapid progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>1.550 (0.735 to 3.270)</td>
<td>0.250</td>
</tr>
<tr>
<td><strong>Prematurity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-term</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Moderately premature</td>
<td>0.513 (0.064 to 4.147)</td>
<td>0.532</td>
</tr>
<tr>
<td>Severely premature</td>
<td>1.320 (0.569 to 4.147)</td>
<td>0.518</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.436 (0.133 to 1.433)</td>
<td>0.172</td>
</tr>
<tr>
<td><strong>First available log_{10} CD4^+ cell count</strong></td>
<td>0.107 (0.016 to 0.713)</td>
<td>0.021</td>
</tr>
<tr>
<td>(per log increase)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First available log_{10} CD8^+ cell count</strong></td>
<td>0.654 (0.081 to 5.256)</td>
<td>0.689</td>
</tr>
<tr>
<td>(per log increase)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First available log_{10} absolute lymphocyte count</strong></td>
<td>0.189 (0.041 to 0.874)</td>
<td>0.033</td>
</tr>
<tr>
<td>(per log increase)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First available CD4^+ percentage</strong></td>
<td>4.158 (1.664 to 10.394)</td>
<td>0.002</td>
</tr>
<tr>
<td>&lt; 20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;= 20%</td>
<td>Reference</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.5 cont: Univariable Cox regression results for rapid progression

<table>
<thead>
<tr>
<th>Factor*</th>
<th>Hazard ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(95% CI)</td>
<td></td>
</tr>
<tr>
<td>Recurrent hepatomegaly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.399 (0.548 to 3.570)</td>
<td>0.482</td>
</tr>
<tr>
<td>Recurrent splenomegaly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.686 (0.371 to 3.694)</td>
<td>0.788</td>
</tr>
<tr>
<td>Recurrent lymphadenopathy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.356 (0.485 to 3.794)</td>
<td>0.562</td>
</tr>
<tr>
<td>Recurrence of enlarged axillary nodes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.362 (0.521 to 3.565)</td>
<td>0.529</td>
</tr>
<tr>
<td>ART ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/ Monotherapy</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Combination</td>
<td>2.624 (0.325 to 21.159)</td>
<td>0.365</td>
</tr>
</tbody>
</table>

* since no rapid progressors received ART prophylaxis for vertical transmission, this could not be assessed

6.3.2 Multivariable Cox regression analysis
Neither gender nor race modified the effects of any of the early laboratory or clinical markers (p > 0.1, all tests for gender, p > 0.4, all tests for race).

From univariable results, first available CD4⁺ cell count, CD4⁺ percentage category and absolute lymphocyte measurements in the first six months of life were considered as candidate terms for the multivariable model for rapid progression. The multivariable model which best predicted rapid progression to serious disease or death included CD4⁺ percentage category and log\(_{10}\) absolute lymphocyte count adjusted for by ART (Table 6.6). A first CD4⁺ percentage below 20%
was associated with more than a six-fold increased risk of progression (HR = 6.62, \( p < 0.001 \)) and, independently, a one log increase in \( \log_{10} \) absolute lymphocyte count decreased risk by 89\% (HR = 0.11, \( p = 0.012 \)). Adjusting for PCP and IVIG prophylaxes did not alter the risk estimates (Appendix 6.2, Table A6.2.5).

### Table 6.6: Multivariable Cox regression results for rapid progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio ( (95% \text{ CI}) )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>First available absolute lymphocyte count (x ( 10^6 ) cells/l) (per log increase)</td>
<td>0.111 (0.020 to 0.613)</td>
<td>0.012</td>
</tr>
<tr>
<td>First available ( \log_{10} ) CD4(^+) percentage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 20%</td>
<td>6.622 (2.350 to 18.661)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>( \geq 20% )</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>ART ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/ Monotherapy</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Combination</td>
<td>3.096 (0.381 to 25.161)</td>
<td>0.291</td>
</tr>
</tbody>
</table>

\(^*\) Model \( \chi^2 = 13.61\); 3 df, \( p = 0.0035 \)

### 6.4 Progression after one year of age

Of the 37 children who progressed to serious disease or death beyond the first year of life, 10 (27\%) acquired encephalopathy, nine (24\%) endured serious recurrent bacterial infections, and 13 (35\%) had opportunistic infections (only three of which were PCP); four progressed with another C-defining illness and one died of unspecified HIV-related causes.

### 6.4.1 Univariable Cox regression analysis

Univariably, gender and race were not significantly associated with long-term disease progression (Table 6.7). Neonatal ART administration did not predict progression to serious
disease beyond one year of age. Of the early recurrent clinical symptoms of HIV infection, neither splenomegaly \( (p = 0.979) \) nor enlarged axillary nodes \( (p = 0.622) \) related to disease progression long-term. None of the first available laboratory values predicted disease progression beyond one year: \( \log_{10} \text{CD4}^+ \) cell count (per log increase) \( (p = 0.558) \); \( \log_{10} \text{CD8}^+ \) cell count (per log increase) \( (p = 0.721) \); \( \log_{10} \) absolute lymphocyte count (per log increase) \( (p = 0.366) \) and \( \text{CD4}^+ \) percentage \( (p = 0.667) \). Factors significantly associated with progression to serious disease or death beyond one year of age in univariable Cox proportional hazards regression were early persistent hepatomegaly \( \text{HR} = 2.22, p = 0.024 \) and early persistent lymphadenopathy \( \text{HR} = 2.16, p = 0.048 \). Moderate prematurity was marginally associated with late disease progression \( \text{HR} = 2.24, p = 0.080 \).
Table 6.7: Univariable Cox regression results for long-term progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>0.959 (0.506 to 1.817)</td>
<td>0.898</td>
</tr>
<tr>
<td><strong>Prematurity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-term</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Moderately premature</td>
<td>2.243 (0.908 to 5.539)</td>
<td>0.080</td>
</tr>
<tr>
<td>Severely premature</td>
<td>1.302 (0.546 to 3.103)</td>
<td>0.552</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.864 (0.370 to 2.018)</td>
<td>0.735</td>
</tr>
<tr>
<td><strong>ART administered neonatally</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.351 (0.046 to 2.662)</td>
<td>0.311</td>
</tr>
<tr>
<td><strong>First available log_{10} CD4^+ cell count (x 10^6 cells/l) (per log increase)</strong></td>
<td>1.654 (0.307 to 8.908)</td>
<td>0.558</td>
</tr>
<tr>
<td><strong>First available log_{10} CD8^+ cell count (x 10^6 cells/l) (per log increase)</strong></td>
<td>1.263 (0.351 to 4.546)</td>
<td>0.721</td>
</tr>
<tr>
<td><strong>First available log_{10} absolute lymphocyte count (x 10^6 cells/l) (per log increase)</strong></td>
<td>0.479 (0.097 to 2.367)</td>
<td>0.366</td>
</tr>
<tr>
<td><strong>First available CD4^+ percentage</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 20%</td>
<td>0.650 (0.091 to 4.626)</td>
<td>0.667</td>
</tr>
<tr>
<td>&gt;= 20%</td>
<td>Reference</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.7 cont: Univariable Cox regression results for long-term progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrent hepatomegaly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2.224 (1.109 to 4.459)</td>
<td>0.024</td>
</tr>
<tr>
<td>Recurrent splenomegaly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.014 (0.352 to 2.920)</td>
<td>0.979</td>
</tr>
<tr>
<td>Recurrent lymphadenopathy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2.155 (1.007 to 4.610)</td>
<td>0.048</td>
</tr>
<tr>
<td>Recurrence of enlarged axillary nodes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.239 (0.528 to 2.908)</td>
<td>0.622</td>
</tr>
<tr>
<td>ART ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.948 (0.983 to 3.860)</td>
<td>0.056</td>
</tr>
<tr>
<td>Combination</td>
<td>0.449 (0.099 to 2.049)</td>
<td>0.302</td>
</tr>
</tbody>
</table>

6.4.2 Multivariable Cox regression analysis
The prognostic ability of early life markers for disease progression after one year of age were not modified by gender or race ($p > 0.2$, all tests for gender, $p > 0.1$, all tests for race).

From univariable analysis, early persistence of hepatomegaly and lymphadenopathy, and prematurity were regarded as potential terms in the multivariable analysis for disease progression beyond one year of age. The appropriate multivariable model for representing prediction of long-term serious disease progression or death included just hepatomegaly with adjustment for ART.
administration (Table 6.8). In multivariable analysis, adjusting for ART, persistent hepatomegaly (HR = 2.15, \( p = 0.040 \)) was the single significant indicator independently associated with disease progression after the first year of life. Separately, persistent lymphadenopathy was associated with a doubling of risk (HR = 2.01), but this did not reach statistical significance (\( p = 0.087 \)). Conclusions were not altered by additional adjustment by PCP prophylaxis and IVIG for the prevention of bacterial infection (Appendix 6.2, Table A6.2.7).

### Table 6.8: Multivariable Cox regression results for long-term progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrent hepatomegaly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2.150 (1.035 to 4.470)</td>
<td>0.040</td>
</tr>
<tr>
<td>ART ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.672 (0.791 to 3.534)</td>
<td>0.179</td>
</tr>
<tr>
<td>Combination</td>
<td>0.369 (0.081 to 1.687)</td>
<td>0.199</td>
</tr>
</tbody>
</table>

\* Model \( \chi^2 = 12.00; 3 \text{ df}; p = 0.0074 \)

### 6.5 Threshold for beyond six months CD4\(^+\) percentage

Using all clinical status and CD4\(^+\) percentage data over the entire follow-up beyond six months of age, classification tree analysis was used to investigate the optimal value for splitting the CD4\(^+\) percentage measurements into two groups in terms of predicting subsequent disease progression. At any age before clinical progression, the threshold value of 10% for CD4\(^+\) percentage was found to best predict progression to serious disease or death, with a misclassification rate of just 12.2%. The corresponding risk of progression in a child with a CD4\(^+\) percentage equal to or below 10% was nearly five times the risk of progression in one with a CD4\(^+\) percentage above 10%. (HR = 4.83, \( p<0.0001 \)).
**Key points: Early-life determinants of clinical disease progression in infected children**

- Overall, early-life measures of CD4⁺ percentage and log absolute lymphocyte counts together with hepatomegaly or lymphadenopathy predict disease progression.
- First CD4⁺ percentage was most informative of the laboratory and clinical indicators.
- Earliest measures of CD4⁺ percentage and log absolute lymphocyte counts in the first six months of life are independently predictive of progression to serious disease or death by one year of age, when progression was mostly due to PCP and other opportunistic infections.
- Early persistence of hepatomegaly predicted serious disease progression beyond the first year of life, when progression was associated with recurrent bacterial infections and encephalopathy.
- Predictive values of clinical and immunological markers did not differ by gender or race.
- A CD4⁺ percentage cut-off of 10% best identified the risk of future disease progression beyond six months of age.
Chapter 7 Conclusions on the analysis of paediatric HIV data

7.1 Conclusions
These findings from the ECS contribute substantially to understanding of vertically-acquired HIV infection. Analyses were based on serial data from birth up to 16 years of consistent follow-up of children vertically-infected with HIV, a longer period of time than other such cohorts (70;71). Since 1985 data were collected in the same centres according to the standard protocol, using the same questionnaires by essentially the same clinicians throughout, while the period saw various changes to the management of HIV infection in children.

The dynamic use of the CDC classification system (72) has revealed a generally better health status in children vertically-infected with HIV than previously recognised, even in those untreated with ART. Since HIV-related disease is rare, side effects of the treatment may become the main clinical concern (146-148); whether these children are at risk of cancers or other diseases at a later age remains unclear.

The estimation of virological and immunological patterns over age made use of longitudinal repeated measurements over 12 to 15 years analysed using mixed effects models formally accounting for within-child correlations, in contrast to the cross-sectional assessments at different ages used by others (43;75;76;78-80;95-97;99). Previous findings of the peak in HIV RNA viral load early in life and subsequent gradual decline thereafter has been reproduced but here also extended to stratified estimation.
Stratification allowed the identification of distinct HIV RNA patterns over age by gender and gender-specific differences by treatment. Results indicate possible gender-specific dynamics of viral replication and that as the benefit of treatment in girls might diminish, it may intensify in boys over age, suggesting a biological explanation for response to HIV infection and treatment in children, even in the pre-pubertal stages. As the current guidelines for the initiation of antiretroviral therapy based on viral load do not allow for gender, there may be a need for consideration of revision. Results here would suggest that cut-off levels at which therapy would be recommended may need to be lower for girls than for boys, after the first year of life. In addition, response to treatment measured through RNA viral load may also need to be interpreted in a different way for boys and girls, with an expected lower initial fall in RNA load in boys after initiation of treatment than in girls. Significant differences in immunological patterns over age by not only gender but also race have been identified for the first time. However, necessity for gender and race specific guidelines based on immune status may not be as important as those for HIV RNA viral load given the small magnitude of differences.

The comparisons of immunological patterns over age in infected and uninfected children born to HIV infected women previously reported (95-97) have been extended to twelve years of age, and although there are clear distinctions overall by infection status, substantial overlap in all lymphocyte measurements has been demonstrated. The overlap of measurements seen in uninfected and infected children limits the use of a single CD4+ cell count measurement to assess the immune and health status of the infected children, further highlighting the limitations of the CDC categorisation (70;141;145) and could indicate the necessity for more extreme limits in predicting serious progression. Immune status assessed by absolute values at each age may
inadequately reflect the ongoing virological activity, and divergence from age-related values may be more clinically relevant.

In addition to virological and immunological differences by gender indicating an underlying hormonal mechanism, immunological differences by race observed in children born to HIV infected mothers may suggest an underlying genetic origin for the levels of markers (86). As well as having implications for clinical decisions about therapeutic management, findings, in particular those on uninfected children (149), may contribute to the broader knowledge necessary for ongoing HIV vaccine development by informing mechanisms of viral activity and host response.

Being the most prognostic of the risk factors for disease progression beyond six months of age, the distinction between above and below 10% for CD4$^+$ percentage is lower than the 15% threshold in the CDC categories below which children are classified as severely immunosuppressed (72), suggesting that reconsideration of category cut-offs should be contemplated. The findings here and elsewhere (150-152) imply that in the absence of specific CD4$^+$ cell assay tools, absolute lymphocyte measurements alone would provide sufficient insight to inform the management of individual children from early life, which could be particularly relevant in less developed countries where laboratory resources may be limited. Findings on the use of clinical as well as immunological factors for predicting disease progression inform care management in such settings where the burden of HIV-related mortality is vast.
The data presented in this thesis largely relate to the long period of follow-up in the era when early ART to delay disease progression was lacking, but, as appropriate adjustment has been made for ART received at a later age, results provide a largely natural account of both short- and long-term paediatric clinical and biological disease progression that remain relevant in the era of HAART.

7.1.1 Pattern of clinical manifestations over age

By using data from this European observational study of children followed from birth, taking a broader and more dynamic approach than that used by others (38;65;70;81) in allocating a CDC category at each visit, thus allowing children to move to a less serious category if their clinical or immunological status improved, the fluctuating pattern of clinical disease over the first 15 years of life became apparent. Generally, infected children - even those surviving previous serious illness - are healthy most of the time. By extending the analyses to 15 years of age, this work adds to the previous findings based on combined data from the ECS and the French paediatric cohort studies six years ago (70). The current estimate for the rate of progression to serious disease is less now due to the inclusion of recently born children who will have received ART at an early stage. The one-year mortality estimate of 8.5% in the ECS described here compares with 5% in an American cohort (71). Part of this difference is likely due to an era effect with the inclusion here of children born in the mid-1980s, not treated early with ART, in contrast with the later establishment of the American study in 1988 as well as the effect of earlier introduction of HAART administration from a young age in the USA.

The birth cohort-stratified analysis of these data shows that with the widespread administration of HAART at an early stage of infection, occurrence of serious disease or death has now become
rare, confirming the previous findings from the Italian paediatric register (48) and results from an American study indicating a reduction in risk of rapid disease progression in children born in recent years (116). It is unfortunate, however, that ART and PCP prophylaxis effects cannot be distinguished, particularly in the context of informing decisions on resource provision in developing countries.

7.1.2 HIV RNA viral load
The levels and patterns of HIV RNA viral load were investigated over nearly 15 years of regular follow-up. This represents almost a two-fold extension of the age range of eight years for HIV RNA viral load data from birth cohorts previously described by others (78;79). Although data spanning the first 18 years of life have been previously reported (80), these were obtained cross-sectionally from clinical trials, involving children selected on the basis of trial criteria, who were more likely to be healthy, while findings here are more representative of the population of all children vertically-infected with HIV.

Findings here of a peak around $10^5$ copies/ml agree with some studies (43;75;110), but are lower than the $10^6$ copies/ml estimates of others (80;81), which is likely due to the previous lack of accounting for repeated measurements within individuals. The continual decline is compatible with some previous findings (77;83;110) but at odds with the reported stabilising or elevation of levels beyond age 6-8 years (44;80). Studies indicating this apparent reconstitution are based on data from populations of children entered into clinical trials, with corresponding differences in characteristics compared with children enrolled and followed from birth in the ECS.
By using fractional polynomial models it was possible to show the actual shape of the HIV RNA viral load pattern over age. It has been demonstrated here that this approach is superior to conventional polynomials, used previously (79), in detecting the early peak. A third approach, the change-point method, which has been appropriately used to model data in HIV infected adults (137), also did not show the magnitude of the peak in children, which was known to be present in the raw data from the running smoother. Furthermore, the linear decline of RNA after peak in the change-point approach was not representative of the true non-linear relationship with age. Lastly, if used as the basis for the model incorporating explanatory factors, difficulty would be involved in identifying strata-specific change-points for the various strata as defined by the four different gender/treatment combinations.

Use of Hughes' method to allow for both within-child correlation of repeated measures and censoring of the assay cut-off value avoids the bias in the parameter estimates (131;132) arising from previous disregard of these values altogether (43;44;75-80;83;86) or use of ad hoc procedures (81). In using sophisticated methods, this work confirms a distinct peak in RNA levels of just below $10^5$ copies/ml around three months of age followed by a decelerating decline thereafter. The current findings of a slowing but continued decline in levels beyond five years of age agrees with the findings from one study (83), but not others (44;77;80) which reported the stabilising of levels at around age five to six years. These earlier results are from studies which did not account for the repeated measurements within children (80), or were on substantially smaller numbers of children with few data on each (44;77). Given the numbers of children and measurements available from the ECS and due account of serial data, inference from the presented results is considerably more reliable.
Although choice of methodology between the two approaches considered for simultaneously
dealing with repeated measures and censoring was not as crucial as that for the basic pattern, the
most complex methodology was more reliable. There was generally close agreement between the
results of the LME approach with mid-point censoring and those of Hughes’ method, but
estimates differed most where heavier censoring of HIV RNA values occurred. This work shows
that the setting of the HIV RNA viral load value to the midpoint within a mixed effects
framework allows the predicted curve to follow a trajectory close to that of the gold standard
Hughes’ method, but where there is heavy censoring, the *ad hoc* method is not as reliable. There
could be a threshold for the proportion of censoring below which the use of Hughes’ method is
not required. The necessity for the complex method is also dependent on the value of the
detection limit. Many of the censored values here have been censored below relatively low
values – less than five percent of the values have been censored above 500 copies/ml. Had there
been more censored values, and a greater number of less precise ones, the sophisticated
methodology of Hughes may have been more critical.

In investigating an age-dependent gender difference by the introduction of an interaction term for
gender by age, this analysis demonstrates for the first time significant and substantial differences
in HIV RNA viral load pattern between boys and girls. This work extended the only previous
analysis of this kind (86) beyond the first two years of life. Since the work has only been
reported as an abstract, methodological comparisons are difficult, but account seems to have
been made for differences in just levels, not patterns over time. In fact, findings here of levels
peaking earlier and significantly more than one \( \log_{10} \) higher at the peak in girls, remaining higher
during the first four years, conflicts with the previous findings of lower levels in girls (86), but
those results were not significant. The finding that levels fall below those in boys beyond five
years of age is consistent with results from adult studies of lower levels in women than men
(87;89;90), and suggests the possible use of lower viral threshold levels in girls than in boys at
older ages, as well as in adults to guide initiation of ART (87).

Simultaneous stratification by treatment administration with gender, not previously considered
within gender differentials, showed that association between treatment and viral load over age
depends on gender. Although observations under treatment were always considerably lower that
those without ART for both sexes, the findings that for boys the difference grew slightly over
age, but for girls it lessened are consistent with findings in adults of a more rapid viral
suppression and more sustained short-term response following HAART in women than in men
with equivalent baseline levels (91;153).

Given the downward direction of viral levels over age regardless of treatment status and the
absence of an interaction of CD4⁺ cell count by age, previous observations of the independence
of the age-related falling off of levels from initiation of ART and immunological status from
age-specific inference (77;81) are confirmed here.

7.1.3 Immunological patterns
The large numbers of available immunological measurements over age allowed modelling taking
in to account their repeated nature, and investigation of the effect of variables that were
potentially related to the level and/or pattern. Although confirming findings from earlier studies
identifying peaks in CD4⁺, CD8⁺ and absolute lymphocytes cell counts in early life of uninfected
children (70;72;98), and the age-related decline (95-97), the distinct double nature of this peak had not been shown before. This may suggest an initial faltering in the development of the immune systems of uninfected children born to HIV infected women, but the precise mechanisms for such dynamics are currently unclear and pose a challenge for immunologists.

With the methodology of smoothers and splines, exact ages at peaks in CD4^+; CD8^+ and absolute lymphocytes have been identified with accuracy not previously achieved (95-97). This knowledge aids the understanding of mechanisms involved in the immunological development of children born to HIV infected women. As reported by others (97;100;101;154), CD8^+ cell counts were significantly higher for infected children than for uninfected children from three months of age onwards indicating ongoing viral activity (28-30). Only 14 of the 186 infected children had been exposed to antiretroviral prophylaxis to prevent vertical transmission in the first weeks of life, and this is unlikely to have substantially influenced the observed initial pattern and level. However, two-thirds of infected children received antiretroviral treatment to delay disease progression at some time, but differences by gender and race remained after adjustment with inclusion in the multivariable models.

Findings here indicate that differences in CD4^+ cell counts by infection status exist from birth, much earlier than the three months to one year previously suggested (95-97); this is likely to be attributable to the abundance of ECS data compared with other studies which lacked power to detect smaller differences. The emerging differences in both CD8^+ cell and absolute lymphocyte counts from three months of age is consistent with findings from previous studies (96;97). Relative to absolute determinations, CD4^+ percentage is stable over age beyond early life and as
such represents a simpler gauge of immunological wellbeing without the need for reference to age.

The findings of a significant effect of maternal immune suppression on levels of CD4+ cell counts in uninfected children could indicate that one cannot generalise from uninfected children born to infected mothers to children in the general population or vice versa (17;97;155). Findings that CD4+ cell counts and absolute lymphocytes were higher for uninfected girls than for uninfected boys, are compatible with adult data (103). Findings here strengthen, and extend to younger ages, previous results suggestive of gender differences in immunological markers reported from the cross-sectional study of immune function in around eight to 12 year-old African American and Latino inner-city children born to uninfected women (104). Together, these findings indicate that infection with HIV may have a more profound effect on the CD4+ cell counts and absolute lymphocytes but less of an impact on CD8+ levels of girls than on that of boys.

Findings of race differences in CD4+ cell counts of uninfected children are at odds with those from two American studies, one on unexposed (104) and the other on uninfected children born to HIV infected women (106), in which lymphocyte levels did not differ by ethnicity. This may be explained by differences in the population characteristics, and also, with just 126 (106) and 152 (104) children respectively, it is possible that the other studies lacked the power to detect differences. Furthermore, as comparison of overall means by ethnicity did not account for any differences over age (104), and as linear regression accounted for any differences in levels but not patterns over age (106), methodology used may have accounted for lack of identified
distinctions. Findings here of the possibility of levels in black children overtaking those of white
children during mid-childhood are compatible with the findings of higher levels in West African
compared with European adults (107). Indirect evidence relating to possible lower CD4+ cell
counts in infected African than in European populations has previously been reported in
comparisons between cohorts of pregnant women included in HIV vertical transmission studies
(150;156-159). However, in those comparisons it was unclear whether the observed differences
were related to the longer duration of infection in the African cohorts, or to a heightened immune
response to other infections more prevalent in the African setting. In the ECS cohort the
observed differences are between black and white children all born in Europe and followed from
birth, and are unlikely to be explained by environmental infectious pressure on the immune
system.

With significantly different CD4+ percentages by gender and race in uninfected children, higher
levels of CD4+ cell counts in uninfected girls and uninfected white children are not fully
explained by a general lymphocytosis in uninfected children. Taken together with previous
findings from an American study of faster CD4+ cell counts declines in infected women and non­
white individuals (90), results on infected children presented here indicate gender and race
differences may increase with age and persist into adulthood.

7.1.4 Determinants of serious clinical disease progression
Findings on determining factors for clinical disease progression extend previous work addressing
the prognostic value of early markers of disease progression and death in HIV infected children
(43;45;78;80;109-112;160), using data available over all ages prior to progression on sufficient
children to draw reliable conclusions. Findings here of early clinical signs being predictive of
overall progression to serious disease are in general agreement with previous studies (71;108).

Independent predictive effects of CD4$^+$ percentage and absolute lymphocytes concur with some findings (83;109), but while other investigators have reported associations with CD4$^+$ cell count (40;80;108), this was not found to persists after inclusion of the important laboratory markers here. Findings of a lack of association of CD8$^+$ levels with progression in vertically-infected individuals are consistent with previous reports (109).

Unlike in many previous studies (40;45;71;80;83), the approach taken here addressed distinct factors contributing to rapid and long-term disease progression. One of the only previous studies to have done so found early presence of lymphadenopathy, hepatomegaly or splenomegaly together with CD4$^+$ percentage to be predictive of rapid progression (109), which is not entirely consistent with findings here. Results from ECS data presented here indicate that although clinical evidence of infection in the first six months of life, such as lymphadenopathy and hepatomegaly, is not associated with subsequent rapid progression before one year of age, it is predictive of the long-term prognosis of the child, while hepatomegaly was the only factor associated with disease progression beyond one year of age when this was assessed separately. The apparent contradiction with findings here could be explained by the fact that unlike with the ECS data where attention was confined to mild symptoms, moderate clinical symptoms were considered, which likely progressed rapidly to severe disease. Furthermore, there were several methodological differences in the approach to the analysis between this other study (109) and the ECS: firstly, rapid progression was defined by them as occurring during the first six months of life only while it was defined in the current analysis as occurring before one year of age; secondly, in some children clinical progression preceding the laboratory determinations was
included in the analysis while we only took progression beyond clinical and laboratory data and thirdly, their statistical methods involved only logistic models with presence or absence of progression as a binary response, without extension to a formal survival analysis.

The finding that in the ECS cohort early immunological factors were predictive of progression before one year of age but not beyond is consistent with earlier observations from the American study indicating levels reflect current status more than long-term survival and wellbeing (108). Although early signs of HIV-related illness were associated with progression by 18 months in this other study, the association was not found beyond that age (108). That they defined the outcome as either first-time or re-emergence of serious disease, may explain apparent discrepancy. Findings of associations of clinical signs at birth with C-defining illness during the first year of life by the French Paediatric HIV Infection Study Group may be explained by the grouping together of several symptoms and the prognostic effect of CD4⁺ cell counts is expected since neither CD4⁺ percentage or absolute lymphocytes were included in the multivariable analysis.

The finding that generally in the ECS cohort, early progression was associated with opportunistic infections whereas progression later in life was due to encephalopathy and serious bacterial infections, is in line with findings from an American birth cohort study which suggested encephalopathy was more likely following development of early symptoms of HIV (161).

Neither the gender nor race of an individual child was associated with progression to serious disease or death. Furthermore, the predictive ability of early-life factors did not depend on the
gender or race of the child. These findings represent the first exploration of this issue. Since the magnitude of the differences by gender and race in levels of immunological markers becomes larger with age during the first few years of life, one could expect to see gender differences in progression for a given CD4+ cell count or race differences in progression for a given CD4+ percentage at two years of age. However, in this dataset there was no difference between boys and girls in the prognostic importance of the values at two years of markers for clinical progression to serious disease, which may have been due to small numbers.

That risk of serious disease progression was not positively associated with exposure to zidovudine prophylaxis to reduce the risk of vertical transmission in their first weeks of life is at odds with other studies (113;114). However, findings from these studies may be explained by the inclusion of children born to women who received zidovudine for clinical indications rather than as vertical transmission prophylaxis, so the effect may be explained by confounding of maternal disease status.

Due to the nature of the data it was not possible to address associations of HIV RNA viral load with subsequent serious disease progression, but had such data been available it is likely that, while there would have been a clear relationship, given the independence of virological and other factors (108), findings here of clinical and immunological effects would have persisted.

### 7.2 Recommendations

On the basis of this work, several recommendations for the analysis of longitudinal clinical and laboratory data can be made. When considering clinical disease patterns it is informative to make use of the CDC classification system in the dynamic way used here. In modelling virological and immunological patterns over age, smoothing techniques (133) and flexible regression methods
such as fractional polynomials (131) and splines (139) reveal of the true structure and suitably
represent of the data. Although this may be of limited practical value, it may be of scientific
importance. As for adult data (137;162), formal account of the repeated measurements by way of
mixed effects models (136) resolves bias in results arising from within-subject correlation, which
is particularly problematic under circumstances where the extent of illness in a child influences
the number of visits it makes, common in such clinic-based cohort studies. To truly represent
patterns of HIV RNA viral levels it is necessary to account for the lower detection level
censoring. The method devised by Hughes (132) offers the most reliable solution, but is complex
to use, and given the relatively low detection values obtainable from assays now, the
methodological precision involved is no longer required. Less sophisticated methods of setting
values midway between zero and detection level can suffice. In elucidating determining factors
for disease progression from birth over childhood, short- and long-term progression should be
considered separately. Finally, treatment administration should be dynamically allowed for in
any analysis of biological data from children vertically-infected with HIV.

7.3 Limitations of analyses

7.3.1 Data
Taken as a whole, the cohort is representative of HIV infected children in Europe over the course
of the epidemic. However, as these data have been collected from a number of centres across
Europe, characteristics of the centre-specific populations of children and clinical practice may
vary and could have influenced findings. However, all centres in the ECS are specialist referral
units and diagnosis of most HIV-related conditions will be in agreement. Analyses have not
directly controlled for any heterogeneity of laboratory methods across sites, but have adjusted for
HIV RNA viral load assay, and as different techniques used for lymphocytes determination are consistent (163), any variation will be small. As data have been gathered over the various eras of paediatric HIV-infection in a changing treatment climate, observations, especially those relating to early life are a mixture of those taken in the earlier as well as more recent years of the study period. Analyses were not adjusted for year of birth per se, but stratification by treatment era has been performed.

As the ECS is an observational study, treatment administration is not random. Treatment has been given to those children exhibiting uncontained HIV viral replication and/or immunologically decline deemed to be at risk of clinically progressing, limiting the scope for assessment of treatment effect on clinical and biological outcomes. Further, treatment policy varies between centres (58), but formal adjustment for or stratification by ART administration has been performed throughout and prophylactic measures have been accounted for when necessary, ensuring results are as reliable as possible.

Although there is a wealth of information available from this long standing cohort of children born to a large number of HIV infected women, some explorations were not possible due to limited numbers of data from infected children. For example, it was not possible to investigate the effects of other factors possibly associated with disease progression in vertically-infected children such as maternal ART use during pregnancy (112-114), timing of transmission (78), maternal viral load (109;115), due to limited data. Small numbers for certain sub-analyses performed, for instance, maternal CD4+ cell count and immunological markers, make it difficult to draw reliable conclusions.
Modelling of laboratory measures over age has been performed on a mixed population of slow and fast progressors. It is possible that drop out of individual children due to death, leading to different study groups at different ages, may distort dynamics of virological and immunological patterns. The levels of CD4^+ cell counts seen at any given age are likely to be representative of those alive at that time, but the slight upturn in CD4^+ cell counts over later ages could be associated with either survivor bias or immunological improvement as a result of effective ART. Indeed, as shown, levels of those who have died were consistently lower compared with those known to be alive. Similarly, the gradual decline in HIV RNA viral levels late in infancy could be explained by the early inclusion but later exclusion of data on those who died — the very infants displaying least viral containment. Since viral enumeration has been standard only in recent years, viral load measurements in the younger years relate to children born in the era of widespread administration of combination therapy early in life. However, as the peak and decline observed occur in the earliest months, preceding ART administration for the prevention of disease progression, its impact is unlikely to play a role in the dynamics, although the effect of AZT prophylaxis for the prevention of vertical transmission cannot be ruled out.

As in any cohort study, loss to follow-up of individuals is a potential source of bias in results. However, ECS sites are all referral centres, and clinicians have strong ties with children (especially infected ones) and their families, maintaining close personal contact. Resulting dropout rate are less than one in ten (9.5%, 18/190) according to records current at time of writing. It is thus unlikely that loss of individual children will have impacted to any great extent on the findings.
Since the proportion of children still alive and in follow-up beyond age ten years is relatively low (53/190, 27.9%), available clinical and biological data become sparse into the pre-pubescent and teenage years. Inference becomes less reliable at these ages and it has not always been possible to draw reliable conclusions. For instance, assessment of race differentials in immunological markers was not feasible at these ages.

Restricted availability of early-life HIV RNA viral load prohibited the analysis of viral load in relation to clinical progression. We were unable to assess the child’s HIV RNA viral load as a marker for disease progression directly here, but the association between clinical indicators such as lymphadenopathy and hepatomegaly and viral activity indicates that early viral load measurements, when available, would be useful prognostic indicators, as has been shown by others (45;81;108). It has been suggested that viral load is the optimal predictor of paediatric HIV progression (108), however, the relative stability of immunological markers, especially CD4\(^+\) percentage, compared to the highly variable HIV RNA viral load levels (164), may make them as clinically relevant in predicting serious disease progression.

7.3.2 Statistical methods

While fulfilling the purposes of the analysis, the application of the EM methodology of Hughes’ to the modelling of HIV RNA viral load data involved the use of software which was not straightforward. An alternative has been demonstrated by Jacqmin-Gadda et al (137), in which model parameters are estimated by direct maximisation of the likelihood, requiring numerical computation of the integral of a multivariate Normal density for each subject with censored observations. However, due to the methods used, the required algorithm only works well for less
than ten censored observations per individual, and since some ECS children had as many as 18 censored HIV RNA observations, it would not fulfil the requirements here.

