Prediction of time to threshold from a repeatedly measured biomarker

Tarylee Reddy

Promotor: Prof. dr. Geert Molenberghs

Co-Promotor: Prof. dr. Marc Aerts

Acknowledgements

This thesis would not have been posssible without the guidance, support and motivation from several individuals. I am most appreciative of my promoter Prof. dr. Geert Molenberghs. Geert, I cannot thank you enough for all your guidance over the past four years. Your commitment to developing others, compassion, and ability to come up with creative solutions to complex problems motivates me to be a better statistician and person everyday. It has been an honour and privilege to work with you. I have thoroughly enjoyed our meetings and our discussions which ranged from statistics to history, politics and languages. I am also grateful to my co-promotor, Prof. dr. Marc Aerts. Marc, thank you so much for your insightful comments and suggestions on this thesis. Prof. dr. Edmund Njeru-Njagi and Prof. dr. Liesbeth Bruckers, I would like to thank you both for your valuable contributions. I have thoroughly enjoyed working with you both. I would also like to extend my gratitude to every member of the jury for taking the time to read an earlier version of this dissertation and for providing valuable comments and suggestions.

I am indebted to the BOF Bilateral Cooperation for funding my PhD. The skills and knowledge I have acquired through attending courses and my interaction with statisticians here, have had an immeasurable effect on my career. I have made a total of five visits to Diepenbeek since the start of my PhD, and during each of these visits Martine Machiels and Marc Thoelen have been most helpful with administration. I thank you both your assistance. I am appreciative of Yimer, Martin, Renata, Tadesse, Carlos, Koen and especially Trishanta for their support and friendship during my stay in Belgium. Trishanta,

you are more a sister than a friend, and I truly appreciate your support over the years. It was so exciting embarking on this journey with you and I have thoroughly enjoyed our adventures together.

There are several statisticians in South Africa who have played a key role by motivating to pursue a PhD in Statistics. They are Prof. Carl Lombard, Ms. Cathy Connolly, Dr. Anneke Grobler, Prof. Henry Mwambi, Dr. Gaetan Kabera and Prof. Samuel Manda. Thank you for being wonderful mentors. For the duration of my PhD my Unit Director at the SAMRC, Prof. Carl Lombard, has been most supportive by offering me the time away to pursue this PhD in Belgium. Thank you so much, Carl. I am also grateful to all my colleagues at the SAMRC Biostatistics Unit for their support and 'holding down the fort' in my absence. Cathy, Tenuja, Renee, Inbarani and Anneke: I thank you all for your friendship and support. Our walks, lunches and tea celebrations have always left me feeling motivated.

I would like to thank my parents, Dr. David Reddy and Sury Reddy, for moulding me into the person that I am today. Thank you for encouraging me to pursue my dreams and for teaching me to see challenges as opportunities. I dedicate this thesis to the memory of my father. This has been the most emotionally trying year, but the desire to make you proud has motivated me to complete this thesis. I am also grateful to my siblings, Cassandra and Sheraton, for being wonderful role models.

Finally, and most importantly, I am indebted to my husband, Yatheen, for encouraging me to pursue this PhD and for his unwavering support.

Tarylee Reddy

Diepenbeek, 2017

List of Publications and Reports

The contents of this thesis are based on the following publications:

Reddy, T., Molenberghs, G., Njeru Njagi, E. and Aerts, M. (2016). A novel approach to estimation of the time to biomarker threshold: applications to HIV. *Pharmaceutical Statistics,* **15**, 541-549.

Reddy, T., Molenberghs, G., Bruckers, L., Njagi, E.N., Aerts, M. and Schurink, G. (2017). Random effects models for estimation of the probability and time to progression of a continuous biomarker. *Revised and resubmitted.*

Reddy, T., Molenberghs, G. and Aerts, M. (2017). Prediction of the time to treatment success from longitudinal viral load measurements in a cohort of HIV positive individuals. *Submitted.*

Reddy, T., Molenberghs and Aerts, M. (2017). Estimating time to remission from repeated ordinal measurements: applications to psychiatry. *To be submitted.*

The author has also contributed to the following publications:

Bantubani, N., Kabera, G., Connolly, C., Rustomjee, R., **Reddy, T**., Cohen, T. and Pym, A. S. (2014). High rates of potentially infectious tuberculosis and multidrug-resistant tuberculosis (MDR-TB) among hospital inpatients in KwaZulu Natal, South Africa indicate risk of nosocomial transmission. *PloS One*, **9**.

Bhorat, I., Bagratee, J. and **Reddy, T**. (2014). Gestational age-adjusted trends and reference intervals of the Modified Myocardial Performance Index (Mod-MPI) and its components, with its interpretation in the context of established cardiac physiological principles. *Prenatal diagnosis,* **34**, 1031-1036.

Bhorat, I. E., Bagratee, J. S., Pillay, M. and **Reddy, T**. (2014). Use of the myocardial performance index as a prognostic indicator of adverse fetal outcome in poorly controlled gestational diabetic pregnancies. *Prenatal diagnosis,* **34**, 1301-1306.

Kilembe, W., Mokgoro, M., Mwaanga, A., Kamusoko, M., **Reddy, T**., Dissen, E., Davitte, J., Hilda, S. P., Brockman, M., Ndungu, T. and Allen, S. (2014). Opportunities for HIV Prevention among Couples in Durban, South Africa. *AIDS Research and Human Retroviruses,* **30**(S1), A111-A112.

Mann, J. K., Chopera, D., Omarjee, S., Kuang, X. T., Le, A. Q., Anmole, G., Danroth, R., Mwimanzi, P., **Reddy, T**., Carlson, J.,. . . and Radebe, M. (2014). Nef-mediated down-regulation of CD4 and HLA class I in HIV-1 subtype C infection: Association with disease progression and influence of immune pressure. *Virology,* **468**, 214-225.

Mokgoro, M., Kilembe, W., Shumba, H., Kamusoko, M., **Reddy, T**., Dissen, E., Davitte, J., Mwaanga, A., Brockman, M., Ndungu, T. and Allen, S. (2014). Knowledge and Perception on HIV Discordance, Transmission and Prevention among Couples in Durban, South Africa. *AIDS Research and Human Retroviruses,* **30**(S1), A274-A274.

Naiker, S., Connolly, C., Wiesner, L., Kellerman, T., **Reddy, T**., Harries, A., McIlleron, H., Lienhardt, C. and Pym, A. (2014). Randomized pharmacokinetic evaluation of different rifabutin doses in African HIV-infected tuberculosis patients on lopinavir/ritonavir-based antiretroviral therapy. *BMC Pharmacology and Toxicology,* **15**, 61.

Abbai, N. S., Moodley, P., **Reddy, T**., Zondi, T. G., Rambaran, S., Naidoo, K. and Ramjee, G. (2015). Clinical Evaluation of the OneStep Gonorrhea RapiCard InstaTest for Detection of Neisseria gonorrhoeae in Symptomatic Patients from KwaZulu-Natal, South Africa. *Journal of Clinical Microbiology,* **53**, 1348-1350.

Bhorat, I. E., Bagratee, J. S., Pillay, M. and **Reddy, T**. (2015). Determination of the myocardial performance index in deteriorating grades of intrauterine growth restriction and its link to adverse outcomes. *Prenatal diagnosis,* **35**, 266-273.

Gounder, K., Padayachi, N., Mann, J. K., Radebe, M., Mokgoro, M., van der Stok, M., Mkhize, L., Mncube, Z., Jaggernath, M., **Reddy, T**.,. . . and Walker, B. D. (2015). High frequency of transmitted HIV-1 Gag HLA class I-driven immune escape variants but minimal immune selection over the first year of clade C infection. *PLoS One,* **10**.

Kilembe, W., Wall, K. M., Mokgoro, M., Mwaanga, A., Dissen, E., Kamusoko, M., Phiri, H., Sakulanda, J., Davitte, J., **Reddy, T**., Brockman, M., Ndungu, T. and Allen, S. (2015). Implementation of couples voluntary HIV counseling and testing services in Durban, South Africa. *BMC Public Health,* **15**, 601.

Kilembe, W., Wall, K. M., Mokgoro, M., Mwaanga, A., Dissen, E., Kamusoko, M., Phiri, H., Sakulanda, J., Davitte, J., **Reddy, T**., Brockman, M., Ndungu, T. and Allen, S. (2015). Knowledge of HIV serodiscordance, transmission, and prevention among couples in Durban, South Africa. *PloS One,* **10**.

Moolla, Y., Moolla, Z., **Reddy, T**. and Magula, N. (2015). The use of readily available biomarkers to predict CD4 cell counts in HIV-infected individuals. *South African Family Practice,* **57**, 293-296.

Naidoo, K., Grobler, A. C., Deghaye, N., **Reddy, T**., Gengiah, S., Gray, A. and Karim, S. A. (2015). Cost-effectiveness of initiating antiretroviral therapy at different points in TB treatment in HIV-TB co-infected ambulatory patients in South Africa. *Journal of Acquired Immune Deficiency Syndromes,* **69**, 576.

Taylor, M., Pretorius, B., Connolly, C., Dukhi, N., Drysdale, R., **Reddy, T**. and Reddy, J. (2015). Cross-sectional survey of dietary diversity among children aged 6-24 months in Kwazulu-natal, South Africa. *Tropical Medicine and International Health,* **20**, 289.

Abbai-Shaik, N. S., **Reddy, T**., Govender, S. and Ramjee, G. (2016). Poor performance of the Chlamydia Rapid Test device for the detection of asymptomatic infections in South African men: A pilot study. *Journal of Sexually Transmitted Diseases*.

Abbai, N. S., **Reddy, T**. and Ramjee, G. (2016). Prevalent bacterial vaginosis infection–a risk factor for incident sexually transmitted infections in women in Durban, South Africa. *International Journal of STD and AIDS,* **27**, 1283-1288.

Hennig, S., Naiker, S., **Reddy, T**., Egan, D., Kellerman, T., Wiesner, L., Owen, A., McIlleron, H. and Pym, A. (2016). Effect of SLCO1B1 polymorphisms on rifabutin pharmacokinetics in African HIV-infected patients with tuberculosis. *Antimicrobial Agents and Chemotherapy,* **60**, 617-620.

Mthiyane, T., Pym, A., Dheda, K., Rustomjee, R., **Reddy, T**. and Manie, S. (2016). Longitudinal assessment of health related quality of life of HIV infected patients treated for tuberculosis and HIV in a high burden setting. *Quality of Life Research,* **25**, 3067-3076.

Skhosana, M., Reddy, S., **Reddy, T**., Ntoyanto, S., Spooner, E., Ramjee, G., Ngomane, N., Coutsoudis, A. and Kiepiela, P. (2016). PIMA point-of-care testing for CD4 counts in predicting antiretroviral initiation in HIV-infected individuals in KwaZulu-Natal, Durban, South Africa. *Southern African Journal of HIV Medicine,* **17**, 1-8.

Street, R. A., **Reddy, T**. and Ramjee, G. (2016). The generational effect on age disparate partnerships and the risk for human immunodeficiency virus and sexually transmitted infections acquisition. *International Journal of STD and AIDS,* **27**, 746-752.

Bhorat, I. E., Bagratee, J. S. and **Reddy, T.** (2017). Assessment of fetal myocardial performance in severe early onset pre-eclampsia (EO-PET) with and without intrauterine growth restriction across deteriorating stages of placental vascular resistance and links to adverse outcomes. *European Journal of Obstetrics and Gynecology and Reproductive Biology,* **210**, 325-333.

Wand, H., **Reddy, T.**, Naidoo, S., Moonsamy, S., Siva, S., Morar, N.S. and Ramjee, G. (2017). A Simple Risk Prediction Algorithm for HIV Transmission: Results from HIV Prevention Trials in KwaZulu Natal, South Africa (2002-2012). *AIDS and Behavior,* 1-12.

Bhorat, I., Pillay, M. and **Reddy, T.** (2017). Determination of the Fetal Myocardial Performance Index in women with Gestational Impaired Glucose Tolerance and to assess whether this parameter is a possible prognostic indicator of adverse fetal outcome. *The Journal of Maternal-Fetal & Neonatal Medicine*, 1-20.

Akilimali, N.A., Chang, C.C., Muema, D.M., **Reddy, T.**, Moosa, M.Y.S., Lewin, S.R., French, M.A. and Ndungu, T. (2017). Plasma But Not Cerebrospinal Fluid Interleukin 7 and Interleukin 5 Levels Pre-Antiretroviral Therapy Commencement Predict Cryptococcosis-Associated Immune Reconstitution Inflammatory Syndrome. *Clinical Infectious Diseases.*

Mabuka, J.M., Dugast, A.S., Muema, D.M., **Reddy, T.**, Ramlakhan, Y., Euler, Z., Ismail, N., Moodley, A., Dong, K.L., Morris, L. and Walker, B.D. (2017). Plasma CXCL13 but not B Cell Frequencies in Acute HIV Infection Predicts Emergence of Cross-Neutralizing Antibodies. *Frontiers in Immunology,* **8**, 1104.

Contents

List of Publications and Reports i

List of Abbreviations xii

List of Figures xiii

List of Tables xv

1 Overview of the Dissertation 1

I Preliminary Materials 5

2 Motivating Examples 7

2.1 HIV/AIDS 8

2.1.1 Background 8

2.1.2 Sinikithemba Cohort Study 9

2.1.3 ACTG 315 Trial 10

2.2 Abdominal Aortic Aneursyms 11

2.2.1 Background 11

2.2.2 AAA Screening Study 12

2.3 Schizophrenia 12

2.3.1 Background 12
2.3.2 Schizophrenia clinical trial 13

3 Longitudinal Data Analysis 15
3.1 Linear Mixed Models 15
3.2 Generalized Linear Mixed Models 17
3.3 Models for Incomplete Longitudinal Data 18

II Contributions 21

4 Estimation of the Time to Two Consecutive Measurements Less Than a Threshold 23
4.1 Introduction 23
4.2 Brief Review of Time to Threshold Modeling 24
4.2.1 Standard Approach to Time to Threshold Modeling 24
4.3 Methodology 25
4.3.1 Expected Time to Attain a Threshold with Persistence Criteria . . 26
4.3.2 Estimation of Time to the First Measurement Less Than a Threshold of Interest 28
4.3.3 Estimation and Inference 28
4.4 Application: Sinikithemba HIV Cohort Study 29
4.4.1 Linear Mixed Model 29
4.4.2 Expected Time to Threshold 31
4.4.3 Sensitivity to Variation in Observation Frequency 36
4.5 Application: Abdominal Aortic Aneurysm (AAA) Study 38
4.5.1 Linear Mixed Model 38
4.5.2 Expected time to threshold 39
4.6 Discussion 44

5 Estimation of Time to Threshold Taking Complex Correlation Structures into Account 47
5.1 Introduction 47

5.2 Methodology . . . 48
5.2.1 Expected Time to Attain a Threshold with Persistence Criteria . . 48
5.2.2 Estimation and Inference . . . 50
5.3 Application: Sinikithemba HIV Cohort Study . . . 51
5.3.1 Linear Mixed Model . . . 51
5.3.2 Expected Time to Threshold . . . 54
5.4 Discussion . . . 57

6 Estimation of Time to Threshold for Biomarkers with Limits of Detection 59
6.1 Introduction . . . 59
6.2 Methodology . . . 60
6.3 Application: ACTG 315 Study . . . 61
6.3.1 Mixed-effects model . . . 61
6.3.2 Expected Time to Threshold . . . 63
6.4 Discussion . . . 69

7 Estimation of Time to Threshold for Ordinal Biomarkers 71
7.1 Introduction . . . 71
7.2 Methodology . . . 72
7.2.1 Expected Time to Attain a Threshold with Persistence Criteria . . 72
7.2.2 Estimation and Inference . . . 73
7.3 Application: Schizophrenia Trial . . . 74
7.3.1 Mixed-effects Model . . . 74
7.3.2 Expected Time to Threshold . . . 75
7.4 Discussion . . . 79

8 Sensitivity Analysis: Missing Data 81
8.1 Introduction . . . 81
8.2 Methodology . . . 82
8.3 Sensitivity Analysis: AAA Study . . . 83
8.3.1 Missing Data in the AAA Study . . . 83
8.3.2 Sensitivity Analysis: Linear Mixed Model Parameter Estimates . . 84

8.3.3 Sensitivity Analysis: Expected Time to Threshold 88
8.4 Discussion 92

9 Concluding Remarks and Future Work **93**

Bibliography **95**

Appendices **103**

A Appendix: Chapter 4 **105**
A.1 Proof of Expressions for C_{ij} and S_{ij} 105
A.2 SAS Program for Time to Threshold Estimation 106

B Appendix: Chapter 5 **111**
B.1 Proof of Expressions for $C_{ij,0}$, $C_{ij,1}$, and S_{ij} 111
B.2 SAS Program for Time to Threshold Estimation for a Model with Serial Correlation 113
B.3 Bias Corrected Accelerated (BCa) Bootstrap Confidence Intervals 120

C Appendix: Chapter 6 **123**
C.1 SAS Program for Time to Threshold Estimation for a Left Censored Outcome 123

D Appendix: Chapter 7 **127**
D.1 SAS Program for Time to Threshold Estimation for Ordinal Outcomes . 127

E Appendix: Chapter 8 **129**
E.1 SAS Program for MNAR Imputation 129

Samenvatting **135**

List of Abbreviations

AAA	Abdominal Aortic Aneursyms
ACTG	AIDS Clinical Trials Group
AIDS	Acquried Immunodeficiency Syndrome
ARV	Antiretroviral
ART	Antiretroviral therapy
BMI	Body Mass Index
CDC	Center for Disease Control
CGI	Clinical Global Impression
CI	Confidence Interval
DL	Direct Likelihood
GLMM	Generalized Linear Mixed Model
HIV	Human Immunodeficiency Virus
IQR	Interquartile Range
LMM	Linear Mixed Model
MAR	Missing at Random

MCAR	Missing Completely at Random
MI	Multiple Imputation
MNAR	Missing Not at Random
SE	Standard Error
VL	Viral Load
WHO	World Health Organization

List of Figures

2.1 *The time course of HIV infection.* 9

2.2 *Sinikithemba CD4 Study: Longitudinal CD4 count measurements for 8 subjects on the square root scale.* 10

2.3 *ACTG 315 Study: Longitudinal viral load measurements for 10 randomly selected patients.* 11

2.4 *AAA Study: Longitudinal abdominal aneursysm measurements.* 13

2.5 *Schizophrenia trial: Longitudinal CGI-S measurements.* 14

4.1 *Sinikithemba CD4 Study: Trajectory of CD4 count in each viral load category* 30

4.2 *Sinikithemba CD4 Study: Longitudinal CD4 count measurements with reference at thresholds 200 and 350 cells/mm^3.* 32

4.3 *Sinikithemba CD4 Study: Estimated probabilities of a single measurement being below the relevant threshold.* 32

4.4 *AAA Study: Longitudinal AAA measurements with reference at 50 mm and 55 mm for selected patients.* 40

4.5 *AAA Study: Estimated probabilities of a single measurement being above the relevant threshold.* 43

5.1 *Sinikithemba CD4 Study: Sample variogram of CD4 count residuals* . . . 52

5.2 *Sinikithemba CD4 Study: Longitudinal CD4 count measurements with reference at 350 cells/mm^3* 54
5.3 *Sinikithemba CD4 Study: Estimated probabilities computed for the selected patients* 55
5.4 *Sinikithemba CD4 Study: Bootstrap sampling distribution of $E(T_i)$ for the independence and AR(1) models* 56

6.1 *ACTG 315 Study: Viral load measurements over time with lowess smooth curve* 62
6.2 *ACTG 315 Study: Observed and fitted profiles* 64
6.3 *ACTG 315 Study: Viral load profiles for six selected patients* 65
6.4 *ACTG 315 Study: Estimated probability of single viral load being less than 1000 log copies/ml* 65
6.5 *ACTG 315 Study: Distribution of estimated time to confirmed treatment success* 66

7.1 *Schizophrenia Trial: Longitudinal CGI-S measurements for selected patients.* 76
7.2 *Schizophrenia Trial: Estimated probabilities, $P(Y_{ij} \leq 3)$, for selected patients.* 76
7.3 *Schizophrenia Trial: Estimated probabilities for 'treatment resistant' patients.* 78
7.4 *Schizophrenia Trial: Longitudinal CGI-S measurements for 'treatment resistant' patients.* 78

8.1 *AAA study: Comparison of mean AAA diameter between patients who did and did not dropout at the subsequent visit.* 84
8.2 *AAA study: Mean AAA diameter for each pattern of dropout.* 85
8.3 *Sensitivity analysis for AAA study: Profiles for patients with differences in $\widehat{T}_i$ exceeding 200%.* 89
8.4 *Sensitivity analysis for AAA study: Observed profiles for the 9 patients in Table 8.4.* 91

List of Tables

4.1 *Sinikithemba CD4 Study: Parameter estimates (standard errors) for the fitted models on each timescale.* 34

4.2 *Sinikithemba CD4 Study: Estimated time to threshold for patients A, B, C, and D.* 35

4.3 *Sinikithemba CD4 Study-Sensitivity analysis: Estimated time to two consecutive measurements less than 350 cells/mm^3 under various scenarios.* . . 37

4.4 *AAA Study: Parameter estimates (standard errors) for the fitted model , where NM, OW and OB denote BMI categories normal, overweight and obese, respectively.* 41

4.5 *AAA Study: Estimated time (years) to threshold for patients 6 selected patients under two definitions.* 42

4.6 *AAA Study: Estimated time (years) to 55 mm for the hypothetical 'median' patient at various ages and BMI categories.* 44

5.1 *Sinikithemba CD4 Study: Fit Statistics for the three models fitted* 52

5.2 *Sinikithemba CD4 Study: Parameter estimates (standard errors) for Model 1 and Model 3* 53

5.3 *Sinikithemba CD4 Study: Estimated time to threshold for patients A, B, C, D and E* 57

6.1 *ACTG 315 Study: Fit Statistics for the four models fitted* 63
6.2 *ACTG 315 Study: Parameter estimates (standard errors) for the standard model with imputed values and direct likelihood model* 67
6.3 *ACTG 315 Study: Estimated time to threshold for the six selected patients* 68

7.1 *Schizophrenia Trial: Parameter estimates (standard errors) for the random intercept (RI) and the random intercept and random slope (RI,RS) mixed effects proportion odds model.* 75
7.2 *Schizophrenia Trial: Estimated time to threshold for six patients.* 77

8.1 *AAA study: Patterns of missing data for AAA diameter.* 83
8.2 *Sensitivity analysis for AAA study: Parameter estimates (standard errors) for the fitted models.* 87
8.3 *Sensitivity analysis for AAA study: Summary of percentage difference in* $\widehat{T}_i$ *for each MNAR scenario.* 88
8.4 *Sensitivity analysis for AAA study:* $\widehat{T}_i$ *for 9 selected patients under direct likelihood (DL) and the four MNAR scenarios.* 90

A.1 *Possible combinations of outcomes which result in continuation after 2, 3, and 4 visits* 110

B.1 *Possible combinations of outcomes which result in termination after 5, 6, and 7 visits* 121

Chapter 1

Overview of the Dissertation

Biomarkers play a key role in medical research, guiding diagnosis and when the physician should intervene. However, several biomarkers are subject to a high degree of measurement error and fluctuation, which raises concerns about the validity of basing clinical decisions on a single measurement. Studies have found that the diagnostic accuracy of consecutive biomarkers performed better than the usage of a single elevated value. Several studies have applied persistence criteria, designating the outcome as the time to the occurrence of two consecutive measurements less than (or greater than) the threshold (Amornkul et al., 2013; Zhang, 2015).

In Chapter 2 we introduce four studies where time to threshold estimation is of key interest, emanating from the fields of HIV/AIDS research, cardiology and psychiatry. In Chapter 3 we discuss the basics of longitudinal data analysis, with a particular emphasis on random effects models. This material serves as the key building block for the novel approach to time to threshold estimation that is proposed in Chapter 4.

In Chapter 4 we propose a method for estimation of the time to threshold in the presence of persistence criteria, using a two-stage approach. In the first stage, a linear mixed model is fitted to the longitudinal measurements, resulting in patient-specific predicted values that are a function of the fixed-effects and empirical Bayes estimates. In the second stage, the probability of experiencing two consecutive measurements less than a relevant threshold k at each time point is computed and substituted into the expression

for the expected time to threshold. Through identification of a recursive relationship of the continuation probabilities at each time point, we show that the computation of the expected times is simple, efficient, and can be implemented using existing software packages. We apply this approach to two studies and conduct sensitivity analysis to determine whether the methodology is robust to deviations from the assumptions made. A possible shortcoming of the approach presented is that it is confined to situations where, conditional on random effects, the residuals are independent (i.e., the assumption of conditional independence).

In Chapter 5 we extend the methodology proposed in Chapter 4 so that serial correlation can be accommodated. Assuming that the Markov property holds, and applying the chain rule of probabilities we prove that the probability of progression at each timepoint can be expressed simply as the product of conditional probabilities. The methodology is applied to a cohort of HIV positive individuals, where the time to reach a CD4 count threshold is estimated. To gauge the impact of erroneously ignoring serial correlation, we compare the estimates from the approach of Chapter 4 to that of the extended approach and find that incorrectly modeling the correlation structure can result in substantial overestimation of the time to threshold.

In Chapter 6 we consider biomarkers that are subject to limits of detection or censoring, and extend the methodology previously developed. After incorporation of the censoring into the likelihood function, we find that the methodology proposed in Chapter 4 can therafter be applied. In this chapter we apply the methodology to viral load measurements from patients enrolled in the ACTG 315 study, and estimate the time to treatment success.

In Chapters 4, 5 and 6 the methodology was developed for continuous biomarkers which is somewhat restrictive. It may be the case that health status is measured as an ordinal variable. as is frequently observed in the field of psychiatry. In Chapter 7 we propose a method, based on the material presented in Chapter 4, to estimate the time to threshold of an ordinal 'biomarker'. This methodology is applied to data from a schizophrenia trial, where time to remission was of interest.

The methodology for time to threshold estimation that was proposed in Chapters 4, 5, 6, and 7, rely on the assumption that the missing data mechanism is ignorable. In Chapter 8 we assess the sensitivity of estimated times to threshold to deviation from the

missing at random assumption.

We conclude by summarizing the key developments in each chapter and discuss areas of further work.

Part I

Preliminary Materials

Chapter 2

Motivating Examples

In this chapter we introduce four longitudinal studies that will be analyzed in various chapters of this thesis. The Sinithemba study (Section 2.1.2) was a study that enrolled HIV positive individuals in South Africa to explore factors that affect HIV progression in individuals that have not been exposed to antiretroviral (ARV) treatment. The ACTG 315 trial (Section 2.1.3), also in the field of HIV, examined individuals' responses to ARV therapy in the first year of treatment initiation. The third study, introduced in Section 2.2.2, is from the field of cardiology and studies the progression of Abdominal Aortic Aneursyms (AAAs) in a cohort of individuals from the Netherlands. The fourth study that will be analysed is a randomized clinical trial that sought to evaluate the efficacy of Risperidone in treating individuals with chronic schizophrenia (Section 2.3.2). A unifying characteristic of the four studies is that in each study a biomarker threshold is of key interest, as it is this threshold upon which decisions regarding therapeutic or surgical interventions are often based. Each study poses a unique challenge, and motivates the various extensions of the standard methodology that follow in subsequent chapters.

The Sinikithemba cohort study (Section 2.1.2) presents several analytical challenges: Firstly, the date of HIV infection is unknown and individuals present at varying stages of the disease, as evidenced by the distribution of CD4 counts at enrolment. Hence careful consideration needs to be given to the timescale on which the data are modelled. Further, in analyses involving time to reach a relevant CD4 count threshold, sufficient thought has

to be given to how to analyse patients who may have already entered the study with a CD4 count below the threshold of interest. This dataset is analyzed again in Chapter 5, where the standard methodology is extended to accommodate serial correlation. The AAA study (Section 2.2.2) is analyzed in Chapter 4 where prediction of time to threshold, based on baseline covariates, was of interest. This study was also characterized by a high degree of dropout, and serves as the case study for sensitivity analysis in Chapter 8. In the second HIV study (Section 2.1.3), the outcome of interest (viral load) is subject to a lower limit of detection, which motivated the extension of the methodology in Chapter 6. Finally, the ordinal outcome in the schizophrenia study (Section 2.3.2) motivated the extension of the methodology from continuous to categorical outcomes that follows in Chapter 7.

2.1 HIV/AIDS

2.1.1 Background

At the end of 2015 it was estimated that 36.7 million people were living with HIV/AIDS globally, 1.8 million of whom were children. The most affected is Sub-Saharan Africa, which accounts for 70% of the global burden of HIV. Surrogate markers for clinical events are particularly useful in the study of HIV progression due to the long natural history of the disease. Peto (1996) lists the key surrogate markers of HIV progression as CD4 count, CD8 count, beta 2 microglobulin, neopterin, plasma HIV RNA load (viral load) and clinical symptoms. CD4 cells, which are responsible for immune response to infections are the primary target of the virus. The CD4 count is cited as the most significant predictor of AIDS related illness and death (Langford et al., 2007). Viral load, which quantifies the level of HIV in blood, is the most widely used surrogate for treatment efficacy. During the acute phase of infection HIV replicates and CD4 cells deplete rapidly after which a steady increase and decrease in viral load and CD4 count, respectively, is observed. The evolution of CD4 count and viral load post HIV infection is presented in Figure 2.1 which was adapted from Fauci et al (1996). Both CD4 count and viral load measurements are subject to a high degree of within-patient variability (Hughes et al., 1994), raising doubt regarding the validity of basing clinical decisions on a single value. A study of variability

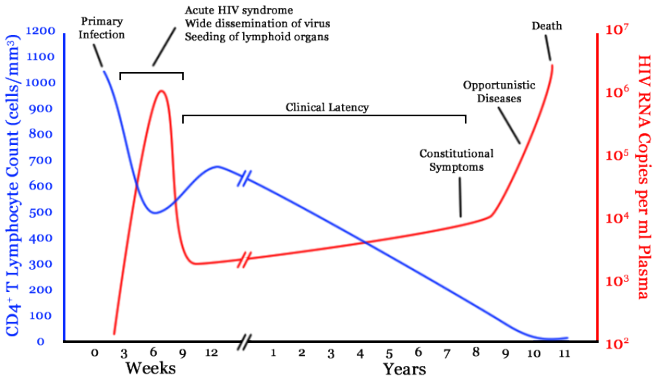


Figure 2.1: *The time course of HIV infection.*

of CD4 count in patients revealed that measurements taken eight weeks apart differed by more than 20 percent (Hughes et al., 1994). Malone et al. (1990) and Crowe et al. (1996) attribute this variability in HIV biomarkers to several factors including diurnal variation, measurement error, psychological and physical stress, diet and the menstrual cycle. This is the motivation behind recent studies on HIV progression defining progression based on *two* consecutive CD4 counts less than 350 cells/mm^3 (Amornkul et al., 2013; Zhang, 2015), and WHO guidelines defining treatment success as the attainment of two consecutive viral load measurements less than 1000 copies/ml (WHO, 2015). In this thesis we examine two datasets from the area of HIV/AIDS. The first dataset (Section 2.1.2) emanates from a cohort study of ARV naïve HIV positive individuals where the outcome of interest was CD4 count measurements. The second study is a clinical study that followed up individuals from the date of treatment initiation, where the outcome of interest was viral load.

2.1.2 Sinikithemba Cohort Study

The Sinikithemba cohort comprises 450 HIV-1 subtype C chronically infected adults enrolled at the McCord Hospital (Durban, South Africa) between August 2003 and 2008. Sociodemographic characteristics, plasma viral load and CD4 count measurements were obtained at baseline. The CD4 count and viral load were measured every 3 and 6 months, respectively, from enrollment. Viral loads were determined using the automated CobasAmplicor HIV-1 Monitor test (version 1.5; Roche Diagnostics). CD4 cells were enumerated

using the Multitest kit (CD4/CD3/CD8/CD45) on a FACSCalibur flow cytometer (Becton Dickinson). In accordance with the national HIV treatment guidelines implemented during the study period, patients were recommended to start ART upon reaching a CD4 count less than 200 cells/mm^3 or WHO stage 3 or 4 symptoms (DOH, 2010). For the purposes of this particular application, 114 patients who had not returned for a subsequent CD4 count measurement after enrolment or who were not confirmed ARV naïve at study entry were excluded from the analysis, resulting in a cohort of 336 patients. The median CD4 count at enrollment was 357 (Inter-quartile range: 259-509) cells/mm^3 and the mean viral load was 4.7 log copies/ml. The overall mean age at enrolment was 33 years and 80% of the patients were female. A variance stabilizing square root transformation was applied to the CD4 count responses. The observed CD4 count trajectories for 8 selected patients is presented in Figure 2.2.

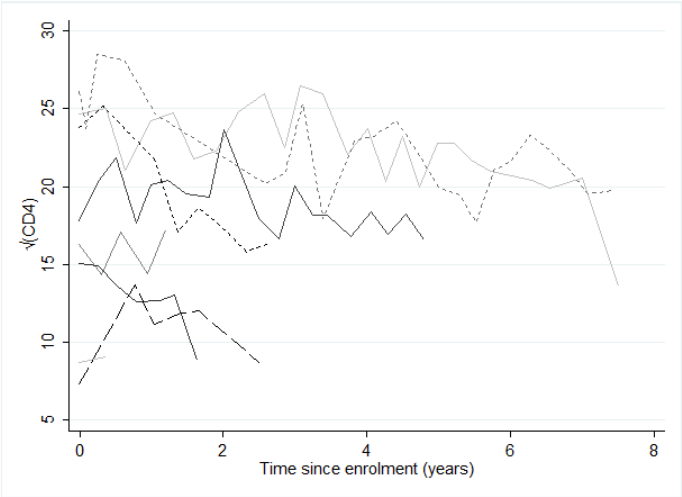


Figure 2.2: *Sinikithemba CD4 Study: Longitudinal CD4 count measurements for 8 subjects on the square root scale.*

2.1.3 ACTG 315 Trial

The ACTG 315 study, conducted by the Aids Clinical Trials Group (ACTG) was designed to evaluate the treatment combination of ritonavir, 3TC and AZT (Wu, 1999). A total of

53 HIV positive patients were treated with the regimen. Viral load was measured on days 0, 2, 7, 10, 14, 21, and 28, and weeks 8, 12, 24, and 48 after initiation of the treatment. The Nucleic Acid Sequence-Based Amplification assay (NASBA) was used to quantify viral load and was subject to a lower detection limit of 100 copies/ml. The viral load trajectories on the log scale for 10 randomly selected patients is presented in Figure 2.3. For exploratory purposes, viral load measurements less than 100 copies/ml were imputed by half the limit of detection i.e. 50 copies/ml≈1.70 log copies/ml. It is clear that the response to treatment is biphasic or possibly triphasic, with the most rapid decline in viral load occurring within the first 14 days of treatment.

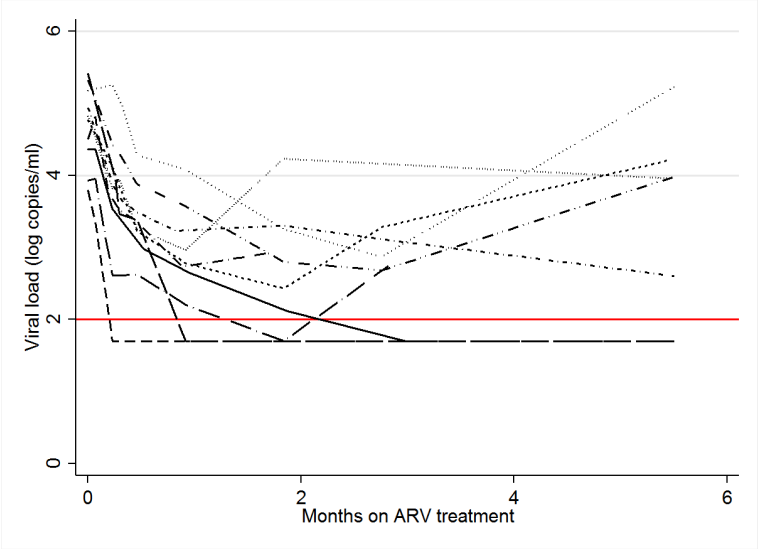


Figure 2.3: *ACTG 315 Study: Longitudinal viral load measurements for 10 randomly selected patients.*

2.2 Abdominal Aortic Aneursyms

2.2.1 Background

An abdominal aortic aneurysm (AAA) is an enlarged area in the lower part of the aorta, which is the vessel that supplies blood to the body, such that the diameter of the aorta is

greater than 30 mm or more than 50% larger than a normal diameter. As the size of the aneurysm increases, so too does the risk of aortic rupture. Since the aorta is the body's key supplier of blood, a ruptured abdominal aortic aneurysm can cause life-threatening bleeding. In abdominal aortic aneurysm (AAA) screening studies, surgery is recommended to patients when the diameter of the aneurysm exceeds 55 mm (Sweeting and Thompson, 2012). The factors that are known to have an effect on AAA diameter include age, smoking status, gender, and weight.

2.2.2 AAA Screening Study

In 2006, Maastricht University Medical Center (MUMC) started a follow-up study in patients with an abdominal aortic aneurysms. Between January 2006 and January 2009, all patients admitted with AAA to the department of Vascular Surgery of the MUMC were invited to participate in the study. Patients that had an aneursym exceeding 55 mm or symptoms of imminent AAA rupture and patients with either an inflammatory or a mycotic aneurysm were excluded from the follow-up study. Patients with an aneurysm diameter between 30 and 55 mm ($n = 110$) were invited to participate in an imaging surveillance program. A total of 100 patients formally entered the follow-up program, and were seen every six months.The median age of patients was 71 years (Interquartile range 65-77) and 80% of the patients were male. Using the CDC classification patients were classified as being normal weight, overweight or obese based on their BMI (body mass index). This resulted in a total of 31, 49 and 20 patients in the categories normal weight, overweight and obese, respectively. The AAA diameter profiles for a random sample of patients are presented in Figure 2.4.

2.3 Schizophrenia

2.3.1 Background

Schizophrenia is a severe mental disorder characterized by abnormal social behavior and affects more than 21 million people worldwide. The disease is characterized by distortions in thinking, perception, emotions, language, sense of self and behaviour. In schizophrenia

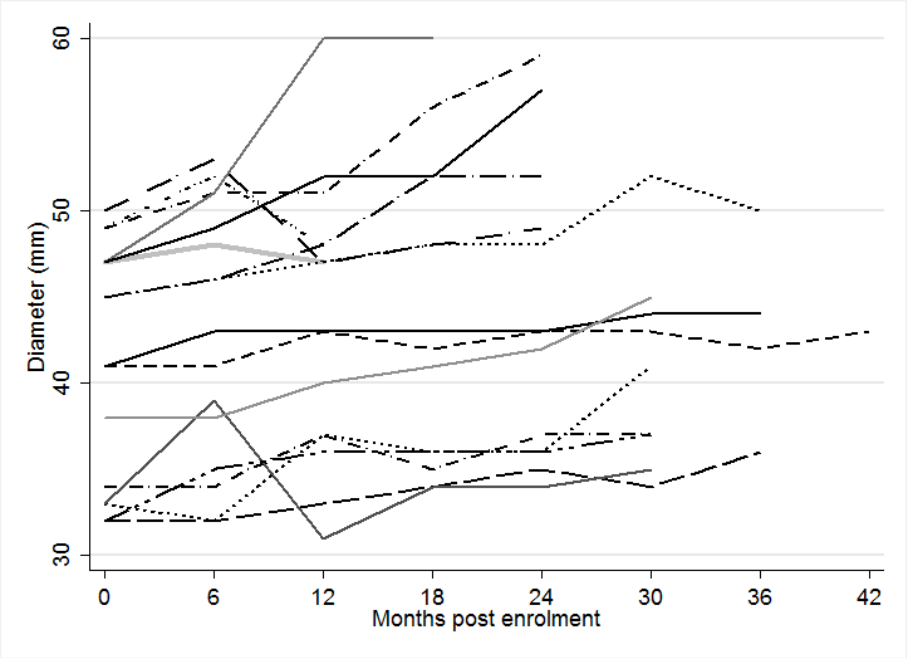


Figure 2.4: *AAA Study: Longitudinal abdominal aneursysm measurements.*

and other mental disorders, the Positive and Negative Syndrome Scale (PANSS), the Brief Psychiatric Rating Scale (BPRS) and the Clinical Global Impression (CGI) are often used as a marker of the patient's condition. For the purposes of this thesis we focus on the CGI score. The CGI-S is the component of the CGI scale that measures severity ranging from 1(not at all ill) to 7(extremely ill).

2.3.2 Schizophrenia clinical trial

A multicentre double-blind trial was conducted to evaluate the efficacy of the drug Risperidone compared to Haloperidol in treating patients with chronic schizophrenia. A total of 1632 patients were randomly assigned to risperidone 1, 4, 8, 12 or 16 mg or haloperidol 10 mg daily for 8 weeks (Peuskens, 1995). We restrict our analysis to the patients who received haloperidol 10 mg or risperidone in doses ranging from 4 to 6 mg, as this is known to be the effective dose in most countries. This resulted in a total of 453 patients being included. PANSS, BPRS and CGI-S measurements were taken at baseline and 1, 2, 4, 6

and 8 weeks post randomization. The median CGI-S score at baseline was 5 (Interquartile range 4-6). The CGI-S scores for a random sample of patients are presented in Figure 2.5.

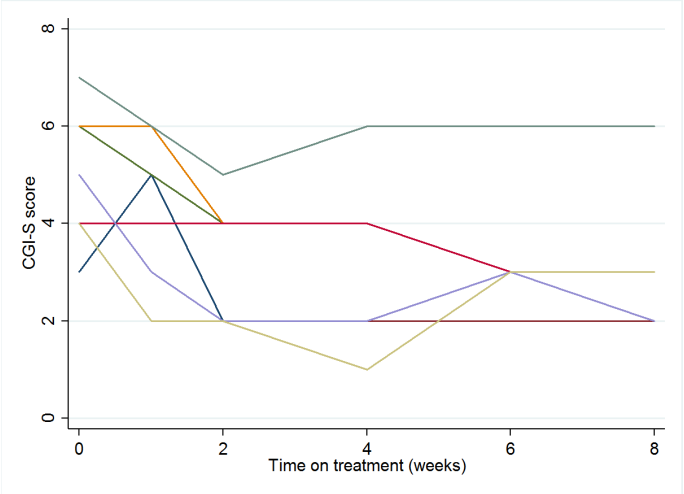


Figure 2.5: *Schizophrenia trial: Longitudinal CGI-S measurements.*

Chapter 3

Longitudinal Data Analysis

In Chapter 2 we introduced four studies, all of which were longitudinal i.e. there were repeated measurements taken on each individual. The exploratory analysis presented illustrated that patients tend to display different evolutions, and that observations within an individual tend to be correlated with fluctuation present. These artifacts can be formally defined as between-subject variability and within-subject variability, respectively. To ensure that valid inference is drawn it is crucial that all sources of variability are taken into account. Mixed effects models have been widely applied in the longitudinal data setting, and form a key 'building block' in the methodology that follows in Chapters 4 to 7. In Section 3.1 we discuss linear mixed models and Section 3.2 we discuss generalized linear mixed models for categorical data. Naturally, an issue that arises in follow-up studies is that patients may have incomplete data possibility due to dropout. In Section 3.3 we briefly discuss models for incomplete data.

3.1 Linear Mixed Models

Letting Y_{ij} denote the outcome observed on individual i at time point j, $i = 1, \ldots, N$, and $j = 1, \ldots, n_i$, and letting $\boldsymbol{Y}_i$ denote the vector of all measurements for subject i,

such that $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$ the general form of the linear mixed model is:

$$\begin{aligned} \boldsymbol{Y}_i &= X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \\ \boldsymbol{b}_i &\sim N(\mathbf{0}, D). \end{aligned} \tag{3.1}$$

The within-patient variation $\boldsymbol{\varepsilon}_i$ can be expressed as a sum of measurement error and serial variation, which will be denoted by $\boldsymbol{\varepsilon}_{i(1)}$ and $\boldsymbol{\varepsilon}_{i(2)}$, respectively.

$$\begin{aligned} \boldsymbol{\varepsilon}_{i(1)} &\sim N(\mathbf{0}, \sigma^2 I_{n_i}), \\ \boldsymbol{\varepsilon}_{i(2)} &\sim N(\mathbf{0}, \tau^2 H_i), \end{aligned}$$

where $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \boldsymbol{\varepsilon}_{1(1)}, \ldots, \boldsymbol{\varepsilon}_{N(1)}, \boldsymbol{\varepsilon}_{1(2)}, \ldots, \boldsymbol{\varepsilon}_{N(2)}$ are independent (Verbeke and Molenberghs, 2009). $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$ represent the fixed and random effects respectively. The $(j,k)^{th}$ element of H_i is $\rho(|t_{ij} - t_{ik}|)$, which is the correlation between $\varepsilon_{ij(2)}$ and $\varepsilon_{ik(2)}$. It follows that

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i, \Sigma_i = \sigma^2 I_{n_i} + \tau^2 H_i). \tag{3.2}$$

The marginal distribution is obtained by integrating over the random effects $f(\boldsymbol{y}_i) = \int f(\boldsymbol{y}_i|\boldsymbol{b}_i) f(\boldsymbol{b}_i) d\boldsymbol{b}_i$, with $f(\boldsymbol{y}_i|\boldsymbol{b}_i)$ the density function of $\boldsymbol{Y}_i$ conditional on $\boldsymbol{b}_i$, and $f(\boldsymbol{b}_i)$ the density function of $\boldsymbol{b}_i$. As a result the marginal distribution of $\boldsymbol{Y}_i$ is given by the density of the n_i-dimensional normal distribution $N(X_i\boldsymbol{\beta}, Z_i D Z_i^{'} + \Sigma_i)$.

When $\Sigma_i = \sigma^2 I_{n_i}$ the model specified in (3.2) is called the conditional independence model, since it implies that conditional on the random effects $\boldsymbol{b}_i$ the measurements on individual i are independent. The parameters that form $V_i = Z_i D Z_i^{'} + \Sigma_i$ are often grouped in a vector $\boldsymbol{\alpha}$ of variance-covariance parameters, such that $\boldsymbol{\theta}^{'} = (\boldsymbol{\beta}^{'}, \boldsymbol{\alpha}^{'})$ denotes the vector of all parameters in the marginal model for $\boldsymbol{Y}_i$. The following marginal likelihood function then needs to be maximized with respect to $\boldsymbol{\theta}$:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} (2\pi)^{-\frac{n_i}{2}} |V_i(\boldsymbol{\alpha})|^{-\frac{1}{2}} \times \exp[-\frac{1}{2}(\boldsymbol{Y}_i - X_i\boldsymbol{\beta})^{'} V_i^{-1}(\boldsymbol{\alpha})(\boldsymbol{Y}_i - X_i\boldsymbol{\beta})].$$

When subject-specific prediction is of key interest, estimates for the random effects $\boldsymbol{b}_i$ are required. In Bayesian terminology, the distribution of $\boldsymbol{b}_i$ is called the prior distribution since it does not depend on the observed data $\boldsymbol{Y}_i$. The posterior distribution of $\boldsymbol{b}_i$, given

the data $\boldsymbol{Y}_i$ is given by

$$f(\boldsymbol{b}_i|\boldsymbol{y}_i) \equiv f(\boldsymbol{b}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i) = \frac{f(\boldsymbol{y}_i|\boldsymbol{b}_i)f(\boldsymbol{b}_i)}{\int f(\boldsymbol{y}_i|\boldsymbol{b}_i)f(\boldsymbol{b}_i)d\boldsymbol{b}_i}.$$

A point estimator for $\boldsymbol{b}_i$ is given by the posterior mean of this posterior density function:

$$\begin{aligned}\hat{\boldsymbol{b}_i}(\boldsymbol{\theta}) &= E[\boldsymbol{b}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i] \\ &= \int \boldsymbol{b}_i f(\boldsymbol{b}_i|\boldsymbol{y}_i)d\boldsymbol{b}_i \\ &= DZ_i^{'}V_i^{-1}(\boldsymbol{\alpha})(\boldsymbol{Y}_i - X_i\boldsymbol{\beta}).\end{aligned}$$

The empirical Bayes (EB) estimates ($\hat{\boldsymbol{b}_i}$) are obtained by replacing the parameters in $\boldsymbol{\theta}$ by their maximum likelihood estimates. Further details on exploratory analysis for longitudinal data, model building strategies and types of covariance structures can be found in Verbeke and Molenberghs (2009) and Diggle (2013).

3.2 Generalized Linear Mixed Models

A key difference between LMMs and GLMMs is that in GLMMs the response variables can come from distributions other than the Gaussian distribution. In addition, rather than modeling the responses directly, a link function is often applied. Generalized linear mixed models (GLMMs) are widely applied in the analysis of longitudinal categorical data. Using the notation as introduced in the previous section, it is assumed that conditional on the random effects, the elements of $\boldsymbol{Y}_i$ are independent and belong to the exponential family density

$$f_i(y_{ij}|\boldsymbol{b}_i, \phi) == \exp\{\phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi)\}.$$

It follows that $E(Y_{ij}|\boldsymbol{b}_i) = \mu_{ij} = \eta^{-1}(x_{ij}^{'}\boldsymbol{\beta} + z_{ij}^{'}\boldsymbol{b}_i)$ where $\eta(.)$ is a known link function.In this formulation x_{ij} and z_{ij} are vectors containing covariate values, ϕ represents a scale parameter and θ_{ij} the natural or canonical parameter(Molenberghs and Verbeke, 2005). As described for the LMM setting in Section 3.1, inference is based on the marginal model for $\boldsymbol{Y}_i$ which is obtained from integrating out the random effects. The likelihood can be

expressed as

$$\begin{aligned} L(\boldsymbol{\beta}, D, \phi) &= \prod_{i=1}^{N} f(\boldsymbol{y}_i|\boldsymbol{b}_i, D, \phi) \\ &= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{b}_i, D, \phi) f(\boldsymbol{b}_i|D) d\boldsymbol{b}_i. \end{aligned}$$

However, in contrast to the LMM setting, in the GLMM setting no analytic expressions are available for the likelihood and numerical approximations are needed. Further details regarding numerical approximation methods can be found in Molenberghs and Verbeke (2005). In this thesis we have a specific interest in models for ordinal data and therefore direct our attention to mixed models for ordinal outcomes for the remainder of the chapter. Proportional odds models are commonly used to analyze ordinal data. The proportional odds model expresses the ordinal responses in C categories in terms of $C-1$ cumulative category comparisons, specifically, $C-1$ cumulative logits (i.e., log odds). The random effects proportional odds model takes the form

$$\text{logit}[P(Y_{ij} \leq k|X_i, Z_i, \boldsymbol{b}_i)] = \alpha_c + x_{ij}^{'}\boldsymbol{\beta} + z_{ij}^{'}\boldsymbol{b}_i, \tag{3.3}$$

where α_c is the cutpoint-specific intercept and $c = 1, 2, \ldots, C-1$. Other models for ordinal data include continuation ratio models and the multigroup logistic model.