As a benchmark for the comparison of immunological data on infected children, age-related reference ranges constructed from data on uninfected children are useful in clinical decision making (98). Use of natural cubic splines in the modelling of immunological markers limited the possibilities for constructing such age-related percentiles. Another downside to both splines and fractional polynomials is that, unlike standard regression models, they do not provide parameters which have meaningful interpretation in the conventional sense. However, since they reflect, with a high degree of accuracy, the course of immunological levels and HIV RNA viral load over age, respectively, their use achieves representation of such longitudinal data not possible by more conventional means, as demonstrated in method comparisons for viral load analysis. Generally, results obtained using fractional polynomial methodology can be sensitive to outlying values, however the large number of observations available for analyses here have produced models which reliably reflected the structure of the data as indicated by the running smoother. Fractional polynomials were found to be the superior of all approaches considered. The change-point model could be enhanced by segmented polynomial regression with integral join-point estimation (165), but would this increase estimation complexity.

The particular spline models identified for immunological markers may be less representative of the data on introduction of explanatory factors, as choice of knots for overall data structure may not be appropriate for gender- and race-specific patterns. In particular, there may be misleading evidence indicating the existence of, or absence of interactions signifying differential patterns as
well as levels. Significance may be under or over estimated, but as the raw data has been considered as well as the formal results in deciding on the importance of effects, conclusions should be as reliable as possible.

Given the departure of smoothed trajectory from the fitted values obtained from the spline models, especially pronounced where there is large between-child variation (happening most markedly in the infected children), running smoothers may not be the best guidance in deciding on the knot positions. It is possible other methods of comparison, such as goodness-of-fit or residuals tests may be more valid, but such analyses would be cumbersome due to the complexity involved.

In determining the optimal cut-off value for separating CD4\(^+\) percentage values for predicting clinical progression, the CART method does not account for the time to onset of disease or the censored nature of these data, which would be taken care of by tree structured survival analysis methodology (TSSA) (166).

Although there are statistical and data-derived limitations to this work, analyses used have been shown to be thorough and relevant to the aim and objectives. Important issues in paediatric HIV have been addressed: the clinical patterns of disease progression have been comprehensively described by use of survival methods and novel use of the CDC classification system revealed the sporadic nature of serious illness which would otherwise have been unobserved. The trajectory of HIV RNA viral load over age has been elucidated by appropriate application of fractional polynomials and methods which simultaneous allowed for repeated measurements and
assay censoring. The levels and patterns of immunological markers for HIV-infection over age have been presented in detail with the use of flexible spline models. Estimates of the patterns in HIV RNA viral load and immunological markers closely follow the structure of the raw data to an extent not possible using other techniques. Since within-child correlation of measurements in both analyses has been rigorously dealt with, estimates obtained are unaffected by bias otherwise arising from disproportionate numbers of observations from especially ill children. The prognostic values of early-life clinical, virological and immunological measurements for prediction of rapid and long-term progression to serious disease or death have been explored in the application of survival methods, with detailed investigation of possible gender and race interactions.

7.4 Recommendations for epidemiological and statistical research

Some methodological aspects of this work could be explored further. A thorough exploration of the issues around apparent delayed suppression of HIV RNA viral load following birth and upturn in CD4\(^+\) cell counts in late childhood involving more detailed evaluation of levels by duration of follow-up would validate these findings. Comparison of the results obtained by fractional polynomials and those from segmented polynomial regression models with estimated join points (165) would verify the optimal approach to the modelling of HIV RNA viral load patterns over age. Assumptions such as Normality, equal variances and proportional hazards and made here could be checked formally using QQ plots, residual examination and tests based on Schoenfeld residuals (167), ensuring choice of models was appropriate for the data available. The results of the application of TSSA (166) in establishment of a cut-off which best separates CD4\(^+\) percentage in terms of risk of future serious disease progression could be compared to those of CART given here to ensure accuracy of findings.
Whether management should be tailored according to gender has been raised previously in adults (87) and remains an important issue in paediatric settings. This, along with whether there should be race-specific guidelines should be investigated further and could be addressed by meta-analysis of combined data from this and other studies. Thorough exploration of the issue of gender and race differentials in the predictive values of laboratory markers for disease progression may be conceivable in the context of meta-analysis of data from various studies with relevant data. Ongoing data collection may permit a formal assessment of the relationship of HIV RNA viral load and risk of disease progression.

In tackling other research questions, the prognostic value of clinical CDC category itself for future disease progression could be investigated and the possibility of identifying viral load "set-points" in children as predictive of progression, as done in adult studies could be explored. Further work on the exploration of the relationship between viral load and CD4+ target cells in children could involve the fitting of a predator-prey model as carried out on adult data (168).

The predictive value of single immunological markers at landmark ages beyond early life for subsequent progression could be assessed in conjunction with investigation of the validity of using values at particular age-points and to what extent the previous trajectory is also important. In relating laboratory data to risk of disease progression, longitudinal viral and immunological measurements could, themselves, be incorporated into models as time-dependent variables, allowing the prediction of immediate risk of disease development (108). By taking, the LMS method (169), for instance, updated age-related percentiles for immunological markers can be constructed.
It is problematic to use an observational study like the ECS to assess benefit or not of ART or prophylaxes for PCP or bacterial infections. Development of methods to measure effectiveness within the setting of a cohort study would further optimise the usefulness of longitudinal data on children vertically-infected with HIV followed from birth.

Attention should be given to those for whom treatment fails; continued focus on the assessment of the clinical impact of ART regimens, and their fine tuning for paediatric HIV infection is a priority. Optimising effectiveness in terms of minimising morbidity, and maintaining immunological buoyancy and viral containment has to be balanced with the priorities of avoidance of resistance to therapy.

Findings on better health than previously appreciated in these children imply that it is likely that a reasonable proportion will reach young adulthood and beyond. The sustained impact of HIV infection and treatment on the health of these children as they advance in life will be of continued interest. Some will embark on parenthood, bringing new challenges and scope for research, thus follow-up should be ongoing for as long as possible.

The future looks to the small, but significant proportion of children who are still being born with HIV infection despite availability of effective intervention regimens. Successful containment of disease development in such children is of paramount importance and their patterns of disease progression should be carefully monitored.
Reference List


(86) Gender differences in immunologic and virologic markers in children born to women infected by HIV. Abstract 513197. 8th Conference on Retroviruses and Opportunistic Infections; Chicago, Illinois; February 4 - February 8.; 2001.


(127) AIDS Program Center for Infectious Diseases. Revision of the CDC surveillance case definition for acquired immunodeficiency syndrome. Council


(152) Badri M, Wood R. Usefulness of total lymphocyte count in monitoring highly active antiretroviral therapy in resource-limited settings. AIDS 2003; 17(4):541-545.


(158) Leroy V, Karon JM, Alioum A, Ekpini AR, Meda N, Greenberg AE et al. Twenty-four month efficacy of a maternal short-course zidovudine regimen to


(160) HIV Paediatric Prognostic Markers Collaborative Study Group. Short-term risk of disease progression in HIV-1 infected children receiving no or minimally effective antiretroviral therapy: estimates according to CD4 percent, viral load, and age. Lancet 2003.


Appendix 2.1  European Collaborative Study data collection forms

MATERNAL INFORMATION AT DELIVERY

PERINATAL INFORMATION

MEDICAL EXAMINATION

ASSESSMENT

LABORATORY INVESTIGATION
MATERNAL INFORMATION AT DELIVERY

Mother's date of birth (day, month, year)
Country of birth ..............................................................................................................

Marital Status
Single (1), Married (2), Divorced, Separated, Widowed (3), Cohabiting (4)

Ethnic Group
Asian (1), White (2), Black (3), Oriental (4), Other (5)
Age when leaving full-time education, years

Obstetric History
Number of previous livebirths ........................................................................................
Number of previous stillbirths ........................................................................................
Number of previous miscarriages ..................................................................................
Number of previous terminations ................................................................................

Mothers Risk Group
History of intravenous Drug Abuse (Y/N)
Trimester of last use: pre-conception (0), 1st (1), 2nd (2), 3rd (3), unknown (9)
Needle sharing? never (1) past (2) present (3) unknown (9)
Sexual partner of Bisexual (Y/N)
Sexual partner of Haemophiliac (Y/N)
Sexual partner of Intravenous Drug Abuser (Y/N)
Sexual partner of Other high risk group (Y/N)
(Specify) ...........................................................................................................................
Other ......................................................................................................................................

Mothers HIV History
Date of first HIV+ test (day, month, year)

Current clinical status
Current HIV staging (CDC) ...............................................................................................
Specify symptoms ............................................................................................................
Date of onset ....................................................................................................................

Details of antiretroviral therapy during pregnancy
Has the woman received any antiretroviral therapy at any time during this pregnancy?
If yes, please give details Y/N

<table>
<thead>
<tr>
<th>Drug</th>
<th>Total daily dose</th>
<th>Date started</th>
<th>Date stopped</th>
<th>Currently taken? (yes/no)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

239
MATERNAL INFORMATION

Laboratory investigations during pregnancy and at delivery:

<table>
<thead>
<tr>
<th>Centre Number</th>
<th>Mothers Study Number</th>
<th>Child Study Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>3-5</td>
<td>6</td>
</tr>
</tbody>
</table>

**Virology**

<table>
<thead>
<tr>
<th></th>
<th>Date:</th>
<th>Date:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV-DNA PCR</td>
<td>Pos / Neg</td>
<td>Pos / Neg</td>
<td>Pos / Neg</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIV-RNA PCR</th>
<th>copies/ml</th>
<th>copies/ml</th>
<th>copies/ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample type</td>
<td>Plasma / Serum</td>
<td>Plasma / Serum</td>
<td>Plasma / Serum</td>
</tr>
<tr>
<td>Assay used</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Other laboratory investigations**

<table>
<thead>
<tr>
<th></th>
<th>Date:</th>
<th>Date:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total lymphocytes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD4 (10^9/litre)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD8 (10^9/litre)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgG (gm/litre)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgA (gm/litre)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IgM (gm/litre)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p24 Ag</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIV Elisa</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### PERINATAL INFORMATION

<table>
<thead>
<tr>
<th>Centre</th>
<th>Mothers Study Number</th>
<th>Child Study Number</th>
<th>Child's date of birth (day, month, year)</th>
<th>Sex (M, F)</th>
<th>Gestational age (weeks)</th>
<th>Birthweight (gm)</th>
<th>OFC (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hospital where delivery took place
Obstetrician (initials)

**Antiretroviral therapy during labour/delivery**

- Y/N
- If yes, which drug? Orally / IV?

**Delivery**

- Caesarean Section: Elective (1), Emergency (2)
- If Caesarean Section, reason:
- Vaginal: Spontaneous (3), vacuum (4), forceps (5)
- Presentation: breech (Y/N)
- Duration of labour 1st stage (if known)
- Duration of labour 2nd stage (if known)
- Time from rupture of membranes to delivery (if known)
- Scalp Electrodes (Y/N)
- Episiotomy or vulvovaginal tear (Y/N)

**Perinatal Problems** (Y/N). Specify Details:

- Hepatomegaly
- Splenomegaly
- Drug Withdrawal Symptoms
- Thrombocytopenic Purpura
- Infection: suspected (1) confirmed (2)
- Transfusion *
- Congenital Abnormalities *
- Other

**Disposition**

- with parents (1) fostered (2) adopted (3)
- remained in hospital (4) other (5) *
- if remained in hospital, say why: *

**Feeding:** breast (1) bottle (2) breast and bottle (3)
- was breast feeding tried and abandoned? Y/N

**Died?** Y/N
- Date of death: (day/month/year)
- Postmortem results, if available *

Take sample required; Please record laboratory results on yellow form
**MEDICAL EXAMINATION**

Please circle or complete as appropriate
Assessment at: 3w, 6w, 3m, 4.5m and 6m

<table>
<thead>
<tr>
<th>Centre</th>
<th>Mothers Study Number</th>
<th>Child Study Number</th>
<th>Date of Examination</th>
<th>Weight (kg)</th>
<th>Height (cm)</th>
<th>OFC (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Recurrent fever** of unknown origin requiring medical attention .......................................................... Y/N

**Chronic or Recurrent diarrhoea** requiring medical attention ............................................................ Y/N

Specify organism ..........................................................................................................................................

**Bacterial infection** ............................................................................................................................... Y/N

If yes, specify:
- Septicaemia, Meningitis, Urinary tract infection, Pneumonia, Other ..................................................

**Communicable Disease** .......................................................................................................................... Y/N

- Measles (1) Mumps (2) Rubella (3) Varicella (4) Zoster (5) Other (6) ..............................................

**Complications** ........................................................................................................................................

**Skin Infection** requiring medical attention .......................................................................................... Y/N

- Staph (1) Strep (2) Herpes (3) Candida (4) Other (5) ........................................................................

**Non-infectious skin eruption** ................................................................................................................. Y/N

- Petechiae/Purpura (1) Eczema (2) Kaposi Sarcoma (3) Other (4) ......................................................

**Palpable Lymph Nodes** ...........................................................................................................................

- Axillary (1) Postoccipital (2) Cervical (3) Inguinal (4) Epitrochlear (5) Other (6) ........................

**Chronic parotid swelling** ...........................................................................................................................

- Chronic otitis media (1) Sinusitis (2) Chronic purulent rhinitis (3) Other (4) .................................

**Upper respiratory tract infection** .............................................................................................................

- Chronic otitis media (1) Sinusitis (2) Chronic purulent rhinitis (3) Other (4) .................................

**Lower respiratory tract disease confirmed by X-ray** ..................................................................................

- Lymphocytic interstitial pneumonitis or Pulmonary lymphoid hyperplasia (1) .................................

- Pneumonia (2) Bronchiolitis (3) Other (4) ...................................................................................

-specify organism, if known .....................................................................................................................

**Opportunistic Infection** ...........................................................................................................................

- PCP (1) CMV (2) Toxo (3) Candida (4) Mycobacterium (5) Other (6) ..............................................

**Hepatomegaly** ...........................................................................................................................................

- ..............................................................................................................................................................

**Splenomegaly** ...........................................................................................................................................

- ..............................................................................................................................................................

242 57-58
**Medical Examination**

Please circle or complete as appropriate

<table>
<thead>
<tr>
<th>Date of Examination</th>
<th>Centre</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mothers Study Number</th>
<th>Child Study Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>3-5</td>
</tr>
</tbody>
</table>

**Neurological abnormality**
- encephalopathy (static/progressive) (1) ............................................................................
- seizures (2) paresis (3) pathologic reflexes (4) increased tone (5) decreased tone (6) abnormal gait (7) other (8) .......................................................

**Other Findings on exam**
- Specify ................................................................................... Y/N

**Developmental Assessment**
- Gross motor Pass (1) Fail (2) Suspicious (3)
- Fine motor/adaptive Pass (1) Fail (2) Suspicious (3)
- Language Pass (1) Fail (2) Suspicious (3)
- Personal/social Pass (1) Fail (2) Suspicious (3)

**Loss of developmental milestones**
- Specify ................................................................................... Y/N

**Neonate**
- Has the baby received any anti-retroviral therapy to reduce the risk of vertical transmission? ............................................................................. Y/N
- If yes: which drug(s)? for how long? .............................................................................

**Treatment**
- Has this child been enrolled in an anti-retroviral treatment trial? ............................................................................. Y/N
- If yes: which trial? .............................................................................
- Current treatment (excluding the above) 
  - IVGG, AZT, DDI, Other .............................................................................

**Hospital Admission(s)**
- (Indicate dates of admission/discharge and diagnoses for each hospitalization) ............................................................................. Y/N

**Immunisations**
- given since last visit ............................................................................. Y/N
- DPT (1) DT(2) Oral Polio (3) Killed Polio (4) Measles (5) MMR (6) 
- Hepatitis B (7) Other (8) .............................................................................
- Abnormal reactions .............................................................................

**Child care**
- mother / father / other relative / fostered / adopted / hospital / institution

**Breast Feeding**
- If stopped, when ............................................................................. Y/N

**Health of Mother**
- Is mother alive / dead?
- If dead, was death HIV-related? cause of death .......................................................
- Mother's current HIV staging (CDC) defining symptoms date of diagnosis current treatment? .....................................................
**ECS.1**

**PROSPECTIVE STUDY OF CHILDREN BORN TO HIV +VE MOTHERS**

**Assessment:** 9, 12, 18 and 24 months; thereafter annually for antibody -ve, uninfected children and 6-monthly for infected children

Please circle or tick as appropriate

<table>
<thead>
<tr>
<th>Centre number</th>
<th>1-3</th>
<th>4-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother Study number</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Child Study number</td>
<td>8-11</td>
<td></td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>12-15</td>
<td></td>
</tr>
<tr>
<td>Height (cm)</td>
<td>16-18</td>
<td></td>
</tr>
<tr>
<td>Head circumference (cm)</td>
<td>For office use only</td>
<td></td>
</tr>
</tbody>
</table>

- Is child alive? Y/N
- Date of assessment (day, month, year)
- Name of paediatrician (initials)
- Is this child HIV infected? Y/N
- Has this child developed AIDS? (CDC def) Y/N
  - if AIDS has been diagnosed since previous report, specify date of diagnosis
  - AIDS indicator disease
- Care
  - mother / father / other relative / fostered / adopted / hospital / institution
- Is mother alive / dead?
  - if dead, was death HIV-related? Y/N
  - cause of death
- Preschooling/Schooling
  - Does this child require special educational provisions Y/N
  - if yes, specify
- Treatment
  - Has this child been enrolled in an anti-retroviral treatment trial Y/N
  - if yes, which trial?
  - Current treatment (excluding the above)
  - Intravenous gammaglobulin/AZT/DDi/other, specify
- Communicable diseases Y/N
  - if yes, specify: measles / whooping cough / varicella / tuberculosis / mumps / zoster
**Oral Candida** since last visit (persisting > 2m or recurring despite therapy)  
if yes, date on onset (day, month, year) ........................................................ 

<table>
<thead>
<tr>
<th>Y / N</th>
<th>1</th>
<th>2-7</th>
</tr>
</thead>
</table>

**Hospital admission(s)**  
(indicate dates of admission/ discharge and diagnoses for each hospitalization)  

<table>
<thead>
<tr>
<th>Y / N</th>
<th>8</th>
<th>9-12</th>
<th>13-16</th>
</tr>
</thead>
</table>

Since the last visit, has this child had any of the following abnormalities:  

<table>
<thead>
<tr>
<th>yes/no</th>
<th>date</th>
<th>clinical presentation</th>
<th>method of diagnosis</th>
<th>diagnosis</th>
</tr>
</thead>
</table>

| renal | Y / N | ............... |
| hepatic | Y / N | ............... |
| gastro-intestinal | Y / N | ............... |
| cardio-vascular | Y / N | ............... |
| central nervous system | Y / N | ............... |
| respiratory | Y / N | ............... |
| malignancy | Y / N | ............... |
| other, specify | Y / N | ............... |

Please complete this section, if the child has died  

| Date of death | ............... |
| Cause of death | ............... |
| Post mortem results | ............... |
## LABORATORY INVESTIGATIONS

### Assessment at:
0-7 days, 3w, 6w, 4.5m, 6m, 9m, 12m, 18m, 24m, and then annually if child presumed not infected, or 6 monthly if infected

### Findings and specify as appropriate:
- **Date blood drawn:** 
  - **Day** / **Month** / **Year**
- **HIV / ELISA**
  - + / - Specify system used
    - antibodies
  - Western blot + / -
  - Virus culture + / - Specify identification system(s)
  - Viral load DNA - PCR
  - RNA - PCR
  - Antigen assay + / - Specify identification system
  - Other tests (eg IVAP, PCR, IgM) Specify method and result + / -

### Laboratory Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG (gm/litre)</td>
<td>36-38</td>
</tr>
<tr>
<td>IgA (gm/litre)</td>
<td>39-41</td>
</tr>
<tr>
<td>IgM (gm/litre)</td>
<td>42-44</td>
</tr>
<tr>
<td>T4 (10^9/litre)</td>
<td>45-48</td>
</tr>
<tr>
<td>T8 (10^9/litre)</td>
<td>49-52</td>
</tr>
<tr>
<td>Absolute lymphocyte (10^9/litre)</td>
<td>53-56</td>
</tr>
<tr>
<td>Neutrophil (10^9/litre)</td>
<td>57-60</td>
</tr>
<tr>
<td>Platelet (10^9/litre)</td>
<td>61-64</td>
</tr>
<tr>
<td>Haemoglobin (gm/dl)</td>
<td>65-67</td>
</tr>
<tr>
<td>Toxo IgG Latex (at 9 months to exclude congenital infection) (+/-)</td>
<td>68</td>
</tr>
<tr>
<td>Tetanus IgG (at least 1 month after third DT/DPT)</td>
<td>69-71</td>
</tr>
<tr>
<td>CMV IgG (+/-)</td>
<td>72</td>
</tr>
</tbody>
</table>
Appendix 2.2  ‘Collaborators in the paediatric centres of the European Collaborative Study

Dr C Giaquinto, Dr O Rampon, Dr F Ebo, Prof R D'Elia and Prof A De Rossi (Università degli Studi di Padova, Italy);
Dr I Grosch-Wörner (Charite Virchow-Klinikum, Berlin, Germany);
Dr J Mok (Royal Hospital for Sick Children, Edinburgh);
Dr I Bates, Dr I de José, Dr F Hawkins, Dr MC Garcia-Rodriguez, Dr C Ladrón de Guevara, Dr J Mª Peña, Dr J Gonzalez Garcia and Dr JR Arribas Lopez (Hospital Infantil La Paz, Madrid);
Prof F Asensi-Botet, Dr MC Otero, Dr D Pérez-Tamarit, Dr A. Orti, Dr M.J San Miguel and Dr R de la Torre (Hospital La Fe, Valencia, Spain);
Dr HJ Scherpbie, ME Kreyenbroek and Dr K Boer, Ms A Hes (Academisch Medisch Centrum, Amsterdam, The Netherlands);
Dr AB Bohlin, Dr E Belfrage, Dr L Navér, Dr S Lindgren (Huddinge and Karolinska University Hospitals, Sweden);
Prof J Levy, Dr P Barlow, Dr M Hainaut, Dr A Peltier, Dr S Wibaut, Dr G Debruyne (Hospital St Pierre, Brussels, Belgium);
Dr A Ferrazin and Prof D Bassetti (Department of Infectious Diseases, University of Genoa, Italy);
Dr A De Maria (Department of Internal Medicine, University of Genoa, Italy) Dr C Gotta (Department of Obstetrics and Gynecology--Neonatology Unit, University of Genoa, Italy);
Dr A Múr, Dr A Payà, Dr M Viñolas, Dr MA López-Vilchez, Dr M Rovira, Dr R Carreras, Dr E Esteban Tores, Dr S Herrero Perez (Hospital del Mar, Universidad Autonoma, Barcelona, Spain);
Appendix 2.3 Publications

1) Fluctuations in Symptoms in HIV-Infected Children: The First 10 Years of Life

2) Level and pattern of HIV-1 RNA viral load by age: differences between girls and boys?


4) Are there gender and race differences in cellular immunity patterns over age in infected and uninfected children born to HIV-infected women?

5) Gender and race do not alter early-life determinants of clinical disease progression in HIV-1 vertically infected children (in press, *AIDS*)
ABSTRACT. Objective. To describe patterns of clinical and immunologic progression in children who are vertically infected with the human immunodeficiency virus.

Methods. Children who were born to mothers who were infected with the human immunodeficiency virus in 11 European centers were enrolled at birth in a prospective study and followed according to a standard protocol. At each visit, a clinical and immunologic class was allocated according to guidelines of the Centers for Disease Control and Prevention (CDC). Progression to serious disease and death was assessed, allowing for available and actual antiretroviral therapy (ART). CDC class at each visit was assessed cross-sectionally.

Results. More than 15% of infected children will have progressed to category C or death by age 1 year and nearly 50% by 10 years. Just under 20% of children will have evidence of severe immunodeficiency by age 1 and 75% by 10 years. In general, immune status poorly reflects clinical condition. Children who were born after 1994, when the recommendation of earlier initiation of more active therapy was introduced, were significantly less likely to progress than those who were born when treatment was not widely available or was largely confined to zidovudine monotherapy. Estimated progression to CDC class C or death initially was faster in untreated than in treated children, but by 10 years estimated cumulative progression was similar in both groups. Treatment started before class C disease was associated with significantly slower progression. Cross-sectional analysis showed that children largely are symptom-free throughout their lives. After 4 years of age, fewer than 25% of infected children had symptoms at any one time, irrespective of ART received.

Conclusion. Vertically infected children are without serious symptoms or signs for most of the time. The prognosis has improved with more widespread availability and use of combination ART. These findings have implications for health, education, and other support-service provision. Pediatrics 2001;108:116–122; HIV infection, clinical patterns, disease progression, antiretroviral therapy, survival, Europe.
mune status over time in treated and untreated children. Although antiretroviral therapy (ART) to delay progression of disease now generally is initiated early in the course of the disease, there remains wide variation between and within countries. The description of the disease pattern in vertically infected children thus has become complicated by the widespread and rapidly changing use and the variation in time of initiation of therapy. To achieve a composite and comprehensive picture of the pattern of clinical disease in children who live with HIV infection, we analyzed data collected prospectively on infected children since 1986 in 11 European centers. We describe fluctuations in clinical and immunologic manifestations in children who either were never treated, or before and after they received antiretroviral treatment. We also investigate the effect of changing treatment policy over time.

METHODS

All children who were born to women who were known to be infected with HIV at the time of delivery were followed according to a standard protocol, with detailed clinical and laboratory information collected regularly in 11 pediatric centers in 7 European countries. Children were seen at birth, 3 and 6 weeks; 3, 4.5, and 6 months; and then at 3-month intervals until 24 months of age. Subsequently, infected children were seen at least twice a year, and clinical and laboratory information and the current treatment were recorded on standard forms. Parental consent was obtained before enrollment in the European Collaborative Study (ECS), and the study was approved by the local ethics committees.

The analyses presented here are based on children who were known to be infected with HIV according to the CDC definition of pediatric HIV infection. The purpose of the study, a CDC class was allocated at each visit on the basis of information regarding current clinical symptoms and signs to enable reclassification with changing clinical status. Clinical manifestations were grouped into 4 categories: N (asymptomatic), A (mildly symptomatic), B (moderately severe symptoms, including lymphoid interstitial pneumonitis [LIP]), and C (severe symptoms). Children who died with HIV-related disease were classified as category D. Infected children were also allocated to a CDC immunologic class at each visit. Normal (category 1), moderate (2), and severe (3) immune suppression was assessed by CD4 cell counts or percentages appropriate for age at the time of assessment.

Statistical Analysis

Kaplan-Meier product-limit analysis was performed of age at progression to each of the CDC categories A (or worse), B (or worse), C (or death), and D. Children were censored at death. Life-table analysis was also performed for progression to immunologic categories 2 and 3. Immunologic and clinical categories were cross-tabulated to assess agreement between the 2 classification systems.

Children were considered in 3 separate cohorts that were defined a priori according to treatment policy at the time of birth. Cohort 1 (1985–1988), no recommendation for treatment; cohort 2 (1989–1994), treatment policy restricted to monotherapy for symptomatic children; and cohort 3 (1995–1999), initiation of combination therapy recommended at an early stage. Progression to serious disease and actual treatment received were assessed for each of these 3 cohorts separately.

Not taking into account vertical transmission prophylaxis, children were identified as ever or never having received ART, and progression to CDC categories A, B, C, and D was examined separately for the 2 groups. Time-dependent Cox regression analysis was used to assess the separate effects of ART, Pneumocystis carinii pneumonia (PCP) prophylaxis, and intravenous immunoglobulin (IVIG). Similar analysis was performed to assess the effect of ART on the risk of progression to severe immunodeficiency (category 3). Progression to category C for children who received prophylactic ZDV in the first weeks of life was compared, with the use of the log rank test, with disease progression in those who had not received prophylaxis.

Cross-sectional analysis was performed with the use of the CDC classification, allocated on the basis of symptoms or signs present at each visit. At 6-month intervals from birth, the numbers of children in each of the CDC classes at that time were ascertained. The intermittence of HIV-related symptoms also was assessed through a look-back exercise for children in follow-up at 5 years of age, providing an example of the clinical and immunologic history of children up to that age.

Analyses were conducted with the use of SAS (Version 6.12; SAS Institute, Inc, Cary, NC) and STATA (STATA Version 6.0; College Station, TX) statistical procedures.

RESULTS

Between 1984 and 1999, 170 infected children were born to 170 mothers. The likely mode of HIV acquisition in the mothers was intravenous drug use (93 women), an intravenous drug-using partner (21), heterosexual transmission (36), blood transfusion (4), or unknown (16). The median number of visits was 15 (range: 1–68). Median length of follow-up was 5.8 years, with a maximum of 15 years. Fifteen children are no longer followed up by ECS clinicians; 4 moved and are now under the care of other clinicians, 8 were living in socially disadvantaged circumstances and their mothers did not return for additional visits, and for 3 the reasons for loss of contact were unclear. The age at which these 15 children were last seen ranged from 5 months to 9 years, and the number of recorded visits ranged from 3 to 26.

Overall Disease Progression in Infected Children

Clinical Status

Forty-five children (26%) died of AIDS, and 2 died of non–HIV-related causes. Including 4 children with LIP (category B), 70 children (41%) progressed to AIDS. More than 15% of infected children will have progressed to category C or death by 1 year of age, rising annually thereafter by 7% to just under 40% by 5 years and then by 2% to just under 50% by 10 years (Fig 1).

Immunologic Status

Overall, progression to moderate or severe immunodeficiency was common and rapid. By 1 year of age, an estimated 50% of children would have progressed to category 2, >90% by 5 years. After 1 year of follow-up, just >20% of infected children will have evidence of severe immunodeficiency (category 3), rising to >50% by 5 years and to approximately 75% by 10 years of age. Immunologic status generally poorly reflected clinical condition.

Progression to Serious Disease by Calendar Period of Birth

The 21 children who were born after 1994 (when treatment policy was to initiate combination therapy at an early stage) were significantly less likely to progress than those who were born at a time when treatment was not widely recommended (N = 47) or recommended for symptomatic children only with the use of monotherapy or combination therapy with 2 nucleoside reverse transcriptase inhibitors (N = 101; Fig 2). For example, by 1 year of age, >25% of
the children in cohort 1 would have progressed to category C compared with 15% in cohort 2 and only 5% in cohort 3 (log rank \( \chi^2 = 9.10; P = .011 \)).

The distinction by birth cohorts on the basis of treatment policy only partially reflected actual treatment received; 53% (25 of 47) of children born into cohort 1, 75% (76 of 101) of children born in cohort 2, and 62% (13 of 21) of children born in cohort 3 actually received ART. However, 80% of children in cohort 1, 75% (57 of 76) in cohort 2, and 54% (7 of 13) in cohort 3 were started on treatment only after they had progressed to serious disease.

Progression in Untreated Children

Of the 170 infected children, 55 never received ART, only 8 of whom were born at a time when early initiation of combination therapy was the recommended policy (22 were born in the pretreatment era and 25 when monotherapy was recommended for symptomatic children). Ten (18%) untreated children remained asymptomatic when last seen between the ages of 1.5 months and 7.6 years (median: 4.5 years). Forty-five children (82%) received a diagnosis of being category A or having more severe HIV-related
symptoms or signs (Table 1); in all but 2 children, this occurred in the first year of life. Overall, 20 (36%) of the 55 children who did not receive treatment died. Progression was rapid, and nearly one third of untreated children would have progressed to serious disease or death by 1 year of age.

Progression in ART-Treated Children

Children were classified as treated, with progression before or after initiation of treatment. At start of treatment, 81 (70%) of 115 treated children received monotherapy (usually ZDV), 33 (29%) received combination therapy without protease inhibitor (PI), and only 1 received combination therapy with PI. The median age by clinical stage at initiation of treatment was 2.5 years for treated children who had category C symptoms, 4 years for those in category B, 7 years for those with mild symptoms, and 6 years for the asymptomatic children. The healthier the child at treatment initiation, the longer they lived: 90% of those children who were asymptomatic at treatment initiation were still alive 5 years later compared with 0% of the children who received ZDV as prophylaxis for vertical transmission (Table 1; log rank = 252; P = .092). However, these results should be interpreted with caution because of the small sample size.

Comparison of Progression in ART-Treated and Untreated Children

The estimated progression to category C and to death initially was more rapid in the untreated children (Table 1), which almost certainly reflects treatment opportunity: these children were enrolled in the early years of the study and died before having had the chance to be treated. By 10 years of follow-up, the estimated cumulative progression to category C is more similar in the 2 groups (61% vs 43%), although the overall patterns were significantly different (log rank $\chi^2 = 4.41; P = .036$).

Progression by Treatment

The immediate risk of progression to category C in children after initiation of ART, PCP prophylaxis, and IVIG separately were investigated by time-dependent Cox regression analysis. In this analysis, any child who had progressed to category C before having received treatment was regarded as untreated. At any time, children who received ART or PCP prophylaxis were less likely to progress to category C than those who did not receive it (relative risk: 0.58 and 0.63; $P = .037$ and .071, respectively). Most children who started ART before progression to category C disease were started on combination therapy. Of the 115 children who received ART, 76 also received PCP prophylaxis; thus, it is difficult to distinguish independent effects of these treatments. There was no evidence of a difference in progression to category C between children did or did not receive IVIG (relative risk: 1.02; $P = .923$).

Risk of progressing to severe immunodeficiency (immunologic category 3) was 64% higher when receiving ART (risk ratio: 1.64; 95% confidence interval: 0.96–2.82) with borderline significance ($P = .070$).

ZDV Vertical Transmission Prophylaxis

A total of 157 (92%) infected children had not received ZDV as prophylaxis for vertical transmission. Nearly 20% of these children would have progressed to category C or death in the first year of life compared with 0% of the children who received prophylaxis. Disease progression patterns were distinct with borderline significance (log rank test $\chi^2 = 2.84; P = .092$). However, these results should be
interpreted in the light of the more widespread use of ART therapy for infected children after 1995. Of the 13 children who received prophylaxis with ZDV, 11 (85%) also subsequently were treated with ART; 91% (10 of 11) were treated with combination therapy.

Presence of HIV-Related Symptoms at 6-Month Age Intervals

At each 6-month visit, most infected children were without serious symptoms or signs (Fig 3). For each of the points before 3 years of life, just more than half of children were asymptomatic, and after 4 years of age, fewer than one quarter of the children had symptoms at any one time. This picture is very similar for both ART-treated and untreated children (data not shown). Infected children who died contribute most of their information in the first 4 years of life; after that age, the picture largely relates to the experience of survivors. However, restricting the analysis to 45 children who died with AIDS showed a similar pattern to that of the overall cohort, with approximately half of the visits reflecting an absence of clinical symptoms, but when there were symptoms, they were likely to be serious.

Clinical Status 5-Year Look-Back

Seventy-six children who were still alive and in follow-up at or beyond age 5 were included in the 5-year look-back analysis (Fig 4). Of the 944 previous visits of the 58 children with no or only minor symptoms at 5 years, 14 had been category C visits. However, the 1 child who had category C symptoms at age 5 primarily had been symptom-free up to that age. The 1 previous visit for category C symptoms was, in fact, the 6-month visit before the 5-year visit when the child had esophageal candida. Conventional cumulative progression of disease thus does not predict clinical status at subsequent visits.

Of these 76 children, 64 had been treated with ART and 12 were untreated. Of the treated children, 46 (72%) were asymptomatic at 5 years of age and had been asymptomatic most (521) of the previous 740 visits, too. One treated child received the diagnosis of having a C-defining illness at the 5-year visit. The 12 never-treated children all were asymptomatic at the last visit closest to 5 years of age, and none of them had ever had a C-defining illness.

With the exception of the 1 child in category C at the 5-year visit, children in clinical category N were most likely to be on combination therapy: 19% (11 of 58) in category N compared with 9% (1 of 11) and none in category B. The immunologic look-back of 5-year-olds gave a similar impression as the clinical exercise.

DISCUSSION

An estimated 15% of infected children will progress to serious disease or die before age 1, but both infected children who had ever and never been treated with ART are well most of the time. Overall, progression of disease is less rapid between the ages of 1 and 5 years and slow between 5 and 10 years. Fewer than 10% of infected children will remain asymptomatic throughout the first 5 years of life. The current analysis extends to 10 years of age and elaborates on the previous one on combined data from the ECS and the French cohort studies, which showed similar rates of disease progression in the first 5 years of life. All 19 children who progressed to moderate disease (class B) because of LIP were born before 1993, and the diagnosis of LIP now is made less frequently. Ten of these 19 died, 3 within days of first diagnosis of LIP, suggesting that the disease may well have been misdiagnosed.

With the more widespread use of ART to delay disease progression in children, it is no longer appropriate to talk in terms of the natural history of vertically acquired pediatric HIV infection. Therefore, we described patterns of disease in both treated and untreated children and grouped children into 3 cohorts according to therapeutic policy recommendations. There is variation in treatment decisions between centers, likely influenced by the experience of the local clinician, and treatment policy changes over time. Our results show that a subset of vertically infected children can do well, even without early initiation of treatment, provided that they survive beyond infancy. Fifty-five infected children who were followed up to 13 years of age were never treated with ART. This group consisted of a few children who were less likely to progress and thus
less likely to be put on ART and a larger group of 22 children who were born in the early years of the study when treatment was recommended only for seriously ill children who progressed so rapidly that treatment could not be initiated in time. Treatment decisions also may be influenced by clinical status, and when treatment was initiated in children before they progressed to serious disease, the risk of further progression was decreased. Treatment decisions also are likely to be influenced by policy and general recommendations. For example, in our cohort, children who were born before 1989 were less likely to be treated early and more likely to be started on monotherapy than those who were born after 1994. Children who were born at a time when combination treatment was becoming more widespread and recommended to be initiated at an early stage were considerably less likely to progress to serious disease than those who were born in the early years of the study when few therapy options were available and when therapy was confined to those with serious disease. This confirms the recent findings from the Italian pediatric register, which suggested that overall survival of vertically infected children improved with the introduction of combination ART.11

Immune status did not reflect clinical stage at any age, and progression of disease is difficult to predict by immune status.12 Immune status assessed by absolute values at each age probably reflects poorly the ongoing virological activity, and divergence from age-related values and centiles may be more clinically relevant. There is no evidence for the existence of a subgroup of infected children who are always severely symptomatic.

In our cohort, progression of disease slowed after the first year of life, and this brings up the issue of when to stop PCP or bacterial infection prophylaxis. Children who received PCP prophylaxis (most commonly trimethoprim-sulfamethoxazole) were less likely to progress to a C-defining illness than those who were not on prophylaxis, but most of these children also were receiving ART. There was no evidence of any beneficial effect of IVIC.13 The small number of infected children who had been exposed to ZVD prophylaxis to reduce the risk of vertical transmission in their first weeks of life were less likely to progress to serious disease than those who were not exposed. However, this may partly reflect the increasing experience in the management of pediatric infection in recent years. The relatively small numbers and nature of the study make it difficult to draw reliable conclusions. In the Italian pediatric register,14 children who were exposed to ZVD in utero were more likely to progress quickly than children who were not exposed, but this analysis also included children who were born to women who received ZVD for clinical indications rather than as vertical transmission prophylaxis.

The approach taken in this analysis is broader than that used by others1,5,15,16 in that a CDC category was allocated at each visit and children were allowed to move to a less serious category as their clinical or immunologic status improved. We thus were able to show that at each 6-month visit, the majority of infected children, whether treated or not, are asymptomatic. This was true also for children who died during the study period. Fluctuations in clinical symptoms mainly were associated with intermittent serious infections, whereas children who received a diagnosis of encephalopathy remained seriously ill or died. These findings have implications for health and education service provision for these children. Adherence to therapy requires not only cooperation from the child and his or her parents but also sustained support from a multidisciplinary team of professionals. In recent years, the emphasis in the care of these children increasingly has included psychosocial support as well as clinical care.

Our findings indicate a generally better-than-anticipated clinical picture for children with vertically acquired HIV infection and may have implications for populations in resource-poor settings. Infants who survive an initial episode of serious opportunistic disease are subsequently likely to become asymptomatic or to have only mild symptoms for prolonged periods, even without aggressive ART.

STUDY PARTICIPANTS

The European Collaborative Study includes the following individuals: Dr C. Giaquinto, Dr E. Ruga, and A. De Rossi (Università degli Studi di Padova, Italy); Dr I. Grosch-Wömer, Dr K. Seel, and Dr A. Schäfer (Charité Virchow-Klinikum, Berlin, Germany); Dr J. Mok (Royal Hospital for Sick Children, Edinburgh); Dr F. Johnstone (Department of Obstetrics, University of Edinburgh, UK); Dr G. García-Rodriguez, Dr I. Botas, Dr I. de José, Dr F. Hawkins, Dr C. Ladrón de Guereva, Dr J. M. Pera, Dr I. Gonzalez Garcia, and Dr J. R. Arribas Lopez (Hospital Infantil La Paz, Madrid); Prof F. Asensio-Botet, Dr M. C. Otero, Dr D. Pérez-Tamarit, Dr S. Ridauru, Dr P. Gregori, and Dr R. de la Torre (Hospital La Fe, Valencia, Spain); Dr H. Scherpberger, M. Kreyenbroek, and Dr K. Boer (Academisch Medisch Centrum, Amsterdam, The Netherlands); Dr A. B. Bohlin, Dr E. Belfrage, and Dr B. Christenson (Huddinge, Karolinska, and Lund University Hospitals, Sweden); Dr P. Barlow, Dr M. Hainaut, Dr A. Feltier, and Dr S. Wibaut (Hospital St. Pierre, Brussels, Belgium); Dr A. Ferrazin, and Prof D. Bassetti (Department of Infectious Diseases, University of Genoa, Italy); Dr A. De Maria (Department of Internal Medicine, University of Genoa, Italy); Dr C. Gotta (Department of Obstetrics and Gynecology-Neonatology Unit, University of Genoa, Italy); Dr A. Múr, Dr A. Payá, Dr M. Viñolas, Dr M. A. López-Vilchez, Dr P. Martínez-Gómez, and Dr R. Carreras (Hospital del Mar, Universitat Autonoma, Barcelona, Spain); and Dr N. H. Valerius (Hvidovre Hospital, Denmark).

ACKNOWLEDGMENTS

We thank K. Bennett and Dr Simona Fiore (London); Prof L. Chieco-Bianchi, Prof F. Zacchello, Dr R. D'Elia, Dr A. M. Laverda, Dr S. Cozzani, Dr C. Cattelan, Dr A. Mazza, Prof B. Grella, Dr A. R. Del Mastro, Dr V. Giacommet, Dr O. Rampon, and S. Oletto (Padua); Dr Cornelia Feiterna and Dr R. Weigel (Berlin); Dr S. Burns, Dr N. Hallam, Dr P. L. Yap, and Dr J. Whitelaw (Edinburgh); Dr B. Sancho and Dr G. Fontan-Casanego (Madrid); Dr F. Asensi, Dr M. C. Otero, Dr A. Perez Tamarit, Dr A. Gonzalez Molina, Dr M. Gobnardo, Dr J. L. Lopez, and Dr J. Cordoba (Valencia); Dr G. Mulder, Mevr T. Kostien, Mevr M. C. A. van Leeuwen, and the participants of the Dutch collaborative study of HIV-infected women and their children (Amsterdam); Dr L. Navé, Dr G. Lidin-Janson, Dr R. Ljung, Dr P. Bolme, and Dr U. Ewald (Sweden); Dr G. Di Siena, Dr E. Pontali, Prof M. F. Pantarotto, G. Mantero, and Dr P. Dignetti (Genoa); and Dr A. Hottard, Dr M. Poncin, Dr S. Sprecher, Dr B. Lejeune, Dr C. Gissiess, and Prof N. Clumeneck (Brussels).
REFERENCES

2. Diaz C, Hanson C, Cooper BK, et al. Disease progression in a cohort of infants with vertically acquired HIV infection observed from birth: the women and infants transmission study (WITS). / Acquir Immune Defic Syndr Hum RetroviroL 1998;18:221-228

THE MERCHANTS OF COOL

Some of the biggest and savviest corporations in the world are drooling over the cash smoldering in the pockets of your teenagers' name-brand jeans. And these commercial commandos are so skillful at extracting it that the poor lambs think each purchase is all their own idea . . .

That's the warning surrounded in a thoughtful, witty report, "The Merchants of Cool," part of the "Front-line" series on PBS . . .

Still, the hunt goes on to woo the money of today's 32 million American teenagers . . . The media watches kids and then sells them an image of themselves. . . And just what is the image? Here's the scariest part. The ever-conspiring spheres of advertising, television and pop music, are serving up a pair of one-size-fits-all archetypes, one male, one female; Mr. Rushkoff [writer and media analyst] calls them the mook and the midriff . . .

The mook is a walking case of arrested development —"crude, loud, obnoxious"—like the comic Tom Green, the guys of "The Man Show," or Howard Stern, all of whom labor merrily in the fertile fields of Viacom. "There is no mook in nature," Mr. Rushkoff says, perhaps too hopefully. "He is a creation designed to capitalize on the testosterone-driven madness of adolescence" . . .

His counterpart, named after a region she is wont to display, is a bundle of old-fashioned stereotypes "repackaged as a new kind of female empowerment." Britney Spears. Get the picture? A creature who understands how to capitalize on her sexuality long before she can spell puberty.


Noted by JFL, MD
Level and pattern of HIV-1-RNA viral load over age: differences between girls and boys?

European Collaborative Study*

Objective: To estimate RNA viral load patterns over age in vertically infected children that account for between- and within-individual variation, treatment and assay cut-off detection level. To investigate possible sex-based differences.

Design: A total of 118 infected children with 894 RNA viral load measurements enrolled in the European Collaborative Study were prospectively followed from birth for up to 15 years.

Methods: Fractional polynomial and mixed effects models with censored data to assess the non-linear pattern of viral load over age, allowing for repeated measures.

Results: The RNA viral load peaked at approximately 3 months of age, and gradually declined thereafter. The sex by age interaction was significant ($\chi^2 = 19.7, P < 0.001$); viral load peaked higher for girls than boys, but after 4 years the RNA load was consistently 0.25–0.5 log_{10} lower for girls than boys. The effects of sex and treatment on viral load vary over age ($\chi^2 = 6.31, P = 0.043$). Sex differences in RNA viral load relating to measurement without treatment were more pronounced than those under treatment. Disease progression was more rapid for girls than for boys up to the age of 4 years, and less rapid thereafter; the overall difference was not statistically significant.

Conclusion: Differences in RNA viral load over age between untreated boys and girls may have implications for policies for the initiation of antiretroviral therapy, but do not seem to translate into differences in progression to serious disease. The findings would suggest underlying biological explanations, which need further investigation.

AIDS 2002, 16:97–104

Keywords: Paediatrics, sex, vertically acquired infection, viral load

Introduction

The clinical progression of disease in HIV-infected children has been well described, and without early antiretroviral treatment, approximately 20% of infected children will have been diagnosed with AIDS or will have died by 1 year of age [1]. Progression is much slower in the remaining infected children, and by 10 years of age an estimated 40% of these will have serious manifestations of HIV disease or will have died [1]. However, most infected children, irrespective of a previous diagnosis of AIDS, have no or only mild symptoms or signs at any given point in time, even when untreated [2].

The pattern of HIV-RNA peripheral viral load in vertically acquired infection has been described in several studies [3–5], mostly on a cross-sectional basis and not always taking into account repeated measurements. Although little is known about viral load patterns in vertically infected children over the first 10 years of life, the dynamics are likely to vary between

Prepared by Marie-Louise Newell, Linsay Gray, Claire Thorne, Mario Cortina-Borja and Catherine Peckham, Institute of Child Health, University College London, UK.

Correspondence to: Professor Marie-Louise Newell, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.

Tel: +44 20 7829 8699; fax: +44 20 7813 8145; e-mail: m.newell@ich.ucl.ac.uk

*for European Collaborative Study collaborators see Appendix 1.

Received: 6 July 2001; revised: 29 August 2001; accepted: 6 September 2001.
and within individual children, with levels in the first 4–6 months of life expected to reflect primary infection. In adults, viral load has been reported to differ according to sex and race in some studies, although the implications of these findings on disease progression remain unclear [6,7].

Viral load is likely to be associated with progression of disease in vertically infected children, as it is in adults, but this relationship is not well documented, especially not on a dynamic basis over a number of years. Progression to moderate or severe immune deficiency, as indicated by CD4 cell measurements, occurs rapidly in children and has been shown to be poorly associated with clinical progression [1,2].

Longitudinal information on a substantial number of infected children enrolled in the European Collaborative Study (ECS) and followed for almost 15 years provides a unique opportunity to investigate the dynamics of vertically acquired infection in an appropriately rigorous manner. It can also be used to confirm the existence, and estimate the extent, of sex differences, and to investigate the implications for the progression of disease and response to treatment. We previously described fluctuations in clinical disease over the first 10 years of life [2], and now investigate virological markers of infection up to 15 years of age.

Materials and methods

The ECS, a prospective study, has been on-going since 1986. Children born to women known to be HIV infected at or before delivery are followed according to a standard protocol, with detailed clinical and laboratory information, including RNA viral load and CD4 cell count, in 11 paediatric centres from eight European countries [8–10]. Children are seen at birth, 3 and 6 weeks, 3, 4.5 and 6 months, and then at 3 monthly intervals until 24 months. Subsequently, infected children are examined at least twice a year, according to the same clinical and laboratory protocol; current treatment is recorded on standard forms. Parental consent is obtained before enrolment in the ECS, and the study was approved by local ethics committees. Information relating to the infected children enrolled in the ECS and entered by 1 April 2001 was included in the analysis.

The analysis presented here is based on children known to be HIV infected according to the Centers for Disease Control and Prevention definition of paediatric HIV infection [11]. A child is classified as infected after the onset of AIDS, or the detection of virus or antigen in at least two blood samples (taken on separate occasions), or the persistence of antibody beyond 18 months of age. A child is presumed uninfected if at least two blood samples are antibody negative and if no virus or antigen has ever been identified.

In the absence of standard treatment regimens for HIV-infected children, decisions about the initiation of treatment are based on individual factors, including the clinical and immunological status of the child and adherence concerns (European Collaborative Study, in preparation), and may change over time. In the early years of the ECS, antiretroviral therapy consisted of zidovudine monotherapy to children with advanced disease, but now usually consists of combination therapy at an earlier stage (European Collaborative Study, in preparation).

Laboratory tests, including HIV-RNA polymerase chain reaction and CD4 cell count measurements, were carried out locally, with the assays used recorded on forms. The HIV-RNA copy number was assessed either by nucleic acid sequence-based amplification (NASBA)/Nuclisens (Organon Teknika, Oss, the Netherlands) or Roche Amplicor Monitor, versions 1.0 and 1.5, (Roche Diagnostic Systems, Basel, Switzerland). CD4 cell counts were based on flow cytometry and expressed as the number of cells per cubic millimetre.

Statistical methods

The values of both HIV-RNA viral load and CD4 cell count were log base 10 transformed. As the assay system and type of treatment could vary over age, both assay and treatment status were introduced in a time-dependent manner. Monotherapy was categorized with no therapy because it is unlikely to be associated with a substantial reduction in viral load, an increase in the CD4 cell count, or delayed clinical progression [12]. Children were thus categorized as either treatment naive/receiving monotherapy or treated with a combination of two or more antiretroviral drugs.

The non-linear pattern of viral load over age and the repeated-measures nature of the data required modelling techniques such as fractional polynomial models [13] and mixed effects models with censored data [14]. Fractional polynomial models are flexible, are able to reveal the 'true' curve shape, and allow for modelling asymptotes. Models are compared for significant improvement in fit using likelihood ratio tests [15]. The data also call for a mixed effects model that can account for left censoring of the RNA viral load measurements. The methods described by Hughes [14] offer a modification of the usual EM estimation procedure for fitting mixed effects models with normal errors by accommodating varying censored observations arising from lower (as well as upper) detection limits. Jaqmin-
Gadda et al. [15] described a way of calculating the likelihood function for methods that include Hughes' as particular cases. As it was not possible to account for mixed effects models with censored data within the framework of fractional polynomial model-fitting, we used a two-step process. Initially, the data were treated as independent uncensored observations and the optimal fractional polynomial model with respect to age was ascertained. This model was then taken forward into the method allowing both repeated measures and censoring.

The model was expanded from that involving only the age terms, through the introduction of stratification by sex, treatment and CD4 cell count to assess the significance of these factors separately. The factors were assessed for significance of interaction with the age terms and, where appropriate, with one another. The significance of individual terms was assessed by comparing nested models using likelihood ratio tests [15]. There were sufficient data to expand the model as far as three-way interaction. Viral load profiles were compared for those who did and those who did not progress to serious disease or death in a descriptive manner.

Progression to serious disease in this cohort was assessed by Kaplan–Meier product limit analysis, and plots were compared by sex using the log-rank test. Regression analysis investigated the differences in log_{10} HIV-RNA viral load before and after the initiation of combination therapy, allowing for other factors.

All analyses were carried out using STATA (STATA Version 6.0; College Station, TX, USA) and S-PLUS 3.4 (Insightful, USA) in a Unix environment, utilizing a modified version of the software provided by Hughes in 1999. We used the software described by Jaqrain-Gadda et al. [15] to compute the multivariate normal integrals in the likelihood function.

**Results**

Of the 178 infected children (8.7% of the total cohort) 41 had died and 19 were no longer in follow-up in an ECS centre after the introduction of routine viral load assays. Viral load measurements were thus available on 118 infected children (Table 1). Of the 894 viral load measurements, 163 were below the assay specific cut-

### Table 1. Infected children in the European Collaborative Study.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Children (n = 118)</th>
<th>Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of measurements</td>
<td>7 (1–23)</td>
<td>894</td>
</tr>
<tr>
<td>Viral load values (copies/ml)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>70 (59.3)</td>
<td>731 (81.8)</td>
</tr>
<tr>
<td>Below detection level</td>
<td>40 (34.7)</td>
<td>163 (18.2)</td>
</tr>
<tr>
<td>Sex n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>60 (50.8)</td>
<td>421 (47.1)</td>
</tr>
<tr>
<td>Male</td>
<td>58 (49.1)</td>
<td>473 (53.9)</td>
</tr>
<tr>
<td>Treatment status n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No or monotherapy</td>
<td>30 (25.4)</td>
<td>272 (30.4)</td>
</tr>
<tr>
<td>Combination therapy</td>
<td>88 (74.6)</td>
<td>622 (69.6)</td>
</tr>
<tr>
<td>CD4 T cell count (× 10^3/l)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td></td>
<td>0.67 (0.001–4.82)</td>
</tr>
<tr>
<td>Assay n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NASBA</td>
<td>167 (20.7)</td>
<td></td>
</tr>
<tr>
<td>Roche</td>
<td>639 (79.3)</td>
<td></td>
</tr>
<tr>
<td>Age at treatment initiation (months)</td>
<td>31.5 (7.6–151.1)</td>
<td></td>
</tr>
<tr>
<td>Age at last visit (months)</td>
<td>93.1 (2.2–179.0)</td>
<td></td>
</tr>
<tr>
<td>CDC immunologically stage at last visit n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Immunologically normal</td>
<td>12 (10.2)</td>
<td></td>
</tr>
<tr>
<td>2 Moderate suppression</td>
<td>50 (42.4)</td>
<td></td>
</tr>
<tr>
<td>3 Severe suppression</td>
<td>50 (42.4)</td>
<td></td>
</tr>
<tr>
<td>Died</td>
<td>6 (5.1)</td>
<td></td>
</tr>
<tr>
<td>CDC clinical stage at last visit n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N No symptoms</td>
<td>15 (12.7)</td>
<td></td>
</tr>
<tr>
<td>A Mildly symptomatic</td>
<td>12 (10.2)</td>
<td></td>
</tr>
<tr>
<td>B Moderately symptomatic</td>
<td>70 (59.3)</td>
<td></td>
</tr>
<tr>
<td>C Severely symptomatic</td>
<td>15 (12.7)</td>
<td></td>
</tr>
<tr>
<td>Died</td>
<td>6 (5.1)</td>
<td></td>
</tr>
</tbody>
</table>

CDC, Centers for Disease Control and Prevention; NASBA, nucleic acid sequence-based amplification.
off value, and the median RNA viral load in the remainder was 14690 copies per millilitre of plasma. Most children (88, 75%) had received combination antiretroviral therapy, with a median age at treatment initiation of 31.5 months. Twenty-one (11 girls and 10 boys) of the 118 children had been diagnosed with serious HIV-related disease (Centers of Disease Control and Prevention classes C or D), of whom six had died.

HIV-RNA viral load over age

The three power fractional polynomial model was not significantly better than the best two power ($\chi^2 = 3.74, P = 0.156$); the optimal model for log viral load and age included an inverse square root of age and a log age term. Any measurements taken on the day of birth were set to day one to resolve numerical indeterminations. The pattern of HIV-RNA viral load over age, in both a non-parametric smoother and the curve predicted from the fractional polynomial model, is seen to peak at approximately 3 months of age, with a gradual decline thereafter (Fig. 1).

A total of 805 observations with information on assay type, age, sex, CD4 cell count and treatment were used in the subsequent modelling. Nearly 80% (638/805) were assessed using a Roche assay and 20% (167/805) using a NASBA assay. The shape of the overall pattern of HIV-RNA viral load over age was similar for RNA measurements assessed by Roche or by NASBA assay.

The $\log_{10}$ RNA viral load did not differ by sex in univariate analysis ($P = 0.971$) (Table 2). A one $\log_{10}$ CD4 cell increase resulted in a significant decrease of 0.895 in $\log_{10}$ viral load ($P < 0.001$). Combination therapy was associated with a 1.28 lower $\log_{10}$ viral load compared with the viral load in the no/monotherapy category ($P < 0.001$). The Roche assay was associated with a 0.404 lower $\log_{10}$ viral load than the NASBA assay ($P = 0.0128$).

**Factors influencing HIV-RNA viral load pattern over age**

Two-way interactions of each of the main effects of sex, treatment and CD4 cell count with age were assessed for significance. The sex by age interaction was significant ($\chi^2 = 19.7, P < 0.001$). The pattern of viral load over age in girls differed from that of boys (Fig. 2). This difference was masked in the previous model, which accounted for only the main effect of sex. Viral load in girls peaks a little earlier (1.1 months) than in boys (2.1 months), and at a higher value and declines more sharply in girls than boys, with a cross-over at approximately 4 years of age. After this age the viral load in girls is consistently lower than that in boys. At the peak the estimated HIV-RNA viral load is more than 1 $\log_{10}$ above the peak level for boys at that age, whereas after the cross-over girls have a predicted viral load 0.25–0.5 $\log$ below that for boys. Interactions of treatment ($\chi^2 = 0.599, P = 0.741$) and CD4 cell count ($\chi^2 = 0.235, P = 0.628$) with age were also investigated. Allowing for an interaction between sex and treatment, there was evidence of a difference in the effect of treatment by sex ($\chi^2 = 5.18, P = 0.023$), suggesting that treatment is associated with a larger decrease in the RNA viral load in boys than in girls.

A full model was fitted to incorporate the main effects of sex, treatment, assay type and CD4 cell count and two-way interactions of age by sex, age by treatment and sex by treatment, and a three-way interaction of age by sex by treatment. Because of the significant three-way interaction of inverse square root age by sex by treatment and $\log_{10}$ age by sex by treatment ($\chi^2 = 6.31, P = 0.043$) the viral load pattern over age is presented separately for the four strata defined by sex and treatment for Roche assay observations, at the median CD4 cell count (Fig. 3). The picture for NASBA observations is similar, but levels are a little lower than for Roche measurements. This model predicts that, for example, at age 12 months the difference in RNA viral load between untreated and treated girls is approximately 2 $\log_{10}$, and 1.5 $\log_{10}$ for treated and untreated boys. Similarly, at 15 years of age, the predicted differences would be 1 $\log_{10}$ for girls and 2 $\log_{10}$ for boys.

**Progression to serious disease**

In the ECS cohort the group of children with available viral load information reflects children enrolled in more recent years, when treatment was also more widely available. Therefore progression is less than would be seen in the total cohort. In the group of 118 infected children with available information about the HIV-RNA viral load, 21 children had progressed to serious HIV disease (15) or death (6). Survival estimates of the progression to CDC class C disease or death indicate that 6% (2–11%) of infected children in this group would have progressed by the age of 1 year, 17% (10–
Table 2. RNA viral load pattern over age by sex, treatment, CD4 cell count and assay in individual analyses.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Univariate coefficient</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/sqrt(Age) increase</td>
<td>-1.16</td>
<td>-1.50 to -0.822</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>ln(Age) increase</td>
<td>-0.496</td>
<td>-0.625 to -0.366</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>-0.006</td>
<td>-0.342 to 0.355</td>
<td>0.971</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No/monotherapy</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combination therapy</td>
<td>-1.28</td>
<td>-1.56 to -0.994</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>1 log_{10} CD4 cell increase</td>
<td>-0.895</td>
<td>-1.14 to -0.651</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Assay</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NASBA</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roche</td>
<td>-0.404</td>
<td>-0.722 to -0.086</td>
<td>0.0128</td>
</tr>
</tbody>
</table>

Cl, Confidence interval; NASBA, nucleic acid sequence-based amplification.

24%) by the age of 5 years, and 22% (13–31%) by 10 years of age. Progression to serious disease or death was similar for boys and girls (log rank test $P = 0.79$). There was no statistically significant association by treatment, and indeed children who had received combination therapy were estimated to have a non-significantly increased rate of clinical progression, reflecting late initiation of treatment.

Only seven of the 21 children were observed to have progressed to serious disease or death by 1 year of age, whereas only five of the 21 who progressed at any age had viral load measurements relating to a time before they had been diagnosed with serious disease. Given the rarity of events, analyses relating peak viral load measurements in the first months of life to the risk of progression were considered to be inappropriate.

The difference in log_{10} HIV-RNA viral load measurements before and after the initiation of combination was evaluated in a regression analysis, allowing for sex, viral load assay, initial viral load, and duration between measurements. A total of 34 children (14 girls and 20 boys) had information available, and the decrease in log viral load after the initiation of treatment was estimated to be 0.48 less in boys than in girls, but this did not reach statistical significance ($P = 0.299$).

Although progression did not differ significantly by sex, progression to serious disease or death was more rapid for girls up to approximately 4 years of age and less rapid than for boys after that age. This trend mirrors that seen for the HIV-RNA viral load pattern over age.

**Discussion**

Using data on the infected children enrolled in the ECS, the level and pattern of HIV-RNA viral load over nearly 15 years was described. The overall picture shows a marked peak in viral load at approximately 3
months of age, probably reflecting primary infection, followed by a rapid decline in the first 5 years of life, slowing thereafter. Our analysis used fractional polynomial models that reflect the actual shape of the viral load pattern over age in vertically infected children. Although the results obtained using fractional polynomial methodology are generally sensitive to outlying values, the large number of observations available for analyses here have produced reliable models. The methodology used allows for both repeated measures within children and the censoring effect of the assay cut-off value. Generally, the assay cut-off value for detection is either set at the assay specific point, or arbitrarily half-way between zero and the cut-off. Both approaches lead to bias in the parameter estimates [13,14]. Furthermore, we were able to use longitudinal repeated measurements over 15 years rather than cross-sectional assessments at different ages.

HIV-RNA viral load has been shown in several studies to be associated with the progression of disease overall [16,17], but there is a lack of information on whether this association differs by sex. In a recent conference presentation [18], average relative and absolute CD4 cell counts were higher in girls than in boys both in infected and uninfected children. In infected children, HIV-RNA levels were lower, but not significantly so, at all time points to 18 months in untreated girls than in boys, but any relationship with the progression of disease was not investigated. Using European data, we show here that allowing for repeated measures within individuals in a model with viral load, age and sex, there was a substantial sex difference in viral load pattern over the first 15 years of life. RNA viral load levels were estimated to peak a little earlier and at substantially higher levels in girls than in boys, with a difference of more than 1 log_{10}. After approximately 5 years of age the levels of RNA in girls are up to half a log_{10} below that of boys. This sex difference in the peripheral viral load pattern over age was reflected in the curve of clinical progression, although overall progression was not statistically significantly associated with sex. This may imply that for a given RNA viral load the progression of disease in boys is slower than for girls.

The data presented here were collected within a prospective cohort study, not a clinical trial and treatment decisions were thus not random. The interactions between age, sex and treatment in the assessment of patterns of viral load are complex. The association between treatment and viral load over age differed between girls and boys. For both boys and girls viral load observations under treatment were consistently and substantially lower. However, whereas the difference in RNA viral load under treatment and not under treatment increased slightly for boys, it narrowed substantially for girls. Without treatment, the measurements of RNA viral load are initially higher for girls than boys, but levels cross over at approximately 3 years of age, and thereafter the measurements relating to untreated girls are consistently lower than in boys. With treatment, viral load measurements for girls started off being higher than for boys, but after approximately 9 months of age they become lower than for boys, with very little difference at later ages. Taken together, these findings indicate possible sex-specific dynamics of viral replication.

In adults, lower HIV-RNA viral loads for women than for men have been described, tentatively suggested to be associated with hormonal levels. Hormonal influences could also play a role in children, even in the pre-pubertal stages. Our findings suggest a biological explanation in response to HIV infection. The progression of disease is generally similar in women and men, and in our study there was no difference in overall progression between boys and girls. However, few studies have successfully associated viral load patterns over a prolonged period of time with clinical progression in men and women separately, and most analyses have been cross-sectional [19]. Guidelines for the initiation of antiretroviral therapy based on viral load do not currently allow for sex, although it has been suggested that, for a similar baseline level, women may achieve viral suppression after highly active antiretroviral therapy at a faster rate than men and have a more sustained response [7].

Conclusion

The findings presented here, if confirmed, may have implications for clinical guidelines for the initiation of antiretroviral therapy in children, and suggest that cutoff levels of RNA viral load at which therapy would be recommended may need to be lower for girls than for boys, after the first year of life. For example, WHO guidelines suggest that for children over 1 year of age the initiation of treatment should be considered if the viral load is over 100,000 copies [20]. In our cohort, 22 girls and 32 boys had an HIV-RNA load above this level over the age of 1 year, with 13 girls and 18 boys being treated at or before the time of high viral load. In addition, response to treatment measured through RNA viral load may also need to be interpreted in a different way for boys and girls, with an expected lower average RNA load for boys after the initiation of treatment than for girls.

Acknowledgements

The authors would like to acknowledge support from Mrs L. Toxle and Dr Simona Fiore (London). They
would also like to thank Prof L. Chieco-Bianchi, Prof F. Zacchello, Dr E. Ruqa, Dr R. D’Elia, Dr A.M. Laverda, Dr S. Cozzani, Dr C. Cattelan, Dr A. Mazza, Prof B. Grella, Dr A.R. Del Mistro and Mrs S. Oletto (Padua); Dr C. Feiterna, and Dr R. Weigel (Berlin); Dr F. Johnstone, Dr S. Burns, Dr N. Hallam, Dr P.L. Yap, and Dr J. Whitelaw (Edinburgh); Dr B. Sancho, and Dr G. Fontan-Casanego (Madrid); Dr F. Asensi, Dr M.C. Otero, Dr A. Perez-Tamarit, Dr A. Gonzalez Molina, Dr M. Gobemandro, Dr J.L. Lopez, and Dr J. Cordoba (Valencia); Dr G. van der Plas (Amsterdam); Dr B. Christensson, Dr P. Bolme, and Dr U. Ewald (Sweden); Dr G. Di Siena, and Dr E. Pontali, Prof M.F. Pantarotto, G. Mantero, and Dr P. Dignetti (Belfrage, Dr L. Navér, Dr A. Ehmst and Prof A. Bates, Dr I. de José, Dr F. Hawkins, Dr J. Gonzalez Garcia and Dr J.R. Arribas Lopez (Hospital Infantil La Paz, Madrid); Prof F. Asensi-Boetet, Dr M.C. Otero, Dr D. Pérez-Tamarit, Dr A. Ortí, Dr M.J. San Miguel and Dr R. de la Torre (Hospital La Fe, Valencia, Spain); Dr H. Scherpierbier, M. Kreyenbroek and Dr K. Boer (Academisch Medisch Centrum, Amsterdam, the Netherlands); Dr A.B. Bohlin, Dr E. Belfrage, Dr L. Navér, Dr A. Ehmst and Prof A. Sönnerborg (Huddinge and Karolinska University Hospitals, and Karolinska Institutet, Sweden); Prof J. Levy, Dr M. Hainaut, Dr A. Peltier, Dr S. Wibaut and Dr P. Dignetti (Huddinge and Karolinska University Hospitals, and Karolinska Institutet, Sweden); Dr C. Giaquinto, Dr O. Rampon, Dr V. Giacomet and A. De Rossi (Universita degli Studi di Padova, Italy); Dr I. Grosch-Wörner (Charite Virchow-Klinikum, Berlin, Germany); Dr J. Mok (Royal Hospital for Sick Children, Edinburgh); Dr I. Bates, Dr I. de José, Dr F. Hawkins, Dr M.C. Garcia-Rodriguez, Dr C. Ladrón de Guevara, Dr J. Mª Pella, Dr J. Gonzalez Garcia and Dr J.R. Arribas Lopez (Hospital Infantil La Paz, Madrid); Prof F. Asensi-Boetet, Dr M.C. Otero, Dr D. Pérez-Tamarit, Dr A. Ortí, Dr M.J. San Miguel and Dr R. de la Torre (Hospital La Fe, Valencia, Spain); Dr H. Scherpierbier, M. Kreyenbroek and Dr K. Boer (Academisch Medisch Centrum, Amsterdam, the Netherlands); Dr A.B. Bohlin, Dr E. Belfrage, Dr L. Navér, Dr A. Ehmst and Prof A. Sönnerborg (Huddinge and Karolinska University Hospitals, and Karolinska Institutet, Sweden); and Dr A. Ferrazin and Prof D. Bassetti, (Department of Infectious Diseases, University of Genoa, Italy); Dr A. De Maria (Department of Internal Medicine, University of Genoa, Italy); and Dr C. Gotta (Department of Obstetrics and Gynecology – Neotoniatrics Unit, University of Genoa, Italy); Dr A. Már, Dr A. Payá, Dr M. Vifiolas, Dr M.A. López-Vilchez, Dr M. Movira, Dr R. Carreras, Dr E. Esteban Tores, Dr S. Herrero Perez (Hospital del Mar, Universidad Autonoma, Barcelona, Spain); and Dr N.H. Valerius (Hvidovre Hospital, Denmark).