3.3 Models for Incomplete Longitudinal Data

Allowing for the possibility of missing data, the full data $\boldsymbol{Y}_i$ vector can be partitioned into the observed ($\boldsymbol{Y}_i^o$) and missing ($\boldsymbol{Y}_i^m$) components, and the dropout indicator D_i for the occasion at which dropout occurs can be defined. The full-data density is then expressed as

$$f(\boldsymbol{y_i}, \boldsymbol{d_i}|\boldsymbol{\theta}, \boldsymbol{\psi}), \tag{3.4}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are parameter vectors associated with the measurement process and the missingness process, respectively. By factorizing this full data density, different modeling frameworks for incomplete longitudinal data emerge. This results in selection, pattern-mixture and shared-parameter models. The selection model framework is based on the

factorization

$$f(\boldsymbol{y_i}, \boldsymbol{d_i}|\boldsymbol{\theta}, \psi) = f(\boldsymbol{y_i}|\boldsymbol{\theta}) f(\boldsymbol{d_i}|\boldsymbol{y_i}, \psi). \tag{3.5}$$

In contrast, pattern-mixture models are based on the reverse factorization of the full-data density:

$$f(\boldsymbol{y_i}, \boldsymbol{d_i}|\boldsymbol{\theta}, \psi) = f(\boldsymbol{y_i}|\boldsymbol{d_i}, \boldsymbol{\theta}) f(\boldsymbol{d_i}|\psi). \tag{3.6}$$

The pattern-mixture model, proposed by Little (1993) specifies a different measurement model for each pattern of missing values, and the full-data density is obtained as the mixture of the models weighted by the probability of each missing value pattern. Shared-parameter models (Wu and Caroll, 1988) assume the existence of latent (unobserved) variables, that are shared between both components of the full-data density. In this setting conditional independence is often assumed, i.e., the measurement process and missingness process are independent conditional on the latent variable.

A non-response process is said to be missing completely at random (MCAR) if missingness is independent of both unobserved and observed data and missing at random (MAR) if, conditional on the observed data, missingness is independent of the unobserved measurements. A process that is neither MCAR nor MAR is termed non-random (MNAR). When missingness is MAR and when $\boldsymbol{\theta}$ and ψ are functionally distinct, the missing data mechanism is said to be ignorable because it does not have to be modeled in order to draw inferences about $\boldsymbol{\theta}$.

Multiple imputation (MI), which was formally introduced by Rubin (1978), is another approach to handling missing data. In MI each missing value is replaced by a list of M plausible values, generated by an imputation model which is estimated based on observed data. Each imputed dataset is then analyzed separately using an analysis model, and results are combined into a single measure using Rubin's rules.

Part II

Contributions

Chapter 4

Estimation of the Time to Two Consecutive Measurements Less Than a Threshold

4.1 Introduction

Biomarkers play a key role in medical research, guiding diagnosis and when the physician should intervene. However, several biomarkers are subject to a high degree of measurement error and fluctuation, which raises concerns about the validity of the usage of a single value to guide decisions. Studies have found that the diagnostic accuracy of consecutive biomarkers performed better than the usage of a single elevated value. Due to the inherent variability of HIV biomarkers, some authors have defined the endpoint in studies using persistence criteria i.e., true decline has occurred when two consecutive measurements below (or above) the threshold of interest are observed (Amornkul et al., 2013; Zhang, 2015). The application of persistence criteria to define true progression or recovery is not limited to the field of HIV. In diabetes screening, diagnosis is only confirmed when a second HbA1c measurement exceeding 6.5% is observed (NIH, 2014). Similarly, in the field of prostate cancer, the diagnosis of treatment failure after radical prostatectomy is

defined by two consecutive measurements of prostate-specific antigen (PSA) greater than 0.20 ng/ml (Heidenreich et al., 2014).

In Section 4.2 we review the standard and model-based methods that are currently used to estimate time to threshold and briefly discuss the shortcomings of these approaches. In Section 4.3 we introduce a novel approach to estimate the time to reach two consecutive measurements less than (or greater than) a threshold. This approach, which was motivated by time to threshold modelling in the HIV setting, is applied in Section 4.4.1 to the Sinikithemba HIV cohort study. We also apply the methodology to the AAA study in Section 4.5, and discuss prediction based on baseline covariates.

4.2 Brief Review of Time to Threshold Modeling

4.2.1 Standard Approach to Time to Threshold Modeling

The time taken for a biomarker to reach specific threshold has been analysed as an outcome in several recent studies. The convention in such studies is to first extract the time to the event, which is analysed in a second stage within the survival analysis framework. Cardeal da Silva et al. (2013) analysed the time to first CD4 count less than 350 cells/mm^3 as the primary outcome in their study, which compared the rate of HIV progression pre and post the introduction of antiretroviral therapy (ART). Amornkul et al. (2013) studied a cohort of recently infected individuals where the effect of HIV subtype on HIV progression was examined. In this study, immunologic progression was defined as time from seroconversion to the first of two consecutive CD4 cell counts less than or equal to 350 cells/mm^3. Imposing persistence criteria such as a 'two consecutive' rule is an improvement on basing clinical decisions on a single CD4 count, but can be unreliable when the time between visits is large. In addition, doubt arises in the classification of patients who enter the study with a CD4 count already below the relevant threshold, which is frequently observed in seroprevalent cohorts. Removing these patients from the analysis results in left truncation and biased inferences.

4.2.1.1 General Model-based Approaches

A criticism of the standard approach discussed above is that it ignores the inherent subject-specific biomarker trajectory, and assumes that the event times are observed without error. Model-based approaches to the time to threshold of a biomarker have recently emerged. One such approach is inverse estimation or calibration. Sweeting and Thompson (2012) examined the time for subjects enrolled in the 'Multicentre aneurysm screening study' to reach an aneurysm diameter of 55 mm. The researchers used the method of inverse estimation in linear and quadratic subject-specific curves from a Bayesian hierarchical model. A limitation of the inverse prediction approach is that complex functions of time cannot easily be included.

In cases where the interest lies in modeling the time to threshold of an ordinal variable, continuous time Markov models have proven to be useful. This involves forming a distinct set of states and computing the mean first hitting time to a particular state. Mandel (2010) added to this methodology by studying the first hitting time to a state followed by a fixed duration of stay in the state. This was applied to a study on multiple sclerosis where sustained progression based on a disability scale was of interest. When the outcome of interest is a continuous biomarker, the construction of discrete states is somewhat arbitrary. Furthermore, due to the high degree of variability of the biomarker, reverse transitions and transitions that skip intermediate states are often observed (Reddy et al., 2011).

4.3 Methodology

In this section we propose an approach to time to threshold modelling that involves two stages. In the first stage, a linear mixed model is fitted to the longitudinal measurements, resulting in patient-specific predicted values that are a function of the fixed-effects and empirical Bayes estimates. In the second stage, the probability of experiencing two consecutive measurements less than a relevant threshold k at each time point is computed and substituted into the expression for the expected time to threshold.

4.3.1 Expected Time to Attain a Threshold with Persistence Criteria

For readability, we will refer to the attainment of two consecutive measurements less than the threshold, as the *event* of interest. In the approach we propose, we consider the individual 'at risk' for the event both prior to and post enrolment. Letting Y_{ij} denote the outcome observed on individual i at time point j , the time to event T_i can be expressed as:

$$T_i = \min\{j \geq 2 : Y_{ij-1} \leq k, Y_{ij} \leq k\}. \tag{4.1}$$

It follows that the expected time for individual i to attain two consecutive measurements less than the threshold k can be expressed as follows:

$$\begin{aligned} E(T_i) &= t_{i2}P(Y_{i1} \leq k, Y_{i2} \leq k) \\ &\quad + t_{i3}P(Y_{i1} > k, Y_{i2} \leq k, Y_{i3} \leq k) \\ &\quad + t_{i4}\left\{\begin{array}{c} P(Y_{i1} > k, Y_{i2} > k, Y_{i3} \leq k, Y_{i4} \leq k) \\ + P(Y_{i1} \leq k, Y_{i2} > k, Y_{i3} \leq k, Y_{i4} \leq k) \end{array}\right\} \\ &\quad + \ldots \\ &= \sum_{j=2}^{\infty} t_{ij}S_{ij}, \end{aligned} \tag{4.2}$$

where t_{ij} represents the time corresponding to the j^{th} visit for individual i, and S_{ij} denotes the probability of individual i experiencing the event, or 'stopping', at t_{ij}. In practice the infinite series may be truncated at a time point considered relevant to the specific application at hand. Possible options for the time at which the series is truncated are the expected lifetime of an individual, or the end of the incubation period of a particular disease. We specify a linear mixed model, which satisfies

$$\begin{aligned} \boldsymbol{Y}_i &= X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i, \\ \boldsymbol{b}_i &\sim N(\mathbf{0}, D), \\ \boldsymbol{\varepsilon}_i &\sim N(\mathbf{0}, \Sigma_i), \end{aligned} \tag{4.3}$$

where $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \varepsilon_i, \ldots, \varepsilon_N$ are independent. $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$ represent the fixed and random effects, respectively (Verbeke and Molenberghs, 2009). It follows that

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i, \Sigma_i).$$

As presented in (4.2), S_{ij} is the sum of several joint probabilities, each of which represents a distinct combination of the values of Y_{ij} that may yield the event at time point j. Assuming conditional independence in (4.3) such that $\Sigma_i = \sigma^2 I_{n_i}$, the joint probabilities that form S_{ij} reduce to the product of the individual probabilities. Hence, S_{ij} may be simplified as follows:

$$\begin{aligned} S_{ij}(X_i, Z_i, \boldsymbol{b}_i, \boldsymbol{\beta}) &= C_{ij-3}P(Y_{ij-2} > k)P(Y_{ij-1} \leq k)P(Y_{ij} \leq k) \\ &= C_{ij-3}[1 - \tilde{\Phi}_{ij-2}(k)][\tilde{\Phi}_{ij-1}(k)][\tilde{\Phi}_{ij}(k)], \end{aligned}$$

where C_{ij-3} denotes the 'continuation probability' at time t_{ij-3} and $\tilde{\Phi}_{ij}(k)$ is a cumulative normal distribution with mean $\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i$ and variance σ^2. It follows that $\tilde{\Phi}_{ij}(k)$ can be expressed as a simple function of the standard univariate normal distribution:

$$\tilde{\Phi}_{ij}(k) = \Phi\left(\frac{k - \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \boldsymbol{z}'_{ij}\boldsymbol{b}_i}{\sigma}\right). \tag{4.4}$$

Therefore, the estimated individual probability $\tilde{\Phi}_{ij}(k)$ is a function of the fixed-effects estimates, empirical Bayes estimates, and measurement error. For a model with a strictly decreasing trend in t_{ij}, at a fixed threshold k, one would expect the probability in (4.4) to decrease with increasing j.

The continuation probability C_{ij} can also be interpreted in the survival analysis framework as the probability of individual i being at risk for the event after time t_{ij}. That is, the probability that individual i has not experienced two consecutive low measurements at, or prior to time point j. It should be evident that as j increases the computation of C_{ij} will become increasingly complex due to the number of combinations considered. Careful examination of the pattern governing the number of combinations that result in continuation at each time point, revealed the recursive relationship

$$C_{ij} = C_{j-2}[1 - \Phi_{ij-1}(k)][\Phi_{ij}(k)] + C_{j-1}[1 - \Phi_{ij}(k)]. \tag{4.5}$$

Further details regarding the proof of (4.5) and its computation can be found in Appendix A.

4.3.2 Estimation of Time to the First Measurement Less Than a Threshold of Interest

For processes that have a relatively 'smooth' evolution or where treatment guidelines are based on a single measurement, estimation of the time to reach a single measurement below a threshold may be of interest. Here the time to event T_i can be expressed as:

$$T_i = \min\{j \geq 1 : Y_{ij} \leq k\}. \tag{4.6}$$

It follows that the expected time for individual i to attain a single measurement less than the threshold k can be expressed as follows

$$\begin{aligned} E(T_i) &= t_{i1}P(Y_{i1} \leq k) + t_{i2}P(Y_{i1} > k, Y_{i2} \leq k) + t_{i3}P(Y_{i1} > k, Y_{i2} > k, Y_{i3} \leq k) \\ &\quad + t_{i4}P(Y_{i1} > k, Y_{i2} > k, Y_{i3} > k, Y_{i4} \leq k) \\ &\quad + \dots \\ &= \sum_{j=1}^{\infty} t_{ij} S_{ij} \end{aligned} \tag{4.7}$$

The computation in this setting is considerably simpler since there is only one possible combination of events that may lead to the event at time t_{ij}. In the case of conditional independence, S_{ij} will simply be the product of cumulative univariate normal probabilities.

4.3.3 Estimation and Inference

It follows from Section 4.3.1 that $E(T_i)$ is a function of the parameters $\boldsymbol{\beta}$, $\boldsymbol{b}_i$, and σ. Hence $\widehat{T}_i$, the estimate of $E(T_i)$, can be computed by substituting each unknown parameter by its corresponding estimate. Further details on inference for fixed effects and empirical Bayes prediction of the random effects in a linear mixed model can be found in Verbeke and Molenberghs (2009). In principle, the Delta method may be used to compute standard errors and 95% confidence intervals for $\widehat{T}_i$. However, the bootstrap offers a more feasible alternative. We propose a conditional version of the non-parametric bootstrap to compute 95% confidence intervals for $\widehat{T}_i$ as follows:

Step 1. Individual i is removed from the full dataset resulting in $N - 1$ cases

Step 2. Sample $N - 1$ subjects with replacement from the dataset in Step 1

Step 3. Append the data of individual i to the bootstrap sample

Step 4. Compute $\widehat{T}_i$

This process is repeated 1000 times. The computation of $\widehat{T}_i$, and the bootstrap confidence interval can be achieved with relatively basic programming. The program to estimate $\widehat{T}_i$, incorporating the recursive formula for continuation probabilities, can be found in Appendix A.2.

4.4 Application: Sinikithemba HIV Cohort Study

The time to CD4 threshold has been analysed as an outcome in several recent studies. The convention in such studies is to first extract the time of the event, which is analysed in a second stage within the survival analysis framework. Cardeal da Silva et al. (2013) analysed the time to first CD4 count less than 350 cells/mm^3 as the primary outcome in their study, which compared the rate of HIV progression pre and post the introduction of antiretroviral therapy (ART). Amornkul et al. (2013) studied a cohort of recently infected individuals where the effect of HIV subtype on HIV progression was examined. In this study, immunologic progression was defined as time from seroconversion to the first of two consecutive CD4 cell counts less than or equal to 350 cells/mm^3. In this section we apply the methodology proposed in Section 4.3 to estimate the time for individuals to reach two consecutive low CD4 counts. We considered two CD4 count thresholds of 200 and 350 cells/mm^3.

4.4.1 Linear Mixed Model

To explore the relationship between baseline viral load and rate of CD4 decline, baseline viral load (VL) was categorized into approximate tertiles as follows: VL $\leq 15,000$, $15,000 <$ VL $\leq 100,000$, and VL $> 100,000$ log copies/ml, which represent low, intermediate, and high viral load, respectively. In Figure 4.1 the fitted loess smooth curves clearly depict that the CD4 count evolution is different between the viral load categories. There were 92 (27%), 117 (35%), and 127 (38%) patients in the low, intermediate, and high viral load categories, respectively. We applied the 'General guidelines for model building'

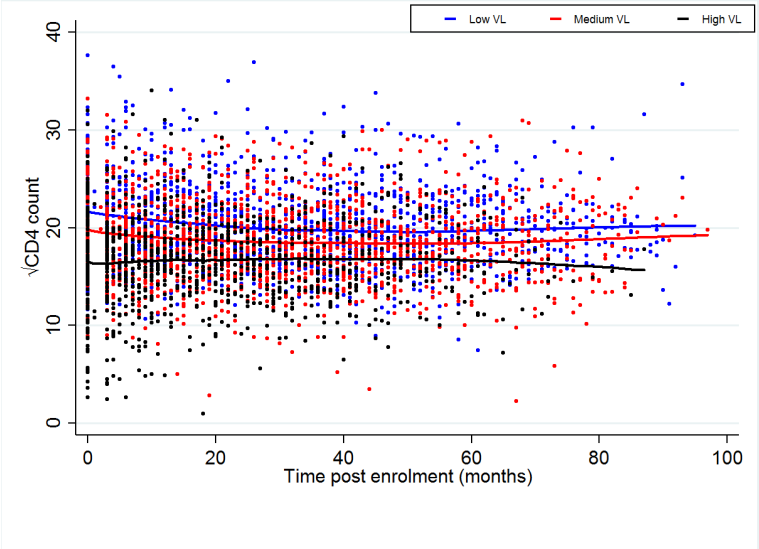


Figure 4.1: *Sinikithemba CD4 Study: Trajectory of CD4 count in each viral load category*

recommended in Verbeke and Molenberghs (2009), commencing with an elaborate mean structure, which included age, gender, baseline viral load, and interaction terms as covariates. Although the subject-specific plots and reduced AIC indicated that a quadratic or cubic model may provide a better fit to the observed data than the linear model, these models would result in implausible predicted trajectories outside of the observation period. By comparing nested models using the likelihood ratio test, the inclusion of age and gender did not significantly improve the model fit ($p = 0.150$). The reduced model was of the form:

$$Y_{ij} = \begin{cases} \beta_{0,L} + b_{0i} + \beta_{1,L} t_{ij} + b_{1i} t_{ij} + \varepsilon_{ij} & \text{if 'low' viral load,} \\ \beta_{0,M} + b_{0i} + \beta_{1,M} t_{ij} + b_{1i} t_{ij} + \varepsilon_{ij} & \text{if 'intermediate' viral load,} \\ \beta_{0,H} + b_{0i} + \beta_{1,H} t_{ij} + b_{1i} t_{ij} + \varepsilon_{ij} & \text{if 'high' viral load.} \end{cases}$$

Since this is a seroprevalent cohort, the date of the last negative HIV test result is unknown and hence the date of seroconversion cannot be estimated without the analysis of additional patients with known dates of infection. For this particular analysis we have examined two possible timescales: date of first contact as time zero and time on study, which is expressed as the difference between the enrollment date and the date at which

the study commenced (1 August 2003). As stated by Sweeting and Thompson (2012), different timescales in hierarchical models can have a strong impact on the predicted random effects due to the shrinkage effect. The REML estimates with standard errors for each of the timescales are presented in Table 4.1. As expected, the model with the time origin as 1 August 2003 resulted in a higher variance of the random intercepts. Through comparison of the AIC and BIC for the two models it is clear that the model with time since enrollment as the timescale provided a better fit to the data. All further analysis was therefore conducted in this timescale, which also facilitates interpretation of the estimated times as times relative to enrollment in the study. We found an overall significant difference in intercepts and slopes between viral load categories ($p < 0.0001$ and 0.0053 respectively). Patients with high viral load displayed a significantly higher rate of decline in CD4 count than patients with low viral load ($p < 0.0001$). More rapid decline in patients with high viral load compared to intermediate viral load was observed; this result was not statistically significant ($p = 0.115$).

4.4.2 Expected Time to Threshold

We allow a 10-year window prior to enrollment where we consider an individual as having the potential to have experienced the threshold. The rationale for this decision is based on the estimated time from seroconversion to death in ART naïve patients which was reported to be approximately 10 years in Sub-Saharan Africa (Van der Paal et al., 2007; Todd et al., 2007). The discrete times that fall outside of the observation period were created in accordance with the study design of three monthly visits. The series was truncated at the visit at which the predicted CD4 count $\widehat{Y_{ij}}$ dropped to zero. Similarly, time t_{i1} was defined as the minimum time at which $\widehat{Y_{ij}} < 1500$ cells/mm^3, which is the upper limit of the CD4 count range for HIV infected individuals. We estimated the time to obtain two consecutive measurements less than the threshold values 200 and 350 cells/mm^3, respectively. For ease of presentation, we have chosen to draw attention to the estimation for four specific patients (Figure 4.2). The estimated probabilities of a single measurement being below the threshold are presented in Figure 4.3. Patient A entered the study with a CD4 count substantially above the 200 and 350 cells/mm^3

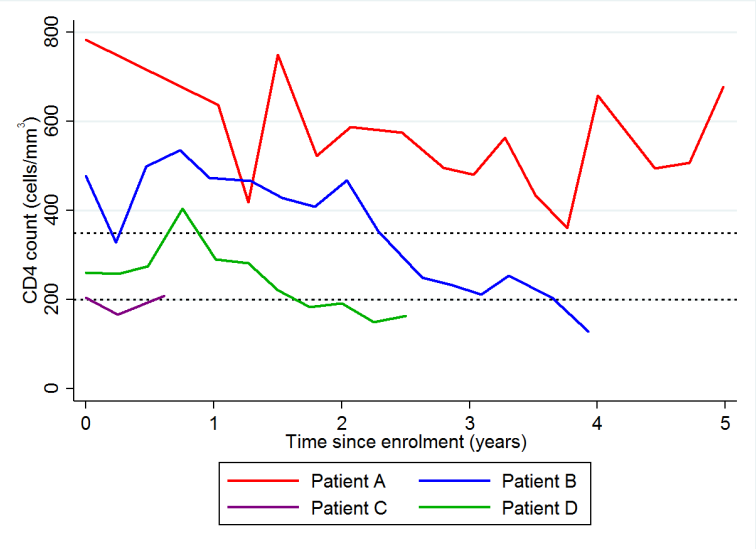
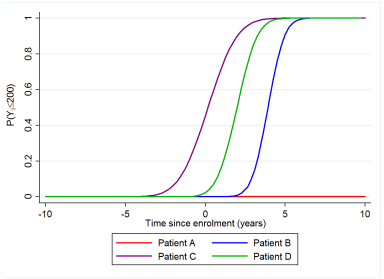
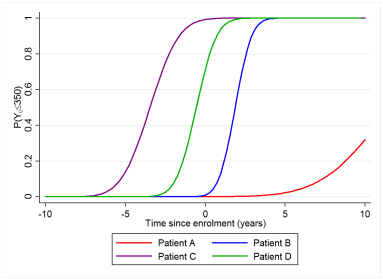


Figure 4.2: *Sinikithemba CD4 Study: Longitudinal CD4 count measurements with reference at thresholds 200 and 350 cells/mm^3.*



(a) Probability of single CD4 count less than 200 cells/mm^3



(b) Probability of single CD4 count less than 350 cells/mm^3

Figure 4.3: *Sinikithemba CD4 Study: Estimated probabilities of a single measurement being below the relevant threshold.*

thresholds and declined at a very slow rate. This is captured by the fitted probabilities where the probability of Patient A obtaining a CD4 count less than 200 cells/mm^3 is zero throughout the period considered. The probability of Patient A experiencing a CD4 count less than 350 cells/mm^3 increases at five years. The estimated time to two consecutive

measurements less than the 200 and 350 cells/mm^3 threshold is presented in Table 4.2.

Patient B, who entered the study with a CD4 count above the 350 cells/mm^3 threshold, exhibited a more rapid rate of CD4 count decline than Patient A. The estimated time for Patient B to reach a threshold of 350 and 200 cells/mm^3 was 2.3 and 4.3 years, respectively. Patients C and D both entered the study with CD4 counts less than 350 but declined at different rates. This is captured by the predicted probabilities in Figure 4.3. Patient C was estimated to have reached the 200 cells/mm^3 threshold 0.38 years post enrollment, and the 350 threshold 3.26 years prior to enrollment. The confidence intervals for the estimated times for patient C reveal that poorer precision is obtained when analyzing individuals with few measurements. Caution should be exercised when interpreting the estimated times for patients who start at a high CD4 count and exhibit a very slow rate of decline. Probabilities of low CD4 count that are zero throughout the period of observation do not pose a problem, but probabilities that increase to greater than zero later in the period can result in estimated times which are sensitive to the frequency and timing of unobserved measurements which are considered. This is discussed further in Section 4.4.3.

Table 4.1: *Sinikithemba CD4 Study: Parameter estimates (standard errors) for the fitted models on each timescale.*

Effect	Parameter	Time-enrolment	Time-calendar origin
Fixed effects estimates (s.e.)			
Intercept	$\beta_{0,L}$	21.2405 (0.4708)	22.0000 (0.5513)
	$\beta_{0,M}$	19.4469 (0.4190)	20.6554 (0.4978)
	$\beta_{0,H}$	16.2821 (0.4057)	17.5021 (0.4909)
Time	$\beta_{1,L}$	-0.5744 (0.1206)	-0.5658 (0.1171)
	$\beta_{1,M}$	-1.0160 (0.1137)	-0.9454 (0.1102)
	$\beta_{1,H}$	-1.3839 (0.1400)	-1.1066 (0.1331)
Covariance parameter estimates (s.e.)			
var(b_{0i})	d_{11}	19.5555 (1.6080)	25.5456(2.2716)
cov(b_{0i}, b_{1i})	d_{12}	-0.4944 (0.3821)	-2.1611 (0.4703)
var(b_{1i})	d_{22}	0.9941 (0.1421)	0.9438 (0.1303)
Measurement error	σ^2	3.1923 (0.0810)	3.2135 (0.0814)
Fit statistics			
AIC		17185.3	17225.9
BIC		17200.5	17241.1
-2 REML log-likelihood		17177.3	17217.9

Table 4.2: *Sinikithemba CD4 Study: Estimated time to threshold for patients A, B, C, and D.*

			≤ 200 cells/mm^3		≤ 350 cells/mm^3	
Patient	VL	Baseline CD4	$\widehat{T}_i$	95% CI	$\widehat{T}_i$	95% CI
A	Low	783	2.92×10^{-5}	(8.88×10^{-6}, 8.24×10^{-5})	3.1552	(2.4946, 3.7362)
B	Low	478	4.2858	(4.2343, 4.3843)	2.3046	(2.2887, 2.3169)
C	High	204	0.3758	(0.0267, 0.5319)	-3.2608	(-5.0051, -2.2874)
D	High	261	2.3335	(2.3005, 2.3763)	-0.2043	(-0.4039, -0.0642)

Several possible analyses using the estimated times can be conducted. We have elected to focus on the estimated probabilities and times themselves as they draw attention to several current issues in the treatment and monitoring of HIV positive patients. There were 30 individuals who had a zero probability of obtaining a single CD4 count less than 200 cells/mm^3 throughout the period considered. These individuals are referred to as long term non-progressors. This contributes additional evidence to the proposition that there are individuals who, possibly due to genetics, are able to control the virus. In addition, the estimated times draw attention to a serious public health concern, namely late presentation for HIV testing. Excluding the individuals who were long term non-progressors, the percentiles of the estimated times were computed. 15% of these patients had already attained a CD4 count less than 200 cells/mm^3 more than six months *prior* to first presentation at the clinic. Hence, patients were choosing to have an HIV test when they were already in the advanced stages of HIV. The ARV treatment guideline in effect during the study recommended treatment initiation at a CD4 count less than 200 cells/mm^3. Therefore, an additional interpretation is that 15% of the patients deviated from the recommended timing of treatment initiation by more than six months. After 2011, the treatment initiation cutoff was raised to 350 cells/mm^3. During our study period, we found that 35% of patients had already attained two consecutive CD4 counts less than 350 cells/mm^3 more than two years prior to enrollment. It is clear that, unless patients present at the clinic earlier for testing, changing treatment guidelines may not have the desired effect. It would be interesting to examine whether health seeking behaviour has changed over time, by studying individuals who first presented at the clinic after 2011.