Appendix

ECS collaborators: Dr C. Giaquinto, Dr O. Rampon, Dr V. Giacomet and A. De Rossi (Università degli Studi di Padova, Italy); Dr I. Grosch-Wörner (Charite Virchow-Klinikum, Berlin, Germany); Dr J. Mok (Royal Hospital for Sick Children, Edinburgh); Dr I. Bates, Dr I. de José, Dr F. Hawkins, Dr M.C. Garcia-Rodriguez, Dr C. Ladrón de Guevara, Dr J. Mª Pella, Dr J. Gonzalez Garcia and Dr J.R. Arribas Lopez (Hospital Infantil La Paz, Madrid); Prof F. Asensi-Boetet, Dr M.C. Otero, Dr D. Pérez-Tamarit, Dr A. Ortí, Dr M.J. San Miguel and Dr R. de la Torre (Hospital La Fe, Valencia, Spain); Dr H. Scherpierbier, M. Kreyenbroek and Dr K. Boer (Academisch Medisch Centrum, Amsterdam, the Netherlands); Dr A.B. Bohlin, Dr E. Belfrage, Dr L. Navér, Dr A. Ehmst and Prof A. Sönnerborg (Huddinge and Karolinska University Hospitals, and Karolinska Institutet, Sweden); Prof J. Levy, Dr M. Hainaut, Dr A. Peltier, Dr S. Wibaut and Dr P. Barlow, (Hospital St Pierre, Brussels, Belgium); Dr A. Ferrazin and Prof D. Bassetti, (Department of Infectious Diseases, University of Genoa, Italy); Dr A. De Maria (Department of Internal Medicine, University of Genoa, Italy); and Dr C. Gotta (Department of Obstetrics and Gynecology – Neonatal Unit, University of Genoa, Italy); Dr A. Már, Dr A. Payá, Dr M. Vifiolas, Dr M.A. López-Vilchez, Dr M. Movira, Dr R. Carreras, Dr E. Esteban Tores, Dr S. Herrero Perez (Hospital del Mar, Universidad Autonoma, Barcelona, Spain); and Dr N.H. Valerius (Hvidovre Hospital, Denmark).

Sponsorship: The European Collaborative Study is a concerted action of the European Commission (Biomed II PL 97 2005 and QLRT-1999-30002). The Medical Research Council (UK) provided support to the coordinating centre. Collaborating centres were supported by grants from the Ministero della Sanita – Istituto Superiore di Sanita, Progetto AIDS (Padua, Genoa); the Medical Research Council (UK); the AIDS Virus Education Research Trust, the Scottish Office Home and Health Department (Edinburgh); Praeventiefonds No. 28-1704 (Amsterdam); Bundesminister fur Gesundheit (Berlin); Fonds Houtman, Office de la Naissance et de L’Enfance, Communauté Francaise de Belgique (Brussels); and the Research Foundations of the Karolinska Institutet (Stockholm).

References


Full title: Modelling HIV RNA viral load in vertically infected children

Short title: Modelling paediatric HIV RNA viral load

Linsay Gray, MSc, Mario Cortina-Borja, PhD, Marie-Louise Newell, PhD, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, University College London, UK.

Author for correspondence: Linsay Gray, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK. Tel: +44 20 7829 8699 Fax: +44 20 7813 8145 Email: L.Gray@ich.ucl.ac.uk
Summary

Human Immunodeficiency Virus (HIV) ribo-nucleic acid (RNA) viral load is a measure of actively replicating virus and is used as a marker of disease progression. For a thorough understanding of the dynamics of the evolution of the virus in the early life of HIV-1 vertically infected children, it is important to elucidate the pattern of HIV RNA viral load over age.

An aspect of assay systems used in the quantification of RNA viral load is that they measure values above particular cut-off values for detection, below which the assays used are not sufficiently sensitive; in this way measurements are potentially left-censored. Recent adult studies suggest that to adequately model RNA pattern over age, it is necessary to account for within-subject correlation, due to repeated measures, and censoring.

The aim of this study therefore was to establish whether it is necessary to use complex methods to allow for repeated measures within individuals and censoring of the HIV RNA viral load in children enrolled in a cohort study.

The approach involved the identification of an appropriate model for the basic pattern of RNA viral load by age and subsequent assessment of various estimation procedures accounting for repeated measures and censoring in different ways. Methods developed by Hughes (1999) involving the expectation-maximisation (EM) algorithm and the Gibbs sampler were taken as the benchmark for comparison of simpler alternatives. Other approaches considered involve linear mixed-effects and ordinary least squares in which censoring is dealt with informally by taking the cut-off value as absolute or taking the mid-point between cut-off and zero.

Fractional polynomials provided a substantially superior approach for modelling the dynamics of viral load over age compared to conventional polynomials or change-point models. Allowing for repeated measures was necessary to improve the power of the likelihood ratio tests required to establish the final model, but methods beyond taking the mid-point for censored values did not further improve the fit. Although Hughes’ methodology is the best approach, its implementation is not necessary for the identification of the optimal model.

Key words: Fractional polynomial, repeated measures, censoring, viral load, HIV-1.
1) Introduction
Vertically infected children have acquired HIV-1 infection during gestation or delivery. RNA viral load indicates actively replicating virus in infected individuals and patterns over age in these children have only recently been elucidated\(^1\). The level varies between and within individuals over time, peaking early in life, reflecting primary infection, and declining gradually thereafter.

Cohort studies provide repeated HIV RNA viral load measurements which are non-independent due to within-subject correlation. Previous descriptions of virological patterns in vertically acquired infection have been mostly on a cross-sectional basis or have not always taken into account repeated measurements\(^2,3\). In determining the pattern, investigators have used simple exploratory methods such as plotting means or medians at different time points\(^4\). These methods do not permit assessment of the effect of explanatory variables. Others have used basic polynomial forms involving quadratic terms which may not account for all the important aspects of the data structure\(^5\). Where within-subject correlation has been addressed\(^6\), methodology was confined to natural cubic splines but in the introduction of various explanatory factors might be difficult with this methodology (as models would involve many parameters).

Furthermore, because HIV-RNA assays have a lower limit below which quantification is not possible, RNA viral load measurements below these assay-specific threshold levels are left-censored. Crude approaches have been taken to allow for either using the threshold values or some arbitrary point, such as the mid-point between the level for detection and zero, or the cut-off for the detection itself\(^7\). Approaches such as these may bias the results in that prediction based on the censored values for the actual values would be systematically higher than prediction based on the true, unknown values below cut-off. This problem is more serious for data ascertained using older generations of assay systems which are less accurate due to higher detection cut-off points in the region of 2000-4000 copies per ml. The necessity of accounting for repeated measurements and censoring in such data has been demonstrated in HIV-infected adults\(^8,9,10,11\), but whether this holds for paediatric data and if so, the degree of complexity required, has not been systematically investigated.

As the pattern of RNA viral load changes non-linearly with respect to age, consideration must be given to the methods available which identify parsimonious models to describe this relationship. Analysing data from cohort studies requires methodology that can also accommodate the censored nature of the repeated measurements. Methods described by Royston and Altman (1994), used extensively for modelling non-linear relationships\(^12\) and Hughes (1999), previously used in the modelling of adult HIV RNA viral load\(^13\), could fulfil these requirements.

The aim of this study was to determine whether complex methods are required to allow for repeated measures within individuals and for censoring of the HIV RNA viral load in the modelling of data from children enrolled in a cohort study. Interest was in the estimation of RNA viral load patterns in vertically infected children, assessing treatment and gender-based differences over age while accounting for assay and CD4\(^+\) cell counts. The analysis comprised a two-step approach: First, conventional polynomials, change-
point methods, and fractional polynomials were compared to identify the most suitable model for the basic pattern of RNA viral load by age. Then, as a benchmark, we incorporated covariates and their interactions, accounting for the repeated measurements structure and the censoring scheme using the sophisticated approach of Hughes’. We assessed the capabilities of four alternative, less complex estimation procedures, by comparison to Hughes’ method.

2) The European Collaborative Study Data
Since 1987, enrolment and follow up of children born to HIV-1 infected mothers has continued according to the standard ECS protocol, in 10 paediatric centres from seven European countries[2-4]. Children are seen at birth, around 3 and 6 weeks, 3, 4.5 and 6 months and then at 3-monthly intervals until 24 months. Subsequently, infected children are seen at least twice a year, and clinical and laboratory information and current treatment recorded on standard forms. Parental consent is obtained before enrolment in the ECS, and the study is approved by the local ethics committees.

Available and acceptable HIV treatment options have varied over time. When given, antiretroviral therapy consists of monotherapy, usually zidovudine (ZDV) or a combination of different anti-retroviral therapies (ARTs). The initiation of administration and adjustment of composition of treatment varies over age for different children.

HIV RNA copy number is assessed by either of two groups of assay systems: NASBA/Nuclisens (Organon Teknika, Oss, The Netherlands) or Roche (Amplicor Monitor, versions 1.0 and 1.5, Roche Diagnostic Systems, Basel, Switzerland).

The values of both HIV-RNA viral load and CD4+ cell count were log base 10 transformed to resolve heteroscedasticity and non-Normality of residuals. Monotherapy was categorised with no therapy since it is unlikely to be associated with a substantial and sustained reduction in RNA viral load, or delayed clinical progression[13]. Measurements were categorised in two groups according to treatment exposure of the child at that time: the first comprising treatment naïve observations and those taken when the child was exposed to only monotherapy, the second corresponding to exposure to combination of two or more ARTs. Any measurements taken on the day of birth were set to day one to overcome numerical indeterminations. As assay system and type of treatment varied over children’s age, both were introduced as time-dependent variables.

Of the 805 observations of HIV RNA viral load available, 147 (18%) were censored; 636 (79%) had been ascertained by Roche and 169 (21%) by NASBA. The detection levels, which vary according to the assay used and its generation, are 4000 (4 measurements), 1000 (1), 800 (2), 500 (13), 400(21), 200(61), 150(1), 100(1), 85(1), 80(1), 50 (30), 40(11). These measurements relate to 118 infected children (58 boys and 60 girls); the median number of measurements per child was 6.5, (range 1 to 23). Most children (88, 75%) had at some point been initiated on combination therapy; the remaining 30 received no therapy or only monotherapy.
3) Modelling changes in HIV RNA viral load over age

The approach to the modelling of HIV RNA viral load over age consisted of a two-step process (Figure 1). First the optimal basic model representing the linear predictor for the overall pattern of HIV RNA viral load as a function of age was chosen among conventional polynomials, change-point models, and fractional polynomials. Then, this function of age was fitted, incorporating important covariate and interaction terms according to the chosen model selection strategy, using, in decreasing order of complexity: 1) Hughes’ Monte Carlo EM algorithm, 2) linear mixed effects (LME) with censored values taken as midway between zero and the cut off values of the assays’ threshold, 3) LME’s with censored values taken as the assays’ cut off values, 4) ordinary least squares (OLS) with censored observations as half the assays’ cut off values, and 5) OLS using the assays’ recorded censoring cut-off values. Hughes’ procedure allows fitting mixed effects models with censored observations, and was taken as the benchmark for comparison with the four simpler estimation procedures considered. The two-stages are elaborated below:

3.1) Establishing the basic model for the overall viral load dynamics over age

To establish the underlying structure of the data, a super smoother algorithm involving automatic span selection by means of cross validation was performed in S-PLUS 2000 (Insightful, USA) in a Windows environment.

3.1.1) Conventional polynomials

We modelled the basic pattern of RNA viral load by age, firstly, using conventional polynomials. Within this framework, models involving terms up to quartic powers were considered to be as elaborate as necessary without loss of parsimony.

3.1.2) Change-point analysis

An alternative is to fit a change-point regression model with an intercept and slope for the initial increase to peak in RNA viral load and another slope for the decline. We considered a change-point model consisting of two straight lines that intersect at age $\tau$, the peak reached by HIV viral load measurements in early ages. The value of $\tau$ was estimated as the age where the running smoother reached its maximum, although other approaches, e.g. segmented regression or partial likelihood could also have been used.

We defined $U_{ij}$, for individual $i$ at the $j$-th examination, as the minimum of $Age_{ij}$ and $\tau$; $T_{ij}$ is defined as the product of $(Age_{ij} - \tau)$ and an indicator function which is 0 for $Age_{ij} < \tau$ and 1 for $Age_{ij} > \tau$. The model is given by:

$$\log_{10}(Viral\ Load_{ij}) = \beta_0 + \beta_1 U_{ij} + \beta_2 T_{ij} + \varepsilon_{ij}$$

3.1.3) Fractional polynomial models

Lastly, the RNA viral load was modelled over age using the approach involving fractional polynomials. These models consist of a group of curves, intermediate between polynomials and curves non-linear in their parameters. A fractional polynomial of degree $d$ can be expressed as $\phi_d(X; \xi, \beta) = \sum_{j=0}^{d} \xi_j H_j(X)$, where for $j=1, \ldots, d$.
$H_j(X)$ is $X^{(p_j)}$ if $p_j \neq p_{j-1}$, and $H_{j-1}(X) \ln(X)$ if $p_j = p_{j-1}$, and $X^{(p)}$ denotes the Box-Tidwell transformation, defined as $X^p$ if $p \neq 0$, and $\ln(X)$ if $p = 0$. The vectors $p = (p_1, \ldots, p_d)$ such that $p_1 < \cdots < p_d$, and $\xi = (\xi_0, \xi_1, \ldots, \xi_d)$ contain the model's powers and coefficients. Usually the power terms are restricted to a predetermined set of integer and non-integer values; the set $\{-2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2, 3\}$ was used. Model selection is determined by changes in the deviance (defined as minus twice the log likelihood); among fractional polynomials with degree $d$ the model with lowest deviance is deemed to have the best fit. In deciding the adequacy of the optimal model with degree $d$ compared to the best fitting model with degree $d+1$, the difference in deviance is compared to $\chi^2_d$ since both an extra power and a regression coefficient are involved. Since the parameter space is discrete, these tests are asymptotically conservative; we did not attempt to modify them. The fractional polynomials were fitted using STATA (STATA Version 6.0; College Station, TX).

The most suitable of the three approaches for modelling the linear predictor of HIV RNA viral load by age was identified by graphical comparison to the super smoother and by evaluation of the Akaike information criterion (AIC).

3.2) Model selection strategy and estimation procedures

3.2.1) Model selection strategy

Having established the most appropriate underlying model for the pattern of RNA viral load by age, the next step was to incorporate relevant covariates and their interactions while accounting for repeated measures and censoring. There was an a priori interest in the differential patterns of HIV-RNA viral load by gender and treatment over age, accounting for CD4$^+$ cell count and assay type. As our main concern was to investigate age-related differences between gender and treatment, those were the only two-way and three-way interactions considered. However, we also adjusted by the main effects of CD4$^+$ cell count and assay; the former because it reflects the child’s health status at any given age, and the latter because it is a centre-specific, rather than a subject-specific measure, which would induce a shift in the corresponding individual log viral load curve. This imposed a model encompassing the main effects of age, gender, treatment, CD4$^+$ cell count and assay type, and two- and three-way interactions of age, gender and treatment as the upper bound for the scope of our model selection strategy.

In the five estimation procedures considered, the decision to fit the three-way interaction age by gender by treatment was determined by the significance of the three corresponding two-way interactions: if at least two of them were significant at the 5% level we would expand the model to include the three-way interaction. If this criterion was satisfied, the three-way interaction would then be assessed adjusting by the main effects of CD4$^+$ cell count and assay. In determining important variables, models were compared using likelihood ratio tests (LRT).
3.2.2) Comparison of the estimation procedures

We start with the most sophisticated model framework, Hughes’ modified Monte Carlo EM algorithm:

The methods described by Hughes\(^8\) offer a modification of the usual EM algorithm for fitting mixed effects models with Normal errors\(^16,17\) by accommodating censored observations arising from lower (as well as upper) detection limits. A general solution that works with any censoring scheme with an arbitrarily complex design matrix for the random effects is to use the Gibbs sampler to maximise the likelihood. This extends the applications proposed by Pettitt\(^18\) for the one-way random-effects model.

The standard mixed effects model\(^17\) can be written as

\[ Y_i = X_i \beta + Z_i \gamma_i + \varepsilon_i, \]

where \( Y_i \) is the vector of \( n_i \) outcomes on the \( i \)th individual \((i = 1, \ldots, m)\), \( \beta \) is a vector of fixed effects, \( \gamma_i \) is a vector of random effects for subject \( i \), \( X_\cdot \) and \( Z_\cdot \) are design matrices, and \( \varepsilon_i \) is a vector of random errors; let \( n = \sum_{i=1}^{m} n_i \) denote the total number of observations.

Assume that \( \gamma_i \) and \( \varepsilon_i \) are independent with \( \gamma_i \sim N(0,\Sigma) \) and \( \varepsilon_i \sim N(0,\sigma^2 I) \), so that \( \text{var}(Y_i) = V_i = Z_i \Sigma Z_i^T + \sigma^2 I \); also let \( W_i = V_i^{-1} \). If the \( m \) individuals are assumed to be independent, the latter matrix has a block diagonal structure.

Instead of completely observed data, \( Y_y \), we observe the pair \( (Q_y, C_y) \) where \( Q_y \) is the (possibly censored) response and \( C_y \) is the censoring indicator. The data are then separated as complete data: \( (Y_i,\gamma_i,\varepsilon_i) \) \( i = 1, \ldots, m \), and observed data: \( (C_i,Q_i) \) \( i = 1,\ldots, m \), where

\[ C_y = -1 \text{ if } Y_y < Q_y \]
\[ C_y = 0 \text{ if } Y_y = Q_y \]
\[ C_y = 1 \text{ if } Y_y > Q_y \]

Let \( \theta = (\beta,\Sigma,\sigma^2) \) be the vector of parameters. As initial values for the EM algorithm we used the OLS estimates for \( \beta \), the \( q \times q \) identity matrix for \( \Sigma \), and 1 for \( \sigma^2 \). Since the data are not completely observed, it is not possible to obtain maximum likelihood (ML) or restricted maximum likelihood (REML) estimates of \( \theta \) with the usual mixed effects framework, as proposed by Laird and Ware\(^17\), and the EM algorithm is a natural way of achieving this. Following Hughes\(^9\), the M-step is:
The Monte Carlo E-step calculates the expected values used in these expressions with respect to the conditional density $f(Y_i | C_i, Q_i, \theta)$ using the Gibbs sampler as follows. Given an initial value of $Y_i$, new values of $Y_i$ can be generated by iteratively sampling from the univariate conditional distributions

$$f(Y_i | Y_{ik}, \theta) = \begin{cases} \Pr[Y_i < Q_y] & \text{if } C_y = -1 \\ \Pr[Y_i > Q_y] & \text{if } C_y = 1 \end{cases}$$

for all $j = 1, \ldots, n_i$, $k \neq j$, and $C_y \neq 0$. This generates values from the conditional distribution $f(Y_i | C_i, Q_i, \theta)$. The probabilities $\Pr[Y_i < Q_y]$ can be easily calculated using properties of the multivariate Normal distribution, since conditionally on $\theta$, $Y_i \sim \text{MVN}(X_i \beta, V_i)$, exactly for MLE’s, which condition on $\theta = (\beta, \Sigma, \sigma^2)$, and approximately for REML’s, which require conditioning on $\theta = (\Sigma, \sigma^2)$ and impose certain prior distribution on $\beta$, as proposed by Dempster et al$^{16}$. We obtained MLE’s.

S-PLUS 6.0 (Insightful, USA) was used in a Unix environment, using a modified version of the Fortran programs provided by Hughes.

On establishing this benchmark model, we then compared models arising from alternative estimation procedures to Hughes’ method, which allow for the repeated measures and left-censoring to varying degrees.

After Hughes’ method, the second most complex estimation procedure used considered LME’s to account for repeated measures, with censored values taken as half the cut-off value of the threshold. Next, we fitted LME models using the assays’ absolute threshold values as response. Then we used OLS to fit models involving only fixed-effects, in which repeated measures are ignored, again taking censored values as midway between zero and cut-off. Finally, the simplest procedure used OLS with censored values taken as absolute. S-PLUS 2000 (Insightful, USA) was used in a Windows environment.

For each procedure, the model selection process followed that of the one for Hughes’, with the significance of all factors of interest being assessed by way of likelihood ratio tests comparing relevant models. In this way, inclusion of terms in the final model for each procedure was determined.
The final model resulting from each procedure was compared to the yardstick model from Hughes' procedure. Approaches were evaluated by the correspondence of covariate and interaction terms in their models to those in the benchmark one. Where exact correspondence occurred, assessment was based on graphical comparison of fitted prediction curves to those created from Hughes' method. Bootstrap resampling was used to construct 95% confidence bands.

4) Application

4.1) Establishing the basic model for the overall pattern of viral load over age

The pattern obtained from the super smoother algorithm (Figure 2) shows a sharp rise in HIV RNA viral load peaking at around 3 months of age with a gradual decline thereafter.

The fitted curve corresponding to the quartic model for the relationship between RNA viral load and age is shown in Figure 2 (AIC = 2536.94). When compared to the running smoother, this does not visually reflect the data. The main drawback of polynomial models is their inability to represent the important rapid peak structure of the data at young ages.

For the change-point model approach, the age at change-point was estimated to occur at \( t_1 = 2.6 \) months, with AIC = 2528.42 (Figure 2).

The most representative one power fractional polynomial model for log RNA viral load and age had a linear age term (deviance = 2584.75). The two power fractional polynomial which best captured the structure had inverse square root of age and log age terms (deviance = 2562.42). The best three power fractional polynomial had inverse square root of age, log age by inverse square root of age and log age by log age by inverse square root of age terms (deviance = 2558.68). The identified two power fractional polynomial was a significantly better fit than the best one power model \( (\chi^2 = \text{Gain} = 22.33, 2 \text{ df}, p < 0.001) \). However, the best fit three power model was not statistically significantly better fitting than the optimal two power one \( (\chi^2 = 3.74, 2 \text{ df}, p = 0.156) \). The optimal fractional polynomial model for log RNA viral load and age was therefore the one with terms for inverse square root of age and log age which had AIC = 2517.24. The predicted curve for this model (Figure 2) shows the peak and gradual decline thereafter, closely following the running smoother.

Comparing the resultant models of the three approaches: the quartic polynomial failed to detect the early peak and the peak of the change-point model did not reflect the magnitude of that of the running smoother. The peak of the fractional polynomial was closest to that of the running smoother. Additionally, formal comparison showed the fractional polynomial, with the lowest AIC, is the most suitable.
4.2) Comparison of the estimation procedures

All the models fitted in the determination of important explanatory terms using Hughes’ method are displayed in Table 1. Adding extra terms to the random effects design matrix did not significantly improve the goodness of fit, therefore the model which accounts only for random intercepts for each child was adequate. The various comparisons of these models to establish significance of individual terms are outlined in Table 2. The two-way interactions of age-by-gender and treatment-by-gender were both significant thus resulting in the fitting of the three-way age-by-gender-by-treatment interaction, which was also found to be significant. The benchmark model was therefore:

$$\log_{10}(\text{Viral Load}_i) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}_i}} + \beta_2 \ln(\text{Age}_i) + \beta_3 \text{Gender}_i + \beta_4 \text{Treatment}_i + \beta_5 \text{CD4}^+_i + \beta_6 \text{Assay}_i$$

$$+ \frac{\beta_7 \text{Treatment}_i}{\sqrt{\text{Age}_i}} + \beta_8 \ln(\text{Age}_i) \times \text{Treatment}_i + \frac{\beta_9 \text{Gender}_i}{\sqrt{\text{Age}_i}} + \beta_{10} \ln(\text{Age}_i) \times \text{Gender}_i$$

$$+ \beta_{11} \text{Gender}_i \times \text{Treatment}_i + \beta_{12} \ln(\text{Age}_i) \times \text{Gender}_i \times \text{Treatment}_i + \gamma_i + \epsilon_i$$

Where $\beta_0, \beta_1, \ldots, \beta_{12}$ denote the fixed effects coefficients for the explanatory terms, $\gamma_i$ is the corresponding vector of random intercepts and $\epsilon_i$ is the error term for child $i$ at visit $j$.

The models constructed using LME’s with midway values approach showed that the two way age-by-gender and gender-by-treatment interactions ($p = 0.015$, $p = 0.046$) were significant whereas the age-by-treatment interaction was not ($p = 0.278$). As in the model arising from Hughes’ approach, the three-way age-by-gender-by-treatment interaction was also significant ($p = 0.038$). Thus, the resulting model included all the terms corresponding to the benchmark model.

To determine how closely this model really represented the gender- and treatment-specific profiles over age, fitted curves were compared graphically (Figure 3). Curves of the model arising from the LME approach with mid-point censoring lie close to those created from the application of Hughes’ method. For each stratum, the 95% confidence bands around the fitted curves of the model from the LME approach contain the fitted curves from Hughes’ method. This showed that the setting of the HIV RNA viral load value to the midpoint within a mixed-effects framework allows the predicted curve to follow the appropriate trajectory. Notably, there was most agreement between the fitted curves at ages and within strata where values of viral load are less likely to be censored.

Models obtained using the mixed-effects procedure with absolute censoring values yielded the two-way interaction of gender by treatment as the only significant one ($p = 0.031$). Since this did not lead to the benchmark model, it inadequately accounted for the HIV RNA viral load censoring, resulting in a loss of power to detect significant interactions.
OLS using both midway and absolute values gave age-by-treatment as the only significant two-way interaction \((p = 0.013, \ p = 0.023, \text{ respectively})\). Again, these approaches do not sufficiently account for censoring or for repeated measures.

5) Discussion

To thoroughly model the pattern of this longitudinal HIV RNA viral load data it is essential to account for the within-subject correlation arising from the repeated measures, and for the left-censoring induced by assay detection thresholds. Simple means for dealing with censoring of setting the value at midpoint between zero and cut-off, are sufficient to identify the model which best describes the age-related dynamics of RNA viral load. Its use permitted the identification of all the relevant explanatory terms involving gender, treatment, assay and CD4\(^+\) cell count and the corresponding trajectories thus is not necessary to utilise the complex methodology of Hughes'.

In using fractional polynomials we have identified a model reflecting the actual form of the RNA viral load pattern over age in children vertically infected with HIV-1. Considered alternatives to the fractional polynomial approach for tackling the non-linear pattern include the change-point method which has problems with constant decay after age \(t_i\), as well as posing difficulties in identifying strata-specific change-points for the various strata as defined by the four different age/treatment combinations. Natural cubic splines provide another possible strategy; this method was not considered as it would rapidly become cumbersome with the introduction of explanatory terms and their interactions.

An alternative to the EM methodology of Hughes' has been demonstrated by Jacqmin-Gadda et al.\(^9\). They estimated model parameters by direct maximisation of the likelihood requiring numerical computation of the integral of a multivariate Normal density for each subject with censored observations. However, the required algorithm works best for integral size less than 10 and thus did not fulfil the requirements here.

Hughes shows that there is a reduction in bias in the estimates of fixed effect and variance components from the analysis of RNA viral load data with a higher degree (38 per cent) of censoring. The smaller censoring proportion involved here could explain why the choice of method for handling censoring between Hughes’ methods and midpoint was not important. Although there was generally close agreement, a degree of departure occurred where values of viral load were more likely to be censored. There may be some threshold censoring proportion below which the use of advanced techniques fails to surmount cruder means. Many of the censored values have been censored below relatively low values – less than 5 per cent (7/147) of the values have been censored above 500 copies. Had more of the censored values been less precise, the sophisticated methodology of Hughes may have been more critical.

We showed that in modelling longitudinal paediatric virological data, it is essential to account for repeated measurements and left-censoring, but with the moderate degree of censoring, methods used need not be complex.
Acknowledgments:

We would like to thank James Hughes (Seattle) for making the software available, Rodolphe Thiébaut (Bordeaux) for helpful discussion and an anonymous referee for his/her suggestions.

ECS collaborators: Dr C Giaquinto, Dr O Rampon, Dr V Giacomet and Prof A De Rossi (Università degli Studi di Padova, Italy); Dr I Grosch-Wörner (Charité Virchow-Klinikum, Berlin, Germany); Dr J Mok (Royal Hospital for Sick Children, Edinburgh); Dr I Bates, Dr I de José, Dr F Hawkins, Dr M Garcia-Rodriguez, Dr C Ladrón de Guevara, Dr J Mª Peña, Dr J Gonzalez Garcia and Dr JR Arribas Lopez (Hospital Infantil La Paz, Madrid); Prof F Asensi-Botet, Dr MC Otero, Dr D Pérez-Tamarit, Dr A. Orti, Dr M J Sna Miguel and Dr R de la Torre (Hospital La Fe, Valencia, Spain); Dr H Scherpber, M Kreyenbroek and Dr K Boer (Academisch Medisch Centrum, Amsterdam, The Netherlands); Dr AB Bohlin, Dr E Belfrage, Dr L Navér, Dr A Ehrnst and Prof A Sönnerborg (Huddinge and Karolinska University Hospitals, and Karolinska Institutet, Sweden); Prof J Levy, Dr M Hainaut, Dr A Peltier, Dr S Wibaut and Dr P Barlow, (Hospital St Pierre, Brussels, Belgium); Dr A Ferrazin and Prof D Bassetti, (Department of Infectious Diseases, University of Genoa, Italy); Dr A De Maria (Department of Internal Medicine, University of Genoa, Italy) Dr C Gotta (Department of Obstetrics and Gynecology—Neonatology Unit, University of Genoa, Italy); Dr A Mûr, Dr A Payà, Dr M Viñolas, Dr MA López-Vilchez, Dr M Rovira, Dr R Carreras, Dr E Esteban Tores, Dr S Herrero Perez (Hospital del Mar, Universidad Autonoma, Barcelona, Spain); and Dr N H Valerius (Hvidovre Hospital, Denmark).
Table 1: Models fitted using Hughes’ method to determine important covariate and interaction terms for the final model. All the models in the table include an intercept term and a random intercept, and ‘Age’ always refers to the two terms specified in the fractional polynomial.

<table>
<thead>
<tr>
<th>Model</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1 Age</td>
<td>-1333.71</td>
</tr>
<tr>
<td>M2 Age + Gender</td>
<td>-1333.70</td>
</tr>
<tr>
<td>M3 Age + Treatment</td>
<td>-1268.62</td>
</tr>
<tr>
<td>M4 Age + Gender + Treatment</td>
<td>-1268.60</td>
</tr>
<tr>
<td>M5 Age + Gender + Age×Gender</td>
<td>-1323.87</td>
</tr>
<tr>
<td>M6 Age + Treatment + Age×Treatment</td>
<td>-1268.32</td>
</tr>
<tr>
<td>M7 Age + Gender + Treatment + Gender×Treatment</td>
<td>-1261.85</td>
</tr>
<tr>
<td>M8 Age + Gender + Treatment + CD4+ Assay + Age×Gender + Age×Treatment + Gender×Treatment</td>
<td>-1211.50</td>
</tr>
<tr>
<td>M9 Age + Gender + Treatment + CD4+ Assay + Age×Gender + Age×Treatment + Gender×Treatment + Age×Gender×Treatment</td>
<td>-1208.34</td>
</tr>
</tbody>
</table>

Table 2: Likelihood ratio tests comparing models in Table 1 for the ascertainment of important interaction terms.

<table>
<thead>
<tr>
<th>Models Comparisons</th>
<th>Interaction Terms</th>
<th>χ²</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2 &amp; M5</td>
<td>Age by Gender</td>
<td>19.70</td>
<td>2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M3 &amp; M6</td>
<td>Age by Treatment</td>
<td>0.59</td>
<td>2</td>
<td>0.741</td>
</tr>
<tr>
<td>M4 &amp; M7</td>
<td>Gender by Treatment</td>
<td>13.50</td>
<td>1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M8 &amp; M9</td>
<td>Age by Gender by Treatment</td>
<td>6.31</td>
<td>2</td>
<td>0.043</td>
</tr>
</tbody>
</table>
Figure 1: Two-step procedure for modelling longitudinal HIV-RNA viral load measurements as a function of age, accounting for gender, treatment, CD4\(^+\) cell count and assay type, accommodating repeated measures and censoring scheme.

Options for the approach to establishing the linear predictor for HIV-RNA viral load as a function of age

- Conventional polynomial?
- Change-point?
- Fractional polynomial?

Alternatives for the procedure for modelling the effects of gender, treatment, CD4\(^+\) cell count and assay type with varying account of repeated measures and censoring

- Hughes' method*
- Linear mixed-effects with midpoint censoring cut-off value?
- Linear mixed-effects with absolute censoring cut-off value?
- Ordinary least squares with midpoint censoring cut-off value?
- Ordinary least squares with absolute censoring cut-off value?

* the established optimal method against which the others are compared
Figure 2: Underlying structure of HIV-RNA viral load by age as shown by the super-smoother (SS) along with conventional polynomial (quartic), change-point model (PW) and fractional polynomial (FP) representations.
Figure 3: Gender and treatment strata-specific fitted curves of the benchmark model from application of Hughes' modified EM algorithm (solid line) and final model (dashed line) with confidence bands (dotted lines) arising from the application of the linear mixed-effects approach with mid-point censored values.
References


Are There Gender and Race Differences in Cellular Immunity Patterns Over Age in Infected and Uninfected Children Born to HIV-Infected Women?