4.4.3 Sensitivity to Variation in Observation Frequency

In Section 4.4.2, we considered an individual to be at risk of experiencing a CD4 count below the thresholds of interest, 10 years prior to and post enrolment, and generated regular three monthly visits for the unobserved period. In this section, we assess the sensitivity of the estimated time to threshold to truncation and the regularity and frequency of visits by simulation.

The following scenarios were considered:

Scenario 1. A period of 10 years prior to and post enrolment was considered, and visits outside the observed period occurred at regular three monthly intervals.

Scenario 2. A period of 5 years prior to and post enrolment was considered. Visits outside the observed period occurred at regular three monthly intervals.

Scenario 3. A period of 10 years prior to and post enrolment was considered and 10% of visits outside the observation period occurred one month later than expected.

Scenario 4. A period of 10 years prior to and post enrolment was considered and 25% of visits outside the observation period occurred one month later than expected.

Scenario 5. A period of 10 years prior to and post enrolment was considered and 10% of visits outside the observation period were missed.

Scenario 6. A period of 10 years prior to and post enrolment was considered and 20% of visits outside the observation period were missed.

Table 4.3: *Sinikithemba CD4 Study-Sensitivity analysis: Estimated time to two consecutive measurements less than 350 cells/mm^3 under various scenarios.*

Patient	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
A	3.1552	0.0050	3.0734	3.1271	3.0219	2.5941
B	2.3046	2.3046	2.2926	2.2926	2.2926	2.2926
C	-3.2608	-3.2056	-3.2056	-3.2073	-3.2119	-2.9572
D	-0.2043	-0.2043	-0.2030	-0.2097	-0.1432	0.0208

From Table 4.3, it is clear that the number of visits considered has an impact on the estimated time to threshold for individuals who enter the study with a high CD4 count and exhibit slow decline. Truncation of the series at 5 years results in an estimated time to threshold of 0.005 years, whereas truncation at 10 years results in an estimated value of 3.155 years for Patient A. Similarly, the number of visits considered outside the period

of observation also has an effect for this type of patient. This is evident from the estimate for Patient A under the assumption of 20% of visits being missed. We found that results were far less sensitive for patients experiencing moderate to rapid decline in CD4 count. As expected, in individuals who are likely to reach the threshold of interest during the observed period, the estimated time is robust to truncation or variation in the time points considered. In all patients studied, results were robust to variation in the timing of visits. Hence, the estimated time to threshold is sensitive to the number of time points for specific patients, but not to the actual values of those time points.

4.5 Application: Abdominal Aortic Aneurysm (AAA) Study

In abdominal aortic aneurysm (AAA) screening studies, surgery is recommended to patients when the diameter of the aneurysm exceeds 55 mm (Sweeting and Thompson, 2012). Due to the high degree of within-patient variability in aortic diameter measurements (Dapunt, 1994), it is of interest to examine the effect of applying persistence criteria in this setting. Estimation of the expected time to threshold for each patient would enable clinicians to identify patients who may be in need of surgery in the near future, and target interventions accordingly. In this section we apply the proposed methodology to estimate the time to reach AAA diameters exceeding 50 mm and 55 mm.

4.5.1 Linear Mixed Model

We examined the effect of various demographic and physiological factors on the aneurysm diameter. We found that increasing age was associated with higher aneurysm diameter at baseline. Baseline aneurysm diameters were also significantly higher in obese individuals compared to normal and overweight individuals ($p < 0.001$). The patient's weight category at baseline had a significant effect on the rate of increase of the aneurysm diameter. Following the guidelines for model building in Verbeke and Molenberghs (2009), the final model, which provided the best fit to the data was a model with random slopes and intercepts. The REML estimates with standard errors for the final model fitted are

presented in Table 4.4.

4.5.2 Expected time to threshold

To facilitate prediction of future time to event for individuals who do not attain the threshold of interest during the follow-up period, a maximum period of 15 years, where we consider the individual to be 'at risk', was used. These discrete visits were equally spaced at six monthly intervals, in accordance with the study design. An aneursym diameter cutoff of 55 mm is of great importance since this is the time at which surgical intervention is recommended. In this section, we also investigate the cutoff of 50 mm, which may prove useful to guide patients and/or surgeons on earlier intervention if needed. 23% of patients entered the study at an advanced stage with aneursym diameters exceeding 50 mm. In order to back-calculate the time at which these individuals may have first reached a value exceeding 50 mm, we created a five year period *prior* to enrolment. Patients dropped out of the study at various points, either due to surgical intervention or for unknown reasons. For the purposes of this analysis we assume ignorability and conduct likelihood-based analysis.

The time taken for individuals to reach two consecutive aneurysm diameter measurements greater than 50 mm, as well as the time to attain the first *single* measurement greater than 55 mm are both of interest. The rationale behind the latter approach, is that the recommendation of surgical intervention is currently based on a single measurement exceeding the threshold. Table 4.5 presents the results for the estimated time to AAA progression, defined as the time to two consecutive measurements exceeding 50 mm, and the time to the first measurement exceeding 55 mm. Patient 1074, who was overweight, entered the study with an aneursym diameter of 52 mm and exhibited a relatively flat profile. The estimated time to attain the first diameter exceeding 55 mm is approximately 1 year, six months after the two consecutive measurements greater than 50 mm are observed. Patient 1241, who had a BMI within the normal range, is expected to attain a measurement greater than 55 mm more than a year after experiencing two consecutive measurements exceeding 50 mm. The precision of the estimated times to threshold were found to be less precise in individuals who had less than three follow-up measurements

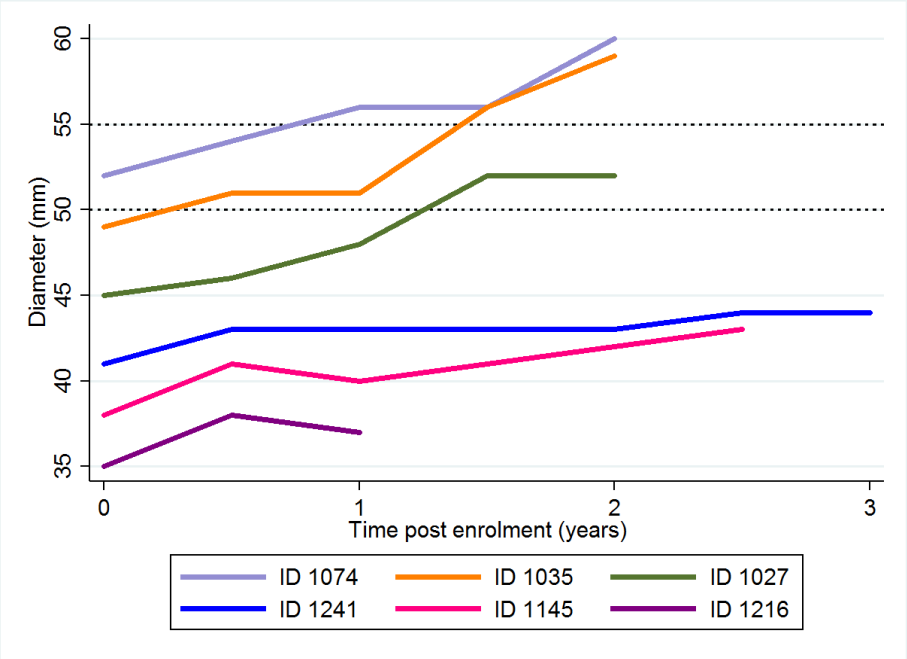


Figure 4.4: *AAA Study: Longitudinal AAA measurements with reference at 50 mm and 55 mm for selected patients.*

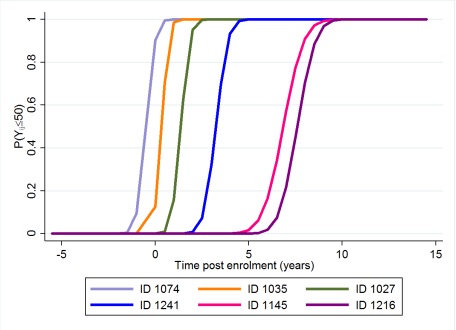
after baseline. This is clear when examining the confidence intervals of the estimates for Patient 1216, who exited the study after one year. The estimated probabilities of a measurement exceeding 50 mm and 55 mm over the period considered is presented in Figure 4.5.

Table 4.4: *AAA Study: Parameter estimates (standard errors) for the fitted model , where NM, OW and OB denote BMI categories normal, overweight and obese, respectively.*

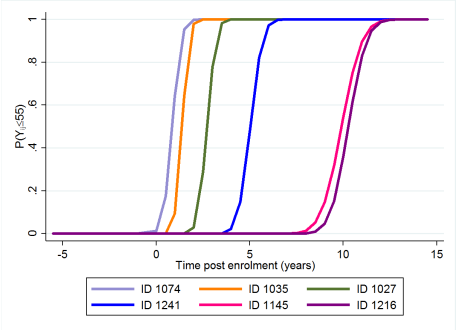
Effect	Parameter	Estimate
Fixed effects estimates (s.e.)		
Age	$\beta_{0,age}$	0.2086 (0.0861)
Intercept	$\beta_{0,NM}$	27.7875 (6.3193)
	$\beta_{0,OW}$	27.4011 (6.1842)
	$\beta_{0,OB}$	30.0097 (6.0743)
Time	$\beta_{1,NM}$	2.0700 (0.3560)
	$\beta_{1,OW}$	2.5381 (0.3124)
	$\beta_{1,OB}$	2.2967 (0.5518)
Covariance parameter estimates (s.e.)		
var(b_{0i})	d_{11}	41.3716 (6.1526)
cov(b_{0i}, b_{1i})	d_{12}	4.0783 (1.6916)
var(b_{1i})	d_{22}	3.4379 (0.8078)
Measurement error	σ^2	1.9986 (0.1968)
Fit statistics		
AIC		1915.4
BIC		1925.8
-2 REML log-likelihood		1907.4

Table 4.5: *AAA Study: Estimated time (years) to threshold for patients 6 selected patients under two definitions.*

				Two consecutive diameters>50 mm		Single diameter>55 mm	
Patient	BMI	Age	Baseline AAA	$\widehat{T}_i$	95% CI	$\widehat{T}_i$	95% CI
1074	Overweight	75	52	0.5098	(0.4420, 0.5661)	1.0538	(1.0155, 1.0868)
1035	Overweight	60	49	1.1144	(1.0594, 1.1623)	1.6168	(1.6018, 1.6458)
1027	Obese	54	45	2.1722	(2.1330, 2.2594)	2.9088	(2.7676, 3.1762)
1241	Normal	75	41	4.0631	(3.6960, 4.9729)	5.1633	(4.6347, 6.5100)
1145	Obese	64	38	7.7206	(7.0330, 8.7258)	9.7853	(8.7456, 11.0864)
1216	Overweight	68	35	8.4627	(7.2313, 10.1325)	10.2385	(8.6929, 12.3832)



(a) Probability of single AAA diameter greater than 50 cells/mm^3



(b) Probability of single AAA diameter greater than 55 cells/mm^3

Figure 4.5: *AAA Study: Estimated probabilities of a single measurement being above the relevant threshold.*

In light of the model results presented in Table 4.4, age and body mass index play a crucial role in the trajectory of AAA diameter. In the following results, we focus on the estimated time to surgery *marginally* for different values of age and BMI status. Using the methods previously presented, and setting the random effects b_i equal to zero to represent a hypothetical 'median' patient, the estimated times to first AAA diameter exceeding 55 mm was calculated (Table 4.6). It is clear that, at the same baseline age, an obese individual will require surgery approximately 1.5 years earlier than an individual of normal weight. It is also of interest that an individual ten years older than another individual in the same weight category, will require surgery approximately one year earlier. These findings are very useful to clinicians, as it enables them to quantify the effect of obesity on AAA's so that they may consequently encourage behavioural changes in patients and plan follow-up visits earlier. Since the estimates produced in Table 4.6, are based only on baseline covariate values and do not incorporate data on baseline or follow-up aneurysm diameter, it is expected that confidence intervals will tend to be wider than those presented previously, for subject-specific estimation.

Table 4.6: *AAA Study: Estimated time (years) to 55 mm for the hypothetical 'median' patient at various ages and BMI categories.*

Age	Weight	$\widehat{T}_i$	95% CI
50	Normal	8.126	(5.277, 11.981)
50	Overweight	6.868	(4.815, 9.499)
50	Obese	6.405	(3.569, 11.283)
60	Normal	7.119	(4.906, 10.227)
60	Overweight	6.046	(4.46, 8.059)
60	Obese	5.497	(3.181, 10.501)
70	Normal	6.111	(4.495, 8.421)
70	Overweight	5.224	(4.005, 6.834)
70	Obese	4.589	(2.729, 8.513)

4.6 Discussion

In this chapter we have proposed and applied a novel approach to estimation of the time to attain two consecutive measurements less than (or greater than) a relevant threshold. This approach takes into account the estimated patient-specific trajectories and measurement error. Through identification of a recursive relationship of the continuation probabilities at each time point, we have displayed that the computation of the expected times is simple, efficient, and can be implemented using existing software packages. The method we have proposed can also accommodate complex functions of time, such as quadratic or cubic terms, in contrast to the inverse estimation framework.

Sensitivity analysis revealed that the estimated times are sensitive to the number of visits considered and the time at which the series is truncated, for patients who exhibit a very slow decline. For other patients, however, we found that results were less sensitive

to the number of visits considered and truncation. Hence, caution should be exercised when interpreting the estimated times for patients who exhibit very slow rates of decline. Another strong assumption that was made for the specific application presented, was that visits prior to enrollment and post dropout occurred at regular, equally spaced time points. In the sensitivity analysis conducted, we found that results were robust to deviation from the regular observation times for all patients.

The proposed methodology rests on the assumption that the residual variability is pure measurement error, which may be violated in certain settings. As discussed by Rizopoulos (2012), extending a linear mixed model by including a more elaborate random effects structure is computationally simpler to implement and can produce practically indistinguishable fits to the data when compared to a model that includes a serial correlation term. However, in some cases, extensive knowledge of the true underlying process which generates the data, may necessitate the inclusion of serial correlation in the model.

Chapter 5

Estimation of Time to Threshold Taking Complex Correlation Structures into Account

5.1 Introduction

In Chapter 4 we proposed a method for estimation of the time to threshold in the presence of persistence criteria, using a two-stage approach. This method, however, was confined to situations where, conditional on random effects, the residuals are independent (i.e., the assumption of conditional independence). In some settings this assumption may be unrealistically simplistic, and at least some part of an individual's profile is a response to time-varying stochastic processes operating within that individual (Verbeke and Molenberghs, 2009). Furthermore, incorrectly modeling the covariance structure can impair inference and estimation in the linear mixed model, as noted by Jacqmin-Gadda et al (2007). In this chapter, we extend the methodology to accommodate serial correlation.

The derivation of the methodology for time to threshold estimation is presented in Section 5.2 and Appendix B. In Section 5.3 the extended methodology is applied to data emanating from a cohort of HIV positive individuals in South Africa.

5.2 Methodology

5.2.1 Expected Time to Attain a Threshold with Persistence Criteria

We return to the notation presented in Section 4.3 where the time to threshold is defined as:

$$T_i = \min\{j \geq 2 : Y_{ij-1} \leq k, Y_{ij} \leq k\},$$

and the expected time to threshold is expressed as the probability weighted sum of the possible event times t_{ij}

$$\begin{aligned} E(T_i) &= t_{i2}P(Y_{i1} \leq k, Y_{i2} \leq k) \\ &\quad + t_{i3}P(Y_{i1} > k, Y_{i2} \leq k, Y_{i3} \leq k) \\ &\quad + t_{i4}\left\{\begin{array}{l} P(Y_{i1} > k, Y_{i2} > k, Y_{i3} \leq k, Y_{i4} \leq k) \\ + P(Y_{i1} \leq k, Y_{i2} > k, Y_{i3} \leq k, Y_{i4} \leq k) \end{array}\right\} \\ &\quad + \ldots \\ &= \sum_{j=2}^{\infty} t_{ij}S_{ij}. \end{aligned}$$

As described in Chapter 3 we specify a linear mixed model of the form

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_{i1} + \boldsymbol{\varepsilon}_{i2}$$

where $\boldsymbol{\varepsilon}_{i(1)}$ and $\boldsymbol{\varepsilon}_{i(2)}$ denote measurement error and serial variation, respectively. It follows that:

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i, \sigma^2 I_{n_i} + \tau^2 H_i), \tag{5.1}$$

where the $(j,k)^{th}$ element of H_i is $\rho(|t_{ij} - t_{ik}|)$, which is the correlation between $\varepsilon_{ij(2)}$ and $\varepsilon_{ik(2)}$. Assuming that the Markov property holds, and using the chain rule of probabilities,

the joint probabilities that form S_{ij} reduce to the product of conditional probabilities. Hence, S_{ij} may be simplified as follows:

$$\begin{aligned} S_{ij}|X_i, Z_i, \boldsymbol{b}_i, \boldsymbol{\beta} &= C_{ij-3,0} \left\{ \begin{array}{l} P(Y_{ij-2} > k|Y_{ij-3} > k)P(Y_{ij-1} \leq k|Y_{ij-2} > k) \\ \times P(Y_{ij} \leq k|Y_{ij-1} \leq k) \end{array} \right\} \\ &\quad + C_{ij-3,1} \left\{ \begin{array}{l} P(Y_{ij-2} > k|Y_{ij-3} \leq k)P(Y_{ij-1} \leq k|Y_{ij-2} > k) \\ \times P(Y_{ij} \leq k|Y_{ij-1} \leq k) \end{array} \right\} \\ &= \left\{ \begin{array}{l} C_{ij-3,0}P(Y_{ij-2} > k|Y_{ij-3} > k) \\ + C_{ij-3,1}P(Y_{ij-2} > k|Y_{ij-3} \leq k) \end{array} \right\} \\ &\quad \times P(Y_{ij-1} \leq k|Y_{ij-2} > k)P(Y_{ij} \leq k|Y_{ij-1} \leq k), \end{aligned} \tag{5.2}$$

where $C_{ij-3,0} + C_{ij-3,1}$ denotes the total 'continuation probability' at time t_{ij-3}, and $C_{ij-3,0}$ and $C_{ij-3,1}$ denote continuation sequences ending with $Y_{ij-3} > k$ and $Y_{ij-3} \leq k$, respectively. The continuation probability can also be interpreted in the survival analysis framework as the probability of individual i being at risk for the event after time t_{ij}. That is, the probability that individual i has not experienced two consecutive low (or high) measurements at, or prior to time point j. It should be evident that as j increases the computation of $C_{ij,0}$ and $C_{ij,1}$ will become increasingly complex due to the number of combinations considered. Careful examination of the continuation sequences revealed that fortunately the following recursive relationships exist:

$$C_{ij,0} = C_{ij-1,0}P(Y_{ij} > k|Y_{ij-1} > k) + C_{ij-1,1}P(Y_{ij-1} > k|Y_{ij-1} \leq k), \tag{5.3}$$

$$C_{ij,1} = C_{ij-1,0}P(Y_{ij} \leq k|Y_{ij-1} > k). \tag{5.4}$$

Further details regarding the proof of (5.3) and (5.4) can be found in Appendix B.1. Using Bayes' theorem, each conditional probability that forms S_{ij} can be expressed as a function of bivariate and univariate normal probability densities such that

$$P(Y_{i\ell} \leq k|Y_{i\ell-1} > k) = \frac{P(Y_{il} \leq k, Y_{i\ell-1} > k)}{P(Y_{i\ell-1} > k)}.$$

Computation of the cumulative bivariate probabilities can be done by integration of the bivariate normal probability density functions. Suppressing i for notational convenience, and denoting the correlation between $Y_{\ell-1}$ and Y_ℓ by $\rho_{\ell-1,\ell}$ the following expressions

exist:

$$\begin{aligned} & P(Y_\ell \leq k, Y_{\ell-1} \leq k) \\ &= \tilde{\Phi}_{\ell-1,\ell}(k,k;\rho_{\ell-1,\ell}) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2_{\ell-1,\ell}}} \int_{-\infty}^{k}\int_{-\infty}^{k} \exp\left(-\frac{\tilde{y}^2_{\ell-1} - 2\rho_{\ell-1,\ell}\tilde{y}_{\ell-1}\tilde{y}_j + \tilde{y}^2_\ell}{2(1-\rho_{\ell-1,\ell})}\right) d\tilde{y}_\ell d\tilde{y}_{\ell-1} \end{aligned} \tag{5.5}$$

where $\tilde{y}_\ell = \dfrac{y_{i\ell} - \boldsymbol{x}'_{i\ell}\boldsymbol{\beta} - \boldsymbol{z}'_{i\ell}\boldsymbol{b}_i}{\sqrt{\sigma^2+\tau^2}}$ and $\tilde{y}_{\ell-1} = \dfrac{y_{i\ell-1} - \boldsymbol{x}'_{i\ell-1}\boldsymbol{\beta} - \boldsymbol{z}'_{i\ell-1}\boldsymbol{b}_i}{\sqrt{\sigma^2+\tau^2}}$ represent the standardized normal transformation of $Y_{i\ell}$ and $Y_{i\ell-1}$ respectively. The three other combinations of probabilities can be expressed as a function of (5.5) and the univariate cumulative distribution of Y_ℓ and $Y_{\ell-1}$ as follows:

$$\begin{aligned} P(Y_{i\ell-1} \leq k, Y_{i\ell} > k) &= \tilde{\Phi}_{\ell-1}(k) - \tilde{\Phi}_{\ell-1,\ell}(k,k;\rho_{\ell-1,\ell}), \\ P(Y_{i\ell-1} > k, Y_{i\ell} \leq k) &= \tilde{\Phi}_{\ell}(k) - \tilde{\Phi}_{\ell-1,\ell}(k,k;\rho_{\ell-1,\ell}), \\ P(Y_{i\ell-1} > k, Y_{i\ell} > k) &= 1 - \tilde{\Phi}_{\ell}(k) - \tilde{\Phi}_{\ell-1}(k) + \tilde{\Phi}_{\ell-1,\ell}(k,k;\rho_{\ell-1,\ell}). \end{aligned}$$

5.2.2 Estimation and Inference

In certain settings the bootstrap sampling distribution may not be symmetric, raising doubt regarding the validity of using percentiles of the distribution as the endpoints for the confidence interval. As noted by previous authors, the coverage error can be substantial if the distribution of the estimate is not symmetric (Carpenter and Bithell, 2000). In such cases, the bias corrected and accelerated (BCa) method (Efron and Tibshirani, 1994), which allows for the lack of symmetry and skewness to vary with changes in the parameter value, offers a more fruitful alternative. Returning to the four steps for bootstrap sampling that were discussed in Section 4.3.3, we follow the same process but for 5000 rather than 1000 iterations. The estimated bias parameter is computed from the resulting distribution. The acceleration parameter, which is a function of the jackknife estimate, is computed using case jacknife resampling where each individual is systematically omitted from the dataset. The endpoints of the confidence intervals are then calculated using the bias and acceleration parameters. Further details regarding the formula for calculating BCa confidence intervals are presented in Appendix B.3. The SAS program to estimate $\widehat{T}_i$ and can be found in Appendix B.2.

5.3 Application: Sinikithemba HIV Cohort Study

In this section we apply the extended methodology proposed in Section 5.2 to the Sinikithemba cohort that was previously analyzed in Chapter 4 assuming that the within-patient variability was purely measurement error and that residuals were uncorrelated. A tool to describe the variance for an irregularly observed process, is the variogram (Diggle, 2013):

$$\gamma(u) = \frac{1}{2}E[(Y(t) - Y(t-u))^2], u \geq 0, \tag{5.6}$$

which is estimated using the squared differences between residuals $v_{ijk} = \frac{1}{2}(r_{ij} - r_{ik})^2$ and the corresponding time difference or lag, $u_{ijk} = t_{ij} - t_{ik}$.The empirical variogram for CD4 measurements is presented in Figure 5.1. From the variogram it is clear that the dominating source of variance is the between-patient variability. There is also strong presence of serial correlation, evidenced by the increase in $v(u)$ with increasing lag. The behavior of the variogram close to zero lag also indicates the presence of measurement error. To ensure that valid estimation and inferences are drawn it is important to take all three sources of variability into account, that is between-subject variability, serial correlation and measurement error.

5.3.1 Linear Mixed Model

As recommended by Verbeke and Molenberghs (2009), several covariance functions were compared using the fixed- and random-effects structure that were fitted to the dataset in Section 4.4. According to information criteria, the structure which provided the best fit was of the first-order autoregressive (AR(1)) type. We draw attention to the following three models below:

Model 1. No serial correlation

Model 2. Common AR(1) correlation for all patients, measurement error

Model 3. Different AR(1) correlation for each viral load category, measurement error.

The number of covariance parameters and fit statistics for each of the models above is presented in Table 5.1.

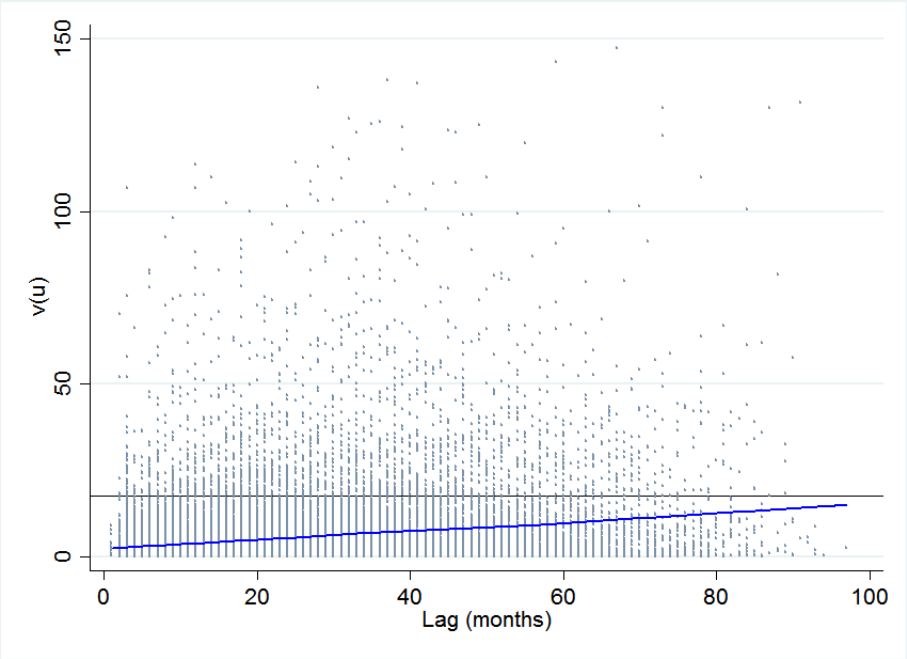


Figure 5.1: *Sinikithemba CD4 Study: Sample variogram of CD4 count residuals*

Table 5.1: *Sinikithemba CD4 Study: Fit Statistics for the three models fitted*

		fit statistic		
Model	#cov par	-2logL	AIC	BIC
1	4	17194.948	17202.9	17218.2
2	6	17194.866	17206.9	17229.8
3	10	17143.852	17163.9	17202.0

Comparing nested models using the likelihood ratio test, the model which included a viral-load-specific serial correlation component (Model 3) resulted in a significantly better fit than Model 1 and Model 2. This is consistent with the findings of Boscardin et al. (1998) and Diggle (2013) who studied the evolution of CD4 counts. As evidenced by the AIC and BIC, a model with a viral-load-specific autoregressive serial correlation function provided the best fit to the data. The REML estimates with standard errors for the model

Table 5.2: *Sinikithemba CD4 Study: Parameter estimates (standard errors) for Model 1 and Model 3*

Effect	Parameter	Model 1	Model 3
Fixed effects estimates (s.e.)			
Intercept	$\beta_{0,L}$	21.2406 (0.4707)	21.2412 (0.4223)
	$\beta_{0,M}$	19.4453 (0.4189)	19.4392 (0.3755)
	$\beta_{0,H}$	16.2826 (0.4056)	16.3238 (0.4753)
Time	$\beta_{1,L}$	-0.0481 (0.0100)	-0.0481 (0.0094)
	$\beta_{1,M}$	-0.0848 (0.0094)	-0.0833 (0.0088)
	$\beta_{1,H}$	-0.1157 (0.0116)	-0.1137 (0.0137)
Covariance parameter estimates (s.e.)			
var(b_{0i})	d_{11}	19.5442 (1.6072)	15.5415 (1.6287)
cov(b_{0i}, b_{1i})	d_{12}	-0.0404 (0.03175)	-0.0388 (0.0275)
var(b_{1i})	d_{22}	0.0068 (0.0009)	0.0059 (0.0009)
Measurement error	σ^2	3.1976 (0.0811)	2.0463 (0.2193)
Serial variation			
Variance	τ_L^2		1.2909 (0.2576)
	τ_M^2		1.0778 (0.2532)
	τ_H^2		12.0392 (3.9736)
Serial parameter	ρ_L		-0.0460 (0.3547)
	ρ_M		-0.0812 (0.2187)
	ρ_H		0.9873 (0.0055)

with no serial component (Model 1) and the model with viral load specific autoregressive serial correlation (Model 3) are presented in Table 5.2. We found an overall significant

difference in intercepts and slopes between viral load categories ($p < 0.0001$ and 0.001, respectively). In addition, patients with high viral load displayed a significantly higher rate of decline in CD4 count than patients with low and intermediate viral load ($p < 0.0001$). In the low and intermediate viral load groups, random intercepts contributed the most to the overall variance, followed by serial correlation and measurement error. The high viral load group contributed the highest variance of the serial correlation component, and the serial correlation in between observations tends to decay at a slower rate compared to those with medium or low viral loads. This is evidenced by the magnitude of the parameter ρ_H.

5.3.2 Expected Time to Threshold

We estimated the time to obtain two consecutive measurements less than the threshold 350 cells/mm^3. For ease of presentation, we have chosen to draw attention to the estimation for five specific patients (Figure 5.2), who entered the study at different CD4 count levels. In Figure 5.3, four estimated probabilities of interest, emanating from the final model (Model 3), are presented for the selected patients.

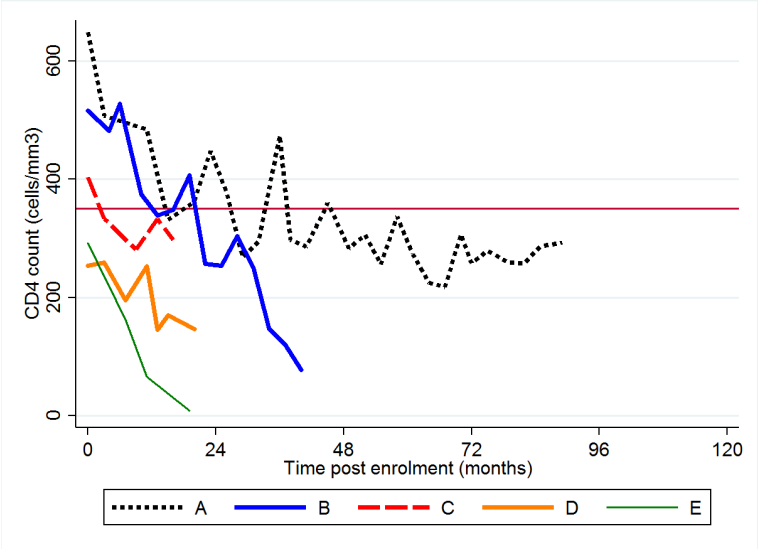
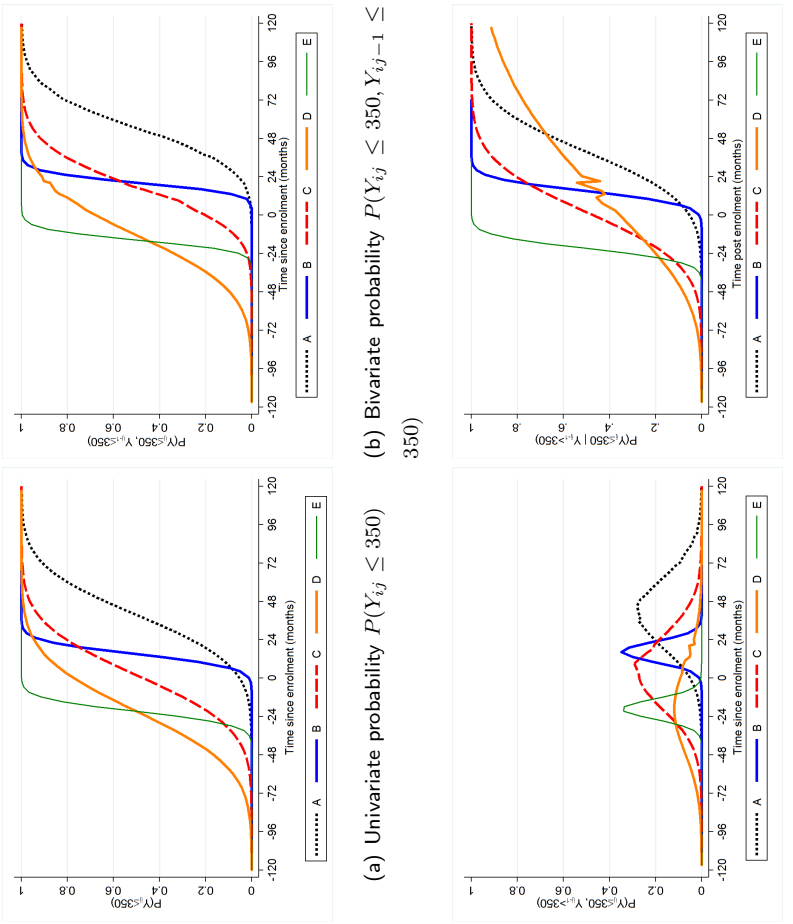


Figure 5.2: *Sinikithemba CD4 Study: Longitudinal CD4 count measurements with reference at 350 cells/mm^3*



(a) Univariate probability $P(Y_{ij} \leq 350)$ (b) Bivariate probability $P(Y_{ij} \leq 350, Y_{ij-1} \leq 350)$

(c) Bivariate probability $P(Y_{ij} \leq 350, Y_{ij-1} > 350)$ (d) Conditional probability $P(Y_{ij} \leq 350 | Y_{ij-1} > 350)$

Figure 5.3: *Sinikithemba CD4 Study: Estimated probabilities computed for the selected patients*

Examination of the bootstrap sampling distributions of the time to threshold, revealed the presence of left skewness and outliers. This is in contrast to the bootstrap sampling distribution which was calculated for the model with uncorrelated residuals. The distributions for the expected time to threshold for individual B are presented in Figure 5.4. In view of the non-normality of the bootstrap distribution for the model with AR(1) se-

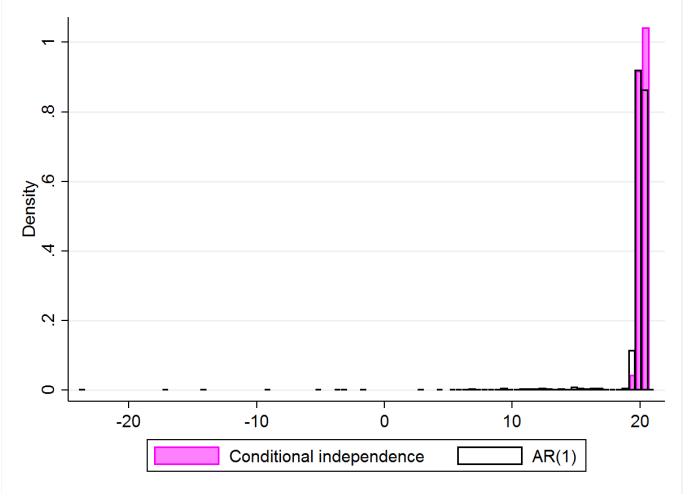


Figure 5.4: *Sinikithemba CD4 Study: Bootstrap sampling distribution of $E(T_i)$ for the independence and AR(1) models*

rial correlation, bias corrected accelerated (BCa) confidence intervals were presented for expected time to threshold emanating from these models. The estimated bias correction (b) and acceleration (a) statistics for individual B, were 0.2106 and 0.0866, respectively. As described in Efron and Tibshirani (1994), the endpoints of the BCa confidence interval can be expressed as a function of these parameters. The resulting 95% BCa interval for individual B was (19.5734, 20.5537). The estimated time to reach consecutive measurements less than 350 cells/mm^3 for all five selected individuals is presented in Table 5.3. To examine the impact of incorrectly assuming that the residuals are independent, we have presented the estimates for both the models that were presented in Table 5.2. It is not unexpected that, after taking into account serial correlation, an individual may be expected to reach the threshold of interest at an earlier time point. The median difference in estimated times between the serial correlation and independence model was -14.3, 0.01

and -0.09 months for the high, medium and low viral load groups, respectively. Hence, in cases where strong serial correlation is present, ignoring this in the model fitting stage can have a large impact on the estimated times to threshold.

Table 5.3: *Sinikithemba CD4 Study: Estimated time to threshold for patients A, B, C, D and E*

			Conditional independence		AR(1) serial correlation	
Patient	VL	Baseline CD4	$\widehat{T}_i$	95% CI	$\widehat{T}_i$	95% CI
A	Low	649	39.6915	(39.1433, 40.1951)	39.4019	(38.8722, 40.2760)
B	Medium	516	20.2437	(19.7295, 20.5117)	20.1297	(19.5734, 20.5537)
C	Low	404	1.4861	(-5.3512, 5.1979)	0.4621	(-8.2944, 5.3382)
D	High	254	-15.4749	(-22.2979, -11.0988)	-31.3803	(-48.9055, -18.1455)
E	Medium	292	-14.0174	(-25.6776, -8.5061)	-15.7942	(-27.3032, -8.5299)

5.4 Discussion

In this Chapter, we extended the methodology so that correlated residuals can be accommodated. This adds to the flexibility of the approach, which was previously found to be computationally efficient and robust to changes in observation times. In the application of the extended methodology, it was revealed that the simple percentile confidence intervals previously applied, no longer sufficed due to the skew distribution of expected times to threshold. This added a degree of computational complexity in the form of jackknife estimation, and a larger number of bootstrap replications required. It was observed that erroneously ignoring the residual correlation when it is strong, may result in substantial overestimation of the time to threshold. It is therefore crucial that careful consideration is given to structure of the linear mixed model fitted in the first stage of our proposed approach.

Chapter 6

Estimation of Time to Threshold for Biomarkers with Limits of Detection

6.1 Introduction

In Chapters 4 and 5 we presented a flexible framework for time to threshold estimation where the biomarker of interest was a continuous, normally distributed variable. The first stage of the approach involved the fitting of a linear mixed model. In the study of biomarkers, particularly in the field of HIV, data can be subject to a lower or upper limit of detection (LOD). A limit of detection can be formally defined as a certain threshold value, below or above which the measurements are not quantifiable. In these cases the crude approach is to impute the 'undetectable' value by half or exactly the limit of detection or to analyze the outcome on the binary scale by simply categorizing a measurement as detectable or not detectable. Studies have shown that simple imputation of the limit of the detection can bias parameter and standard error estimates (Rose, 2015). Ideally, one should incorporate the partial information provided by the censored values, into the likelihood function. In this chapter, we discuss estimation of parameters in random ef-

fects models, taking into account censoring, and apply the methodology of Chapter 4 to estimate times to threshold. The methodology is applied in Section 6.3 to the viral load data from the ACTG 315 trial.

6.2 Methodology

In the presence of left censoring, the observed value of the response Y_{ij} can be written as (Y_{ij}^*, C_{ij}), where Y_{ij}^* is the observed value and C_{ij} is the censoring indicator such that Y_{ij} is observed if $C_{ij} = 0$ and is left censored if $C_{ij} = 1$. Hence

$$Y_{ij} \begin{cases} = Y_{ij}^* & \text{if } C_{ij} = 0 \\ \leq d & \text{if } C_{ij} = 1. \end{cases} \tag{6.1}$$

where d is the lower limit of detection of the assay. As defined previously $f(\boldsymbol{y}_i|\boldsymbol{b}_i)$ denotes the density function of a mixed effects model, given random effects b_i, we assume that conditional on the random effects, the observations $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ are independent then we can define the likelihood for the observed data $(\boldsymbol{Y}_{i*}, \boldsymbol{C}_j)$ as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \int \left\{ \prod_{j=1}^{n_i} (f(y_{ij}|\boldsymbol{b}_i))^{1-C_{ij}} (F(d|\boldsymbol{b}_i))^{C_{ij}} \right\} \times f(\boldsymbol{b}_i) d\boldsymbol{b}_i. \tag{6.2}$$

That is, conditioning of the random effects, an undetectable measurement contributes the cumulative normal density,

$$F(d|\boldsymbol{b}_i) \equiv P(Y_{ij} < d|\boldsymbol{b}_i)$$

to the likelihood function. Unlike the likelihood for the standard linear mixed model discussed in Section 3.1 the expression in (6.2) does not have a closed form expression and numerical integration is necessary. After estimating the parameters for the model, the same methodology that assumes conditional independence (Chapter 4) can be applied to estimate the probabilities of interest and the expected time to threshold:

$$\begin{aligned} S_{ij}(X_i, Z_i, \boldsymbol{b}_i, \boldsymbol{\beta}) &= C_{ij-3} P(Y_{ij-2} > k) P(Y_{ij-1} \leq k) P(Y_{ij} \leq k) \\ &= C_{ij-3} [1 - \tilde{\Phi}_{ij-2}(k)][\tilde{\Phi}_{ij-1}(k)][\tilde{\Phi}_{ij}(k)], \end{aligned}$$

where C_{ij-3} denotes the 'continuation probability' at time t_{ij-3} and $\tilde{\Phi}_{ij}(k)$ is a cumulative normal distribution with mean $\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i$ and variance σ^2. The recursive relationship of continuation probabilities, derived in Appendix A also holds in this framework:

$$C_{ij} = C_{j-2}[1 - \Phi_{ij-1}(k)][\Phi_{ij}(k)] + C_{j-1}[1 - \Phi_{ij}(k)].$$

6.3 Application: ACTG 315 Study

In this section, we apply the methodology presented above, in conjunction with the time to threshold methodology discussed in Chapter 4, to estimate the time to treatment success in HIV positive individuals on ARV therapy. We analyse the viral load data described in Section 2.1.3 from the 46 patients in the ACTG 315 study. For the purpose of exploratory analysis the censored observations were imputed by half the limit of detection (i.e. 1.69 on the log scale). The distribution of viral load response over time is depicted in Figure 6.1. Examining Figure 6.1 and the individual profiles of patients that were presented in Section 2.1.3, it is clear that the most rapid decline in viral load occurs in the first two weeks after initiation of treatment, after which the viral load continues to decay at a slower rate until two months. Approximately two months after treatment initiation the viral load remains relatively stable with a tendency to slightly increase. It is clear that the conventional linear model with a single slope would not fit the empirical data at hand.

6.3.1 Mixed-effects model

We propose a piecewise linear mixed model with a different slope prior to, and after the critical time points of 14 and 60 days post treatment initiation, such that

$$Y_{ij} = \begin{cases} \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i} t_{ij} + \varepsilon_{ij} & \text{if } t_{ij} \leq 14, \\ \beta_0 + b_{0i} + (\beta_1 + \beta_2) t_{ij} + (b_{1i} + b_{2i}) t_{ij} + \varepsilon_{ij} & \text{if } 14 < t_{ij} \leq 60, \\ \beta_0 + b_{0i} + (\beta_1 + \beta_2 + \beta_3) t_{ij} + (b_{1i} + b_{2i} + b_{3i}) t_{ij} + \varepsilon_{ij} & \text{if } t_{ij} > 60. \end{cases}$$

To inform the decision regarding the random effects and covariance structure in the model that incorporates left censoring into the likelihood, initial model building was performed using the standard mixed model on the data with imputed values. The fit statistics for

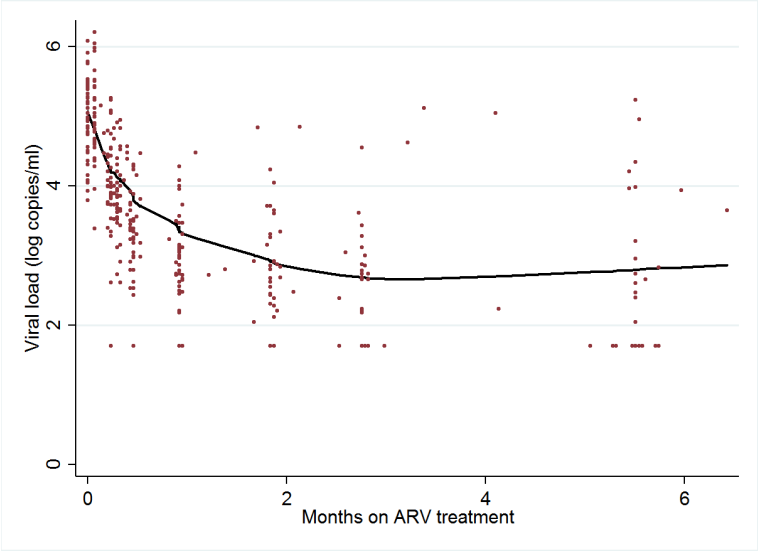


Figure 6.1: *ACTG 315 Study: Viral load measurements over time with lowess smooth curve*

four models with different random effects and residual correlation structure is presented in Table 6.1. The timescale considered for modeling purposes was expressed as years post ARV treatment initiation.

Model 1. Random effects $b_{0i}, b_{1i}, b_{2i}, b_{3i}$ and uncorrelated residuals

Model 2. Random effects b_{0i}, b_{2i}, b_{3i} and uncorrelated residuals

Model 3. Random effects b_{0i}, b_{2i}, b_{3i} and AR(1) correlation structure

Model 4. Random effects b_{0i}, b_{2i}, b_{3i} and compound symmetry correlation structure

A comparison of the fit statistics for Models 1 and 2 indicated that a the random effect corresponding to the first slope was not necessary in the model. In Models 3 and 4 we incorporated correlation into the residuals, and compared the fit statistics to that of Model 2. Using the likelihood ratio test, and comparing the BIC values it is clear that the random effects parameters and the correlation structure of the random effects sufficiently capture the correlation between responses. Using the likelihood presented in (6.2) and fitting a model with the random effects structure of Model 2, the model parameters were estimated. Adaptive Gaussian quadrature with 10 points was used to integrate over

Table 6.1: *ACTG 315 Study: Fit Statistics for the four models fitted*

		fit statistic		
Model	#cov par	-2logL	AIC	BIC
1	11	545.5	567.5	587.6
2	7	551.8	565.8	578.6
3	8	549.4	565.4	580.8
4	8	551.8	567.8	582.5

random effects. The SAS NLMIXED code for fitting the model, as well as the code for time to threshold estimation is presented in Appendix C. These results, as well as those emanating from the mixed model with imputed values are presented in Table 6.2. The magnitude of fixed effects estimates β_1, β_2 draw attention to the rapid response to treatment in the first two weeks after treatment, and the subsequent rapid rebound of viral load six weeks later. After 8 weeks on treatment the viral load exhibits a stabler trend with a tendency to slightly increase. The highest variance was observed in the random slope parameter (d_{44}) that captures the trajectory after 8 weeks on treatment. Significant negative correlation was observed between the two random slopes cov(b_{2i}, b_{3i}), confirming that individuals who experience more stable viral loads between week 2 and 8 are more likely to experience virological rebound after 8 weeks. We observe substantial differences in covariance parameter estimates between the standard model with imputed values and the model that incorporates censoring into the likelihood. This concurs with the results from the simulation study conducted by Thiébaut (2004). Subject-specific prediction was performed using the fixed-effects and random-effects estimates from the final model. The observed and fitted profiles for 6 selected individuals are presented in Figure 6.2.

6.3.2 Expected Time to Threshold

In this section, we estimate the time to confirmed treatment success, defined as two consecutive viral load measurements less than 1000 copies/ml. The maximum duration of

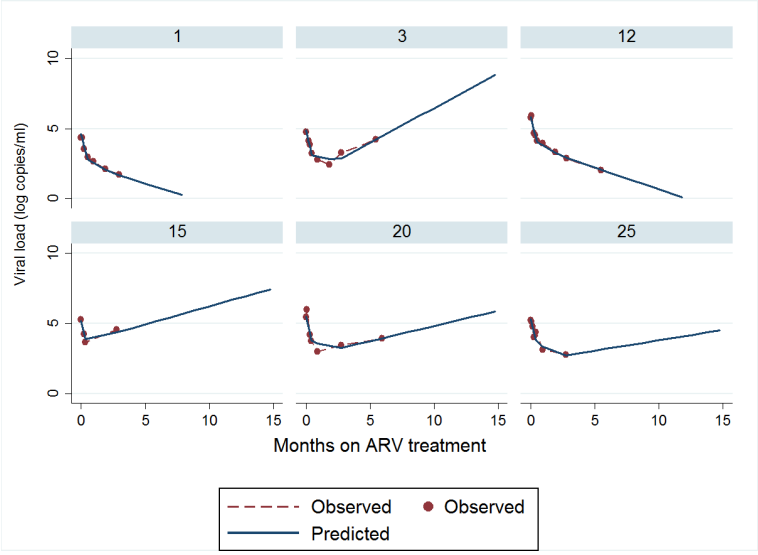


Figure 6.2: *ACTG 315 Study: Observed and fitted profiles*

follow-up in the ACTG 315 study was 6 months. To facilitate prediction of treatment success that may have occurred after the six month period, we allow a maximum period of 15 months where we consider the individual to have the potential for treatment success. The visits outside of the observation period were created at monthly intervals, in accordance with the study design. For illustrative purposes the estimates are presented for same six selected patients presented in Figure 6.2. Figure 6.3 depicts the observed profiles of these patients, with reference line at 1000 copies/ml. Patient 1 entered the study with a viral load load of 4.36 on the log scale, which declined for the duration of follow-up. Patient 3, in contrast, entered the study with a high viral load of 4.8 and experienced a rapid decline in the first month post treatment, but exhibited an increasing viral load thereafter, indicative of treatment failure. A similar trend was observed for Patients 15 and 20. Patient 12, similar to patient 1, was able to maintain a low viral load for the period considered. In Figure 6.4 we present predicted probability of single measurement being less than 1000 copies/ml. These predicted probabilities adequately capture the features of the observed profiles. Patient 1, as expected, had the highest probability of reaching a low viral load in the first month of treatment. Patient 12 had a high probability of experiencing a low

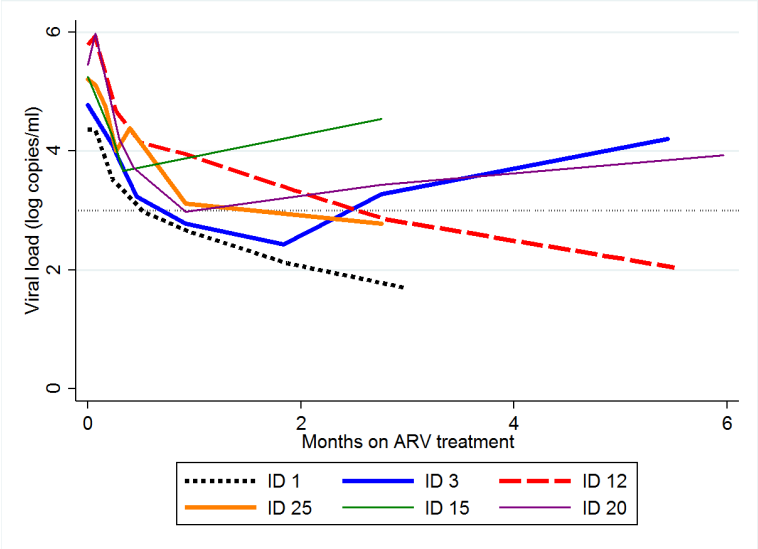


Figure 6.3: *ACTG 315 Study: Viral load profiles for six selected patients*

viral load after approximately 4 months. Patient 15 had a zero probability of experiencing a measurement less than 1000 throughout the period considered. The estimated time

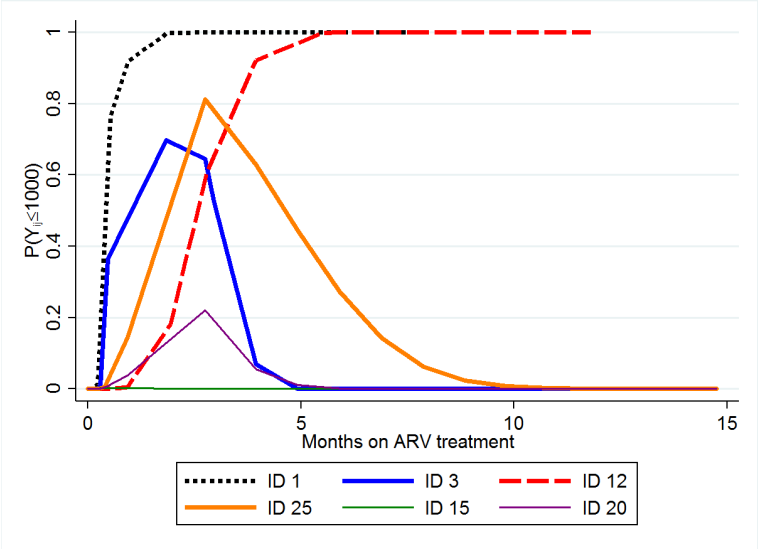


Figure 6.4: *ACTG 315 Study: Estimated probability of single viral load being less than 1000 log copies/ml*

to confirmed treatment success, with 95% confidence intervals is presented for the six patients in Table 6.3. Overall, the precision of the time to threshold estimates were high, with narrow confidence observed in patients who had more than three months of follow-up. The confidence interval for $\widehat{T}_i$ for Patient 25, who was under follow-up until day 84 and contributed 6 viral load measurements, is considerably wider. This is understandable, considering the form of the model fitted, where a separate slope is fitted for the period $t_{ij} > 60$ days. In the sensitivity analysis that was conducted in Section 4.4.3 we noted that estimated times to threshold should not be interpreted in the absence of the predicted probabilities for individuals who exhibit a consistently 'high' flat profile. This is observed for Patient 20 who maintained a high viral load for the duration of follow-up. Although this individual had a probability less than 0.20 of reaching a threshold less than 1000 copies/ml for the entire period considered (Figure 6.4), the estimated threshold is not exactly zero ($\widehat{T}_i$=0.071). We therefore recommend that all estimated times less than 0.50 years be examined in conjunction with the predicted probabilities. The distribution of the estimated time to confirmed treatment success is presented in Figure 6.5. Out of the 46 patients, 5 patients were 'treatment resistant' and had predicted times to threshold of 0 years. Approximately 80% of patients were expected to have experienced treatment success before 3 months.