European Collaborative Study

Institute of Child Health, University College London, UK

Summary: This study investigated whether age-related patterns of immunologic markers in 1488 uninfected (9789 measurements) and 186 infected (3414 measurements) children differed by gender and race. CD4⁺, CD8⁺, and absolute lymphocytes by HIV infection status, gender, and race were assessed using linear mixed-effects natural cubic spline models, allowing for prematurity and maternal CD4⁺ cell count. In uninfected children, levels of all 3 markers peaked twice in the first few months of life, declining to adult levels by around 8 years of age; uninfected boys and uninfected black children had significantly reduced CD4⁺ and absolute lymphocyte counts; the gender difference was especially pronounced in black children. Infected children had substantially lower levels and distinctly different patterns; with, e.g., by age 6 months CD4⁺ cell counts nearly 1200 per mm³ lower than in uninfected infants. Levels also significantly differed by gender and race for infected children, although for gender in the opposite direction. The gender and race differences in CD4⁺ levels were not explained by a general lymphocytosis nor were they confounded by treatment. These substantial differences in immunologic markers may reflect underlying genetic influence on the cellular immune system and may have implications for clinical decisions about therapeutic management. Key Words: pediatric, gender, race, vertically acquired infection, Europe

Prepared by: Linsay Gray, Marie-Louise Newell, Mario Cortina-Boija, Claire Thorne, and Catherine Peckham

Address correspondence and reprint requests to M.L. Newell, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK. E-mail: m.newell@ich.ucl.ac.uk

ECS collaborators: Dr. C. Giaquinto, Dr. O. Rampon, Dr. V. Giacomini, and Prof. A. De Rossi (Università degli Studi di Padova, Italy); Dr. I. Grosch-Wörner (Charité Virchow-Klinikum, Berlin, Germany); Dr. J. Mok (Royal Hospital for Sick Children, Edinburgh); Dr. I. Bates, Dr. I. de José, Dr. F. Hawkins, Dr. M.C. García-Rodríguez, Dr. C. Ladroño de Guevara, Dr. J. Mª Pela, Dr. J. Gonzalez Garcia, and Dr. J.R. Arribas Lopez (Hospital Infantil La Paz, Madrid); Prof. F. Asensi-Botet, Dr. M.C. Otero, Dr. D. Pérez-Tamarit, Dr. A. Ortí, Dr. M.J. San Miguel and Dr.R.de la Torre (Hospital La Fe, Valencia, Spain); Dr. H. Scherpriber, M. Kreyenbroek, and Dr. K. Boer (Academisch Medisch Centrum, Amsterdam, The Netherlands); Dr. A.B. Bohlin, Dr. E. Belfrage, Dr. L. Navér (Huddinge and Karolinska University Hospitals, Sweden); Prof. J. Levy, Dr. M. Hainaut, Dr. A. Pelletier, Dr. S. Wibust, and Dr. P. Barlow (Hospital St. Pierre, Brussela, Belgium); Dr. A. Perazza and Prof. D. Bassetti (Department of Infectious Diseases, University of Genoa, Italy); Dr. A. De Maria (Department of Internal Medicine, University of Genoa, Italy); Dr. C. Cots (Department of Obstetrics and Gynecology–Neonatology Unit, University of Genoa, Italy); Dr. A. Mür, Dr. A. Payá, Dr. M. Válolas, Dr. M.A. López-Vilchez, Dr. M. Rovira, Dr. R. Carreras, Dr. E. Esteban Tores, Dr. S. Herrero Perez (Hospital del Mar, Universidad Autonoma, Barcelona, Spain); and Dr. N.H. Valerius (Hvidovre Hospital, Denmark).

The European Collaborative Study is a concerted action of the European Commission (Biomed II PL 97 2005 and QLRT-1999-30002). The Medical Research Council (UK) provides support to the coordinating center. Collaborating centers were supported at various times by grants from the Ministero della Sanità–Istituto Superiore di Sanità, Progetto AIDS (Padua, Genoa); the Medical Research Council (UK), the AIDS Virus Education Research Trust, the Scottish Office Home and Health Department (Edinburgh); Preventiefonds No. 28-1704 (Amsterdam); Bundesminister für Gesundheit (Berlin); Fonds Houtman, Office de la Naissance et de L'Enfance, Communauté Francaise de Belgique (Brussels); and the Research Foundations of Karolinska Institutet (Stockholm).

Manuscript received November 18, 2002; accepted May 13, 2003.
The developing immune system in children may be influenced by chronic maternal infections such as HIV or by factors such as gender or race, but methodological differences between studies hinder interpretation of the limited information about the immunologic pattern over age in uninfected children born to HIV-infected mothers. Without reliable population-based age-related standards, it is difficult to interpret immunologic markers in vertically infected children and to explore possibly associated explanatory variables. The CD4\(^+\) cell-based immunologic classification of disease progression for infected children is only weakly associated with clinical disease progression, especially in the first years of life. This may be due to inadequate accounting for age-related variation in the dynamics of infection. There is thus a need for reliable age-related patterns of CD4\(^+\) and CD8\(^-\) cell counts and absolute lymphocyte measurements, for uninfected children of infected mothers that could subsequently be used in the management of infected children.

We previously reported distinct differences between girls and boys in HIV RNA levels and patterns over age in the European Collaborative Study (ECS). Data collected on both infected and uninfected children born to HIV-infected women provided an opportunity to assess patterns of CD4\(^+\) and CD8\(^+\) cell counts and absolute lymphocyte measurements in the first 12 years of life and to further investigate possible associations with gender, race, prematurity, and maternal HIV progression.

**METHODS**

The ECS, a prospective study, has been ongoing since 1986. Children born to HIV-infected women are followed regularly from birth, using a standard protocol, with detailed clinical and laboratory information, in 11 pediatric centers from 8 European countries. Parental consent is obtained, and the study is approved by local ethics committees. HIV infection status was identified as previously described. Laboratory tests including CD4\(^+\), CD8\(^-\) cell count, and absolute lymphocyte measurements, were carried out locally according to standard procedures that do not vary between centers. Venous blood specimens were anticoagulated using ethylenediamine tetra-acetic acid or heparin and processed within 24 hours following collection. An automated hemocytometer was used to obtain absolute lymphocyte counts and the subtypes of white blood cell. Tests were based on flow cytometry (FACSCAN) with Becton-Dickinson antibodies. For the analyses CD4\(^+\), CD8\(^-\), and absolute lymphocyte measurements were log 10 transformed.

**Statistical Methods**

The structure of the data was visualized and modeling was informed by supersmoothers. Natural cubic spline models that reflected the structure of the mean function were fitted. As these models are nonparametric, the magnitudes of their coefficients cannot be interpreted in a way that is meaningful in the conventional quantitative sense. The repeated measurement nature of the data was allowed for in linear mixed-effects models with a general form for the random effects covariance. Effects of gestational age, gender, maternal race, and maternal CD4\(^+\) cell count at delivery were assessed using likelihood ratio tests. Running percentiles over age from the data from uninfected children were used in the comparison with data for infected children. P values comparing individuals' predictions for combined data on uninfected and infected children at particular ages were derived from 2 sample t tests of interpolated predicted values. Analyses were carried out using S-PLUS 2000 (Insightful, Seattle, WA) in a Windows environment.

**RESULTS**

A total of 13,203 records with at least one of CD4\(^+\) cell, CD8\(^-\) cell, or absolute lymphocyte count measurement were available for all 1674 enrolled children between birth and 12 years of age (Table 1): 9789 lymphocyte counts related to 1488 uninfected and 3414 to 186 infected children. Overall, maternal CD4\(^+\) cell counts were <200 cells in 68 of 477 mothers with available data. Only 8.3% of mothers had been diagnosed with AIDS. Two-thirds of infected children received combination antiretroviral treatment to delay disease progression, with a median age of 3.2 years at initiation.

**Patterns Over Age in Uninfected Children**

The running smoothers plots show the underlying patterns for the observed data on CD4\(^+\), CD8\(^-\) cell, and absolute lymphocyte counts of uninfected children (Fig. 1A–C). CD4\(^+\) cell counts peaked at 3 weeks of age, then dipped before peaking again at 6 months and declining gradually thereafter. The pattern for CD8\(^-\) cell counts was similar although with a more protracted second peak, while for the absolute lymphocytes both peaks were smoothed out.

The modeled patterns over age are shown in Figure 2A. CD4\(^+\) cell counts approach adult values some time after age 6 years. For instance, the predicted CD4\(^+\) cell counts peak at 3238 per mm\(^3\) at 3 weeks, then again at 3009 at 6 months, dropping to 2597 at 1 year of age and to 1096 by age 5. For CD8\(^-\) cell counts, the predicted values peak at 1343 per mm\(^3\) at 2 weeks and then at 11 months at 1219; the subsequent estimated decrease in values is slow, reaching 739 at 5 years. The predicted absolute lymphocyte values peak at 6470 per mm\(^3\) at 3 weeks, again at 6702 at 7 months; there was a subsequent gradual decline to 6149 at 1 year, falling to 3016 by 5 years. CD4\(^+\) cells as a percentage of absolute lymphocyte counts fell rapidly from around 55% at birth, leveling off...
Estimated levels of CD4+ cell counts in the first year of life were similar for premature and full-term uninfected infants (P = 0.615). On the other hand, estimated CD4 levels, but not patterns, were associated with gender and race. Levels for girls were systematically higher than for boys (P = 0.003), with differences increasing after about 4 years of age (Fig. 3). CD4 cell counts for white children were consistently higher than those for black children until around age 9 (P < 0.001), after which data were sparse and the pattern became difficult to interpret (Fig. 3). Gender effects were not significantly different for white and black children (P = 0.638).

Contrary to what was seen for CD4+ cell counts, estimated levels of CD8+ cell counts did not vary by gender (P = 0.448), although levels of absolute lymphocytes did (P = 0.0042). The direction of the association with race and gender was similar to that seen for CD4+ cell counts, with levels for white children higher than those for black children for both CD8+ and absolute lymphocyte cell counts (P < 0.001). Predicted CD4% was higher for girls (P = 0.0004) and white children (P = 0.003), indicating that the gender and race differences seen in CD4 cell counts were not due to a general increase in absolute lymphocytes. Allowing for interactions between gender and race for CD8, absolute lymphocyte counts and CD4% did not statistically significantly improve the models (P = 0.951, P = 0.999, P = 0.396, respectively).

FIGURE 1D–F shows the raw data and the running smooths for CD4+ cell, CD8+ cell, and absolute lymphocyte counts for infected children. Observed CD4+ cell counts fell quickly below CD8+ cell counts before recovering slightly at around 8 years of age.

CD4+ predicted values peak at 7 weeks at 2334 per mm3, falling to 1391, approximately half the estimated value for uninfected children, at 1 year, and reaching a low point of 278, about 1/3 of that for uninfected children at around 6 years. Similar to predictions for uninfected children, the peak of 1723 per mm3 for the predicted level of CD8+ occurs at 4 months, falling to 1639 at 1 year, reaching a nadir of 719 at just after 8 years. The
peak for absolute lymphocytes occurs at 2 months at 6,738 per mm$^3$, declining to 5,098 at 1 year and falling to the lowest value of 1,718 at 8.6 years. The spline representing CD4 as a percentage of total lymphocytes had a sharp fall from around 55% at birth to around 25% by 2 years, falling more gradually thereafter before leveling off at 20% after 4 years of age.

CD4$^+$ cell count patterns of infected children do not vary by gestational prematurity in the first year of life. In contrast to the uninfected children, CD4$^+$ cell counts of infected girls were initially lower than those for infected boys, becoming equal at around 6 years of age ($P < 0.0001$). Similar to the models for uninfected children, CD4$^+$ cell levels for white children are generally higher than those for black children ($P < 0.0001$), and both these and gender differences persisted after adjustment by antiretroviral treatment. The patterns by gender did not vary according to race ($P = 0.204$).
For CD8\(^+\) cell counts and absolute lymphocytes, levels for boys were higher than those for girls and, for both, levels for white children were higher until around 3 years of age, falling below those of black children thereafter. Levels of CD4\(^+\) differed for boys and girls, but not in any consistent way, with the pattern for girls fluctuating above and below that for boys. Levels differed by race from 1 year, with white children having consistently higher values than black children.

Comparison of Infected and Uninfected Children

Although the CD4\(^+\) cell count measurements were consistently lower for infected children, there was considerable overlap with measurements for uninfected children. For instance, of the 2781 measurements on infected children, 1511 (54.3\%) were above the running 5th percentile for uninfected children. Overlap was especially pronounced at young ages (73.9\% at <1 year) but decreased thereafter (56.8\% at 1-5 years and 38.8\% at >5 years). On the other hand, 73.5\% of measurements for uninfected children were below the 95th percentiles for infected children.

An alternative way of quantifying differences in levels for uninfected and infected children is through the fitting of a spline-based model for data on all children with terms for interactions with infection status. This model included knots at 1 and 3 weeks; 3, 6, and 10 months; and 3, 6, and 10 years and indicated substantial and significant differences in estimated CD4 cell counts between infected and uninfected children from birth onwards (Table 2). The largest absolute difference of 1486 per mm\(^2\) was at 10 months, relating to the measurements of 993 uninfected and 145 infected children. This difference subsequently decreased to 509 cells at age 10 years, although the number of children tested at that age was small, especially among uninfected children. Infected children who died consistently had significantly lower CD4 cell counts than infected children who survived (\(P < 0.0001\)), and those tested at later ages were thus a selected population of survivors. Nearly 1/3 (14/47) of children who died did so during the first year of life. Similar analyses revealed significantly lower values for infected children throughout the 12 years of follow-up for CD4\(^+\) (\(P < 0.001\)) and from 3 months onward for CD8\(^+\) cell and absolute lymphocyte counts (\(P < 0.001\) and \(P < 0.001\)).

DISCUSSION

In this European cohort measurements on immunologic parameters were available for uninfected and infected children born to HIV-infected mothers for up to 12 years. In uninfected children, levels of CD4\(^+\) cell counts peaked early and did not approach adult values until after age 6 years; estimated patterns for CD8\(^+\) and absolute lymphocytes were similar to those of CD4\(^+\) cell counts. Patterns for infected children were substantially different, with CD4\(^+\) values peaking at 7 weeks to a lower value than in uninfected children. In infected, but not in uninfected, children CD8\(^+\) cell counts were higher than CD4\(^+\) cell counts. There was considerable overlap in measurements for uninfected and infected children, although estimated CD4\(^+\) cell counts differed by as much as almost 1500 per mm\(^2\) at 10 months of age.

The large number of available measurements over age were modeled allowing for their repeated nature and the effect of variables that might be related to the level or pattern. The double peak in predicted CD4\(^+\) cell counts for uninfected children at 3 weeks and 3 months, with similar double peaks in CD8\(^+\) and absolute lymphocytes, highlights the complex and dynamic nature of a developing immune system. Although findings from previous studies have identified an early peak, the double nature of this peak has not previously been reported.\(^1,5\)

Based on almost 10,000 measurements on nearly 1500 uninfected children, we found that although immuno-

<table>
<thead>
<tr>
<th>Knot for age</th>
<th>Log CD4(^+) cell counts for uninfected (absolute) n</th>
<th>Log CD4(^+) cell counts for infected (absolute) n</th>
<th>Absolute cell count difference</th>
<th>(P) value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, age = 0</td>
<td>3.253 (1791) 90</td>
<td>3.191 (1552) 15</td>
<td>239</td>
<td>0.0152</td>
</tr>
<tr>
<td>1 week</td>
<td>3.453 (2836) 251</td>
<td>3.345 (2215) 50</td>
<td>621</td>
<td>0.0001</td>
</tr>
<tr>
<td>3 weeks</td>
<td>3.501 (3168) 351</td>
<td>3.406 (2545) 48</td>
<td>623</td>
<td>0.0003</td>
</tr>
<tr>
<td>3 months</td>
<td>3.447 (2798) 801</td>
<td>3.331 (2142) 105</td>
<td>656</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>6 months</td>
<td>3.479 (3010) 1022</td>
<td>3.265 (1840) 138</td>
<td>1170</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>10 months</td>
<td>3.444 (2783) 993</td>
<td>3.172 (1486) 145</td>
<td>1296</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>3 years</td>
<td>3.160 (1444) 460</td>
<td>2.788 (614) 112</td>
<td>830</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>6 years</td>
<td>3.007 (1016) 117</td>
<td>2.401 (252) 79</td>
<td>764</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>10 years</td>
<td>2.906 (805) 18</td>
<td>2.472 (296) 35</td>
<td>509</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* From \(t\) tests.
logic patterns did not differ by gender or race, levels did: CD4+ cell counts and absolute lymphocytes were higher for girls than for boys and for white children compared with black children. White children also had higher CD8+ cell counts than black children. It is possible, but unlikely, that maternal immune status confounds these observed race differences. Maternal CD4+ and child CD4+ are related (data not shown), but, in this relatively healthy cohort, there are not many women with low CD4+ levels, and although median counts differed for black and white women, this was not substantial. Further, one would expect race-specific effects to result in a natural correlation between mothers and children. Gender and race differences were not explained by a general increase in absolute lymphocytes. Gender differences in CD4+, CD8+, and absolute lymphocytes levels in infected children were marked, but in an opposite direction than seen in uninfected children. Contrary to our findings, in a study of 126 uninfected children born to HIV-infected mothers in the United States, with <200 measurements in total, CD4+ cell counts did not differ significantly by ethnicity.1 This study differed from ours not only in the population characteristics but also in its cross-sectional nature, which makes interpretation difficult. In a different context, some gender and race differences in immunologic markers were reported from a cross-sectional study of immune function in about 150 healthy African-American and Latino inner-city children, unexposed to HIV infection, between the ages of 8–12 years; the within-child pattern over age was not assessed.

The differences by gender and race observed in uninfected children born to HIV-infected mothers may indicate an underlying genetic origin for the levels of immunologic markers. Indirect evidence relating to possible lower CD4+ levels in African than European populations has previously been limited to comparisons between cohorts of pregnant women included in HIV vertical transmission studies.13–15 However, in those comparisons it was unclear whether the observed differences were related to the longer duration of infection in the African cohorts or to a heightened immune response to other infections more prevalent in the African context. In our cohort of uninfected children, the observed differences are between black and white children, all born in Europe and followed from birth, and are unlikely to be explained by environmental infectious pressure on the immune system, or by HIV infection.

As a mother's HIV infection may impact on the developing immune system of her infected infant, one cannot generalize from uninfected children born to infected mothers to children in the general population.16–17 We found some effect of maternal immune suppression on levels of CD4 cell counts in uninfected children. Based on only a subset of 438 women, levels of CD4+ cell counts were lower in uninfected children born to women with low CD4+ count at delivery, especially in the first 2 years of life, although differences at ages after about 6 years were difficult to assess as data in these subgroups were sparse (data not shown). Further, children in our cohort differ from those in the general population in that most were not breastfed; this too could have consequences for the development of cellular immunity and levels of CD4+ and absolute lymphocyte counts. However, the levels of absolute lymphocytes in the uninfected children in our cohort were similar to those of children between the ages of 8–12 years in an HIV-unexposed American study,2 which suggests that even if there is an effect of maternal infection initially this is no longer apparent at later ages.

We found substantial differences not only in the levels but also in the patterns of CD4+, CD8+, and absolute lymphocytes by HIV infection status of the child. The gender difference observed was in an opposite direction to that seen for uninfected children, with higher levels in CD4+ counts, CD8+ cell counts, and absolute lymphocytes in boys. As reported by others,17–20 CD8+ cell counts were significantly higher for infected children than for uninfected children from 3 months of age onwards. Only 14 of the 186 infected children had been exposed to antiretroviral prophylaxis to prevent vertical transmission in the first weeks of life, and this is unlikely to have substantially influenced the observed initial pattern and level. However, 2/3 of infected children received antiretroviral treatment to delay disease progression at some time, but differences by gender and race remained after adjustment. Levels of those who have died were consistently lower compared with those alive (data not shown). The pattern and levels seen in the first several years of life are thus likely to reflect the natural course of disease, but the slight upturn in CD4 counts later on could possibly be associated with survivor bias and treatment effect, which needs further investigation.

The overlap between measurements for uninfected and infected children, with more than half the measurements for infected children was substantial, exceeding the 5th percentile for uninfected children. Nearly 10% of CD4+ measurements in uninfected children fell into the Centers for Disease Control (CDC) moderately or severely immunosuppressed categories (data not shown) and 45% of those in infected children. The inadequacy of the CDC categorization has been highlighted previously.3,4 The overlap of measurements seen in uninfected and infected children limits the use of a single CD4+ measurement to
assess the immune and health status of the infected children. Thus our findings have implications for guidelines for treatment initiation and change and suggest further investigation into tailorung management according to gender and race.

Our findings of a possible genetic basis for the level of CD4+, CD8+, and absolute lymphocytes in uninfected children born to HIV-infected mothers may contribute to the broader knowledge necessary for ongoing HIV vaccine development and further understanding of the development of patterns of immunologic markers in the first decade of life.

Statement: Predicted CD4+ cell counts over age and further details on methodology used can be made available on request from the corresponding author.

Acknowledgments: The authors thank Mrs. L. Tottle and Dr. Simona Fiore (London); Prof. L. Chieco-Bianchi, Prof. F. Zacchello, Dr. E. Rugs, Dr. R. D'Elia, Dr. A. M. Laverda, Dr. S. Cozzani, Dr. C. Cattelan, Dr. A. Mazza, Prof. B. Grella, Dr. AR. Del Mistro, and Mrs. S. Oletto (Padua); Dr. Cornelia Feitema and Dr. Ralf Weigel (Berlin); Dr. S. Burns, Dr. N. Hallam, Dr. P.L. Yap, and Dr. J. Whitelaw (Edinburgh); Dr. B. Sancho and Dr. G. Fontan-Casanego (Madrid); Dr. A. Gonzalez Molina, Dr. M. Gobernado, Dr. JL. Lopez, and Dr. J. Cordoba (Valencia); A. van der Plas (Amsterdam); Dr. B. Christensson, Dr. G. Eriksson, and Prof. U. Ewald (Sweden); Dr. G. Di Siena, Dr. E. Postali, Prof. M.F. Pancaro, G. Mantero, and Dr. P. Dignetti (Genoa); and Dr. A. Hottard, Dr. M. Poncin, Dr. S. Sprecher, Dr. B. Lejeune, Dr. G. Zississ, and Prof. N. Clumeck (Brussels); and Dr. M. Guxens, Dr. P. Martinez (Barcelona).

REFERENCES
Gender and race do not alter early-life determinants of clinical disease progression in HIV-1 vertically infected children

European Collaborative Study

Prepared by: Linsay Gray, MSc, Marie-Louise Newell, PhD, Mario Cortina-Borja, PhD, Claire Thorne, PhD, Institute of Child Health, University College London, UK.

Author for correspondence: Prof ML Newell, Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, 30 Guilford Street, London WCIN 1EH, UK. Tel: +44 20 7829 8699 Fax: +44 20 7813 8145 Email: m.newell@ich.ucl.ac.uk

Abstract: 247 words

Text: 3,183 words

ECS collaborators: Dr C Giaquinto, Dr O Rampon, Dr F Ebo, Prof R D'Elia and Prof A De Rossi (Università degli Studi di Padova, Italy); Dr I Grosch-Wörner (Charité Virchow-Klinikum, Berlin, Germany); Dr J Mok (Royal Hospital for Sick Children, Edinburgh); Dr J Bates, Dr I de José, Dr F Hawkins, Dr MC Garcia-Rodriguez, Dr C Ladrón de Guevara, Dr J Mª Peña, Dr J Gonzalez Garcia and Dr JR Arribas Lopez (Hospital Infantil La Paz, Madrid); Prof F Asensi-Botet, Dr MC Otero, Dr D Pérez-Tamarit, Dr A. Orti, Dr M.J San Miguel and Dr R de la Torre (Hospital La Fe, Valencia, Spain); Dr HJ Scherpbier, ME Kreyenbroek and Dr K Boer, Ms A Hes (Academisch Medisch Centrum, Amsterdam, The Netherlands); Dr AB Bohlin, Dr E Belfrage, Dr L Navér, Dr S Lindgren (Huddinge and Karolinska University Hospitals, Sweden); Prof J Levy, Dr P Barlow, Dr M Hainaut, Dr A Peltier, Dr S Wibaut, Dr G Debruyne (Hospital St Pierre, Brussels, Belgium); Dr A Ferrazin and Prof D Bassetti, (Department of Infectious Diseases, University of Genoa, Italy); Dr A De Maria (Department of Internal Medicine, University of Genoa, Italy) Dr C Gotta (Department of Obstetrics and Gynecology--Neonatology Unit, University of Genoa, Italy); Dr A Már, Dr A Payà, Dr M Viñolas, Dr MA López-Vilchez, Dr M Rovira, Dr R Carreras, Dr E Esteban Tores, Dr S Herrero Perez (Hospital del Mar, Universidad Autonoma, Barcelona, Spain); Dr N H Valerius (Hvidovre Hospital, Denmark); Dr T Niemieç and Dr M Marczynska (Centrum Diagnostyki I Terapii AIDS, Warsaw, Poland).

289
Abstract

Objective: To identify early life predictors of clinical progression before and beyond the first year of life.

Design: Prospective follow-up of 161 vertically HIV infected children in the ongoing European Collaborative Study provided data from birth over 16 years.

Methods: Kaplan Meier and Cox regression procedures were used to assess the predictive value of first available laboratory and clinical markers for progression defined as serious disease or death. We investigate gender and race effects on associations and the optimal threshold for longitudinal CD4+ percentage measurements after age 6 months for predicting disease progression.

Results: Earliest (during the first six months) measurements of CD4+ percentage below 20% (three-fold increased risk (p=0.041)) and absolute lymphocytes (AL) (reduction of risk of three-quarters for a one log increase (p=0.014)) were independently associated with overall and rapid disease progression during the first year. Persistent lymphadenopathy (or hepatomegaly) in early life was also additionally associated with overall disease progression, and after age one year (greater than doubling of risk, (p=0.040)), but not with rapid progression. Associations were not significantly dependent on gender or race. CD4+ percentage of 10% was the best prognostic cut-off.

Conclusions: Early clinical markers are strongly predictive of disease progression after one year of age into adolescence. However, rapid progression is less straightforward to predict, probably due largely to early progression during the first few months in such individuals. The independently predictive value of AL measurements suggest they could be used alone in the management of children in resource-poor settings.

Keywords: paediatric, determinants, disease progression, gender, race, vertically-acquired infection, Europe
Introduction

Without effective antiretroviral therapy (ART), a quarter of vertically infected children born in developed countries progress to serious disease or death in the first year of life, rising to half by 5 years of age (1). Immunological and virological factors (2,3) are associated with progression, as are clinical symptoms such as hepatomegaly, splenomegaly, and lymphadenopathy in early life (2,4). Prematurity may also be associated with rapid disease progression in the first year of life (5). Whether factors associated with early disease progression differ from those associated with disease progression after age one year is unknown.

Although age-related virological and immunological levels and patterns in children differed by gender and race in the ECS (6,7), reported rates of clinical progression in vertically infected children are similar for boys and girls (8). Whether the risk of progression for specific values of laboratory markers varies by gender or race remains unclear.

Prospective data from the European Collaborative Study (ECS) on children born to HIV-infected women, provide a unique opportunity to identify early life predictors of clinical progression in the first year and thereafter over 16 years, and to address whether differences in laboratory markers by gender and race result in differential risks of progression to serious disease.

Methods

The ECS is an ongoing prospective study with almost 17 years follow-up of a representative cohort of children born to HIV-infected women. Detailed clinical and laboratory information from 11 paediatric centres in eight European countries is collected according to a standard protocol. Infected children are seen at birth, 3 and 6 weeks, 3, 4.5, 6, 9, 12, 18 and 24 months and at least twice yearly thereafter (9). Parental consent is obtained; the study is approved by local ethics committees. HIV infection status was identified by the onset of AIDS, detection of virus or antigen in at least two separate blood samples and/or persistence of antibody beyond 18 months of age.

Severe disease progression is defined by CDC category C, or death (11); rapid disease progression is defined as occurring within the first year of life. Children were classified as severely premature if born at or before 34 weeks of gestation, moderately premature between 34 and 37 weeks and full-term beyond 37 weeks. Antiretroviral treatment (ART) was categorised into monotherapy or combination (two or more) ART. Categories were updated at each visit, allowing children to be re-assigned if the number of drugs prescribed increased. Black children in this cohort are born to mothers from sub-Saharan Africa.

Laboratory tests were carried out locally. Venous blood specimens were anticoagulated using EDTA or heparin and processed within 24 hours. An automated haemocytometer was used to obtain absolute lymphocyte (AL) counts and subtypes of white blood cells. Tests were based on flow cytometry (FACSCAN) with Becton-Dickinson antibodies. CD4+, CD8+ cell count and AL measurements were log base 10 transformed. CD4+ cell counts were also expressed as a percentage of AL count. From exploratory analysis of the
distribution of CD4+ percentage, values above or below 20% gave the best categorisation in the first six months, with lower values being appropriate for later in life.

With the exceptions of an investigation of gender and race specific effects and assessment of a threshold for CD4+ percentage over all ages preceding disease progression, this work chiefly explores the early-life determinants of serious clinical disease progression or death. The predictive values of first available, pre-disease progression, log CD4+ cell count, AL count and CD4+ percentage, presence of hepatomegaly, splenomegaly and lymphadenopathy (and specifically axillary node enlargement) at two or more visits during the first six months of life were estimated. The main effects of gender, prematurity, race, neonatal prophylactic ART on risk of progression, and modification by gender and race on the effects of the early life biological indicators were assessed. Further, as the difference in CD4+ count levels by gender increases from age one year onward, measurements at age two (above or below 500 cells (11)) were assessed for effects on disease progression specific to boys and to girls; similarly, as levels of CD4+ percent reportedly differ consistently for black and white children after early life, measurements above or below 15% (11) were assessed for race-specific effects. In this observational study, the effect of ART given before disease progression could not be assessed, but use was accounted for throughout.

With widespread use of neonatal zidovudine and initiation of ART early in life (for prevention of vertical transmission and disease progression respectively) in recent years, disease progression and HIV-related death in the ECS has been almost entirely restricted to those born before 1997 (1). We therefore consider only children born before 1997 but include follow-up information until September 2002. HIV-RNA viral load assays have become routinely available only since 1997, and consequently, there are few available measurements early in life for those who have progressed.

**Statistical methods**

The overall survivor function was obtained by Kaplan-Meier analysis. Cox proportional hazards modelling (12) was performed to estimate survival allowing for covariates. ART treatment was included as a time-dependent variable. Candidate terms for the multivariable model included ART treatment and factors reaching significance at the 15% level in univariate models. Comparisons between models were performed using likelihood ratio tests to identify the optimum model. The relative predictive capabilities of the various terms in the final model were determined by changes in the deviance, excluding each in turn, where a large difference indicates substantial contribution to the model. All survival analyses were performed using STATA (STATA Version 7.0; College Station, TX).

In a separate analysis, classification trees (13) were used to detect the most predictive threshold for longitudinal CD4+ percentage measurements at any age after the first six months of life prior to disease progression. This was carried out using S-PLUS 2000 (Insightful, Seattle, USA) in a Windows environment.
Results

Laboratory measurements and clinical information from 5209 visits were available for all 161 infected children in the ECS born before 1997 (Table 1). By one year of age, an estimated 17.7% (95% confidence interval (CI) 12.6% to 24.6%) of children will have progressed to serious disease or death; increasing to 36.3% (95% CI: 29.2% to 44.4%) by age five, and 44.1% (95% CI: 36.3% to 52.7%) by 10 years. Overall, similar numbers of girls (33/76, (43%)) and boys (32/85 (38%)) progressed. There were some differences by race with 52/113 (46%) white, 10/32 (31%) black and only 3/16 (23%) of the other (mainly Asian) children progressing (p = 0.065), which reflects later enrolment of children of non-white ethnicity when ART use was increasingly widespread (9).

Overall disease progression and early immunological and clinical measurements

First laboratory measurements were taken at a mean age of 1.5 months (range: birth – 5.8 months). The median first CD4+ percentage for rapid progressors was 35.0% (interquartile range (IQR) 24.6 to 43.5), substantially lower than the 43.1% (36.5%-54.5%) for children progressing to serious disease or death beyond one year of age. Univariably, first available log AL count and first CD4+ percentage (Figure 1) predicted subsequent progression at any point during follow-up, but first available log CD4+ and CD8+ cell count did not (Table 2). The risk of progression to serious disease or death was associated with hepatomegaly and lymphadenopathy observed at two or more occasions, but not with axillary nodes enlargement or with splenomegaly. Progression did not significantly differ by gender, race, prematurity or neonatal ART prophylaxis (Table 2).

As maternal CD4+ cell counts were only recorded routinely in participating centres since 1995, there was a limited subset (24/161 (15%)) of mothers with available measurements at delivery and multivariable analysis was thus not possible. In univariable analysis a log increase in maternal CD4+ count was compatible with a 75% reduction in risk of progression for the child (p = 0.171).

Gender- and race-specific effects of clinical and immunological markers

In analyses of models including terms representing interactions of gender and race with each of the clinical and immunological measurements, effects on disease progression of early immunological markers and clinical indicators were not modified by gender or race (p > 0.3, all tests for gender, p > 0.5, all tests for race). In the subset of 22 children who had not progressed to serious disease or death by age 2 years and had available CD4+ measurements at that age, a CD4+ count below 500 cells per mm³ at age 2 years was associated with a greater risk of clinical progression for boys than for girls although the difference was not statistically significant. (hazard ratio (HR) = 3.65 and 2.62, respectively, p = 0.521). A CD4+ percent of less than 15% was associated with greater progression risk in black children than white, but again, not statistically significantly so (HR = 8.88 and 1.31, respectively, p = 0.213).

Simultaneous predictive value of early clinical and immunological variables

In multivariable analysis allowing for ART (Table 2), children with a first CD4+ percentage below 20% had nearly a three-fold increased risk of progression to serious disease or death compared with those with a value above 20% (p = 0.041). A log (ten-fold) increase in first AL count independently reduced risk of progression by 77% (p =
0.014) and lymphadenopathy at two or more visits within the first six months of life increased risk by 88% (p = 0.045). A CD4+ percentage above or below 20% was substantially more predictive of disease progression than AL count or lymphadenopathy (deviance differences = 81.4, 5.2 and 3.2 respectively). For example, keeping other variables constant (no early persistent lymphadenopathy and an AL count of 6000 cells (3.78 in log units), say) and accounting for treatment, compared to a child with an early CD4+ percentage above 20% (with a risk of progression by age 5 years of 25%), one with a CD4+ percentage value below 20% were 2.79 times more likely to progress (with a risk of progression by age 5 years of 70% (2.79 x 25%). Hepatomegaly (HR = 1.885, p = 0.035; Model $\chi^2 = 17.32$; p = 0.0039) and lymphadenopathy had similar and interchangeable, but not independent predictive effects. Allowing for TMX-SMP PCP prophylaxis and for the use of intravenous immunoglobulin before disease progression did not significantly alter estimates.

Threshold for CD4+ percentage
Using all clinical status and CD4+ percentage data beyond six months of age, classification tree analysis was used to investigate the optimal value for splitting the CD4+ percentage measurements into two groups in terms of predicting subsequent disease progression. At any age, the threshold value of 10% for CD4+ percentage (HR = 4.83 (95% CI, 2.63 to 8.86, p<0.0001) was found to best predict progression to serious disease or death.