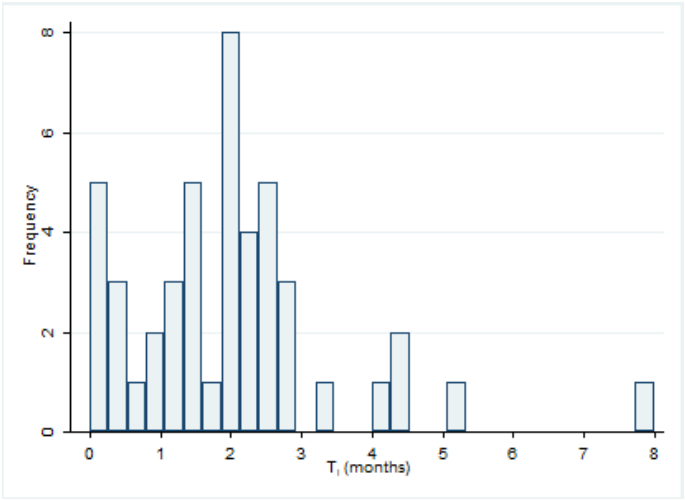


Figure 6.5: *ACTG 315 Study: Distribution of estimated time to confirmed treatment success*

Table 6.2: *ACTG 315 Study: Parameter estimates (standard errors) for the standard model with imputed values and direct likelihood model*

Effect	Parameter	Likelihood with imputed values	Direct likelihood
	Fixed effects estimates (s.e.)		
Intercept	β_0	5.073 (0.087)	5.071 (0.085)
Time (years)	β_1	-47.065 (1.517)	-46.773 (1.478)
	β_2	42.5615 (1.943)	41.709 (1.99)
	β_3	6.491 (1.558)	5.157 (2.124)
	Covariance parameter estimates (s.e.)		
Measurement error	σ^2	0.106 (0.010)	0.099 (0.010)
var(b_{0i})	d_{11}	0.283 (0.064)	0.525 (0.059)
var(b_{2i})	d_{33}	28.017 (7.361)	5.997 (0.861)
var(b_{3i})	d_{44}	75.516 (23.810)	10.882 (2.144)
cov(b_{0i}, b_{2i})	d_{13}	-0.698 (0.494)	-0.407 (0.583)
cov(b_{0i}, b_{3i})	d_{14}	1.105 (0.873)	1.006 (1.267)
cov(b_{2i}, b_{3i})	d_{34}	-37.216 (12.096)	-39.989 (16.878)
	Fit statistics		
-2 REML log-likelihood		551.8	507.9
AIC		565.8	529.9
BIC		578.6	550

Table 6.3: *ACTG 315 Study: Estimated time to threshold for the six selected patients*

Patient	Baseline viral load	$\widehat{T}_i$	95% CI
1	4.362	1.309	(1.207, 1.413)
3	4.771	1.474	(1.251, 1.675)
12	5.785	4.516	(4.268, 4.628)
15	5.255	7.95×10^{-8}	$(2.57 \times 10^{-9}, 1.46 \times 10^{-5})$
20	5.447	0.071	(0.031,0.118)
25	5.204	2.624	(0.974, 4.571)

6.4 Discussion

In this Chapter, we discussed parameter estimation in a mixed model for a continuous biomarker subject to left censoring. The previously proposed methodology of Chapter 4 was adapted and applied to the ACTG 315 study. Unlike the models with a single slope that were sufficient for the AAA and Sinikithemba cohort study, the ACTG 315 study warranted that a model with a more complex piecewise-linear structure, with three random effects, was applied. We observed that even in this extended setting with censoring, our methodology proved to be computationally efficient. A drawback of fitting the model with separate slopes was that individuals who may have dropped out prematurely may not have sufficient data after the last critical point to accurately predict time to threshold. In this application we focused on estimating the time to confirmed treatment success. Similarly, further analysis can be done to estimate the time to treatment failure as defined by the WHO WHO (2015) as two consecutive viral loads exceeding 1000 after 3 months on treatment. Unfortunately, the ACTG 315 study would not be suitable for this purpose since the most intensive period of follow-up was in the first three months of treatment, with 83% of the observations falling into that time frame.

Chapter 7

Estimation of Time to Threshold for Ordinal Biomarkers

7.1 Introduction

In the preceeding chapters, the methodology that was proposed was developed for the analysis of a continuous biomarker. In the field of psychiatry, established scales are often used to categorize a patient's health status, and it is these scales that are often used as the endpoints in clinical trials. Studies have found that applying a continuous model to ordinal data can yield correlated residuals as the continuous model would not take into account the ceiling and floor effects of an ordinal outcome. This can result in biased estimates of coefficients. Secondly, for the purposes of time to threshold modeling, where prediction is of key interest, a continuous model can yield predicted values outside of the range of the ordinal variable. Hence, in order to accurately estimate the time to threshold of an ordinal outcome it is crucial that a model for ordinal responses is fitted. In Section 7.2 we extend the approach presented in Chapter 4 to handle ordinal outcomes and propose new expressions for the functions S_{ij} and C_{ij} that were previously introduced.

This methodology is applied to the data from the Schizophrenia trial in Section 7.3, where time to threshold estimates and confidence intervals are presented for selected patients.

7.2 Methodology

7.2.1 Expected Time to Attain a Threshold with Persistence Criteria

Letting Y_{ij} denote the ordinal response on individual i and timepoint j, and $c = 1, 2, \ldots, C$ denote the response categories, the time to threshold can be expressed as:

$$T_i = \min\{j \geq 2 : Y_{ij-1} \leq c, Y_{ij} \leq c\}. \tag{7.1}$$

It follows that the expected time for individual i to attain two consecutive measurements less than the threshold c can be expressed as follows:

$$\begin{aligned} E(T_i) &= t_{i2}P(Y_{i1} \leq c, Y_{i2} \leq c) \\ &\quad + t_{i3}P(Y_{i1} > c, Y_{i2} \leq c, Y_{i3} \leq c) \\ &\quad + t_{i4}\left\{\begin{array}{l} P(Y_{i1} > c, Y_{i2} > c, Y_{i3} \leq c, Y_{i4} \leq c) \\ + P(Y_{i1} \leq c, Y_{i2} > c, Y_{i3} \leq c, Y_{i4} \leq c) \end{array}\right\} \\ &\quad + \ldots \\ &= \sum_{j=2}^{\infty} t_{ij}S_{ij}, \end{aligned} \tag{7.2}$$

Following the notation presented in Section 3.2, we specify a mixed-effects proportional odds logistic regression model which satisfies

$$\begin{aligned} \lambda_{ij,c} &= \text{logit}[P(Y_{ij} \leq c|X_i, Z_i)] \\ &= \alpha_c + x_{ij}^{'}\boldsymbol{\beta} + z_{ij}^{'}\boldsymbol{b}_i. \end{aligned}$$

It follows that

$$P(Y_{ij} \leq c|X_i, Z_i, \boldsymbol{b}_i) = \frac{1}{1 + \exp(-\lambda_{ij,c})}$$

for $C - 1$ strictly increasing model thresholds α_c. Since conditional independence is assumed in the mixed effects ordinal regression, it follows that the relationships that were

derived in Section 4.3 and Appendix A, hold true in this setting, the stopping probability S_{ij}, conditional on the fixed and random-effects, can be expressed as

$$\begin{aligned} S_{ij}|X_i, Z_i, \boldsymbol{b}_i, \boldsymbol{\beta}) &= C_{ij-3}P(Y_{ij-2} > c)P(Y_{ij-1} \leq c)P(Y_{ij} \leq c) \\ &= C_{ij-3}\left(\frac{\exp(-\lambda_{ij-2,c})}{1+\exp(-\lambda_{ij-2,c})}\right) \\ &\quad \times \left(\frac{1}{1+\exp(-\lambda_{ij-1,c})}\right)\left(\frac{1}{1+\exp(-\lambda_{ij,c})}\right), \end{aligned}$$

where C_{ij-3} denotes the 'continuation probability' at time t_{ij-3}. The continuation probability C_{ij} can also be expressed as a function of it's predesessors C_{ij-1} and C_{ij-2} such that

$$\begin{aligned} C_{ij} &= C_{ij-2}P(Y_{ij-1} > c)P(Y_{ij} < c) + C_{ij-1}P(Y_{ij} > c) \\ &= C_{ij-2}\left(\frac{\exp(-\lambda_{ij-1,c})}{1+\exp(-\lambda_{ij-1,c})}\right)\left(\frac{1}{1+\exp(-\lambda_{ij,c})}\right) \\ &\quad + C_{ij-1}\left(\frac{\exp(-\lambda_{ij,c})}{1+\exp(-\lambda_{ij,c})}\right) \end{aligned}$$

7.2.2 Estimation and Inference

It follows from Section 7.2 that $E(T_i)$ is a function of the parameters $\boldsymbol{\beta}$, $\boldsymbol{b}_i$, and α_c. Hence $\widehat{T}_i$, the estimate of $E(T_i)$, can be computed by substituting each unknown parameter by its corresponding estimate. We again propose a conditional version of the non-parametric bootstrap to compute 95% confidence intervals for $\widehat{T}_i$ as follows:

Step 1. Individual i is removed from the full dataset resulting in $N-1$ cases

Step 2. Sample $N-1$ subjects with replacement from the dataset in Step 1

Step 3. Append the data of individual i to the bootstrap sample

Step 4. Compute $\widehat{T}_i$

This process is repeated 1000 times. The computation of $\widehat{T}_i$, and the bootstrap confidence interval can be achieved with relatively basic programming following the estimation of the model parameters using PROC GLIMMIX or PROC NLMIXED. The program to estimate $\widehat{T}_i$, incorporating the recursive formula for continuation probabilities is the same as that presented in Appendix A.2 with the exception of the first stage of model fitting.

7.3 Application: Schizophrenia Trial

In Section 2.3.2 we discussed the CGI-S (Clinical Global Impressions Severity) scale which is measured on a 7 point ordinal scale:

$$\text{CGI-S} = \begin{cases} 1 & \text{Normal} \\ 2 & \text{Borderline ill} \\ 3 & \text{Mildly ill} \\ 4 & \text{Moderately ill} \\ 5 & \text{Markedly ill} \\ 6 & \text{Severely ill} \\ 7 & \text{Extremely ill} \end{cases}$$

In a study conducted by (Dunayevich, 2006), remission was defined as consecutive Clinical Global Impressions-Severity (CGI-S) scores ≤ 3. In this section we apply this definition of remission to the Schizophrenia trial to gain insight into the time to remission of patients in the study.

7.3.1 Mixed-effects Model

We considered a model of the form

$$\text{logit}[P(Y_{ij} \leq c|X_i, Z_i)] = \begin{cases} \alpha_c + b_{0i} + \beta_{1,H} t_{ij} + b_{1i} t_{ij} & \text{if treated with Haloperidol,} \\ \alpha_c + b_{0i} + \beta_{1,R} t_{ij} + b_{1i} t_{ij} & \text{if treated with Risperidone.} \end{cases}$$

The parameter estimates and fit statistics for the model with random intercept only (RI) and random intercept and slope (RI, RS) are presented in Table 7.1. It is clear that random intercepts contribute the highest variance but also that the model that includes random slope provides a significantly better fit to the data. There was a significant effect of treatment, with patients on Risperidone having a higher chance of experiencing lower CGI-S scores over time, than patients on the active control Haloperidol.

Table 7.1: *Schizophrenia Trial: Parameter estimates (standard errors) for the random intercept (RI) and the random intercept and random slope (RI,RS) mixed effects proportion odds model.*

Effect	Parameter	Model with RI	Model with RI and RS
Fixed effects estimates (s.e.)			
Intercept	α_1	-10.955 (0.465)	-11.579 (0.522)
	α_2	-6.913 (0.252)	-7.242 (0.258)
	α_3	-3.833 (0.208)	-3.682 (0.171)
	α_4	-0.301 (0.188)	-0.076 (0.143)
	α_5	2.57 (0.201)	2.828 (0.168)
	α_6	8.138 (0.38)	10.359 (0.623)
Time	$\beta_{1,H}$	0.313 (0.023)	0.368 (0.046)
	$\beta_{1,R}$	0.350 (0.023)	0.412 (0.045)
Covariance parameter estimates (s.e.)			
var(b_{0i})	d_{11}	12.539 (1.104)	5.934 (0.42)
cov(b_{0i}, b_{1i})	d_{12}		0.170 (0.068)
var(b_{1i})	d_{22}		0.268 (0.049)
Fit statistics			
-2 log-likelihood		5568.76	5492.17
AIC		5586.76	5514.17
BIC		5623.8	5559.44

7.3.2 Expected Time to Threshold

In the Schizophrenia trial patients were followed up for a maximum of 8 weeks. To allow a window for estimation of times to thresholds that may be reached after 8 weeks, we created additional visits up to a maximum of 20 weeks. We now, using the parameter

estimates from the random intercepts and slope model, predict the probability of an individual experiencing CGI-S$\leq$ 3 at every time point. The longitudinal profiles for six selected patients is presented in Figure 7.1. The predicted probabilities $P(Y_{ij} \leq 3)$, for these selected patients are presented in Figure 7.2. We observed high precision in

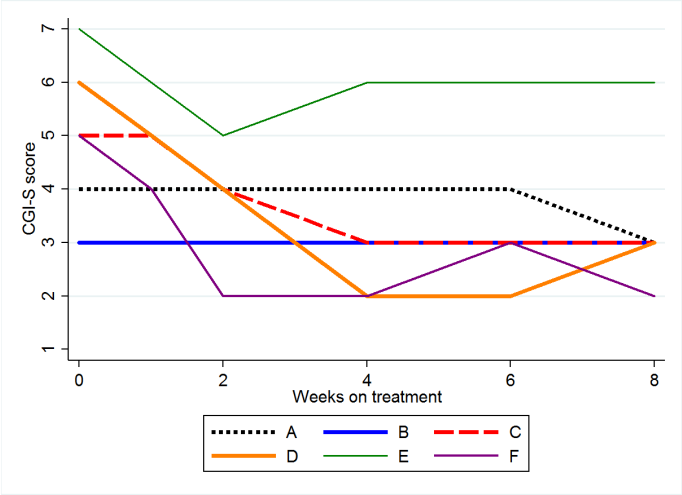


Figure 7.1: *Schizophrenia Trial: Longitudinal CGI-S measurements for selected patients.*

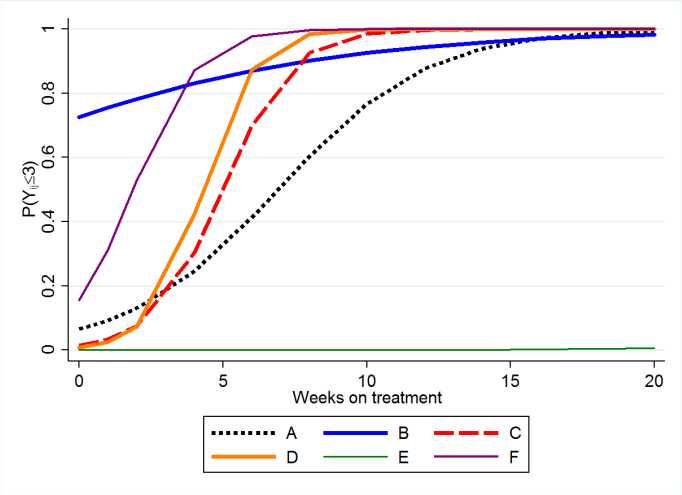


Figure 7.2: *Schizophrenia Trial: Estimated probabilities, $P(Y_{ij} \leq 3)$, for selected patients.*

nearly all estimates, since 94% of patients in the study had at least three measurements

recorded. The estimated time to two consecutive CGI-S≤ 3 is presented in Table 7.2. The rapid decline in CGI-S observed for patient F is clearly captured by both the predicted probabilities $P(Y_{ij} \leq 3)$ and the estimated time to remission of 3.16 weeks. Patient A, who had a baseline CGI-S score of 4, is only expected to reach remission after 10 weeks. We now draw attention to patient E who had a zero probability of reaching a low CGI-S score for the entire period. Unlike in the three other datasets, this type of patient is not an anomaly in the Schizophrenia trial. We observed that 114 of the total 453 patients were expected to never reach remission. In Figure 7.3 we present the predicted probabilities for

Table 7.2: *Schizophrenia Trial: Estimated time to threshold for six patients.*

Patient	Baseline CGI-S	Treatment	$\widehat{T}_i$	95% CI
A	4	Risperidone	10.3456	(10.1283, 10.5934)
B	3	Haloperidol	2.4768	(2.2095, 2.7629)
C	5	Haloperidol	8.2966	(8.0889 8.5333)
D	6	Haloperidol	7.4633	(7.2484, 7.6866)
E	7	Risperidone	0.0005	(0.0001, 0.0025)
F	5	Risperidone	3.1601	(4.4503, 5.1372)

all these patients and the observed profiles for a selected group in Figure 7.4. For ease of inspection the observed profile curves are presented on separate axes and not overlayed on the same graph. This observation does raise doubt regarding the validity of using an absolute cutoff value to determine remission in patients with schizophrenia, and suggests that 'change' scores like the CGI-I, which measures relative improvement since baseline, may be more appropriate.

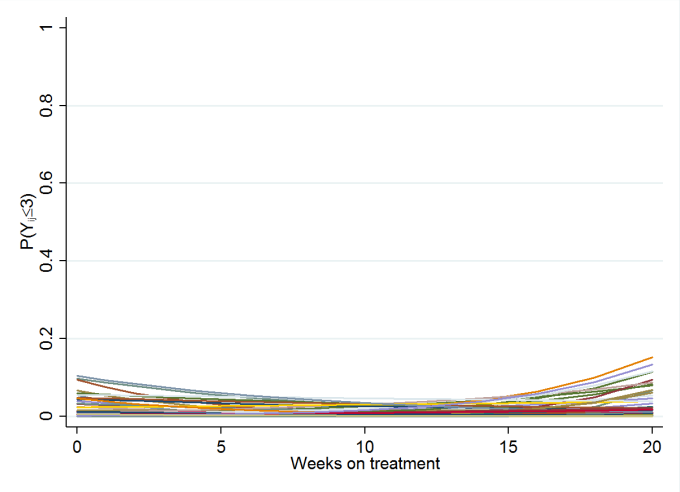


Figure 7.3: *Schizophrenia Trial: Estimated probabilities for 'treatment resistant' patients.*

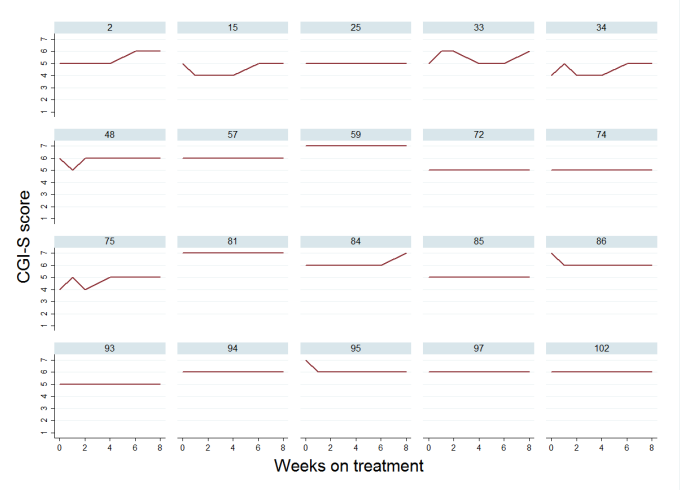


Figure 7.4: *Schizophrenia Trial: Longitudinal CGI-S measurements for 'treatment resistant' patients.*

7.4 Discussion

In this chapter, we extended the methodology, that was developed for time to threshold estimation for a continuous biomarker, so that time to threshold may also be estimated for ordinal outcomes. Through the observation that the recursive relationships still hold in this setting, a high degree of computational efficiency was once again achieved. The methodology was applied to the Schizophrenia dataset and was able to draw attention to the low rate of remission in the 8 week period that was observed. As further work the same methodology can be applied to analyse the CGI-I (Clinical Global Impression Improvement) scale, designating a value 2 (Much improved) as the endpoint of interest. The model we fitted assumed that the proportional odds assumption holds, but it would be possible to fit alternate models that relax this assumption. As a starting point separate intercepts can be modelled for each treatment and implemented in PROC NLMIXED.

Chapter 8

Sensitivity Analysis: Missing Data

8.1 Introduction

The methodology for time to threshold estimation that was proposed in Chapters 4, 5, 6, and 7, rely on the assumption that the missing data mechanism is ignorable. That is, data were missing at random (MAR) and the parameters governing the missing data mechanism were distinct from the parameters in the models that were estimated. The MAR assumption implies that individuals who dropout would have similar outcomes to individuals who continue in the study, given their observed data prior to dropout. To explore the impact of deviations from the missing at random (MAR) assumption on the estimated time to threshold, it is advisable to conduct a sensitivity analysis, within which missing not at random (MNAR) models play a major role (Molenberghs and Kenward, 2007).

In this Chapter we return to the AAA data that were introduced in Section 2.2.2 and analyzed in Chapter 4, and examine the effect of departures from the MAR assumption on the estimated time to threshold.

8.2 Methodology

A straightforward sensitivity analysis for the MAR assumption in multiple imputation is based on the pattern-mixture model approach that was discussed in Section 3.3. Molenberghs and Kenward (2017) outline the sequence of steps for pattern-mixture-based multiple imputation as follows:

Step 1. Fit a model to the pattern t-specific identifiable densities: $f_t(y_1, \ldots, y_t)$,

Step 2. Select an identification method of choice (ACMV, CCMV or NCMV),

Step 3. Using the identification method in Step 2, determine the conditional distributions of the unobserved outcomes, given the observed outcomes:

$$f_t(y_{t+1}, \ldots, y_T | y_1, \ldots, y_t).$$

Step 4. Using MI methodology, draw M multiple imputations for the unobserved outcomes, given the observed outcomes and the correct pattern-specific density,

Step 5. Analyze each of the M multiply-imputed datasets using the method of choice,

Step 6. Conduct inferences using Rubin's rules.

It is also possible to modify the models in Step 3 or imputations in specific, explicit ways that capture MNAR features. This is formally defined as 'controlled imputation'. A common feature that all controlled imputation methods have is the ability to construct MNAR models using components from an MAR model, with the possible addition of fixed, sensitivity parameters (Kenward, 2015). In imputation with delta adjustment, after dropout in a given pattern, subjects may be made to shift with an amount δ, relative to the MAR-based prediction. Other available methods include tipping point analysis and control-based imputation.