Rapid progression
Of the 65 children who progressed to serious disease or died, 28 (43%) did so within the first year of life. The majority (19/28,68%) had opportunistic infections, mostly Pneumocystis carinii pneumonia (PCP) (12/19,63%); four had encephalopathy; two serious bacterial infections and the remaining three died at home probably of an undiagnosed opportunistic infection. The predictive value of a log increase in first CD4+ cell count was significant and reduced rapid progression risk by 89% (HR = 0.11, 95% CI, 0.02 to 0.71, p = 0.021); a first CD4+ percentage below 20% was associated with a fourfold increase in risk (HR = 4.16, 95% CI, 1.66 to 10.39, p = 0.002). The reduction in risk of rapid progression with a log increase in first AL measurement was 81% (HR = 0.19, 0.04 to 0.87, p = 0.003). Prematurity was not associated with rapid disease progression (p = 0.813). In a multivariable model also including ART, first CD4+ percentage below 20% and log AL count were independently predictive of rapid disease progression or death (HR = 6.62, 95% CI, 2.35 to 18.66, p < 0.001 and HR = 0.11, 95% CI, 0.39 to 0.98, p = 0.012, respectively). Neither gender nor race modified effects of any early laboratory or clinical markers. Adjusting for TMP-SMX and IVIG prophylaxis did not alter risk estimates.

Progression to serious disease after one year of age
The remaining 37 (57%) children progressed to serious disease after age one year: 10 (27%) with encephalopathy, nine (24%) with serious recurrent bacterial infections, 13 (35%) with opportunistic infections (only three of which were PCP); four with another C-defining illness and one died of unspecified HIV-related causes. In univariable analysis, early persistent hepatomegaly (HR = 2.22, 95% CI, 1.11 to 4.46, p = 0.024) and early persistent lymphadenopathy (HR = 2.16, 95% CI, 1.01 to 4.61, p = 0.048) were
associated with progression beyond the first year. Severe prematurity was marginally associated with late disease progression (HR = 2.24, 95% CI, 0.91 to 5.54, p = 0.080). In multivariable analysis, adjusting for ART, persistent hepatomegaly (HR = 2.15, 95% CI, 1.03 to 4.47, p = 0.040) was the single significant independent indicator of disease progression after age one year. Separately, persistent lymphadeopathy was associated with a doubled, but not statistically significant risk (HR = 2.01, 95% CI, 0.90 to 4.46, p = 0.087). Adjustment for TMP-SMX and IVIG did not impact on estimates.

Discussion

In this cohort of children born and followed up in Europe, early life measures of CD4+ percentage below 20% and a ten-fold increase in AL counts were independently associated with clinical progression throughout childhood, and with rapid disease progression or death within the first year of life. Early persistence of lymphadenopathy and hepatomegaly additionally predicted subsequent overall progression to serious disease and progression beyond one year of life, but not rapid progression. CD4+ percentage was more informative than any other laboratory or clinical indicator. However, for a given value of clinical or immunological marker early in life or around age two years, disease progression did not differ by gender or race. Less than half of this cohort received combination therapy before disease progression, but differentiating between double and other combination therapies in the treatment adjustment had no bearing on the results (data not shown). With adjustment for ART, results inform knowledge about underlying mechanisms of vertically-acquired disease progression and thus remain relevant in the HAART era.

These findings enhance previous work addressing the prognostic value of early markers of disease progression and death in the early years in life. However, earlier studies used measurements relatively close to progression (10,14), or cross-sectional data from clinical trials (15-17). Studies with longitudinal data from birth have been limited to shorter follow-up (3,18,19), have focused on limited laboratory determinations (2),(20-23) or have involved too few children to draw reliable conclusions (22). Whether predictive values of given markers vary by gender or race has not been previously explored. We have been able to consider an extensive period of follow-up, providing a largely natural account of both short- and long-term paediatric disease progression and predictors.

Although clinical evidence of infection in the first six months of life, such as lymphadenopathy and hepatomegaly, are not associated with subsequent rapid progression before age one year, they are predictive of long-term prognosis. Hepatomegaly was the only factor associated with disease progression beyond age one year when assessed separately. These findings appear to contradict those of Rich et al (2) who found early presence of lymphadenopathy, hepatomegaly or splenomegaly with CD4+ percentage predicted rapid progression. However, their approach differed methodologically: firstly, rapid progression was defined as occurring during the first six months of life only; secondly, clinical progression preceding laboratory determinations in some children was nonetheless included and thirdly, their statistical methods involved only logistic models with presence or absence of progression as a binary response. That early immunological factors were predictive of progression before one year of age but not
beyond in our cohort is consistent with earlier observations of levels reflecting current status more than long-term survival and wellbeing (3).

We recently described significant differences in immunological patterns over 12 years by gender and race (6). However, these differences were small in the first year, which agrees with our findings here that effects of early life markers on overall disease progression are not dependent on gender or race. Despite magnitude of differences in immunological patterns becoming larger as children get older (7), we did not find gender differences in progression for a given CD4+ count or race differences in progression for a given CD4+ percentage at age two years. This may have been due to small numbers.

A CD4+ percentage cut-off of 10% best identified the risk of disease progression beyond six months of age, which is lower than the 15% threshold in the CDC categories below which children are classified as severely immunosuppressed (11). The CDC clinical and immunological classification systems have been shown to be in poor agreement (1,25). The high degree of overlap in lymphocyte measurements in uninfected and infected children (6) could explain the necessity for more extreme limits in predicting serious progression.

Because of limited numbers of infected children and of mothers with relevant data, we were unable to investigate effects of other factors possibly associated with disease progression in vertically infected children such as maternal ART use (5,20,26), timing of transmission (23), p24 antigen (17), maternal viral load (27,2), and vitamin A (2). Our findings of higher maternal CD4+ lymphocytes associated with lower risk of disease progression, although not significant, are compatible with findings of others (2,28).

We were also unable to assess the child’s HIV RNA viral load as a marker for disease progression directly here, but the association between clinical indicators such as lymphadenopathy and hepatomegaly and viral activity indicates that early viral load measurements, when available, would be useful prognostic indicators, as has been shown by others (3,17,19). It has been suggested that virus load is the optimal predictor of paediatric HIV progression (3), however, the relative stability of immunological markers, especially CD4+ percentage, compared to the highly variable HIV RNA viral load levels (7), may make them as clinically relevant in predicting serious disease progression. We were unable to confirm an association of either overall or rapid disease progression with prematurity (5) which could be due to small numbers of infants born very prematurely.

The findings here and elsewhere (29,30,31) imply that in the absence of specific CD4+ cell assays, AL measurements alone would provide sufficient insight to inform management of individual children after the first few months of life, which could be particularly relevant in less developed countries where laboratory resources may be limited.

Generally in this cohort, early progression was due to opportunistic infections whereas progression later in life was more dominated by encephalopathy and bacterially-related illnesses. Our findings indicate that CD4+ percentages and AL counts could inform prevention management of PCP and similar morbidities, whereas occurrence of early
persistent hepatomegaly could alert the need for measures to prevent bacterial infections and encephalopathy. This is in line with findings from an American birth cohort study which suggested encephalopathy was more likely following development of early symptoms of HIV (32).

This extension of knowledge about the associations of early-life clinical and immunological factors with rapid and long-term progression to serious disease or death informs the understanding of the dynamics of vertically-acquired HIV infection.

Acknowledgement
We gratefully acknowledge the support from Mrs L Toxtle and Dr Simona Fiore (London). We thank Prof L Chieco-Bianchi, Prof F Zacchello, Dr A Mazza, Dr E Ruga, Dr A Laverda, Dr AM Del Mistro, and Mrs S Oletto (Padua); Dr C Feiterna, and Dr Ralf Weigel (Berlin); Dr S Burns, Dr N Hallam, Dr PL Yap, and Dr J Whitelaw (Edinburgh); Dra B Sancho, and Dr G Fontan-Casanego (Madrid); Dr A Gonzalez Molina, Dr M Gobenaro, Dr JL Lopez, and Dr J Cordoba (Valencia); A van der Plas (Amsterdam); Dr B Christensson, Dr P Bolme, Dr U Ewald (Sweden); Dr G Di Siena, Prof M F Pantarotto, G Mantero, and Dr P Dignetti, Dr M Camera, Dr R Rosso, Dr B Ciravegna (Genoa); and Dr A Hottard, Dr M Poncin, Dr S Sprecher, Dr B Lejeune, Dr G Zississ, and Prof N Clumeck (Brussels), Dr M Guxens, Dr P Martinez (Barcelona); Dr S Pisa, Dr B Martinez de Tejada, Dr L Zamora, (Barcelona); Dr M Casellas Caro (Barcelona); Dr Y Canet (Barcelona); Dr G Zucotti (Milan); Dr M Carla Re, Dr V Venturi (Bologna); Dr C Christini, Dr F Castelli, Dr A Rodella, Dr E Prati, Prof. M. Duse (Brescia); Dr G Scaravelli, Dr M Stegagno (Rome); Dr I Quinti, Prof A Pachi (Rome); Dr G Noia (Rome); Dr M De Santis (Rome); Prof PA Tovo, Dr C Gabiano, Dr N Ziarati (Turin); Dr AE Semprini (Milan), Dr F Ravagni Probizer (Pavia); Dr G Ferraris, Dr A Buceri (Milan); Dr L Rancilio (Milan); Dr J Jimenez (Madrid); Dr A Horban (Warsaw); Dr E Pagliaro, Dr M T Melisi (Naples), The Regional Health Office and RePuNaRC (Naples).

The European Collaborative Study is a concerted action of the European Commission (Biomed II PL 97 2005 and QLRT-1999-30002). The Medical Research Council (UK) provides support to the coordinating centre. Collaborating centres were supported at various times by grants from the Ministero della Sanita - Istituto Superiore di Sanita, Progetto AIDS (Padua, Genoa); the Medical Research Council (UK), the AIDS Virus Education Research Trust, the Scottish Office Home and Health Department (Edinburgh); Praeventiefonds No. 28-1704 (Amsterdam); Bundesminister fur Gesundheit (Berlin); Fonds Houtman, Office de la Naissance et de L’Enfance, Communaute Francaise de Belgique (Brussels); and the Research Foundations of Karolinska Institutet (Stockholm).
References


(10) HIV Paediatric Prognostic Markers Collaborative Study Group. Short-term risk of disease progression in HIV-1 infected children receiving no antiretroviral therapy or zidovudine monotherapy: estimates according to CD4 percent, viral load, and age. Submitted 2003.

(11) Centers For Disease Control. 1994 Revised classification system for human immunodeficiency virus infection in children less than 13 years of age. MMWR 1994; 43(RR12):1-10.


(14) CASCADE Collaboration. Short-term risk of AIDS according to the current CD4 count and viral load in antiretroviral naive individuals and those treated in the monotherapy era. Submitted 2003.


Table 1: Laboratory and clinical data on the first six months of life for 161 infected children born before 1997

<table>
<thead>
<tr>
<th>Factor</th>
<th>Progressors n = 65 (40%)</th>
<th>Non- Progressors n = 96 (60%)</th>
<th>Total n = (161)</th>
</tr>
</thead>
<tbody>
<tr>
<td>first available log CD4+ cell count median (IQR)*</td>
<td>3.32 (3.18 to 3.51)</td>
<td>3.41 (3.28 to 3.5)</td>
<td>3.39 (3.25 to 3.51)</td>
</tr>
<tr>
<td>first available log CD8+ cell count median (IQR)</td>
<td>3.21 (3.08 to 3.29)</td>
<td>3.15 (2.97 to 3.32)</td>
<td>3.19 (2.99 to 3.30)</td>
</tr>
<tr>
<td>first available absolute lymphocyte count median (IQR)</td>
<td>3.74 (3.66 to 3.88)</td>
<td>3.81 (3.70 to 3.90)</td>
<td>3.78 (3.68 to 3.89)</td>
</tr>
<tr>
<td>first available CD4 percentage median (IQR)</td>
<td>41.7 (29.0 to 49.9)</td>
<td>41.4 (32.8 to 50)</td>
<td>41.4 (30.2 to 50.0)</td>
</tr>
<tr>
<td>Hepatomegaly (observed at 2 or more visits within first six months)</td>
<td>49 (75.4%)</td>
<td>83 (86.5%)</td>
<td>132 (82.0%)</td>
</tr>
<tr>
<td>Yes</td>
<td>16 (24.6%)</td>
<td>13 (13.5%)</td>
<td>29 (18.0%)</td>
</tr>
<tr>
<td>Splenomegaly (observed at 2 or more visits within first six months)</td>
<td>58 (89.2%)</td>
<td>85 (88.5%)</td>
<td>143 (88.8%)</td>
</tr>
<tr>
<td>No</td>
<td>7 (10.8%)</td>
<td>11 (11.5%)</td>
<td>18 (11.2%)</td>
</tr>
<tr>
<td>Yes</td>
<td>53 (81.5%)</td>
<td>85 (88.5%)</td>
<td>138 (85.7%)</td>
</tr>
<tr>
<td>Lymphadenopathy (observed at 2 or more visits within first six months)</td>
<td>12 (18.5%)</td>
<td>11 (11.5%)</td>
<td>23 (14.3%)</td>
</tr>
<tr>
<td>No</td>
<td>53 (81.5%)</td>
<td>79 (82.3%)</td>
<td>132 (82.0%)</td>
</tr>
<tr>
<td>Yes</td>
<td>12 (18.5%)</td>
<td>17 (17.7%)</td>
<td>29 (18.0%)</td>
</tr>
<tr>
<td>ART Treatment ever before progression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>46 (70.8%)</td>
<td>22 (22.9%)</td>
<td>68 (42.2%)</td>
</tr>
<tr>
<td>Monotherapy</td>
<td>16 (24.6%)</td>
<td>7 (7.3%)</td>
<td>23 (14.3%)</td>
</tr>
<tr>
<td>Combination</td>
<td>3 (4.6%)</td>
<td>67 (69.8%)</td>
<td>70 (43.5%)</td>
</tr>
<tr>
<td>Median age (years) at last visit (IQR)</td>
<td>3.7 (1.4 to 7.1)</td>
<td>9.5 (6.4 to 11.7)</td>
<td>7.1 (3.1 to 11.0)</td>
</tr>
<tr>
<td>Median number of visits (IQR)</td>
<td>13 (7 to 80)</td>
<td>20 (13 to 69)</td>
<td>18 (11 to 80)</td>
</tr>
</tbody>
</table>

* IQR, inter-quartile range
<table>
<thead>
<tr>
<th>Factor</th>
<th>Univariable</th>
<th>Multivariable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hazard ratio (95% CI)**</td>
<td>p-value</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td>Reference</td>
<td>1.18 (0.73, 1.91)</td>
</tr>
<tr>
<td>Boys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prematurity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-term</td>
<td>Reference</td>
<td>1.32 (0.71, 2.43)</td>
</tr>
<tr>
<td>Moderately premature</td>
<td></td>
<td>1.44 (0.67, 3.09)</td>
</tr>
<tr>
<td>Severely premature</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>Reference</td>
<td>0.67 (0.34, 1.31)</td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AZT administered neonatally***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.19 (0.03, 1.37)</td>
<td>0.100</td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>first available log CD4+ cell count (per log increase)</td>
<td>0.44 (0.11, 1.79)</td>
<td>0.253</td>
</tr>
<tr>
<td>first available log CD8+ cell count (per log increase)</td>
<td>0.96 (0.29, 3.18)</td>
<td>0.953</td>
</tr>
<tr>
<td>first available log absolute lymphocyte count (per log increase)</td>
<td>0.31 (0.10, 0.99)</td>
<td>0.049</td>
</tr>
<tr>
<td>first available CD4 percentage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20%</td>
<td>2.15 (0.85, 5.48)</td>
<td>0.107</td>
</tr>
<tr>
<td>&gt;= 20%</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Hepatomegaly (observed at 2 or more visits) ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>1.83 (1.07, 3.12)</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenomegaly (observed at 2 or more visits)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.05 (0.48, 2.31)</td>
<td>0.896</td>
</tr>
</tbody>
</table>
Lymphadenopathy (observed at 2 or more visits)

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>p</th>
<th>Reference</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.76 (0.99, 3.13)</td>
<td>0.052</td>
<td>1.883 (1.015, 3.493)</td>
<td>0.045</td>
</tr>
</tbody>
</table>

Enlarged axillary nodes (observed at 2 or more visits)

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.27 (0.67, 2.39)</td>
<td>0.460</td>
</tr>
</tbody>
</table>

Treatment

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>p</th>
<th>Reference</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.75 (0.93, 3.32)</td>
<td>0.084</td>
<td>1.495 (0.732, 3.054)</td>
<td>0.270</td>
</tr>
<tr>
<td>Combination</td>
<td>0.61 (0.16, 2.29)</td>
<td>0.466</td>
<td>0.504 (0.121, 2.091)</td>
<td>0.345</td>
</tr>
</tbody>
</table>

Maternal CD4+ count at delivery (per log increase)

<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.274 (0.043, 1.746)</td>
<td>0.171</td>
</tr>
</tbody>
</table>

* Final model: $\chi^2 = 17.51$, $p = 0.0036$
** CI, confidence interval
*** candidate for final model but not subsequently significant

**Figure 1**: Kaplan-Meier survival plot of progression to CDC clinical category C or death by CD4+ percentage category in the first six months of life.
Appendix 2.4  Clinical and immunological CDC categories for children with HIV infection

A. Clinical categories

<table>
<thead>
<tr>
<th>CATEGORY N: NOT SYMPTOMATIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children who have no signs or symptoms considered to be the result of HIV infection or who have only one of the conditions listed in Category A.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CATEGORY A: MILDLY SYMPTOMATIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children with two or more of the conditions listed below but none of the conditions listed in Categories B and C.</td>
</tr>
<tr>
<td>- Lymphadenopathy (0.5 cm at more than two sites; bilateral = one site)</td>
</tr>
<tr>
<td>- Hepatomegaly</td>
</tr>
<tr>
<td>- Splenomegaly</td>
</tr>
<tr>
<td>- Dermatitis</td>
</tr>
<tr>
<td>- Parotitis</td>
</tr>
<tr>
<td>- Recurrent or persistent upper respiratory infection, sinusitis, or otitis media</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CATEGORY B: MODERATELY SYMPTOMATIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children who have symptomatic conditions other than those listed for Category A or C that are attributed to HIV infection. Examples of conditions in clinical Category B include but are not limited to:</td>
</tr>
<tr>
<td>- Anemia (&lt;8 gm/dL), neutropenia (&lt;1,000/mm³), or thrombocytopenia (&lt;100,000/mm³) persisting 30 days</td>
</tr>
<tr>
<td>- Bacterial meningitis, pneumonia, or sepsis (single episode)</td>
</tr>
<tr>
<td>- Candidiasis, oropharyngeal (thrush), persisting (&gt;2 months) in children &gt;6 months of age</td>
</tr>
<tr>
<td>- Cardiomyopathy</td>
</tr>
<tr>
<td>- Cytomegalovirus infection, with onset before 1 month of age</td>
</tr>
<tr>
<td>- Diarrhea, recurrent or chronic</td>
</tr>
<tr>
<td>- Hepatitis</td>
</tr>
<tr>
<td>- Herpes simplex virus (HSV) stomatitis, recurrent (more than two episodes within 1 year)</td>
</tr>
<tr>
<td>- HSV bronchitis, pneumonitis, or esophagitis with onset before 1 month of age</td>
</tr>
<tr>
<td>- Herpes zoster (shingles) involving at least two distinct episodes or more than one dermatome</td>
</tr>
<tr>
<td>- Leiomyosarcoma</td>
</tr>
<tr>
<td>- Lymphoid interstitial pneumonia (LIP) or pulmonary lymphoid hyperplasia complex</td>
</tr>
<tr>
<td>- Nephropathy</td>
</tr>
<tr>
<td>- Nocardiosis</td>
</tr>
<tr>
<td>- Persistent fever (lasting &gt;1 month)</td>
</tr>
<tr>
<td>- Toxoplasmosis, onset before 1 month of age</td>
</tr>
<tr>
<td>- Varicella, disseminated (complicated chickenpox)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CATEGORY C: SEVERELY SYMPTOMATIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children who have any condition listed in the 1987 surveillance case definition for acquired immunodeficiency syndrome, with the exception of LIP.</td>
</tr>
</tbody>
</table>
B. Immunological categories based on age-specific CD4^+ T-lymphocyte counts and percent of total lymphocytes

<table>
<thead>
<tr>
<th>Immunologic category</th>
<th>Age of child</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;12 months</td>
</tr>
<tr>
<td></td>
<td>(x 10^6 cells/l) (%)</td>
</tr>
<tr>
<td>1: No evidence of suppression</td>
<td>≥1,500 (≥25)</td>
</tr>
<tr>
<td>2: Evidence of moderate suppression</td>
<td>750–1,499 (15–24)</td>
</tr>
<tr>
<td>3: Severe suppression</td>
<td>&lt;750 (&lt;15)</td>
</tr>
</tbody>
</table>
Appendix 2.5 Fractional polynomial models

A fractional polynomial of degree $m$ for a single covariate $X$, subject to the restriction $X > 0$, is defined as:

$$\Phi_m(X; \xi, p) = \xi_0 + \sum_{j=1}^{m} \xi_j X^{(p_j)}$$

(1)

where $p = (p_1, \ldots, p_m)$ is a real-valued vector of powers with $p_1 < \ldots < p_m$ and $\xi = (\xi_0, \xi_1, \ldots, \xi_m)$ are real-valued coefficients. The power transformation $X^{(p_j)}$ is defined as the Box-Tidwell transformation, $\log X$ if $p_j = 0$ and $X^{p_j}$ otherwise. In general it is required that $X > 0$.

For arbitrary powers $p_1 \leq \ldots \leq p_m$, set $H_0(X) = 1, p_0 = 0$, then

$$\Phi_m(X; \xi, p) = \sum_{j=1}^{m} \xi_j H_j(X),$$

(2)

where for $j = 1, \ldots, m$

$$H_j(X) = \begin{cases} X^{(p_j)} & \text{if } p_j \neq p_{j-1}, \\ H_{j-1}(X) \ln X & \text{if } p_j = p_{j-1}. \end{cases}$$

Usually the power terms are restricted to a predetermined set of integer and non-integer values; the set $\{-2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2, 3\}$ was used (131). Given a set of data points for a dependent random variable and a number of predictor variables, the deviance is a measure of the total variability left after fitting the predictor variables and is equal to $-2 \log(L)$ where $L$ is the likelihood. Among fractional polynomials with degree $m$ the model with lowest deviance is
deemed to have the best fit. In deciding the adequacy of the optimal model with degree $m$
compared to the best fitting model with degree $m+1$, the difference in deviance is compared
to $\chi^2_{(2)}$ since both an extra power and a regression coefficient are involved. Since the parameter
space is discrete, these tests are asymptotically conservative (131); we did not attempt to modify
them.
Appendix 2.6 Change-point models

A change-point model consisting of two straight lines that intersect at age $\tau$, the peak reached by HIV viral load measurements in early ages was considered. The value of $\tau$ was estimated as the age where the running smoother reached its maximum, although other approaches, e.g. segmented regression or partial likelihood could also have been used. $U_{ij}$, for individual $i$ at the $j$-th examination, is defined as the minimum of $Age_{ij}$ and $\tau$; $T_{ij}$ is defined as the product of $(Age_{ij} - \tau)$ and an indicator function which is 0 for $Age_{ij} < \tau$ and 1 for $Age_{ij} > \tau$. The model is given by:

$$\log_{10}(\text{Viral Load}_{ij}) = \beta_0 + \beta_1 U_{ij} + \beta_2 T_{ij} + \epsilon_{ij},$$

where $\epsilon_{ij} \sim \text{iid } N(0, \sigma^2)$. 


Appendix 2.7  Hughes' modified Monte Carlo EM algorithm

This extends the applications proposed by Pettitt in 1986 (170) for the one-way random effects model.

The standard mixed effects model of Laird & Ware (136) can be written as:

\[ Y_i = X_i \alpha + Z_i \beta_i + e_i, \]

where \( Y_i \) is the vector of \( n_i \) outcomes on the \( i \)th individual \((i = 1, \ldots, m)\), \( \alpha \) is a vector of fixed effects, \( \beta_i \) is a vector of random effects for subject \( i \), \( X_i \) and \( Z_i \) are design matrices, and \( e_i \) is a vector of random errors. Use of these quantities without the subscript \( i \) will denote the respective vector or matrix for the entire sample of \( m \) individuals. Assume that \( \beta_i \) and \( e_i \) are independent with

\[ \beta_i \sim N(0, \Sigma) \quad \text{and} \quad e_i \sim N(0, \sigma^2 I), \]

so that \( \text{var}(Y_i) = Z_i \Sigma Z_i^T + \sigma^2 I. \)

Instead of completely observed data, \( Y_i \), we observe the pair \((Q_y, C_y)\) where \( Q_y \) is the (possibly censored) response and \( C_y \) is the censoring indicator.

Complete data: \((Y_i, \beta_i, e_i)\) \(i = 1, \ldots, m\)

Observed data: \((C_i, Q_i)\) \(i = 1, \ldots, m\), where

\[ Y_y < Q_y \quad \text{if} \quad C_y = -1 \]
\[ Y_y = Q_y \quad \text{if} \quad C_y = 0 \]
\[ Y_y > Q_y \quad \text{if} \quad C_y = 1. \]

Let \( \theta = (\beta, \Sigma, \sigma^2) \) be the vector of parameters. As initial values for the EM algorithm we used the OLS estimates for \( \beta \), the \( q \times q \) identity matrix for \( \Sigma \), and 1 for \( \sigma^2 \). Since the data are not
completely observed, it is not possible to obtain maximum likelihood (ML) or restricted
maximum likelihood (REML) estimates of $\theta$ with the usual mixed effects framework, as
proposed by Laird and Ware (136), and the EM algorithm is a natural way of achieving this.
Following Hughes (132), the M-step is:

$$\hat{\alpha} = (X^TWX)^{-1}X^TWE(Y|C, Q, \theta)$$
$$\hat{\Sigma}^2 = \frac{1}{n} \sum_{m=1}^{n} E(\gamma_i, \gamma_j^T | C, Q, \theta)$$
$$\sigma^2 = \frac{1}{n} \sum_{m=1}^{n} E(e_i, e_j | C, Q, \theta)$$

The Monte Carlo E-step calculates the expected values used in these expressions with respect to
the conditional density $f(Y_i|C, Q, \theta)$ using the Gibbs sampler as follows. Given an initial
value of $Y_i$, new values of $Y_i$ can be generated by iteratively sampling for the univariable
conditional distributions

$$f(Y_{ij} | Y_{ik}, \theta) = \begin{cases} 
\text{Pr}[Y_{ij} < Q_{ij}] & \text{if } C_{ij} = -1 \\
\text{Pr}[Y_{ij} > Q_{ij}] & \text{if } C_{ij} = 1 
\end{cases}$$

for all $j = 1, ..., n_i$, $k \neq j$, and $C_{ij} \neq 0$. This generates values from the conditional distribution
$f(Y_i|C, Q, \theta)$. The probabilities $\text{Pr}[Y_{ij} < Q_{ij}]$ can be easily calculated using properties of the
multivariate Normal distribution, since conditionally on $\theta$, $Y_i \sim \text{MVN}(X, \beta, V_i)$, exactly for
MLE's, which condition on $\theta = (\beta, \Sigma, \sigma^2)$, and approximately for REML's, which require
conditioning on $\theta = (\Sigma, \sigma^2)$ and impose certain prior distribution on $\beta$, as proposed by
Dempster et al. We obtained MLE's.
Appendix 2.8  \textbf{Natural cubic splines}

Let \( t_1, \ldots, t_k \) be real numbers on the interval \([a,b]\), satisfying \( a < t_1 < t_2 < \ldots < t_k < b \). A function \( g \) defined on \([a,b]\) is a cubic spline if two conditions are satisfied. First, on each of the intervals \((a, t_i), (t_i, t_{i+1}), \ldots, (t_k, b)\), \( g \) is a cubic polynomial; second, the polynomial pieces fit together at the points \( t_i \) in such a way that \( g \) itself and its first and second derivatives are continuous at each \( t_i \), and hence on the whole of \([a,b]\). The points \( t_i \) represent the knots. A natural cubic spline is a cubic spline with zero second and third derivatives in \([a,b]\).
Appendix 2.9  Programmes used for selected analyses

Appendix 2.9.1: STATA programmes for fractional polynomials for HIV RNA viral load

* * Fractional polynomials *

clear
cd c:\Ecsdatabase\ViralLoad
insheet using VL2AprilApp.txt
g logvl=log10(vl)
graph logvl age
fracpoly regress logvl age
predict yfit2
graph logvl yfit2 age, xla yla c(.l) s(o1) sort
* get degree-3 fit and compare models
fracpoly regress logvl age, degree(3) compare
predict yfit3
* get degree-1 fit
fracpoly regress logvl age, degree(1)
predict yfit1
graph logvl yfit1 yfit2 yfit3 age, xla yla c(.ll) s(oii) sort
Appendix 2.9.2: STATA programmes for conventional polynomials for HIV RNA viral load

cd c:\LinsayPhD\ViralLoad\ViralLoad
insheet using VL2AprilApp.txt
g logvload=log10(vl)
g monthsat=age

* one power Polynomial - LINEAR
regr logvload monthsat
predict yfitlin
graph logvload yfitlin monthsat, xla yla c(.1) s(oi) sort

* two power Polynomial - QUADRATIC
g agesq= monthsat^2
regr logvload monthsat agesq
predict yfitquad
graph logvload yfitquad monthsat, xla yla c(.ll) s(oiii) sort

* three power Polynomial - CUBIC
g agecu= monthsat^3
regr logvload monthsat agesq agecu
predict yfitcu
graph logvload yfitcu monthsat, xla yla c(.l) s(oi) sort

* four power Polynomial - QUARTIC
g agequart= monthsat^4
regr logvload monthsat agesq agecu agequart
drop yfitquart
predict yfitquart
graph logvload yfitquart monthsat, xla yla c(.1) s(oi) sort
Appendix 2.9.3: S-PLUS programme for change-point model for HIV RNA viral load

attach(viral9Ea)

# assume data.frame viral9Ea is attached
# spl<- supsmu(Age, LogVL)
tl<- spl$x[spl$y == max(spl$y)]

# tl is 2.59 months
detach()

viral9EaPW<- viral9Ea

viral9EaPW$W1<- pmin(viral9EaPW$Age, rep(tl, length(viral9EaPW$Age)))
viral9EaPW$W2<- (viral9EaPW$Age - tl) * ifelse( viral9EaPW$Age >= tl, 1, 0)

viral9EaPW.glm<- glm( LogVL ~ W1 + W2, data=viral9EaPW)
summary( viral9EaPW.glm, cor=F)
Appendix 2.9.4: S-PLUS programme for Hughes' method for HIV RNA viral load

```r
viral9a$Age<- ifelse(viral9a$Age ==0, 0.03, viral9a$Age)
viral9a$T4<- ifelse(viral9a$T4 ==0, 0.001, viral9a$T4)
viral9a$T4a<- ifelse(viral9a$T4==26, 2.6, viral9a$T4)
viral9a$Trt3<- ifelse(viral9a$TrtType >=2, 2, viral9a$TrtType)
viral9a$Trt2<- ifelse(viral9a$Trt3==2, 2, 1)

viral9a$Trt3<- as.factor(viral9a$Trt3)
viral9a$Trt2<- as.factor(viral9a$Trt2)
attr(viral9a$Trt3, 'levels')<- c('No', 'Mono', 'Comb')
attr(viral9a$Trt2, 'levels')<- c('No-Mono', 'Comb')

viral9a$IC<- as.factor(viral9a$IC)

viral9a$LogVL<- log(viral9a$VL,10)
viral9a$LogT4<- log(viral9a$T4,10)
viral9a$LogT4a<-log(viral9a$T4a,10)
viral9a$X0 <- rep(1, 735)
viral9a$X1<-1/sqrt(viral9a$Age)
viral9a$X2<-log(viral9a$Age)
viral9a$X3<-viral9a$LogT4a

viral9a$C0<-as.numeric(viral9a$Sex) -1
viral9a$C1<-ifelse(viral9a$TrtType=='Mono', 1, 0)
viral9a$C2<-ifelse(viral9a$TrtType=='Comb', 1, 0)
viral9a$C3<-ifelse(viral9a$IC=='2', 1, 0)
viral9a$C4<-ifelse(viral9a$IC=='3', 1, 0)
viral9a$C5<-as.numeric(viral9a$Trt2) -1
viral9a$X1C0<-viral9a$X1*viral9a$C0
viral9a$X2C0<-viral9a$X2*viral9a$C0
viral9a$X1X3<-viral9a$X1*viral9a$X3
viral9a$X1X3C0<-viral9a$X1*viral9a$X3*viral9a$C0
viral9a$X1C1<-viral9a$X1*viral9a$C1
viral9a$X1C2<-viral9a$X1*viral9a$C2
viral9a$X1C3<-viral9a$X1*viral9a$C3
viral9a$X1C4<-viral9a$X1*viral9a$C4
viral9a$X1C5<-viral9a$X1*viral9a$C5
viral9a$X2C5<-viral9a$X2*viral9a$C5
viral9a$X1C0C1<-viral9a$X1*viral9a$C0*viral9a$C1
viral9a$X1C0C2<-viral9a$X1*viral9a$C0*viral9a$C2
viral9a$X1C0C3<-viral9a$X1*viral9a$C0*viral9a$C3
viral9a$X1C0C4<-viral9a$X1*viral9a$C0*viral9a$C4
viral9a$X1C0C5<-viral9a$X1*viral9a$C0*viral9a$C5
viral9a$X1C0C1<-viral9a$X1*viral9a$C0*viral9a$C1
viral9a$X1C0C2<-viral9a$X1*viral9a$C0*viral9a$C2
viral9a$X1C0C3<-viral9a$X1*viral9a$C0*viral9a$C3
viral9a$X1C0C4<-viral9a$X1*viral9a$C0*viral9a$C4
viral9a$X1C0C5<-viral9a$X1*viral9a$C0*viral9a$C5
viral9a$X1C0C1<-viral9a$X1*viral9a$C0*viral9a$C1
viral9a$X1C0C2<-viral9a$X1*viral9a$C0*viral9a$C2
viral9a$X1C0C3<-viral9a$X1*viral9a$C0*viral9a$C3
viral9a$X1C0C4<-viral9a$X1*viral9a$C0*viral9a$C4
viral9a$X1C0C5<-viral9a$X1*viral9a$C0*viral9a$C5
```

dyn.load('censre2.o')

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + Z0')
virall0a.crla<- censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X = cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X1),
Z=matrix(virall0a$ZO,ncol=1),
D=1, sigma=1, iseed=-10101 )
# alpha=c(6.92, 0.09, -0.3),
attr(virall0a.crla, 'model')<-'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + Z0'

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Assay + Z0')
virall0a.class<- censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X = cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X1,virall0a$X2,virall0a$X10),
Z=matrix(virall0a$ZO,ncol=1),
# alpha=c(6.92, 0.09, -0.3, 0.5), D=1,
D=1, sigma=1, iseed=-10101)
attr(virall0a.class, 'model')<-'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Assay + Z0'

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + Z0')
virall0a.crlb<- censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X = cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X10,virall0a$X2,virall0a$X10),
Z=matrix(virall0a$ZO,ncol=1),
# alpha=c(6.92, 0.09, -0.3, 0.5), D=1,
D=1, sigma=1, iseed=-10101)
attr(virall0a.crlb, 'model')<-'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + Z0'

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Trt2Comb + Z0')
virall0a.crlc<- censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X = cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X10,virall0a$X2,virall0a$X10),
Z=matrix(virall0a$ZO,ncol=1),
# alpha=c(6.92, 0.09, -0.3, 2, 4), D=1,
D=1, sigma=1, iseed=-10101)
attr(virall0a.crlc, 'model')<-'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Trt2Comb + Z0'

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 T4 + Z0')
virall0a.crid<- censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X = cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X10,virall0a$X2,virall0a$X10),
Z=matrix(virall0a$ZO,ncol=1),
#alpha=c(6.92, 0.09, -0.3, 2), D=l, sigma=l, iseed=-10101)
attr(virall0a$cr1bc, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 T4a + Z0'

censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL, X=
$cbind(virall0a$X0, virall0a$X1, virall0a$X2, virall0a$X0hist, virall0a$X1hist, virall0a$X2hist),
Z=matrix(virall0a$ZO, ncol=1),
#alpha=c(6.92, 0.09, -0.3, 2, 4), D=1,
D=1, sigma=1, iseed=-10101)
attr(virall0a$cr2b, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + a4 Trt2Comb + Z0'

sink("virall0a.reduced.sex")

sink()
virall0a.cr2c<-
censre2(id=viral10a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X =
cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$C5,virall0a$X1C5,virall0a$X2C5),
Z=matrix(virall0a$Z0,ncol=1),
 alpha=c(6.92, 0.09, -0.3, 2, 4, -0.6, -1.2, -0.6, -1.2), D=1,
 D=1, sigma=1, iseed=-10101)
attr(virall0a.cr2c, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Trt2Comb + a4
Trt2Comb:1/sqrt(Age) + a5 Trt2Comb:log(Age) + Z0'

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 T4 + a4 T4:1/sqrt(Age) +
a5 T4:log(Age) + Z0')
virall0a.cr2d<-
censre2(id=viral10a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X = cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X3,virall0a$X1X3,
virall0a$X2X3),
Z=matrix(virall0a$Z0,ncol=1),
 alpha=c(6.92, 0.09, -0.3, 2, 0.1, 0.1), D=1,
 D=1, sigma=1, iseed=-10101)
attr(virall0a.cr2d, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 T4 + a4 T4:1/sqrt(Age) + a5
T4:log(Age) + Z0'

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + a4 Trt2Comb + a5
Sex:Trt2Comb + Z0')
virall0a.cr2g<-
censre2(id=viral10a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X = cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$C0,virall0a$C5,
virall0a$C0C5),
Z=matrix(virall0a$Z0,ncol=1),
 alpha=c(6.92, 0.09, -0.3, 2, 0.5, 0.25, 0.25, 2, 4), D=1,
 D=1, sigma=1, iseed=-10101)
attr(virall0a.cr2g, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + a4 Trt2Comb +
a5 Sex:Trt2Comb + Z0'

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Assay + a4
1/sqrt(Age):Assay + a5 log(Age):Assay + Z0')
virall0a.cr2ass<-
censre2(id=viral10a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X =
cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$C10,virall0a$X1C10,virall0a
a$X2C10),
Z=matrix(virall0a$Z0,ncol=1),
 alpha=c(6.92, 0.09, -0.3, 0.5), D=1,
 D=1, sigma=1, iseed=-10101)
attr(virall0a.cr2ass, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Assay + a4 1/sqrt(Age):Assay +
a5 log(Age):Assay + Z0'
sink('viral9Ea.sexassay')
print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + a4 Sex:1/sqrt(A) +a5 Sex:log(Age) + a6 Assay + Z0')
viral9Ea.sexassay<- censre2(id=viral9Ea$Serial, C=viral9Ea$CutOff, Q=viral9Ea$LogVL,
X=cbind(viral9Ea$X0,viral9Ea$X1,viral9Ea$X2,viral9Ea$C0,viral9Ea$X1C0,viral9Ea$X2C0,viral9Ea$C10),
Z=matrix(viral9Ea$Z0, ncol=1),
D=1, sigma=1, iseed=-10101)
attr(viral9Ea.sexassay, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + a4 Sex:1/sqrt(A) +a5 Sex:log(Age) + a6 Assay + Z0'
sink()

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Trt2Comb + a4 Trt2Comb:1/sqrt(Age) + a5 Trt2Comb:log(Age) + a6 Assay + Z0')
viral10a.cr2cA<- censre2(id=viral10a$Serial, C=viral10a$CutOff, Q=viral10a$LogVL,
X=cbind(viral10a$X0,viral10a$X1,viral10a$X2,viral10a$C5,viral10a$X1C5,viral10a$X2C5,viral10a$C10),
Z=matrix(viral10a$Z0, ncol=1),
#alpha=c(6.92, 0.09, -0.3, 2, 4, -0.6, -1.2, -0.6, -1.2), D=1,
D=1, sigma=1, iseed=-10101)
attr(viral10a.cr2cA, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Trt2Comb + a4 Trt2Comb:1/sqrt(Age) + a5 Trt2Comb:log(Age) + a6 Assay + Z0'

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 T4 + a4 T4:1/sqrt(Age) +a5 T4:log(Age) + a6 Assay + Z0')
viral10a.cr2dA<- censre2(id=viral10a$Serial, C=viral10a$CutOff, Q=viral10a$LogVL,
X=cbind(viral10a$X0,viral10a$X1,viral10a$X2,viral10a$X3,viral10a$X1X3,viral10a$X2X3,viral10a$C10),
Z=matrix(viral10a$Z0, ncol=1),
#alpha=c(6.92, 0.09, -0.3, 2, 0.1, 0.1, 0.1), D=1,
D=1, sigma=1, iseed=-10101)
attr(viral10a.cr2dA, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 T4 + a4 T4:1/sqrt(Age) +a5 T4:log(Age) + a6 Assay + Z0'

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + a4 Trt2Comb + a5 Sex:Trt2Comb + a6 Assay + Z0')
viral10a.cr2gA<- censre2(id=viral10a$Serial, C=viral10a$CutOff, Q=viral10a$LogVL,
X=cbind(viral10a$X0,viral10a$X1,viral10a$X2,viral10a$C0,viral10a$C5,viral10a$X1C0,viral10a$C0C5,viral10a$C10),
Z=matrix(viral10a$Z0, ncol=1),
#alpha=c(6.92, 0.09, -0.3, 2, 0.5, 0.25, 0.25, 2, 4), D=1,
D=1, sigma=1, iseed=-10101
attr(virall0a.cr2gA, 'model')<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + a4 Trt2Comb +
a5 Sex:Trt2Comb + a6 Assay + Z0'

print( 'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + a4 Trt2Comb + a5
1/sqrt(Age):Sex + a6 log(Age):Sex + a7 1/sqrt(Age):Trt2Comb + a8
log(Age):Trt2Comb + a9 Sex:Trt2Comb + a10 1/sqrt(Age):Sex:Trt2Comb + all
log(Age):Sex:Trt2Comb + Z0')

virall0a.cr3ab<-
censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X = cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X3,virall0a$X4,
virall0a$X1C0,virall0a$X1C1,virall0a$X1C2,virall0a$X1C3,
virall0a$X1C4,virall0a$X1C5,
virall0a$X2C0,virall0a$X2C1,virall0a$X2C2,virall0a$X2C3,
virall0a$X2C4,virall0a$X2C5,
\mbox{with covariates:} \alpha=c(6.22, -1,-0.25,-0.83,-0.29,-0.77, -1.6, 0.1, 0.525, 0.59, -0.026,
1.2, 0.51, -4.87, -0.55,0.01, 0.01, 0.01 ),
D=1, sigma=1, iseed=-10101)
attr(virall0a.cr3ab, 'model'))<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 Sex + a4 Trt2Comb + a5
1/sqrt(Age):Sex + a6 log(Age):Sex + a7 1/sqrt(Age):Trt2Comb + a8
log(Age):Trt2Comb + a9 Sex:Trt2Comb + a10 1/sqrt(Age):Sex:Trt2Comb + all
log(Age):Sex:Trt2Comb + Z0')

virall0a.cr3aca<-
censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL,
X = cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X3,virall0a$X4,
virall0a$X1C0,virall0a$X1C1,virall0a$X1C2,virall0a$X1C3,
virall0a$X1C4,virall0a$X1C5,
virall0a$X2C0,virall0a$X2C1,virall0a$X2C2,virall0a$X2C3,
virall0a$X2C4,virall0a$X2C5,
\mbox{with covariates:} \alpha=c(6.22, -1,-0.25,-0.83,-0.29,-0.77, -1.6, 0.1, 0.525, 0.59, -0.026,
1.2, 0.51, -4.87, -0.55,0.01, 0.01, 0.01 ),
D=1, sigma=1, iseed=-10101)
attr(virall0a.cr3aca, 'model'))<-
'LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 T4a + a4 Sex + a5 Trt2Comb + a6
1/sqrt(Age):Sex + a7 log(Age):Sex + a8 1/sqrt(Age):Trt2Comb + a9
log(Age):Trt2Comb + a10 Sex:Trt2Comb + all 1/sqrt(Age):Sex:Trt2Comb + a12
log(Age):Sex:Trt2Comb + Z0')

print('LogVL ~ a0 + a1/sqrt(Age) + a2 log(Age) + a3 T4a + a4 Sex + a5 Trt2Comb + a6
1/sqrt(Age):Sex + a7 log(Age):Sex + a8 1/sqrt(Age):Trt2Comb + a9
log(Age):Trt2Comb + a10 Sex:Trt2Comb + all 1/sqrt(Age):Sex:Trt2Comb + a12
log(Age):Sex:Trt2Comb + Z0')
virall0a.crs3acacompA<-  
censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL, 
X=  
cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X3,virall0a$C0,virall0a$C5 
', 
virall0a$X1C0,virall0a$X2C0,virall0a$X1C5,virall0a$X2C5, 
virall0a$C0C5, virall0a$C10),  
Z=matrix(virall0a$ZO,ncol=1),  
#alpha=c(6.22, -1,-0.25,-0.83,-0.29,-0.77, -1.6, 0.1, 0.525, 0.59, -0.026, 
1.2, 0.51, -4.87, -0.55,0.01, 0.01, 0.01 ),  
D=1, sigma=1, iseed=-10101)  
attr(virall0a.crs3acacompA, 'model')<-  
'LogVL ~ a O + a1/sqrt(Age) + a2 log (Age) + a3 T4a + a4 Sex + a5 Trt2Comb + 
a6 1/sqrt(Age):Sex + a7 log(Age):Sex + a8 1/sqrt(Age):Trt2Comb + a9 
log(Age):Trt2Comb + a10 Sex:Trt2Comb + ZO'  
print  
('LogVL ~ a O + a1/sqrt(Age) + a2 log (Age) + a3 T4a + a4 Sex + a5 Trt2Comb + 
a6 1/sqrt(Age):Sex + a7 log(Age):Sex + a8 1/sqrt(Age):Trt2Comb + a9 
log(Age):Trt2Comb + a10 Sex:Trt2Comb + ZO')  
virall0a.crs3acacomp<-  
censre2(id=virall0a$Serial, C=virall0a$CutOff, Q=virall0a$LogVL, 
X=  
cbind(virall0a$X0,virall0a$X1,virall0a$X2,virall0a$X3,virall0a$C0,virall0a$C5 
', 
virall0a$X1C0,virall0a$X2C0,virall0a$X1C5,virall0a$X2C5, 
virall0a$C0C5, virall0a$C10),  
Z=matrix(virall0a$ZO,ncol=1),  
#alpha=c(6.22, -1,-0.25,-0.83,-0.29,-0.77, -1.6, 0.1, 0.525, 0.59, -0.026, 
1.2, 0.51, -4.87, -0.55,0.01, 0.01, 0.01 ),  
D=1, sigma=1, iseed=-10101)  
attr(virall0a.crs3acacomp, 'model')<-  
'LogVL ~ a O + a1/sqrt(Age) + a2 log (Age) + a3 T4a + a4 Sex + a5 Trt2Comb + 
a6 1/sqrt(Age):Sex + a7 log(Age):Sex + a8 1/sqrt(Age):Trt2Comb + a9 
log(Age):Trt2Comb + a10 Sex:Trt2Comb + ZO'
Appendix 2.9.5: S-PLUS programme for mixed-linear with mid-point censoring for HIV RNA viral load

####################################
#create mid-point of LogVL, LogVLMP
####################################

viral9Ea$LogVLMP <- ifelse(viral9Ea$CutOff==-1, viral9Ea$LogVL/2, viral9Ea$LogVL)

#check LogVLMP

list(viral9Ea$LogVL[1:10], viral9Ea$LogVLMP[1:10], viral9Ea$CutOff[1:10])

#################################################
#group the data by LogVLMP in preparation for lme
#################################################

viral9EaGMP<- groupedData(LogVLMP ~ Age | Serial, data = viral9Ea)

#now run the different models - lme

#########################################################################
viral9EaGMP.Imela<- lme( LogVLMP ~ I(1/sqrt(Age)) + log(Age), data=viral9EaGMP, random= ~1 | Serial, method='ML')
summary( viral9EaGMP.Imela)tTable
logLik(viral9EaGMP.Imela)

#########################################################################
viral9EaGMP.Imelb<- lme( LogVLMP ~ (I(1/sqrt(Age)) + log(Age) ) + Sex, data=viral9EaGMP, random= ~1 | Serial, method='ML')
summary( viral9EaGMP.Imelb)tTable
logLik(viral9EaGMP.Imelb)

#########################################################################
viral9EaGMP.Imelc<- lme( LogVLMP ~ I(1/sqrt(Age)) + log(Age) + Trt2 , data=viral9EaGMP, random= ~1 | Serial, method='ML')
summary( viral9EaGMP.Imelc)tTable
logLik(viral9EaGMP.Imelc)

#########################################################################
viral9EaGMP.Imeld<- lme( LogVLMP ~ I(1/sqrt(Age)) + log(Age) + LogT4, data=viral9EaGMP, random= ~1 | Serial, method='ML')
summary( viral9EaGMP.Imeld)tTable
logLik(viral9EaGMP.Imeld)

#########################################################################
viral9EaGMP.lmele<- lme( LogVLMP ~ I(1/sqrt(Age)) + log(Age) + Assay, data=viral9EaGMP, random= ~1 | Serial, method='ML')
summary( viral9EaGMP.lmele)tTable
```
logLik(viral9EaGMP.lme1e)
#########################################################################
viral9EaGMP.lme1f<- lme( LogVLMP ~ I(1/sqrt(Age)) + log(Age) + Sex + Trt2,
data=viral9EaGMP, random= ~ 1 | Serial, method='ML')
summary( viral9EaGMP.lme1f)$tTable
logLik(viral9EaGMP.lme1f)
#########################################################################
viral9EaGMP.lme2b<- lme( LogVLMP ~ Sex*(I(1/sqrt(Age)) + log(Age) ),
data=viral9EaGMP, random= ~1 | Serial, method='ML')
summary( viral9EaGMP.lme2b)$tTable
logLik(viral9EaGMP.lme2b)
#########################################################################
viral9EaGMP.lme2c<- lme( LogVLMP ~ Trt2*(I(1/sqrt(Age)) + log(Age) ),
data=viral9EaGMP, random= ~1 | Serial, method='ML')
summary( viral9EaGMP.lme2c)$tTable
logLik(viral9EaGMP.lme2c)
#########################################################################
viral9EaGMP.lme2d<- lme( LogVLMP ~ LogT4*(I(1/sqrt(Age)) + log(Age) ),
data=viral9EaGMP, random= ~1 | Serial, method='ML')
summary( viral9EaGMP.lme2d)$tTable
logLik(viral9EaGMP.lme2d)
#########################################################################
# viral9EaGMP.lme2e<- lme( LogVLMP ~ Assay*(I(1/sqrt(Age)) + log(Age) ) +
data=viral9EaGMP, random= ~1 + I(1/sqrt(Age)) + log(Age) | Serial,
# method='ML')
# summary( viral9EaGMP.lme2e)$tTable
# logLik(viral9EaGMP.lme2e)
#########################################################################
viral9EaGMP.lme3acomp<-- lme( LogVLMP ~ Sex*(I(1/sqrt(Age)) + log(Age) ) +
Trt2*(I(1/sqrt(Age)) + log(Age) ) + LogT4 + Assay,
data=viral9EaGMP, random= ~1 | Serial, method='ML')
summary( viral9EaGMP.lme3acomp)$tTable
logLik(viral9EaGMP.lme3acomp)
#########################################################################
viral9EaGMP.lme3aca<-- lme( LogVLMP ~ Sex*Trt2*(I(1/sqrt(Age)) + log(Age) ) +
LogT4 + Assay,
data=viral9EaGMP, random= ~1 | Serial, method='ML')
```
summary(viral9EaGMP.lme3aca)$tTable
logLik(viral9EaGMP.lme3aca)
Appendix 2.9.6: S-PLUS programmes for splines for immunological markers

# fitting splines to log10 CD4+ for uninfected children
#
detach(1)
attach(\"C:\\linsay\\CD4\\Data\",1)
bl<- kidsft$inf=="1" & kidsft$age >= 0
c1<-is.na(kidsft$age) & !is.na(kidsft$logCD4)
a12<- kidsft$age <= 144

# generate number sequence for prediction
xPP<- seq(0/12, 144/12, length=length(kidsft$ageyrs[bl]))

kidsft.new <- groupedData( logCD4 ~ age | Serial, data = kidsft,
    labels = list( x = "Age", y = "log 10 CD4 cell count"),
    units = list( x = "(months)"))

# CD4 Spline with knots at 1 & 3 weeks, 1, 3, 6 months & 10 years
# Accounting for Repeated Measures
#
splineCD4.1.lme<-lme(logCD4 ~ ns(ageyrs,knots=c(0.25, 0.75, 1, 3, 6, 120)/12),
    random = ~ 1,
    na.action=na.omit, data=kidsft.new, subset=bl, method='ML')

# generate predicted values
predCD4.1.lme<-predict(splineCD4.1.lme, newdata=data.frame(ageyrs=xPP),
    level=0)

# CD4 Spline with knots: 3 weeks, 6 months, 4 years, 10 years
# Accounting for Repeated Measures
#
splineCD4.3.lme<- lme(logCD4 ~ ns(ageyrs,knots=c(0.75/12, 6/12, 3/12, 120/12)), random=1,
    data=kidsft.new, subset=bl, na.action=na.omit, method='ML')
predCD4.3.lme<-predict(splineCD4.3.lme, newdata=data.frame(ageyrs=xPP),
    level=0)

#
```r
splineCD4.4.lme <- lme(logCD4 ~ ns(ageyrs, knots=c(0.75/12, 6/12, 48/12, 120/12)), random=-1,
data=kidsft.new, subset=bl, na.action=na.omit, method='ML')
predCD4.4.lme <- predict(splineCD4.4.lme, newdata=data.frame(ageyrs=xPP), level=0)
#
# CD4 Spline with knots: 3 weeks, 1 month, 6 months, 10 years
# Accounting for Repeated Measures
#
splineCD4.5.lme <- lme(logCD4 ~ ns(ageyrs, knots=c(0.75/12, 1/12, 6/12, 120/12)), random=-1,
data=kidsft.new, subset=bl, na.action=na.omit, method='ML')
predCD4.5.lme <- predict(splineCD4.5.lme, newdata=data.frame(ageyrs=xPP), level=0)
#
# CD4 Spline with knots: 1 & 3 weeks, 3, 6 months & 10 years
# Accounting for Repeated Measures
#
splineCD4.6.lme <- lme(logCD4 ~ ns(ageyrs, knots=c(0.25/12, 0.75/12, 3/12, 6/12, 120/12)), random=-1,
data=kidsft.new, subset=bl, na.action=na.omit, method='ML')
predCD4.6.lme <- predict(splineCD4.6.lme, newdata=data.frame(ageyrs=xPP), level=0)
#
# CD4 Spline with knots: 1 & 3 weeks, 3, 8 months & 10 years
# Accounting for Repeated Measures
#
splineCD4.7.lme <- lme(logCD4 ~ ns(ageyrs, knots=c(0.25/12, 0.75/12, 3/12, 8/12, 120/12)), random=-1,
data=kidsft.new, subset=bl, na.action=na.omit, method='ML')
predCD4.7.lme <- predict(splineCD4.7.lme, newdata=data.frame(ageyrs=xPP), level=0)
#
# CD4 Spline with knots: 1 & 3 weeks, 3, 9 months & 10 years
# Accounting for Repeated Measures
#
splineCD4.8.lme <- lme(logCD4 ~ ns(ageyrs, knots=c(0.25/12, 0.75/12, 3/12, 9/12, 120/12)), random=-1,
data=kidsft.new, subset=bl, na.action=na.omit, method='ML')
```

326
predCD4.8.lme<-predict(splineCD4.8.lme, newdata=data.frame(ageyrs=xPP), level=0)

# # CD4 Spline with knots: 1 & 3weeks, 3, 10 months & 10 years
# Accounting for Repeated Measures
#
splineCD4.9.lme<- lme(logCD4 ~ ns(ageyrs,knots=c(0.25/12, 0.75/12, 3/12,10/12,120/12)), random=~1, 
data=kidsft.new, subset=b1, na.action=na.omit, method='ML')
predCD4.9.lme<-predict(splineCD4.9.lme, newdata=data.frame(ageyrs=xPP), level=0)

# # CD4 Spline with knots: 1 & 3weeks, 10 months & 10 years
# Accounting for Repeated Measures
#
splineCD4.10.lme<- lme(logCD4 ~ ns(ageyrs,knots=c(0.25/12, 0.75/12, 10/12,120/12)), random=~1, 
data=kidsft.new, subset=b1, na.action=na.omit, method='ML')
predCD4.10.lme<-predict(splineCD4.10.lme, newdata=data.frame(ageyrs=xPP), level=0)

# # CD4 Spline with knots: 1 & 3weeks, 3, 6 and 10 months & 10 years
# Accounting for Repeated Measures
#
splineCD4.11.lme<- lme(logCD4 ~ ns(ageyrs,knots=c(0.25/12, 0.75/12, 3/12, 6/12, 10/12,120/12)), random=~1, data=kidsft.new, subset=b1, na.action=na.omit, method='ML')
predCD4.11.lme<-predict(splineCD4.11.lme, newdata=data.frame(ageyrs=xPP), level=0)

# # testing main effects and interaction term for explanatory factors for
# log10 CD4+ for uninfected children
#
# General programming objects
c1<-.is.na(kidsft$age) & !is.na(kidsft$logCD4)
a12<-.kidsft$age <= 144
r1<-.is.na(kidsft$Race2)
s1<-.is.na(kidsft$Sex)
p1<-.is.na(kidsft$Prem)
p3<-is.na(kidsft$Prem3)
m1<-is.na(kidsftmcd$Matcat)
t1<-is.na(kidsft$trt)
al<- kidsft$age <= 12

# Programming objects for Uninfected children analysis
b1<- kidsft$inf=="1" & kidsft$age >= 0
xPP<- seq(0/12, 144/12, length=length(kidsft$ageyrs[b1]))

# Programming objects for Infected children analysis
b2<- kidsft$inf=="2" & kidsft$age >= 0
xIPP<- seq(0/12, 144/12, length=length(kidsft$ageyrs[b2]))

# Testing significance of Gender
splineCD4.9.lme.Sex<- lme(logCD4 ~ ns(ageyrs,knots=c(0.25/12, 0.75/12, 3/12, 10/12, 120/12))*Sex, random=~1, data=kidsft.Sex, subset=bl, na.action=na.omit, method = 'ML')

# run comparison model for where Sex is defined
splineCD4.9s1.lme<- lme(logCD4 ~ ns(ageyrs,knots=c(0.25/12, 0.75/12, 3/12,10/12,120/12)), random=~1, data=kidsft.new, subset=bl&s1, na.action=na.omit, method='ML')

splineCD4.9s1S.lme<- lme(logCD4 ~ ns(ageyrs,knots=c(0.25/12, 0.75/12, 3/12,10/12,120/12))+Sex, random=~1, data=kidsft.new, subset=bl&s1, na.action=na.omit, method='ML')

anova(splineCD4.9s1S.lme, splineCD4.9s1.lme)
anova(splineCD4.9.lme.Sex, splineCD4.9s1S.lme)

# Testing significance of Race as White vs Black
splineCD4.9.lme.Race<- lme(logCD4 ~ ns(ageyrs,knots=c(0.25/12, 0.75/12, 3/12, 10/12,120/12))*Race, random=~1, data=kidsft.Race, subset=bl, na.action=na.omit, method = 'ML')

splineCD4.9r1.lme<- lme(logCD4 ~ ns(ageyrs,knots=c(0.25/12, 0.75/12, 3/12,10/12,120/12)), random=~1, data=kidsft.new, subset=bl&r1, na.action=na.omit, method='ML')

splineCD4.9r1R.lme<- lme(logCD4 ~ ns(ageyrs,knots=c(0.25/12, 0.75/12, 3/12,10/12,120/12))+Race2, random=~1, data=kidsft.Race, subset=bl&r1, na.action=na.omit, method='ML')

328
anova(splineCD4.9rlR.lme, splineCD4.9rlR.lme)
anova(splineCD4.9.lme.Race2, splineCD4.9rlR.lme)

# fitting interaction of gender with race for CD4

spline\text{logCD4.4 lme.Sex.Race}<- \text{lme(logCD4} ~ \text{ns(ageyrs, knots=c(0.25, 3, 6, 24, 120)/12)}*\text{Sex*Race2, random=-1, data=kidsft.logCD4.Sex.Race, subset=bl, na.action=na.omit, method = 'ML')}

spline\text{logCD4.4 lme.Sex.Race.Com}<- \text{lme(logCD4} ~ \text{ns(ageyrs, knots=c(0.25, 3, 6, 24, 120)/12)}*\text{Sex*Race2 - ns(ageyrs, knots=c(0.25, 3, 6, 24, 120)/12):Sex:Race2, random=-1, data=kidsft.logCD4.Sex.Race, subset=bl, na.action=na.omit, method = 'ML')}


# Testing significance of Prematurity in three:
# "severely premature" <=34 weeks - 1
# "moderately premature" 34,37 - 2
# "full-term" >37 weeks - 3
#
options(contrasts= c ('conts. treatment', 'conts.poly'))
splineCD4.9.lme.Prem3<- \text{lme(logCD4} ~ \text{ns(ageyrs, knots=c(0.25/12, 0.75/12, 3/12, 10/12,120/12)})*\text{Prem3, random=-1, data=kidsft.new, subset=bl, na.action=na.omit, method='ML')}

# run comparison model for where Prem3 is defined

spline\text{CD4.9p3.lme}<- \text{lme(logCD4} ~ \text{ns(ageyrs, knots=c(0.25/12, 0.75/12, 3/12, 10/12,120/12)), random=-1, data=kidsft.new, subset=bl&p3, na.action=na.omit, method='ML')
spline\text{CD4.9p3P.lme}<- \text{lme(logCD4} ~ \text{ns(ageyrs, knots=c(0.25/12, 0.75/12, 3/12, 10/12,120/12))} +\text{Prem3, random=-1, data=kidsft.new, subset=bl&p3, na.action=na.omit, method='ML'})
anova(splineCD4.9p3.lme, splineCD4.9p3P.lme)
anova(splineCD4.9p3P.lme, splineCD4.9.lme.Prem3)

# CD4 run comparison model for where Matcat is defined

#
ml<-.is.na(kidsft$Matcat)

splineCD4.9.lme.Matcat<- lme(logCD4 ~ ns(ageyrs, knots=c(0.25/12, 0.75/12, 3/12, 10/12,120/12))*Matcat, random=~1, data=kidsft.new, subset=bl, na.action=na.omit, method='ML')

#*****************************************************************************
# run comparison model for where Matcat is defined
#*****************************************************************************

splineCD4.9ml.lme<- lme(logCD4 ~ ns(ageyrs, knots=c(0.25/12, 0.75/12, 3/12, 10/12,120/12)), random=~1, data=kidsft.new, subset=bl&ml, na.action=na.omit, method='ML')

splineCD4.9mlM.lme<- lme(logCD4 ~ ns(ageyrs, knots=c(0.25/12, 0.75/12, 3/12, 10/12,120/12))+Matcat, random=~1, data=kidsft.new, subset=bl&ml, na.action=na.omit, method='ML')

anova(splineCD4.9ml.lme, splineCD4.9mlM.lme)
anova(splineCD4.9mlM.lme, splineCD4.9.lme.Matcat)
Appendix 2.9.7: STATA programmes for Cox proportional hazards regression for determinants of disease progression

* OVERALL DISEASE PROGRESSION
*

\texttt{cd c:\linsay\DiseaseProgression\Stata}
\texttt{clear}
\texttt{insheet using TD973.txt}

* create the age in years variable
\texttt{sort serial agetd}
\texttt{g ageyears=agtd/12}
\texttt{summarize ageyears}

* grouping >2 treatments in "numtrt" into one category
\texttt{g conttrt = numtrt}
\texttt{recode conttrt 2/max=2}

* Restrict "race" to comparison of white and black children
* Want "racewb", Race as a comparison of just Black (2) and white (1)
\texttt{g racewb = race}
\texttt{recode racewb 1=. 2=1 3=2 4=. 5=. 9=.}

* STSET THE DATA FOR SURVIVAL ANALYSIS
\texttt{stset ageyears, failure(cord) id(serial)}

* Overall survival
\texttt{sts graph, b2title("Age (years)") l2title("Kaplan-Meier survival estimate")}
\texttt{t2title(" ")}
\texttt{sts list, at(1 5 10)}

* Univariate analysis:
* try the "healthiest" as the reference category where applicable
* (other than for treatment - take "None"/"No or Mono" as reference

* Sex
  * with "girls" as the reference category
\texttt{sts graph, by(sex)}
\texttt{stcox sex, robust hr nolog}

* Gestational age
\texttt{stcox gest, robust hr nolog}

* Prematurity
  * with full-term as the reference category
\texttt{sts graph, by(prem)}
\texttt{char prem [omit] 3}
\texttt{xi: stcox i.prem, robust hr nolog}

* Neo-natal AZT
  * with "No" as the reference category
sts graph, by(azt_neo)
stcox azt_neo, robust hr nolog

* Race White vs Black
  * with "white" as the reference category
sts graph, by(racewb)
char race [omit] 2
xi: stcox i.racewb, robust hr nolog

* Hepatomegaly twice within first 6 months
  * with "No" as the reference category
sts graph, by(hep62) b2title("Age (years)") l2title("Kaplan-Meier survival estimate")
stcox hep62, robust hr nolog

* Splenomegaly twice within first 6 months
  * with "No" as the reference category
sts graph, by(splen62)
stcox splen62, robust hr nolog

* Lymphadenopathy twice within first 6 months
  * with "No" as the reference category
sts graph, by(lymph62)
stcox lymph62, robust hr nolog

* Axillary nodes twice within first 6 months
  * with "No" as the reference category
sts graph, by(axil62)
stcox axil62, robust hr nolog

* First log CD4 cell count within first 6 months - continuous
stcox lcd46f, robust hr nolog

* First log CD4 cell count within first 6 months - categorical
  * with ">=log10(1500)" as the reference category
char lcd46fc [omit] 3
sts graph, by(lcd46fc)
xi: stcox i.lcd46fc, robust hr nolog

* First log CD8 cell count within first 6 months - continuous
stcox lcd86f, robust hr nolog

* First CD4 percentage within first 6 months - continuous
stcox cd4pc6f, robust hr nolog

* First CD4 percentage within first 6 months - categorical
  * with ">=25%" as the reference category
char cd4pc6fc [omit] 4
sts graph, by(cd4pc6fc)
xi: stcox i.cd4pc6fc, robust hr nolog

* CD4 percentage ever below 10%
  * with "No" as the reference category
sts graph, by(b10)
stcox b10, robust hr nolog
* First log absolute lymphocyte count within first 6 months - continuous
  stcox lal6f, robust hr nolog

* no, mono or Combination treatment
  * with "no" as the reference category
  sts graph, by(conttrt)
  xi: stcox i.conttrt, robust hr nolog

table cd4pc6fc cd4pcf20
tabulate cord cd4pcf20, row col
char cd4pcf20 [omit] 2
*sts graph, by(cd4pcf20) b2title("Age (years)") l2title("Probability of progression to serious disease or death") t ("") xlab(0, 5, 10, 15, 20)
s([-] [ ])
  xi: stcox cd4pcf20, robust hr nolog
sort cd4pcf20

* when stratified at 15, there are too few observations in one category -
* even although not significant at the 5% level, use cd4pcf20 in multivariate
* analysis.