Table 8.1: *AAA study: Patterns of missing data for AAA diameter.*

	Visit							
Pattern	Baseline	6 months	12 months	18 months	24 months	30 months	36 months	Total patients
1	X	X	X	X	X	X	X	7
2	X	X	X	X	X	X		12
3	X	X	X	X	X			27
4	X	X	X	X				7
5	X	X	X					21
6	X	X						16
7	X							10

8.3 Sensitivity Analysis: AAA Study

8.3.1 Missing Data in the AAA Study

In the AAA study, where the maximum duration of follow-up was 3 years, only 7(7%) patients had complete data for all 7 visits. The different patterns of missing data and the number of patients in each of these patterns are presented in Table 8.1. The pattern of missing data is clearly monotone, that is, if a particular observation is missing on an individual, so too are all subsequent measurements. By the fourth visit, scheduled to occur two years post enrolment, 53% of patients had already left the study. Departure from the study could be due to several reasons, such as : the patient underwent aortic repair surgery, the patient died, or that the patient was unable to attend further visits for reasons related or unrelated to the study. In the data that we were provided, no information regarding the reason for dropout was provided. Figure 8.1 presents the observed mean AAA diameter at each of the timepoints for the participants who dropped out at the subsequent visit and for those who did not drop out at the subsequent visit. There was no significant difference in the mean diameter between patients who dropped out and those who did not, at visits 1, 2, 3, 4 and 6. However, at visit 5, the mean diameter of patients who

subsequently dropped out was significantly higher than the mean observed for those who did not dropout (49.96 vs 43.73, p-value 0.008). The mean AAA trajectory over time, for each dropout pattern described in Table 8.1, is presented in Figure 8.2.

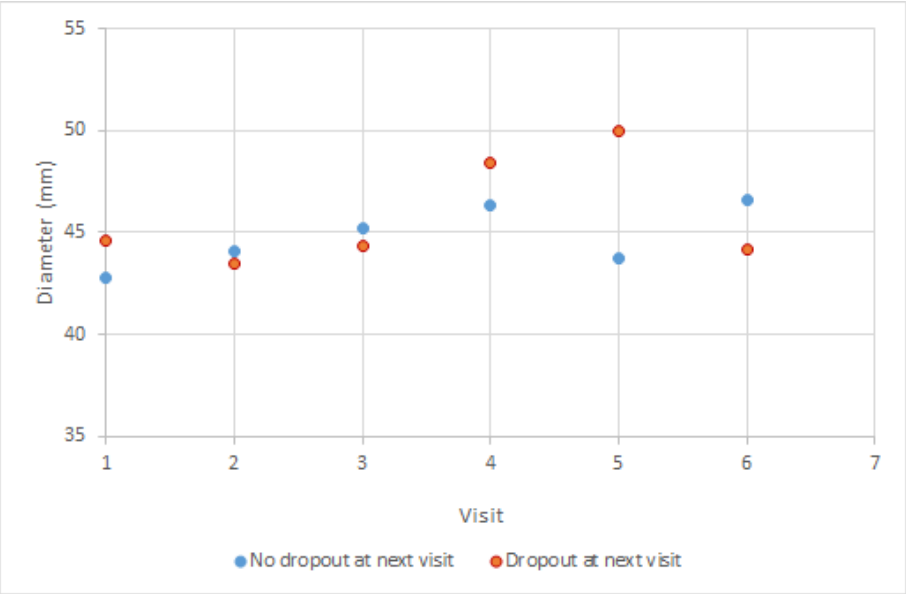


Figure 8.1: *AAA study: Comparison of mean AAA diameter between patients who did and did not dropout at the subsequent visit.*

8.3.2 Sensitivity Analysis: Linear Mixed Model Parameter Estimates

Since the pattern of missing data in the AAA study is clearly monotone, we apply the monotone regression method for imputation. In this method a regression model is fitted for each variable with missing values, with the previous variables as covariates. Naturally a problem arises if the number of covariates exceeds the number of observations for a particular variable. In the AAA data, age, BMI and all 7 AAA measurements were initially considered in the imputation model. Unfortunately, due to the limited number of observations for the last visit (visit 7), we were unable to include this measurement in the imputation model. Ten datasets were imputed under each of the following four MNAR

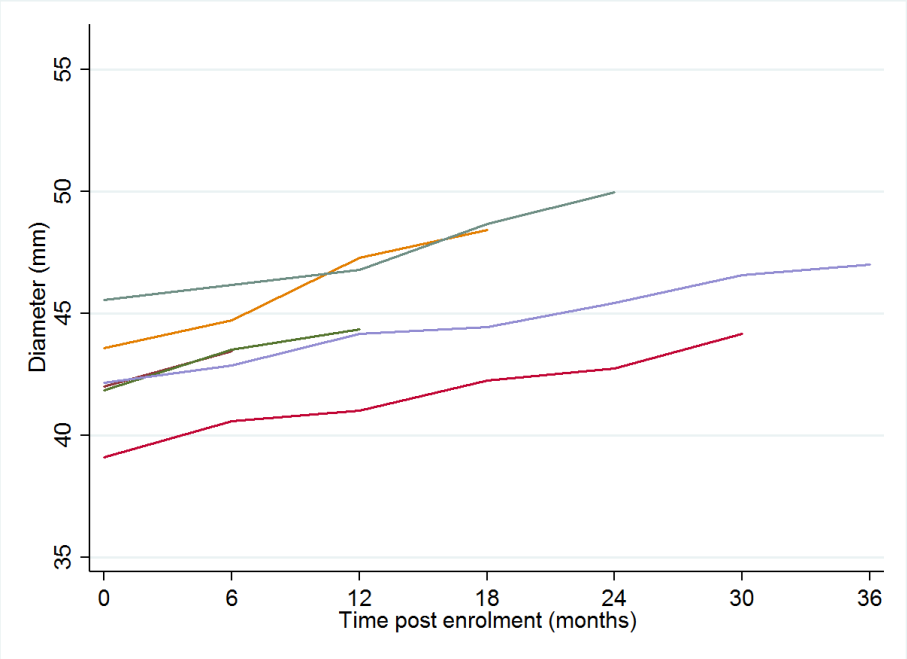


Figure 8.2: *AAA study: Mean AAA diameter for each pattern of dropout.*

scenarios:

$\delta = -1$**:** Missing values at visits 2 and 3 are shifted down by 1 mm,

$\delta = 1$**:** Missing values at all visits are shifted up by 1 mm,

$\delta = 2$**:** Missing values at all visits are shifted up by 2 mm,

$\delta = 3$**:** Missing values at all visits are shifted up by 3 mm.

The rationale behind applying a negative shift only to earlier visits, is that a negative shift at all visits would result in model with a negative slope since imputations are generated sequentially. In the case of the evolution of AAA diameters over time, a negative slope is implausible. The estimated parameters under the ignorable or direct likelihood (DL) approach excluding the last measurement, and the four MNAR scenarios, are presented in Table 8.2. Negligible differences were observed in parameter estimates under the DL approach that included all measurements (Table 4.4), and the DL approach that excluded

the last measurement. As expected, the model with a negative shift parameter, resulted in lower values for all time-effects ($\beta_{1,NM}$, $\beta_{1,OW}$, $\beta_{1,OB}$) , compared to the DL approach and MNAR scenarios with positive shift parameters. Marked differences were also observed in the estimated fixed intercepts under all scenarios. The SAS code for implementation of the four MNAR based imputation scenarios is presented in Appendix E.1.

Table 8.2: *Sensitivity analysis for AAA study: Parameter estimates (standard errors) for the fitted models.*

			MNAR Imputation			
Effect	Parameter	Direct Likelihood	$\delta = -1$	$\delta = 1$	$\delta = 2$	$\delta = 3$
		Fixed effects estimates (s.e.)				
Age	$\beta_{0,age}$	0.213 (0.086)	0.197 (0.099)	0.219 (0.098)	0.232 (0.095)	0.240 (0.094)
Intercept	$\beta_{0,NM}$	27.2 (6.345)	28.787 (7.277)	26.770 (7.265)	25.395 (7.163)	24.489 (7.100)
	$\beta_{0,OW}$	27.181 (6.205)	28.188 (7.172)	26.474 (7.117)	25.275 (6.979)	24.447 (6.891)
	$\beta_{0,OB}$	29.797 (6.095)	30.294 (6.982)	29.035 (6.907)	28.237 (6.763)	27.711 (6.674)
Time	$\beta_{1,NM}$	1.965 (0.384)	1.943 (0.388)	2.861 (0.463)	3.596 (0.578)	4.309 (0.708)
	$\beta_{1,OW}$	2.66 (0.335)	2.377 (0.398)	3.313 (0.491)	4.110 (0.625)	4.913 (0.778)
	$\beta_{1,OB}$	2.307 (0.551)	1.835 (0.677)	2.831 (0.735)	3.698 (0.859)	4.586 (1.020)
		Covariance parameter estimates (s.e.)				
var(b_{0i})	d_{11}	41.683 (6.211)	40.427 (6.292)	40.665 (6.327)	40.698 (6.363)	40.589 (6.410)
cov(b_{0i}, b_{1i})	d_{12}	4.101 (1.698)	4.830 (1.987)	5.026 (2.106)	5.191 (2.324)	5.337 (2.668)
var(b_{1i})	d_{22}	3.44 (0.812)	3.296 (1.786)	3.899 (1.891)	5.387 (2.546)	7.723 (3.625)
Measurement error	σ^2	1.978 (0.198)	2.867 (1.147)	2.943 (1.159)	3.425 (1.334)	4.250 (1.658)

8.3.3 Sensitivity Analysis: Expected Time to Threshold

In this section we estimate time taken for individuals to reach two consecutive aneurysm diameter measurements greater than 50 mm. Under each of the MNAR scenarios, we computed $\widehat{T}_i$ for all patients in each imputed dataset. The values for all 10 imputations were thereafter averaged, resulting in a single $\widehat{T}_i$ value for each patient, under each scenario. To quantify the difference in $\widehat{T}_i$ between the DL approach and the MNAR scenarios, we calculated the percentage difference:

$$\frac{\widehat{T}_i\text{MNAR} - \widehat{T}_i\text{DL}}{\widehat{T}_i\text{DL}} \times 100$$

These results are summarized in Table 8.3. The median differences observed were all less than 10% in absolute value. As the value of δ increases, resulting in a steeper increase in AAA diameter, the time taken to reach the threshold of 50 mm decreases. Differences greater than 200% in absolute value were examined in detail. These large differences were observed in patients with estimated times to threshold close to 0, that is, patients who are expected to never reach the threshold 50 mm. The observed profiles for these patients are presented in Figure 8.3. Unsurprisingly, all these patients entered the study with small AAA diameters, and progressed at a very slow rate. In these cases, even a minor departure from MAR can have a significant effect on the estimated time to threshold. In Table

Table 8.3: *Sensitivity analysis for AAA study: Summary of percentage difference in $\widehat{T}_i$ for each MNAR scenario.*

MNAR scenario	Median %Difference	IQR %Difference	Min %Difference	Max % Difference
$\delta = -1$	2.53	-29.18;13.05	-473.15	7.48×10^{11}
$\delta = 1$	-3.06	-21.08;8.01	-683.39	1.07×10^{12}
$\delta = 2$	-8.30	-31.67;11.59	-604.08	1.28×10^{12}
$\delta = 3$	-9.78	-38.97;15.20	-729.95	1.02×10^{12}

8.4 the estimated time to threshold for 9 selected patients under the direct likelihood approach and the four MNAR approaches are presented. The observed profiles for these

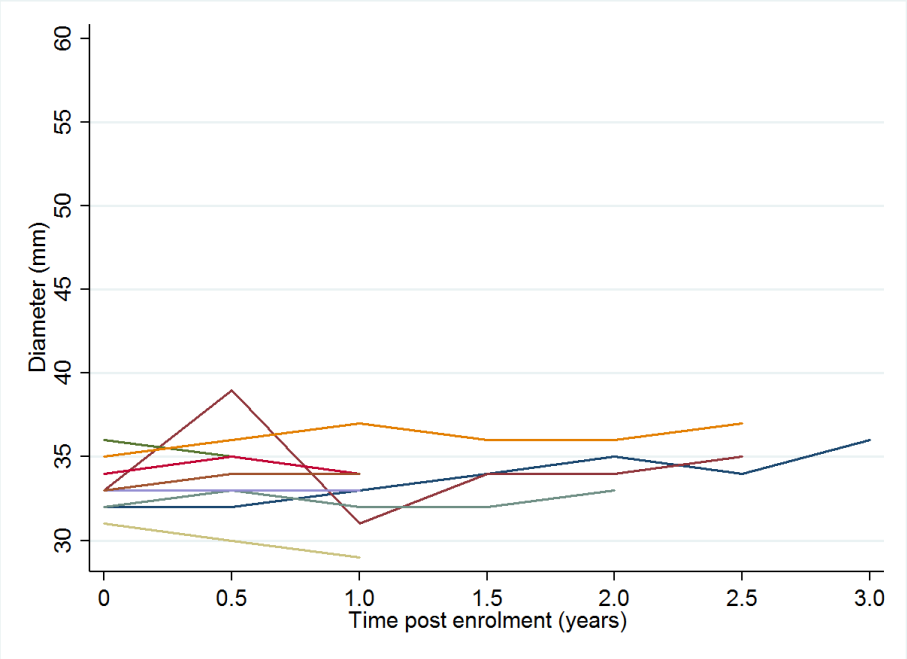


Figure 8.3: *Sensitivity analysis for AAA study: Profiles for patients with differences in $\widehat{T}_i$ exceeding 200%.*

9 patients is in Figure 8.4. In patients with at least 4 observations, the estimated times to threshold were robust to departures from the MAR assumption.

Table 8.4: *Sensitivity analysis for AAA study: $\widehat{T}_i$ for 9 selected patients under direct likelihood (DL) and the four MNAR scenarios.*

			MNAR Imputation			
Patient	Total visits	DL	$\delta = -1$	$\delta = 1$	$\delta = 2$	$\delta = 3$
1027	5	2.171	2.390	2.283	2.214	2.163
1035	5	1.117	1.151	1.194	1.231	1.268
1038	6	5.877	6.494	5.977	5.748	5.630
1041	4	0.911	0.812	0.881	0.935	0.980
1074	5	0.513	0.529	0.605	0.668	0.734
1080	2	2.300	2.791	1.783	1.543	1.411
1145	6	7.720	8.249	7.484	7.158	6.965
1164	4	3.393	4.017	3.189	2.781	2.526
1216	3	8.229	5.153	5.924	6.160	4.710

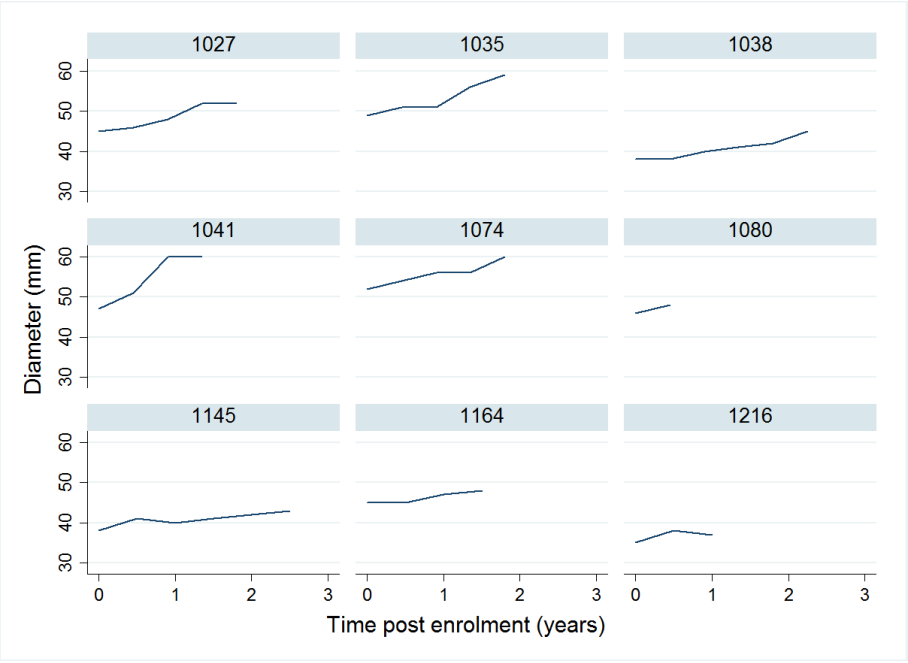


Figure 8.4: *Sensitivity analysis for AAA study: Observed profiles for the 9 patients in Table 8.4.*

8.4 Discussion

In this Chapter, we applied multiple imputation with delta adjustment, to assess the extent to which conclusions may differ if data were missing not at random. We observed that even minor shifts can have a substantial impact on the estimated time to threshold for individuals with a relatively 'flat' profile or fewer than three measurements. This concurs with our earlier findings in Section 4.4.3, when we assessed robustness of results to variations in the observation frequency. In the 'general' patient, with an increasing or decreasing profile and at least four measurements, results were relatively robust to deviation from the MAR assumption, with the highest adjustment resulting in a maximum change of 0.80 years. Multiple imputation under the MNAR assumption with delta adjustment has proven to be a powerful tool for conducting sensitivity analysis in the time to threshold framework. The approach can also easily be applied to ordinal outcomes, such as the CGI-S scores in the Schizophrenia data, by applying adjustments to the predicted probabilities.

Chapter 9

Concluding Remarks and Future Work

In this thesis we proposed and applied a novel approach to estimation of the time to attain two consecutive measurements less than (or greater than) a relevant threshold. This approach takes into account the estimated patient-specific trajectories and measurement error. Through identification of a recursive relationship of the continuation probabilities at each time point, we displayed that the computation of the expected times is simple, efficient, and can be implemented using existing software packages. Sensitivity analysis revealed that the estimated times are sensitive to the number of visits considered and the time at which the series is truncated, for patients who exhibit a very slow decline. For other patients, however, we found that results were less sensitive to the number of visits considered and truncation. Hence, caution should be exercised when interpreting the estimated times for patients who exhibit very slow rates of decline.

The proposed methodology was extended in Chapter 5 so that complex residual correlation structures can be incorporated into the estimation of times to threshold. Through the derivation of recursive relationships of the probabilities in this more complex setting, computational efficiency was achieved. We observed that erroneously ignoring serial correlation may bias estimated times to threshold and recommend that careful consideration

is given to the model fitted in the first stage of the approach.

In Chapter 6, the methodology was extended to to accommodate outcomes that may be subject to censoring due to limits of detection. After incorporating the censoring into the likelihood for the model for the outcome, we found that our standard methodology can easily be applied to estimate time to threshold. In Chapter 7 we shifted our focus from continuous outcomes to ordinal outcomes and presented methodology for time to threshold estimation for ordinal biomarkers.

The robustness of estimated times to threshold to deviations from the MAR assumption, was assessed through pattern-mixture based multiple imputation under the MNAR assumption. We observed that even minor deviations can have a substantial impact on the estimated time to threshold for individuals with a relatively 'flat' profile or fewer than three measurements. In the 'general' patient, with an increasing or decreasing profile and at least four measurements, results were relatively robust to deviation from the MAR assumption.

Although this thesis provided a flexible framework for time to threshold estimation, that can accommodate complex features of biomarkers, further developments are necessary. With particular reference to examining HIV treatment success, it would be interesting to examine the performance of our approach when a mixed model with random change points for the viral load trajectory is used. In certain settings clinical decisions are based on cutoffs of more than one biomarker. It is therefore of interest to extend our approach to multivariate outcomes. Persistence criteria may also be more restrictive, where decisions are based on *three* consecutive measurements rather than two. It would be of value to extend our methodology to the 'three consecutive' framework, and examine whether the recursive relationships of probabilities continue to hold.

Bibliography

Amornkul, P. N., Karita, E., Kamali, A., Rida, W. N., Sanders, E. J., Lakhi, S., Price, M. A., Kilembe, W., Cormier, E., Anzala, O., et al. (2013). Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-saharan africa. *AIDS,* **27,** 2775–2786.

Boscardin, W. J., Taylor, J. M., and Law, N. (1998). Longitudinal models for AIDS marker data. *Statistical Methods in Medical Research,* **7,** 13–27.

Burzykowski, T., Molenberghs, G., and Buyse, M. (2005). *The evaluation of surrogate endpoints*. New York: Springer.

Cardeal da Silva, D., Casseb, J., Mirzazadeh, A., Arruda, L. B., and Rutherford, G. W. (2013). Is the rate of CD4 cell decline changing over time in antiretroviral-naive patients? *AIDS Patient Care and STDs,* **27,** 69–70.

Carpenter, J. and Bithell, J. (2000). Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in medicine,* **19,** 1141–1164.

Crowe, S., Hoy, J., and Mills, J. (1996). Management of the HIV-infected patient. CUP Archive.

Dapunt, O .E. (1994). The natural history of thoracic aortic aneurysm. *The Journal of thoracic and cardiovascular surgery,* **107,** 1323–1332.

Davies, M. A., Moultrie, H., Eley, B., Rabie, H., Van Cutsem, G., Giddy, J., Wood, R., Technau, K., Keiser, O., Egger, M., Boulle, A., and International Epidemiologic Databases to Evaluate AIDS Southern Africa (IeDEA-SA) Collaboration. (2014). Virologic failure and second-line antiretroviral therapy in children in South Africa–the IeDEA Southern Africa collaboration. *Journal of acquired immune deficiency syndromes,* **56,** 270–278.

DHHS (2015). Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. H 2–H 3.

SADOH (2010). National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults. *Department of Health, Republic of South Africa*

Diggle, P. (2002). *Analysis of Longitudinal Data*. Oxford Statistical Science Series. OUP Oxford.

Diggle, P., Heagerty, P., Liang, K. Y., and Zeger, S. (2013) *Analysis of longitudinal data.*

Dunayevich, E., Sethuraman, G., Enerson, M., Taylor, C.C. and Lin, D. (2014). Characteristics of two alternative schizophrenia remission definitions: relationship to clinical and quality of life outcomes. *Schizophrenia research,* **86,** 300–308.

Efron, B. and Tibshirani, R.J. (1994). *An introduction to the bootstrap*. CRC press.

Fauci, A. S., Pantaleo, G., Stanley, S. and Weissman, D. (1996). Immunopathogenic Mechanisms of HIV Infection. *Annals of Internal Medicine,* **124,** 654–663.

Jacqmin-Gadda, H., and Sibillot, S., Proust, C., Molina, J. and Thiébaut, R. (2008). Robustness of the linear mixed model to misspecified error distribution. *Computational Statistics & Data Analysis,* **51,** 5142–5154.

Jacqmin-Gadda, H., Thiébaut, R., Chêne, G., and Commenges, D. (2000). Analysis of left-censored longitudinal data with application to viral load in hiv infection. *Biostatistics,* **1,** 355–368.

Heidenreich, A., Bastian, P. J., Bellmunt, J., Bolla, M.,Joniau, S., van der Kwast, T., Mason, M., Matveev, V., Wiegel, T., Zattoni, F., and Mottet, N. (2014). EAU Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent–Update 2013. *European urology,* **65,** 124–137.

Heitjan, D. and Rubin, D. (1991). Ignorability and coarse data. *The Annals of Statistics,* **19,** 2244–2253.

Hughes, M. D., Stein, D. S., Gundacker, H. M., Valentine, F. T., Phair, J. P., and Volberding, P. A. (1994). Within-subject variation in CD4 lymphocyte count in asymptomatic human immunodeficiency virus infection: implications for patient monitoring. *Journal of Infectious Diseases,* **169,** 28–36.

Hughes, Michael D. (2010). Analysis and design issues for studies using censored biomarker measurements with an example of viral load measurements in HIV clinical trials. *Statistics in medicine,* **19,** 3171–3191.

Kent, K.C. (2014). Clinical practice: Abdominal aortic aneurysms. *The New England Journal of Medicine,* **371,** 2101–2108.

Kenward, M.G., (2015). Controlled multiple imputation methods for sensitivity analyses in longitudinal clinical trials with dropout and protocol deviation. *Clinical Investigation,* **5,** 311–320.

Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association,* **88,** 125–134.

LaFleur, B., Lee, W., Billhiemer, D., Lockhart, C., Liu, J. and Merchant, N. (2011). Statistical methods for assays with limits of detection: Serum bile acid as a differentiator between patients with normal colons, adenomas, and colorectal cancer. *Journal of carcinogenesis,* **10,** 12.

Langford, S. E., Ananworanich, J., and Cooper, D. A. (2007). Predictors of disease progression in HIV infection: a review. *AIDS Research and Therapy,* **4,** 11.

Mandel, M. (2010). Estimating disease progression using panel data. *Biostatistics,* **175,** 304–316.

Mandel, M., Gauthier, S.A., Guttmann, C., Weiner, H.L. and Betensky, R.A. (2007). Estimating Time to Event From Longitudinal Categorical Data: An Analysis of Multiple Sclerosis Progression. *Journal of the American Statistical Association,* **102,** 1254–1266.

Malone, J. L., Simms, T. E., Gray, G. C., Wagner, K. F., Bruge, R. J., and Burke, D. S. (1990). Sources of Variability in Repeated T-helper Lymphocyte counts from Human Immunodeficiency Virus Type 1-Infected patients: Total Lymphocyte Count Fluctuations and Diurnal Cycle are important. *Journal of Acquired Immune Deficiency Syndromes,* **3,** 144–151.

Mellors, J., MuÃśoz, A., Giorgi, J., Margoiick, J., Tassoni, C., Gupta, P., Kingsley, L., John A. Todd, P., Saah, A. J., Detels, R., Phair, J. P., and Jr, C. R. R. (1997). Plasma Viral Load and CD4 Lymphocytes as Prognostic Markers of HIV-1 Infection. *Annals of internal medicine,* **126,** 946–954.

Mofenson, L. M., Korelitz, J., Meyer, W. A., Bethel, J., Rich, K., Pahwa, S., Moye, J., Nugent, R., and Read, J. (1997). The Relationship between Serum Human Immunodeficiency Virus Type 1 (HIV-1) RNA Level, CD4 Lymphocyte Percent, and Long-Term Mortality Risk in HIV-1 Infected Children. *The Journal of Infectious Diseases,* **175,** 1029–1038.

NIH (2013). Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

Molenberghs, G. and Kenward, M. (2007). *Missing data in clinical studies*. New York: John Wiley & Sons.