**********************************************************************
* what about gender interactions?
**********************************************************************
* Gestational age * sex
  xi: stcox i.sex*gest, dead(cord) robust hr nolog
  lrtest, saving(0) force
  xi: stcox gest sex, dead(cord) robust hr nolog
  lrtest, force

* Prematurity * sex
  * with full-term as the reference category
  xi: stcox i.sex*prem, dead(cord) robust hr nolog
  lrtest, saving(0) force
  xi: stcox i.prem sex, dead(cord) robust hr nolog
  lrtest, force

* Neo-natal AZT * sex
  * with "No" as the reference category
  xi: stcox i.sex*azt_neo, dead(cord) robust hr nolog
  lrtest, saving(0) force
  xi: stcox i.azt_neo i.sex, dead(cord) robust hr nolog
  lrtest, force

* Race White vs Black * sex
  * with "white" as the reference category
  char race [omit] 2
  xi: stcox i.sex*racewb, dead(cord) robust hr nolog
  lrtest, saving(0) force
  xi: stcox i.racewb sex, dead(cord) robust hr nolog
  lrtest, force

* Hepatomegaly twice within first 6 months * sex
  * with "No" as the reference category
* Splenomegaly twice within first 6 months * sex
  * with "No" as the reference category
    xi:stcox i.sex*splen62, dead(cord) robust hr nolog
    lrtest, saving(0) force
    stcox splen62 sex, dead(cord) robust hr nolog
    lrtest, force

* Lymphadenopathy twice within first 6 months * sex
  * with "No" as the reference category
    xi:stcox i.sex*lymph62, dead(cord) robust hr nolog
    lrtest, saving(0) force
    stcox lymph62 sex, dead(cord) robust hr nolog
    lrtest, force

* Axillary nodes twice within first 6 months * sex
  * with "No" as the reference category
    xi:stcox i.sex*axil62, dead(cord) robust hr nolog
    lrtest, saving(0) force
    stcox axil62 sex, dead(cord) robust hr nolog
    lrtest, force

* First log CD4 cell count within first 6 months - continuous - * sex
  xi:stcox i.sex*lcd46f, dead(cord) robust hr nolog
  lrtest, saving(0) force
  stcox lcd46f sex, dead(cord) robust hr nolog
  lrtest, force

* First log CD8 cell count within first 6 months - continuous - * sex
  xi:stcox i.sex*lcd86f, dead(cord) robust hr nolog
  lrtest, saving(0) force
  stcox lcd86f sex, dead(cord) robust hr nolog
  lrtest, force

* First log CD4 cell count within first 6 months - categorical - * sex
  * with ">=log10(1500)" as the reference category
    char lcd46fc [omit] 3
    xi:stcox i.sex*lcd46fc, dead(cord) robust hr nolog
    lrtest, saving(0) force
    stcox lcd46fc sex, dead(cord) robust hr nolog
    lrtest, force

* First CD4 percentage within first 6 months - continuous - * sex
  xi:stcox i.sex*cd4pc6f, dead(cord) robust hr nolog
  lrtest, saving(0) force
  stcox sex cd4pc6f, dead(cord) robust hr nolog
  lrtest, force

* First CD4 percentage within first 6 months - categorical - * sex
  * with ">=20%" as the reference category
    char cd4pcf20 [omit] 2
    xi:stcox i.sex*cd4pcf20, dead(cord) robust hr nolog
    lrtest, saving(0) force
stcox cd4pcf20 sex, dead(cord) robust hr nolog
lrtest, force

* First log absolute lymphocyte count within first 6 months - continuous - *
  xi: stcox i.sex*la16f, dead(cord) robust hr nolog
  lrtest, saving(0) force
  stcox la16f sex, dead(cord) robust hr nolog
  lrtest, force

* no, mono or Combination treatment * sex
  * with "no" as the reference category
  xi: stcox i.sex*conttrt, dead(cord) robust hr nolog
  lrtest, saving(0) force
  xi: stcox i.conttrt i.sex, dead(cord) robust hr nolog
  lrtest, force

****************************************************************************************

* what about racewb interactions?
****************************************************************************************

* Gestational age * racewb
  xi: stcox i.racewb*gest, dead(cord) robust hr nolog
  lrtest, saving(0) force
  xi: stcox gest racewb, dead(cord) robust hr nolog
  lrtest, force

* Prematurity * sex
  * with full-term as the reference category
  xi: stcox i.racewb*prem, dead(cord) robust hr nolog
  lrtest, saving(0) force
  xi: stcox i.prem racewb, dead(cord) robust hr nolog
  lrtest, force

* Neo-natal AZT * racewb
  * with "No" as the reference category
  xi: stcox i.racewb*azt_neo, dead(cord) robust hr nolog
  lrtest, saving(0) force
  xi: stcox i.azt_neo i.racewb, dead(cord) robust hr nolog
  lrtest, force

* Race White vs Black * racewb
  * with "white" as the reference category
  char race [omit] 2
  xi: stcox i.sex*racewb, dead(cord) robust hr nolog
  lrtest, saving(0) force
  xi: stcox i.racewb sex, dead(cord) robust hr nolog
  lrtest, force

* Hepatomegaly twice within first 6 months * racewb
  * with "No" as the reference category
  xi: stcox i.racewb*hep62, robust hr nolog
  lrtest, saving(0) force
  xi: stcox hep62 racewb, robust hr nolog
  lrtest, force

* Splenomegaly twice within first 6 months * racewb
* with "No" as the reference category
\texttt{xi:stcox i.racewb*splen62, dead(cord) robust hr nolog}
\texttt{lrtest, saving(0) force}
\texttt{stcox splen62 racewb, dead(cord) robust hr nolog}
\texttt{lrtest, force}

* Lymphadenopathy twice within first 6 months * racewb
* with "No" as the reference category
\texttt{xi:stcox i.racewb*lymph62, robust hr nolog}
\texttt{lrtest, saving(0) force}
\texttt{stcox lymph62 racewb, robust hr nolog}
\texttt{lrtest, force}

* Axillary nodes twice within first 6 months * racewb
* with "No" as the reference category
\texttt{xi:stcox i.racewb*axil62, dead(cord) robust hr nolog}
\texttt{lrtest, saving(0) force}
\texttt{stcox axil62 racewb, dead(cord) robust hr nolog}
\texttt{lrtest, force}

* First log CD4 cell count within first 6 months - continuous - * racewb
\texttt{xi:stcox i.racewb*lcd46f, dead(cord) robust hr nolog}
\texttt{lrtest, saving(0) force}
\texttt{stcox lcd46f racewb, dead(cord) robust hr nolog}
\texttt{lrtest, force}

* First log CD8 cell count within first 6 months - continuous - * racewb
\texttt{xi:stcox i.racewb*lcd86f, dead(cord) robust hr nolog}
\texttt{lrtest, saving(0) force}
\texttt{stcox lcd86f racewb, dead(cord) robust hr nolog}
\texttt{lrtest, force}

* First log CD4 cell count within first 6 months - categorical - * racewb
* with ">=\log_{10}(1500)" as the reference category
\texttt{char lcd46fc [omit] 3}
\texttt{xi: stcox i.racewb*lcd46fc, robust hr nolog}
\texttt{lrtest, saving(0) force}
\texttt{stcox lcd46fc racewb, robust hr nolog}
\texttt{lrtest, force}

* First CD4 percentage within first 6 months - continuous - * racewb
\texttt{xi: stcox i.racewb*cd4pc6f, dead(cord) robust hr nolog}
\texttt{lrtest, saving(0) force}
\texttt{stcox racewb cd4pc6f, dead(cord) robust hr nolog}
\texttt{lrtest, force}

* First CD4 percentage within first 6 months - categorical - * racewb
* with ">=20\%" as the reference category
\texttt{char cd4pcf20 [omit] 2}
\texttt{xi: stcox i.racewb*cd4pcf20, robust hr nolog}
\texttt{lrtest, saving(0) force}
\texttt{stcox cd4pcf20 racewb, robust hr nolog}
\texttt{lrtest, force}

* First log absolute lymphocyte count within first 6 months - continuous - * racewb
\texttt{xi: stcox i.racewb*1a16f, robust hr nolog}
lrtest, saving(0), force
stcox lal6f racewb, robust hr nolog
lrtest, force

* no, mono or Combination treatment * racewb

* with "no" as the reference category
xi: stcox i.racewb*conttrt, dead(cord) robust hr nolog
lrtest, saving(0), force
xi: stcox i.conttrt i.racewb, dead(cord) robust hr nolog
lrtest, force

* Multivariable models
*

drop if lcd46f==.
drop if cd4pcf20==.
drop if lal6f==.

* Final model/s:

****** With first CD4 count and AL SIMULTANEOUSLY ******
xi: stcox sex azt_neo racewb hep62 lymph62 i.conttrt lcd46f lal6f, robust hr nolog

* Take out sex
xi: stcox azt_neo racewb hep62 lymph62 i.conttrt lcd46f lal6f, robust hr nolog

* Take out racewb
xi: stcox azt_neo hep62 lymph62 i.conttrt lcd46f lal6f, robust hr nolog

* Take out azt_neo
xi: stcox hep62 lymph62 i.conttrt lcd46f lal6f, robust hr nolog

* Take out hep62
xi: stcox lymph62 i.conttrt lcd46f lal6f, robust hr nolog

* Compared to model with only lcd46f
xi: stcox lymph62 i.conttrt lcd46f, robust hr nolog

* Compared to model with only lal6f
xi: stcox lymph62 i.conttrt lal6f, robust hr nolog

* Take out lymph62 instead
xi: stcox hep62 i.conttrt lcd46f lal6f, robust hr nolog

* Compared to model with only lcd46f
xi: stcox hep62 i.conttrt lcd46f, robust hr nolog

* Compared to model with only lal6f
xi: stcox hep62 i.conttrt lal6f, robust hr nolog

* Take out hep62 and lymph62
xi: stcox i.conttrt lcd46f lal6f, robust hr nolog
* Compared to model with only lcd46f
  `xi: stcox i.conttrt lcd46f, robust hr nolog`

* Compared to model with only la16f
  `xi: stcox i.conttrt la16f, robust hr nolog`

* Final model/s:

  ****** With first CD4 % (as categorical) and AL SIMULTANEOUSLY ******
  `xi: stcox sex azt_neo racewb hep62 lymph62 i.conttrt cd4pcf20 la16f, robust
  hr nolog`

  * Take out sex
  `xi: stcox azt_neo racewb hep62 lymph62 i.conttrt cd4pcf20 la16f, robust hr
  nolog`

  * Take out racewb
  `xi: stcox azt_neo hep62 lymph62 i.conttrt cd4pcf20 la16f, robust hr nolog`

  * Take out azt_neo
  `xi: stcox hep62 lymph62 i.conttrt cd4pcf20 la16f, robust hr nolog`

  * Take out hep62
  `xi: stcox lymph62 i.conttrt cd4pcf20 la16f, robust hr nolog`

  * Compared to model with only lcd46f
  `xi: stcox lymph62 i.conttrt cd4pcf20, robust hr nolog`

  * Compared to model with only la16f
  `xi: stcox lymph62 i.conttrt la16f, robust hr nolog`

  * Take out lymph62 instead
  `xi: stcox hep62 i.conttrt cd4pcf20 la16f, robust hr nolog`

  * Compared to model with only cd4pcf20
  `xi: stcox hep62 i.conttrt cd4pcf20, robust hr nolog`

  * Compared to model with only la16f
  `xi: stcox hep62 i.conttrt la16f, robust hr nolog`

  * Take out hep62 and lymph62
  `xi: stcox i.conttrt cd4pcf20 la16f, robust hr nolog`

  * Compared to model with only lcd46f
  `xi: stcox i.conttrt cd4pcf20, robust hr nolog`

  * Compared to model with only la16f
  `xi: stcox i.conttrt la16f, robust hr nolog`

  * Final model/s:

  ****** With first CD4 % (as categorical, AL and CD8 SIMULTANEOUSLY ******
  `xi: stcox lymph62 i.conttrt cd4pcf20 la16f la86f, robust hr nolog`
* Final model/s: 
***** With first CD8 (as continuous) and AL SIMULTANEOUSLY *****
 xi: stcox lymph62 i.conttrt lal6f lcd86f, robust hr nolog

* Final model/s: 
***** With first CD4 % (as categorical) and CD8 (as continuous) SIMULTANEOUSLY *****
 xi: stcox lymph62 i.conttrt cd4pcf20 lcd86f, robust hr nolog
Appendix 4.1  Fitted one-, two- and three-power fractional polynomials

Figure A4.1A: Fitted one-, two- and three-power fractional polynomials

![Graph showing fitted one-, two- and three-power fractional polynomials.](image-url)
### Appendix 4.2  
**Expressions of models fitted using Hughes' method to determine important covariate and interaction terms for the final model**

<table>
<thead>
<tr>
<th>Model Number</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>$\log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + Z_i \gamma_i + \varepsilon</em>{ij}$</td>
</tr>
<tr>
<td>A2</td>
<td>$\log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{Gender}</em>{ij} + Z_i \gamma_i + \varepsilon_{ij}$</td>
</tr>
<tr>
<td>A3</td>
<td>$\log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{Treatment}</em>{ij} + Z_i \gamma_i + \varepsilon_{ij}$</td>
</tr>
<tr>
<td>A4</td>
<td>$\log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{CD4}</em>{ij} + Z_i \gamma_i + \varepsilon_{ij}$</td>
</tr>
<tr>
<td>A5</td>
<td>$\log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{Assay}</em>{ij} + Z_i \gamma_i + \varepsilon_{ij}$</td>
</tr>
<tr>
<td>A6</td>
<td>$\log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{Gender}</em>{ij} + \beta_4 \text{Treatment}_{ij}$</td>
</tr>
<tr>
<td>A7</td>
<td>\begin{align*} \log_{10}(\text{Viral Load}<em>{ij}) &amp;= \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{Gender}</em>{ij} \ &amp;+ \frac{\beta_4 \text{Gender}<em>{ij}}{\sqrt{\text{Age}</em>{ij}}} + \beta_5 \ln(\text{Age}<em>{ij}) \cdot \text{Gender}</em>{ij} + Z_i \gamma_i + \varepsilon_{ij} \end{align*}</td>
</tr>
<tr>
<td>A8</td>
<td>\begin{align*} \log_{10}(\text{Viral Load}<em>{ij}) &amp;= \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{Treatment}</em>{ij} \ &amp;+ \frac{\beta_4 \text{Treatment}<em>{ij}}{\sqrt{\text{Age}</em>{ij}}} + \beta_5 \ln(\text{Age}<em>{ij}) \cdot \text{Treatment}</em>{ij} + Z_i \gamma_i + \varepsilon_{ij} \end{align*}</td>
</tr>
</tbody>
</table>
Appendix 4.2 cont. Expressions of models fitted using Hughes’ method to determine important covariate and interaction terms for the final model

<table>
<thead>
<tr>
<th>A9</th>
<th>[ \log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 CD4</em>{ij} + \frac{\beta_4}{\sqrt{\text{Age}<em>{ij}}} + \beta_5 \ln(\text{Age}</em>{ij}) \cdot CD4_{ij} + Z_{11} \gamma_i + \epsilon_{ij} ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A10</td>
<td>[ \log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{Gender}</em>{ij} + \beta_4 \text{Treatment}<em>{ij} + \beta_5 \ln(\text{Gender}</em>{ij}) \cdot \text{Treatment}<em>{ij} + Z</em>{11} \gamma_i + \epsilon_{ij} ]</td>
</tr>
<tr>
<td>A11</td>
<td>[ \log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{Gender}</em>{ij} + \beta_4 \text{Treatment}<em>{ij} + \beta_5 CD4</em>{ij} + \beta_6 \text{Assay}<em>{ij} + \frac{\beta_7}{\sqrt{\text{Age}</em>{ij}}} + \frac{\beta_8}{\sqrt{\text{Age}<em>{ij}}} \cdot \text{Gender}</em>{ij} + \beta_9 \ln(\text{Age}<em>{ij}) \cdot \text{Treatment}</em>{ij} + Z_{11} \gamma_i + \epsilon_{ij} ]</td>
</tr>
<tr>
<td>A12</td>
<td>[ \log_{10}(\text{Viral Load}<em>{ij}) = \beta_0 + \frac{\beta_1}{\sqrt{\text{Age}</em>{ij}}} + \beta_2 \ln(\text{Age}<em>{ij}) + \beta_3 \text{Gender}</em>{ij} + \beta_4 \text{Treatment}<em>{ij} + \beta_5 CD4</em>{ij} + \beta_6 \text{Assay}<em>{ij} + \frac{\beta_7}{\sqrt{\text{Age}</em>{ij}}} + \frac{\beta_8}{\sqrt{\text{Age}<em>{ij}}} \cdot \text{Gender}</em>{ij} + \beta_9 \ln(\text{Age}<em>{ij}) \cdot \text{Treatment}</em>{ij} + \beta_{10} \ln(\text{Age}<em>{ij}) \cdot \text{Treatment}</em>{ij} + Z_{11} \gamma_i + \epsilon_{ij} ]</td>
</tr>
</tbody>
</table>

Note: "*" denotes interaction terms.
Appendix 5.1 Example of establishing choice of knots for spline model for CD4⁺ cell count of uninfected children

Figure A5.1A: Example of establishing choice of knots for spline model for CD4⁺ cell count of uninfected children
Appendix 5.2 Super smoothers for selected immunological markers by gender and race

Figure A5.1: Super smoothers of log10 CD8+ cell counts over age by gender in uninfected children
Figure A5.2: Super smoothers of CD4$^+$ percentage over age by race in uninfected children
Figure A5.3: Super smoothers of $\log_{10}$ CD4$^+$ cell counts over age by gender in infected children
Figure A5.4: Super smoothers of $\log_{10}$ CD4$^+$ cell counts over age by race in infected children.
Figure A5.5: Fitted values of $\log_{10}$ CD4$^+$ cell counts over age by vital status at last follow-up in infected children
Appendix 5.3  · Immunological markers by gender and race over age in infected children adjusted for by ART

Table A5.1: Gender, race and CD4⁺ cell counts over age in infected children adjusted for by ART

<table>
<thead>
<tr>
<th>Factor</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$-value</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1.97</td>
<td>1</td>
<td>0.1601</td>
<td>27.78</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Race</td>
<td>0.0064</td>
<td>1</td>
<td>0.9361</td>
<td>19.13</td>
<td>1</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

Table A5.2: Gender, race and CD8⁺ cell counts over age in infected children adjusted for by ART

<table>
<thead>
<tr>
<th>Factor</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$-value</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>8.27</td>
<td>1</td>
<td>0.004</td>
<td>53.20</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Race</td>
<td>1.65</td>
<td>1</td>
<td>0.1984</td>
<td>14.41</td>
<td>1</td>
<td>0.0132</td>
</tr>
</tbody>
</table>

Table A5.3: Gender, race and absolute lymphocyte counts over age in infected children adjusted for by ART

<table>
<thead>
<tr>
<th>Factor</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$-value</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>6.91</td>
<td>1</td>
<td>0.0086</td>
<td>40.99</td>
<td>1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Race</td>
<td>0.098</td>
<td>1</td>
<td>0.7545</td>
<td>10.23</td>
<td>1</td>
<td>0.0367</td>
</tr>
</tbody>
</table>
Table A5.4: Gender, race and percentage of CD4\(^+\) cell counts over age in infected children adjusted for by ART

<table>
<thead>
<tr>
<th>Factor</th>
<th>Main effect</th>
<th>Interaction with age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\chi^2)</td>
<td>df</td>
</tr>
<tr>
<td>Gender</td>
<td>0.0088</td>
<td>1</td>
</tr>
<tr>
<td>Race</td>
<td>0.0694</td>
<td>1</td>
</tr>
</tbody>
</table>
### Appendix 6.1  Considered multivariable models

Table A6.1.1: Considered multivariable models for predicting overall progression

<table>
<thead>
<tr>
<th>Model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$ test of $\beta_1=...\beta_p = 0$</td>
<td>19.58</td>
<td>20.56</td>
<td>17.51</td>
<td>9.80</td>
<td>13.02</td>
</tr>
<tr>
<td>df</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0066</td>
<td>0.0022</td>
<td>0.0036</td>
<td>0.0440</td>
<td>0.0112</td>
</tr>
<tr>
<td>(AZT) administered neonatally</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.439</td>
<td>(0.468)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first available CD4(^+)% below/above 20</td>
<td>0.352</td>
<td>(0.029)</td>
<td>0.340</td>
<td>(0.023)</td>
<td>0.358</td>
</tr>
<tr>
<td>first available log(_{10}) CD8(^+) count (x 10(^6) cells/l) (per log increase)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first available log(_{10}) absolute lymphocyte count (x 10(^6) cells/l) (per log increase)</td>
<td>0.244</td>
<td>(0.019)</td>
<td>0.243</td>
<td>(0.018)</td>
<td>0.233</td>
</tr>
<tr>
<td>Hepatomegaly (2 or more episodes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.514</td>
<td>(0.291)</td>
<td>1.572</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymphadenopathy (2 or more episodes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.427</td>
<td>(0.378)</td>
<td>1.449</td>
<td>(0.363)</td>
<td>1.883</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.408</td>
<td>(0.372)</td>
<td>1.384</td>
<td>(0.393)</td>
<td>1.495</td>
</tr>
<tr>
<td>Combination</td>
<td>0.565</td>
<td>(0.501)</td>
<td>0.470</td>
<td>(0.301)</td>
<td>0.504</td>
</tr>
</tbody>
</table>
Table A6.1.1 *cont.* Considered multivariable models for predicting overall progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$ test of $\beta_1=\ldots\beta_p=0$</td>
<td>17.32</td>
<td>12.73</td>
<td>11.87</td>
<td>6.07</td>
<td>6.32</td>
</tr>
<tr>
<td>df</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.0039</td>
<td>0.0127</td>
<td>0.0183</td>
<td>0.1085</td>
<td>0.0968</td>
</tr>
<tr>
<td>(AZT) administered neonatally</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first available CD4&lt;sup&gt;+&lt;/sup&gt;% below/above 20</td>
<td>0.331</td>
<td>0.447</td>
<td>0.482</td>
<td>(0.018)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>first available log&lt;sub&gt;10&lt;/sub&gt; CD8&lt;sup&gt;+&lt;/sup&gt; count (x 10&lt;sup&gt;6&lt;/sup&gt; cells/l) (per log increase)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first available log&lt;sub&gt;10&lt;/sub&gt; absolute lymphocyte count (x 10&lt;sup&gt;6&lt;/sup&gt; cells/l) (per log increase)</td>
<td>0.253</td>
<td>0.345</td>
<td>0.337</td>
<td>(0.026)</td>
<td>(0.073)</td>
</tr>
<tr>
<td>Hepatomegaly (2 or more episodes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.885</td>
<td>1.194</td>
<td>1.816</td>
<td>(0.035)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymphadenopathy (2 or more episodes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>No</td>
<td>1.426</td>
<td>1.480</td>
<td>1.481</td>
<td>1.674</td>
<td>1.720</td>
</tr>
<tr>
<td>Monotherapy</td>
<td>(0.348)</td>
<td>(0.300)</td>
<td>(0.309)</td>
<td>(0.156)</td>
<td>(0.126)</td>
</tr>
<tr>
<td>Combination</td>
<td>0.307</td>
<td>0.523</td>
<td>0.506</td>
<td>0.462</td>
<td>0.577</td>
</tr>
<tr>
<td></td>
<td>(0.213)</td>
<td>(0.366)</td>
<td>(0.336)</td>
<td>(0.501)</td>
<td>(0.427)</td>
</tr>
</tbody>
</table>

352
Table A6.1.1 *cont.* Considered multivariable models for predicting overall progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Model 11</th>
<th>Model 12</th>
<th>Model 13</th>
<th>Model 14</th>
<th>Model 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$ test of $\beta_1=\ldots=\beta_p = 0$</td>
<td>14.55</td>
<td>13.61</td>
<td>20.22</td>
<td>16.32</td>
<td>9.97</td>
</tr>
<tr>
<td>df</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.0125</td>
<td>0.0183</td>
<td>0.0025</td>
<td>0.0060</td>
<td>0.0761</td>
</tr>
<tr>
<td>(AZT) administered neonatally</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>first available CD4 %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>below/above 20</td>
<td>0.464</td>
<td>0.360</td>
<td>0.446</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.043)</td>
<td>(0.105)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>first available log_{10} CD8^+ count</td>
<td>1.479</td>
<td>1.635</td>
<td>0.602</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(x 10^6 cells/l) (per log increase)</td>
<td>(0.558)</td>
<td>(0.461)</td>
<td>(0.431)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>first available log_{10} absolute lymphocyte count</td>
<td>0.331</td>
<td>0.179</td>
<td>0.234</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(x 10^6 cells/l) (per log increase)</td>
<td>(0.055)</td>
<td>(0.006)</td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatomegaly (2 or more episodes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.471</td>
<td>1.661</td>
<td>1.452</td>
<td>1.401</td>
<td>1.446</td>
</tr>
<tr>
<td></td>
<td>(0.336)</td>
<td>(0.182)</td>
<td>(0.327)</td>
<td>(0.377)</td>
<td>(0.330)</td>
</tr>
<tr>
<td>Yes</td>
<td>1.531</td>
<td>1.341</td>
<td>1.393</td>
<td>1.401</td>
<td>1.446</td>
</tr>
<tr>
<td></td>
<td>(0.299)</td>
<td>(0.474)</td>
<td>(0.375)</td>
<td>(0.377)</td>
<td>(0.330)</td>
</tr>
<tr>
<td>Lymphadenopathy (2 or more episodes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.414</td>
<td>1.452</td>
<td>1.393</td>
<td>1.401</td>
<td>1.446</td>
</tr>
<tr>
<td></td>
<td>(0.374)</td>
<td>(0.327)</td>
<td>(0.375)</td>
<td>(0.377)</td>
<td>(0.330)</td>
</tr>
<tr>
<td>Combination</td>
<td>0.488</td>
<td>0.520</td>
<td>0.554</td>
<td>0.565</td>
<td>0.562</td>
</tr>
<tr>
<td></td>
<td>(0.320)</td>
<td>(0.366)</td>
<td>(0.421)</td>
<td>(0.431)</td>
<td>(0.428)</td>
</tr>
</tbody>
</table>

353
Table A6.1.2: Considered multivariable models for predicting rapid progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>138</td>
<td>138</td>
<td>138</td>
<td>138</td>
<td>138</td>
<td>138</td>
</tr>
<tr>
<td>( \chi^2 ) test of ( \beta_1=...\beta_p = 0 )</td>
<td>6.96</td>
<td>13.61</td>
<td>11.72</td>
<td>6.28</td>
<td>10.18</td>
<td>3.98</td>
</tr>
<tr>
<td>df</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>p-value</td>
<td>0.0732</td>
<td>0.0035</td>
<td>0.0084</td>
<td>0.0433</td>
<td>0.0062</td>
<td>0.1369</td>
</tr>
<tr>
<td>first available log_{10} absolute lymphocyte count (x 10^6 cells/l) (per log increase)</td>
<td>0.679 (0.733)</td>
<td>0.111 (0.012)</td>
<td>0.0084</td>
<td>0.0433</td>
<td>0.0062</td>
<td>0.1369</td>
</tr>
<tr>
<td>first available CD4 percentage</td>
<td>2.401 (0.420)</td>
<td>3.096 (0.291)</td>
<td>2.748 (0.348)</td>
<td>2.358 (0.431)</td>
<td>2.826 (0.328)</td>
<td>2.382 (0.417)</td>
</tr>
<tr>
<td>Treatment</td>
<td>No/Monotherapy</td>
<td>Combination</td>
<td>No/Monotherapy</td>
<td>Combination</td>
<td>No/Monotherapy</td>
<td>Combination</td>
</tr>
</tbody>
</table>
### Table A6.1.3: Considered multivariable models for predicting long-term progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>237</td>
<td>237</td>
<td>237</td>
<td>237</td>
<td>237</td>
</tr>
<tr>
<td>( \chi^2 ) test of ( \beta_1 = \ldots \beta_p = 0 )</td>
<td>16.31</td>
<td>12.43</td>
<td>11.91</td>
<td>16.99</td>
<td>9.98</td>
</tr>
<tr>
<td>df</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>( p )-value</td>
<td>0.0224</td>
<td>0.0145</td>
<td>0.0641</td>
<td>0.0093</td>
<td>0.0187</td>
</tr>
<tr>
<td>Lymphadenopathy (2 or more episodes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.514 (0.376)</td>
<td>1.365 (0.533)</td>
<td>2.093 (0.075)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatomegaly (2 or more episodes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.775 (0.158)</td>
<td>1.847 (0.170)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prematurity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-term</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately premature</td>
<td>0.642 (0.424)</td>
<td></td>
<td>0.634 (0.438)</td>
<td>0.667 (0.456)</td>
<td></td>
</tr>
<tr>
<td>Severe prematurity</td>
<td>0.464 (0.090)</td>
<td>0.456 (0.105)</td>
<td>0.494 (0.110)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.645 (0.212)</td>
<td>1.615 (0.224)</td>
<td>1.780 (0.132)</td>
<td>1.708 (0.165)</td>
<td>1.732 (0.143)</td>
</tr>
<tr>
<td>Monotherapy</td>
<td>0.375 (0.213)</td>
<td>0.358 (0.191)</td>
<td>0.411 (0.261)</td>
<td>0.390 (0.226)</td>
<td>0.395 (0.238)</td>
</tr>
<tr>
<td>Combination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table A6.1.3: Considered multivariable models for predicting long-term progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Model 6</th>
<th>Model 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>237</td>
<td>237</td>
</tr>
<tr>
<td>$\chi^2$ test of $\beta_1=...\beta_p = 0$</td>
<td>12.00</td>
<td>7.29</td>
</tr>
<tr>
<td>df</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.0074</td>
<td>0.2001</td>
</tr>
<tr>
<td>Lymphadenopathy (2 or more episodes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hepatomegaly (2 or more episodes)</td>
<td>2.150</td>
<td>(0.040)</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prematurity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-term</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately premature</td>
<td>0.655</td>
<td>(0.463)</td>
</tr>
<tr>
<td>Severely premature</td>
<td>0.478</td>
<td>(0.119)</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.672</td>
<td>1.967</td>
</tr>
<tr>
<td>Combination</td>
<td>(0.179)</td>
<td>(0.059)</td>
</tr>
<tr>
<td></td>
<td>0.369</td>
<td>0.473</td>
</tr>
<tr>
<td></td>
<td>(0.199)</td>
<td>(0.334)</td>
</tr>
</tbody>
</table>
Appendix 6.2  Multivariable models for determinants analysis adjusted by PCP prophylaxis and IVIG ever before progression

Table A6.2.1: Timing of progression by PCP prophylaxis and IVIG ever before progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Rapid Progressors</th>
<th>Long-term Progressors</th>
<th>Progressors</th>
<th>Non-Progressors</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 28 (43.1%)</td>
<td>n = 37 (56.9%)</td>
<td>n = 65 (40.4%)</td>
<td>n = 96 (59.6%)</td>
<td>n = (161)</td>
</tr>
<tr>
<td>PCP prophylaxis ever before progression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>20 (71.4)</td>
<td>15 (40.5)</td>
<td>35 (53.8)</td>
<td>48 (50.0)</td>
<td>83 (51.6)</td>
</tr>
<tr>
<td>Yes</td>
<td>8 (28.6)</td>
<td>22 (59.5)</td>
<td>30 (46.2)</td>
<td>48 (50.0)</td>
<td>78 (48.4)</td>
</tr>
<tr>
<td>IVIG ever before progression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>20 (71.4)</td>
<td>14 (37.8)</td>
<td>34 (52.3)</td>
<td>59 (61.5)</td>
<td>93 (57.8)</td>
</tr>
<tr>
<td>Yes</td>
<td>8 (28.6)</td>
<td>23 (62.2)</td>
<td>31 (47.7)</td>
<td>37 (38.5)</td>
<td>68 (42.2)</td>
</tr>
</tbody>
</table>

Table A6.2.2: Univariable results for overall progression to serious disease or death by PCP prophylaxis and IVIG ever before progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)*</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCP prophylaxis ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.702 (0.435 to 1.135)</td>
<td>0.149</td>
</tr>
<tr>
<td>IVIG ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.051 (0.651 to 1.696)</td>
<td>0.838</td>
</tr>
</tbody>
</table>
Figure A6.2: Kaplan-Meier survival plot of progression to CDC clinical category C or death by receipt of PCP prophylaxis. The red line presents survival in those administered PCP prophylaxis and the green line those not administered PCP prophylaxis.
Figure A6.3: Kaplan-Meier survival plot of progression to CDC clinical category C or death by receipt IVIG as prophylaxis for bacterial infections. The red line presents survival in those administered IVIG and the green line those not administered IVIG.
Table A6.2.3: Multivariable results for overall progression to serious disease or death adjusted for PCP and bacterial infection prophylaxes

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>first available log&lt;sub&gt;10&lt;/sub&gt; absolute lymphocyte count (x 10&lt;sup&gt;6&lt;/sup&gt; cells/l) (per log increase)</td>
<td>0.237 (0.079 to 0.710)</td>
<td>0.010</td>
</tr>
<tr>
<td>first available CD4 percentage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20%</td>
<td>2.60 (0.994 to 6.84)</td>
<td>0.052</td>
</tr>
<tr>
<td>&gt;= 20%</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Lymphadenopathy (2 or more episodes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.759 (0.964 to 3.211)</td>
<td>0.066</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.778 (0.838 to 3.776)</td>
<td>0.134</td>
</tr>
<tr>
<td>Combination</td>
<td>0.592 (0.149 to 2.349)</td>
<td>0.456</td>
</tr>
<tr>
<td>PCP prophylaxis ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.592 (0.340 to 1.033)</td>
<td>0.065</td>
</tr>
<tr>
<td>IVIG ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.110 (0.611 to 2.018)</td>
<td>0.732</td>
</tr>
</tbody>
</table>
Table A6.2.4: Univariable results for rapid progression to serious disease or death by PCP prophylaxis and IVIG ever before progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)*</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCP prophylaxis ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.369 (0.164 to 0.828)</td>
<td>0.016</td>
</tr>
<tr>
<td>IVIG ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.489 (0.218 to 1.097)</td>
<td>0.083</td>
</tr>
</tbody>
</table>

Table A6.2.5: Multivariable results for rapid progression to serious disease or death adjusted for PCP and bacterial infection prophylaxes

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>first available log_{10} absolute lymphocyte count (x 10^6 cells/l) (per log increase)</td>
<td>0.108 (0.022 to 0.531)</td>
<td>0.006</td>
</tr>
<tr>
<td>first available CD4 percentage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20%</td>
<td>6.234 (2.254 to 17.24)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>&gt;= 20%</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference*</td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Combination</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PCP prophylaxis ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.441 (0.189 to 1.029)</td>
<td>0.058</td>
</tr>
<tr>
<td>IVIG ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.896 (0.361 to 2.223)</td>
<td>0.813</td>
</tr>
</tbody>
</table>

* Estimates unstable due to collinearity with PCP prophylaxis and IVIG
Table A6.2.6: Univariable results for long term progression to serious disease or death by PCP prophylaxis and IVIG ever before progression

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)*</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCP prophylaxis ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.101 (0.573 to 2.116)</td>
<td>0.772</td>
</tr>
<tr>
<td>IVIG ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.813 (0.934 to 3.518)</td>
<td>0.079</td>
</tr>
</tbody>
</table>

Table A6.2.7: Multivariable results for long-term progression to serious disease or death adjusted for PCP and bacterial infection prophylaxes

<table>
<thead>
<tr>
<th>Factor</th>
<th>Hazard ratio (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatomegaly (2 or more episodes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2.359 (1.073 to 5.185)</td>
<td>0.033</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Monotherapy</td>
<td>1.637 (0.708 to 3.788)</td>
<td>0.249</td>
</tr>
<tr>
<td>Combination</td>
<td>0.383 (0.825 to 1.777)</td>
<td>0.220</td>
</tr>
<tr>
<td>PCP prophylaxis ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0.648 (0.281 to 1.492)</td>
<td>0.308</td>
</tr>
<tr>
<td>IVIG ever before progression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.646 (0.778 to 3.483)</td>
<td>0.192</td>
</tr>
</tbody>
</table>