Molenberghs, G. and Kenward, M.G. (2017). *Analysis of incomplete data. In: Dmitrienko, A. and Koch, G. (eds.) Analysis of Clinical Trials Using SAS (Second edition).*. Cary, NC: SAS Press.

NIH National Institute of Diabetes and Digestive and Kidney diseases (2014). Diagnosis of Diabetes and Prediabetes.

O'Connor, J., Smith, C., Lampe, F., Johnson, M., Sabin, C., and Phillips, A. (2014). Rate of viral load failure over time in people on ART in the UK Collaborative HIV Cohort (CHIC) study. Abstracts of the HIV Drug Therapy Glasgow Congress 2014 *Journal of the International AIDS Society,* **17,** 19527.

Pantazis, N., Touloumi, G., Walker, A. S., and Babiker, A. G. (2005). Bivariate modelling of longitudinal measurements of two human immunodeficiency type 1 disease progression markers in the presence of informative drop-out. *Journal of the Royal Statistical Society,* **54,** 405–403.

Peuskens, J. (1995). Risperidone in the treatment of patients with chronic schizophrenia: a multi-national, multi-centre, double-blind, parallel-group study versus haloperidol.. *The British Journal of Psychiatry,* **166,** 712–726.

Peto, T. (1996). Surrogate markers in HIV disease. *Journal of Antimicrobial Chemotherapy,* **37,** 161–170.

Reddy, T., Molenberghs, G., Njeru Njagi, E. and Aerts, M. (2016). A novel approach to estimation of the time to biomarker threshold: applications to HIV. *Pharmaceutical statistics,* **15,** 541–549

Reddy, T., Mwambi, H., and Ndung'u, T. (2011). Modelling HIV progression using multistate markov models. *South African Statistical Journal Proceedings: Proceedings of the 53rd Annual Conference of the South African Statistical Association for 2011 (SASA 2011): Congress 1*, pp. 100–117. Sabinet Online.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data*. Boca Raton: Chapman and Hall/CRC.

Rose, C.E., Gardner, L., Craw, J., Girde, S., Wawrzyniak, A.J., Drainoni, M.L., Davila, J., DeHovitz, J., Keruly, J.C., Westfall, A.O. and Marks, G. (2015). A comparison of methods for analyzing viral load data in studies of HIV patients. *PLoS ONE,* **10,** 6.

Rubin, D.B. (1978). Multiple imputations in sample surveys âĂŞ a phenomenological Bayesian approach to nonresponse. In: Imputation and Editing of Faulty or Missing Survey Data. Washington, DC: U.S. Department of Commerce, 1–23.

Segal, M. R., James, I. R., French, M. A., and Mallal, S. A. (1995). Statistical issues in the evaluation of markers of HIV progression. *International Statistical Review,* **63,** 179–197.

Sweeting, M. and Thompson, S. (2012). Making predictions from complex longitudinal data, with application to planning monitoring intervals in a national screening programme. *Journal of the Royal Statistical Society, Series A (Statistics in Society),* **11,** 569–586.

Taffe, P. and May, M. (2008). A joint back calculation model for the imputation of the date of HIV infection in a prevalent cohort. *Statistics in Medicine,* **27,** 4835–4853.

Thiébaut, R. and Jacqmin-Gadda, H. (2004). Mixed models for longitudinal left-censored repeated measures. Computer methods and programs in biomedicine. *PLoS ONE,* **74,** 255–260.

Todd, J., Glynn, J. R., Marston, M., Lutalo, T., Biraro, S., Mwita, W., Suriyanon, V., Rangsin, R., Nelson, K. E., Sonnenberg, P., et al. (2007). Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low and middle-income countries before highly active antiretroviral therapy. *AIDS,* **21,** S55–S63.

Van der Paal, L., Shafer, L. A., Todd, J., Mayanja, B. N., Whitworth, J. A., and Grosskurth, H. (2007). HIV-1 disease progression and mortality before the introduction of highly active antiretroviral therapy in rural Uganda. *AIDS,* **21,** S21–S29.

Verbeke, G. and Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Verbeke, G. and Molenberghs, G. (2010). Arbitrariness of models for augmented and coarse data, with emphasis on incomplete data and random effects models. *Statistical Modelling,* **10,** 391–419.

Zhang, Y., Makvandi-Nejad, S., Qin, L., Zhao, Y., Zhang, T., Wang, L. and Wu, H. (2015). Interferon-induced transmembrane protein-3 rs12252-C is associated with rapid progression of acute HIV-1 infection in Chinese MSM cohort. *AIDS* **29,** 889–894.

WHO (2015). Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV. *World Health Organization*

Wu, H. and Ding, A.A (1999). Population HIV-1 Dynamics In Vivo: Applicable Models and Inferential Tools for Virological Data from AIDS Clinical Trials. *Biometrics,* **55,** 410-418.

Appendices

Appendix A

Appendix: Chapter 4

A.1 Proof of Expressions for C_{ij} and S_{ij}

For ease of notation, the index i is suppressed in the equations that follow. We introduce E_j which denotes the outcome indicator at time point j such that

$$E_j = \begin{cases} 0 & \text{if } Y_j > k \\ 1 & \text{if } Y_j \leq k. \end{cases}$$

Continuation after j visits can be defined in terms of the combination of outcomes observed, such that the event of two consecutive low CD4 count outcomes {1, 1} has not occurred at, or prior to the j^{th} visit. The possible combinations which lead to continuation after 2, 3, and 4 visits, respectively, are presented in Table A.1. There are 3, 5, and 8 combinations of outcomes that lead to continuation after time points 2, 3, and 4, respectively.

The number of combinations which result in continuation after each visit follows a Fibonacci sequence $\{1, 1, 2, 3, 5, 8, 13, \ldots\}$, where each term is defined as the sum of its two predecessors. Specifically, the number of outcome combinations that result in continuation after a sequence of j visits is the $j + 1^{th}$ Fibonacci number, f_{j+1}. Continuation at visit j can be expressed as a function of continuation at visit $j - 1$ and $j - 2$. A continuation sequence should end in either (A): {0}, which is the union of {0,0} and {1,0}

or (B): {0,1}. This implies that a continuation sequence of length j can be constructed uniquely from (A): a continuation sequence of length $j-1$, followed by $\{0\}$ and (B): a continuation sequence of length $j-2$, followed by $\{0, 1\}$. Letting C_j denote the continuation probability at visit j, and assuming that the outcomes at each visit are independent, this recursive relationship can be presented as follows:

$$C_j = C_{j-2} \times P(E_{j-1} = 0) \times P(E_j = 1) + C_{j-1} \times P(E_j = 0).$$

This relationship is illustrated in Table A.1 for $j = 4$. Assuming that the process 'stops' when two consecutive $\{1, 1\}$ outcomes are observed for the first time, the number of combinations which result in a 'stop' at sequence of j visits is f_{j-2}. For a 'stop' to be observed at any $j \geq 3$, the last three outcomes in the sequence are confined to be of the form $\{0, 1, 1\}$. Hence, the stopping probability S_j is

$$S_j = C_{j-3} \times P(E_{j-2} = 0) \times P(E_{j-1} = 1) \times P(E_j = 1).$$

A.2 SAS Program for Time to Threshold Estimation

```
********** Stage 1 : Fit LMM ******************
ods output CovParms=covp;
proc mixed data=aaa method=reml covtest ;
class id bmicat;
model diam=  bmicat bmicat*timeyears age/ noint outp=predlin solution;
random intercept timeyears/  type=un  subject=id;
repeated / subject=id r;
run;
ods output close;


/***  Create dataset of predictions and residual variance****/
data covp2 (drop=CovParm estimate subject); set covp;
  if CovParm='Residual' then residvar=estimate;
   if residvar eq . then delete;
   link=1;
run;
data preds;set predlin;link=1;run;
data predres;merge preds covp2;by link;run;
```

```
/************************************************/
******** Calculate cumulative probabilities and Cij and Sij ***/

data prob;set predres;p=cdf('NORMAL',k,pred,sqrt(residvar));
proc sort data=prob;by id timemonth;run;
data probc2 ;
   set prob ;
   by id day;
   length visit c lagp lag2p q lagq lag2q lagc lag2c 8;
   if first.id then call missing(of visit lag:);
   visit+1;
   q=1-p;
   select (visit);
     when (1) c=.;
     when (2) c=lagp*p + lagq*p + lagp*q ;
     when (3) c=lag2p*lagp*p
                          + lag2q*lagp*p + lag2p*lagp*q + lag2q*lagp*q + lag2p*lagq*p;
     otherwise c=lag2c*lagp*q + lagc*p ;
   end;
   output;
   lag2p=lagp;
   lagp=p;
   lag2q=lagq;
   lagq=q;
   lag2c=lagc;
   lagc=c;
   retain lag: ;
run;

/******** Compute stopping probabilities Sij *************/
data forsij;set probc2;by id; lag3c=lag3(c);lag3p=lag3(p);lag3q=lag3(q);
if visit eq 2 then sij=(lagq*q);
if visit eq 3 then sij=(lag2p*lagq*q);
if visit eq 4 then sij=(lag3p*lag2p*lagq*q) + (lag3q*lag2p*lagq*q);
if visit gt 4 then  sij=lag3c*lag2p*lagq*q;
t_ij_sij= timemonth*sij;if t_ij_sij eq . then t_ij_sij=0;
run;
/******** Compute expected time ***********************/
data e_ti ;
  set forsij ;
  by id;
  if first.id then
   do;
    e_ti = 0;
   end;
```

```
  e_ti + t_ij_sij;
run;

/*****************************************************/
/***********     Non-parametric case bootstrap sampling     *****/
/*****************************************************/
/*For example, if we require the 95% CI for patient 25:*/

data justid;set aaa.justid;if id eq 25 then delete;run;
data aaa_excl;set aaa;if id eq 25 then delete;run;
data aaa_pt1;set aaa;if id=25;BootID=2000;run;
options nonotes nosource nosource2 errors=0;
%boot(1000)
data test;set work.bootsum;where id eq 25;run;

proc univariate data=test;
  var e_ti;
  output out=forid25 pctlpre=P_ pctlpts= 2.5,97.5;run;


/*************************************************/
/*************** Macros *************************/
%macro boot(B);
%do i=1 %to &B;
*** Select random sample from patient data frames ***;
%bootselect(justid,work.select1,&i*7,1000);
*** Merge random sample with longitudinal data ***;
%bootmerge(work.select1,work.aaa_excl,work.mergedboot,id);
/******** Add back that patient to the dataset*********/
proc append base=mergedboot data=aaa_apt1;run;
run;
%analyze(work.mergedboot,work.results);
%bootsave(work.results,work.bootsum);
%end;
%mend;

%macro bootselect(indata,outdata,seed,start);
data &outdata;
choice=ceil(ranuni(&seed)*n); * Creates Random Variable from 1 to n *;
 set &indata point=choice nobs=n;
  BootID=&start+_N_; * ID for future analysis *;
    if _N_>n then stop;
   *proc print data=work.select&bootno(obs=10);
   *title "Frame &bootno with Seed &Seed";
  run;
%mend;
```

```
*** Merges data from &Bootdata with the longitudinal data &Longdata ***;
*** ID is the subject ID on the original datasets                    ***;
*** The merged dataset will have a new ID: BootID                    ***;
%macro bootmerge(bootdata,longdata,outdata,id);
proc sql;
  create table &outdata
    as select *
    from &bootdata as l left join &longdata as r
      on l.&id=r.&id
    order by BootID;
      quit;
%mend;

*** Saves the results from &Indata into &Outdata ***;
*** &Indata will ususally have one record        ***;
%macro bootsave(indata,outdata);
   %if &I=1 %then %do; * Saves results in &outdata *;
     data &outdata;
       set &indata;
         run;
   %end;
   %else %do;          * Appends results to &outdata *;
     proc append base=&outdata new=&indata force;
       run;
   %end;
%mend;
```

Table A.1: *Possible combinations of outcomes which result in continuation after 2, 3, and 4 visits*

E_1	E_2	E_3	E_4
$j \leq 2$			
1	**0**		
0	**1**		
0	**0**		
$j \leq 3$			
1	**0**	**1**	
1	**0**	**0**	
0	**1**	**0**	
0	**0**	**1**	
0	**0**	**0**	
$j \leq 4$			
1	**0**	0	1
0	**1**	0	1
0	**0**	0	1
1	**0**	**1**	0
1	**0**	**0**	0
0	**1**	**0**	0
0	**0**	**1**	0
0	**0**	**0**	0

Appendix B

Appendix: Chapter 5

B.1 Proof of Expressions for $C_{ij,0}$, $C_{ij,1}$, and S_{ij}

Continuation after j visits can be defined in terms of the combination of outcomes observed, such that the event of two consecutive low CD4 count outcomes $\{1, 1\}$ has not occurred at, or prior to the j^{th} visit. There are 3, 5, and 8 combinations of outcomes that lead to continuation after time points 2, 3, and 4, respectively.

The number of combinations which result in continuation after each visit follows a Fibonacci sequence $\{1, 1, 2, 3, 5, 8, 13, \ldots\}$, where each term is defined as the sum of its two predecessors. Specifically, the number of outcome combinations that result in continuation after a sequence of j visits is the $j + 1^{st}$ Fibonacci number, f_{j+1}. Continuation at visit j can be expressed as a function of continuation at visit $j - 1$ and $j - 2$. A continuation sequence should can end in either 0 or 1. Letting C_j denote the continuation probability at visit j, and assuming that the outcomes at each visit are independent, the recursive relationship can be presented as follows:

$$C_j = C_{j,0} + C_{j,1}, \tag{B.1}$$

where $C_{j,0}$ denotes the continuation probability for sequences ending with outcome $E_j = 0$, and $C_{j,1}$ denotes the continuation probability for sequences ending with outcome $E_j = 1$.

By the chain rule, the joint distribution of a set of random variables can be expressed as a product of conditional probabilities. That is,

$$P(E_n, E_{n-1} \dots E_1) = P(E_n \mid E_{n-1}, \dots, E_1) P(E_{n-1}, E_{n-2} \dots E_1).$$

Repeating this process yields the expression:

$$P(\bigcap_{j=1}^{n} E_j) = \prod_{j=1}^{n} P(E_n \mid \bigcap_{k=1}^{n-1} E_k).$$

By the Markov assumption

$$P(E_n \mid \bigcap_{k=1}^{n-1} E_k) = P(E_n \mid E_{n-1}).$$

Assuming that the Markov assumption holds,

$$\begin{aligned} C_{j,0} &= C_{j-1,0} P(E_j = 0 | E_{j-1} = 0) + C_{j-1,1} P(E_j = 0 | E_{j-1} = 1), \\ C_{j,1} &= C_{j-1,0} P(E_j = 1 | E_{j-1} = 0). \end{aligned}$$

All that is now needed is the specification of the initial continuation probabilities, $C_{2,(0)}$ and $C_{2,(1)}$.

$$\begin{aligned} C_{2,(0)} &= P(E_1 = 0) P(E_2 = 0 | E_1 = 0) + P(E_1 = 1) P(E_2 = 0 | E_1 = 1) \\ &= P(E_2 = 0), \\ C_{1,(1)} &= P(E_1 = 0) P(E_2 = 1 | E_1 = 0) \\ &= P(E_1 = 0, E_2 = 1). \end{aligned}$$

Assuming that the process 'stops' when two consecutive $\{1, 1\}$ outcomes are observed for the first time, the number of combinations which result in a 'stop' at sequence of j visits is f_{j-2}. For a 'stop' to be observed at any $j \geq 3$, the last three outcomes in the sequence are confined to be of the form $\{0, 1, 1\}$. Hence, the stopping probability S_j is

$$\begin{aligned} S_j &= C_{j-3,0} P(E_{j-2} = 0 | E_{j-3} = 0) P(E_{j-1} = 1 | E_{j-2} = 0) P(E_j = 1 | E_{j-1} = 1) \\ &\quad + C_{j-3,1} P(E_{j-2} = 0 | E_{j-3} = 1) P(E_{j-1} = 1 | E_{j-2} = 0) P(E_j = 1 | E_{j-1} = 1) \\ &= [C_{j-3,0} P(E_{j-2} = 0 | E_{j-3} = 0) + C_{j-3,1} P(E_{j-2} = 0 | E_{j-3} = 1)] \\ &\quad \times P(E_{j-1} = 1 | E_{j-2} = 0) P(E_j = 1 | E_{j-1} = 1). \end{aligned} \tag{B.2}$$

B.2 SAS Program for Time to Threshold Estimation for a Model with Serial Correlation

```
proc sort data=sk2;by id timeyrs;run;
ods exclude all;
proc mixed data=sk2 method=reml covtest;
class id baselinevlcat timemonthcls;
model sqrtcd  = baselinevlcat  baselinevlcat*timemonth/ noint  outpm=predlinspexp;
random intercept timemonth/  type=un subject=id solution;
repeated timemonthcls/type=AR(1) local group=baselinevlcat subject=id rcorr=1,14,10;
CONTRAST 'test equal slopes' baselinevlcat*timemonth 1 -1  0,
                             baselinevlcat*timemonth 0  1 -1 ;
ods output CovParms=covp;
ods output SolutionR=rand1;
run;
 ods exclude none;
data rand1; set rand1;
  keep id effect estimate;
run;
 data rand12; set rand1; by id;
  retain ranint ranslope;
  if effect='Intercept' then ranint=estimate;
  if effect='timemonth' then ranslope=estimate;
  if last.id then do;
     output; ranint=.; ranslope=.;
  end;
  drop effect estimate;
run;
proc sort data=rand12;by id;run;

proc sort data=predlinspexp;by id timemonth;run;
data allpred;merge rand12 predlinspexp;by id;run;
data allpred;set allpred;manpred=pred+ranint+(ranslope*timemonth);run;

data covp1; set covp;
  if CovParm='Residual' then residvar=estimate;
if residvar=. then delete;
keep residvar;
run;
proc print data=covp;run;
data covp2 (keep=serialpar baselinevlcat); set covp;where  CovParm='AR(1)';
if group='baselinevlcat 1'  then baselinevlcat=1;
if group='baselinevlcat 2'  then baselinevlcat=2;
if group='baselinevlcat 3'  then baselinevlcat=3;
```

```
rename estimate=serialpar;
run;
data covp3 (keep=estimate baselinevlcat); set covp;where  CovParm='Variance';
if group='baselinevlcat 1'  then baselinevlcat=1;
if group='baselinevlcat 2'  then baselinevlcat=2;
if group='baselinevlcat 3'  then baselinevlcat=3;
run;
data covp3;set covp3;rename estimate=serialvar;run;

data allcovpars;merge covp1 covp2 covp3;
lagresidvar=lag(residvar);
if lagresidvar=. then lagresidvar=lag(residvar);
if residvar=. then residvar=lagresidvar;
drop lagresidvar;
run;

proc sort data=allpred;by baselinevlcat id;run;
proc sort data=allcovpars;by baselinevlcat ;run;

data preds;set allpred;;if manpred<0 then delete;run;
data predres;merge preds allcovpars;by baselinevlcat;run;

/*************** Now to Calculate the probabilities ***************/
/**** Generate the lag times and correlations by patient ****/
proc sort data=predres;by id timemonth;run;
data predresx;set predres;
by id;
lagtime = lag(timemonth);
if first.id then lagtime = .;
currmean=manpred;
prevmean=lag(manpred);
if first.id then prevmean= .;
d_pc=abs(lagtime-timemonth);
p_pc=serialpar**d_pc;
corr=(serialvar*p_pc)/(serialvar+residvar);
currstd=(18.70-currmean)/(sqrt(residvar+serialvar));
prevstd= (18.70-prevmean)/(sqrt(residvar+serialvar));
run;

/** Calculate all the univariate,bivariate and conditional probabilities we need**/
data proby;set predresx;
lc=probnorm(currstd);
hc=1-lc;
lp=probnorm(prevstd);
hp=1-lp;
lclp=probbnrm(currstd,prevstd,corr);
```

```
hclp=lp-lclp;
lchp=lc-lclp;
hchp=hc-(lp-lclp);
hc_hp=(hchp)/hp;
hc_lp=(hclp)/lp;
lc_hp=(lchp)/hp;
lc_lp=(lclp)/lp;
run;
/************** Calculate Cij using recursive relationship*****************/
data proby2;set proby;by id;if first.id then visit=1;else visit+1;run;
data probc2 ;
   set proby2 ;
   by id timemonth;
      length  lag3hc lag2hc lag3hc  lag3lc lag2lc  laglc  lag2hc_lp laghc_lp  lag2lc_hp laglc_hp
 lag2lc_lp laglc_lp   lag2hc_hp laghc_hp c0 c1 lag2c1 lag2c0 lagc1 lagc0 8;
   if first.id then call missing(of lag:);
   if  visit eq 2 then c1= lchp;
   if visit eq 2 then c0=hclp + hchp ;
     if visit eq 3 then c0=(hc_hp*laghc_lp*lag2lc) + (hc_hp*laghc_hp*lag2hc)
+ (hc_lp*laglc_hp*lag2hc);
if visit eq 3 then c1=(lc_hp*laghc_lp*lag2lc) +(lc_hp*laghc_hp*lag2hc);
     if visit eq 4 then c0=(hc_hp*laghc_hp*lag2hc_lp*lag3lc)
        +(hc_hp*laghc_hp*lag2hc_hp*lag3hc)
+(hc_hp*laghc_lp*lag2lc_hp*lag3hc)
+(hc_lp*laglc_hp*lag2hc_lp*lag3lc)
+(hc_lp*laglc_hp*lag2hc_hp*lag3hc);
if visit eq 4 then  c1=(lc_hp*laghc_hp*lag2hc_lp*lag3lc)
+(lc_hp*laghc_hp*lag2hc_hp*lag3hc)
+(lc_hp*laghc_lp*lag2lc_hp*lag3hc);

if visit gt 4 then c0=(lagc0*hc_hp) + (lagc1*hc_lp); lagc0 = lag(c0);
if visit gt 4 then  c1=(lag2c0*lc_hp*laghc_hp)+(lag2c1*lc_hp*laghc_lp);
output;
lag3hc=lag2hc;
lag2hc=laghc;
laghc=hc;

lag3lc=lag2lc;
lag2lc=laglc;
laglc=lc;

lag2hc_lp= laghc_lp;
laghc_lp=hc_lp;

lag2lc_hp=laglc_hp;
laglc_hp= lc_hp;
```

```
lag2hc_hp= laghc_hp;
laghc_hp=hc_hp;

lag2lc_lp= laglc_lp;
laglc_lp=lc_lp;

lag2c0= lagc0;
lagc0=c0;
lag2c1= lagc1;
lagc1=c1;
retain lag: ;
run;


data forsij;set probc2;by id; lag3c0=lag(lag2c0);
lag3c1=lag(lag2c1);
if first.id then lag3c1=.;
if visit eq 2 then sij=lclp;
if visit eq 3 then sij=lc_lp*laglc_hp*lag2hc;
if visit eq 4 then sij=(lc_lp*laglc_hp*lag2hc_hp*lag3hc)
+(lc_lp*laglc_hp*lag2hc_lp*lag3lc);
if visit gt 4 then sij=(lag3c0*lag2hc_hp*laglc_hp*lc_lp)
+ (lag3c1*lag2hc_lp*laglc_hp*lc_lp);
t_ij_sij= timemonth*sij;if t_ij_sij eq . then t_ij_sij=0;
run;


/******* Manually check Si5 **********/
data check;set forsij;by id;lag4lc=lag(lag3lc);lag4hc=lag(lag3hc);
if visit=5 then testsij=(lc_lp*laglc_hp*lag2hc_hp*lag3hc_lp*lag4lc)
+ (lc_lp*laglc_hp*lag2hc_hp*lag3hc_hp*lag4hc)
+ (lc_lp*laglc_hp*lag2hc_lp*lag3lc_hp*lag4hc);run;



data e_ti ;
  set forsij ;
  by id;
  IF first.id THEN
   DO;
    e_ti = 0;
   END;
  e_ti + t_ij_sij;
run;
```

```
/*********************************************************************/
/******************** Estimate bias for ID 1 ************************/

data test;set work.bootsum;where id eq 1 and reason eq 'Convergence criteria met.';
sample = _N_; run;
data bootorig;
set sel_orig (in=a)
test;
if a then sample=0;
run;
data bias;set bootorig end=last;
retain orig;
if sample=0 then orig=e_ti;
if e_ti lt orig then lessthan=1;
else lessthan=0;
retain nless 0;
if sample ne 0 then nless=nless+lessthan;
propless=nless/sample;
bias=probit(propless);
if last then do;
output bias;
end;
run;

/*********************************************************/
/***************** Jacknife **********************************/

data justid;set paper1.justid;if id eq 1 then delete;drop patient;run;
data sk2_excl;set work.sk2;if id eq 1 then delete;run;
data sk2_pt1;set work.sk2;if id=1;run;
data origjack; /* create a new data set which contains observation */
set justid end=eof; /* numbers 1 to &nobs (no. obs in data set) */
obsnum=_n_;
if eof then call symput('nobs', put(obsnum, 2.));
run;
%macro jack(J);
%do i=1 %to &J;
*** Select random sample from patient data frames ***;
%jackselect(origjack,work.select1);
*** Merge  with longitudinal data ***;
%jackmerge(work.select1,work.sk2_excl,work.mergedjack);
/******** Add back that patient to the dataset*********/
proc append base=mergedjack data=sk2_pt1;run;
%jackanalyze(work.mergedjack,work.results);
%jacksave(work.results,work.jacksum);
%end;
```

```
%mend;
%jack(335);
*** Saves the results from &Indata into &Outdata ***;
*** &Indata will ususally have one record        ***;
%macro jacksave(indata,outdata);
   %if &I=1 %then %do; * Saves results in &outdata *;
     data &outdata;
       set &indata;
         run;
   %end;
   %else %do;          * Appends results to &outdata *;
     proc append base=&outdata new=&indata force;
       run;
   %end;
%mend;


%macro jackselect(indata,outdata);
data &outdata;
        set &indata;
     where obsnum ne &i;
  run;
%mend;

*** Merges data from &Bootdata with the longitudinal data &Longdata ***;
*** ID is the subject ID on the original datasets                    ***;
*** The merged dataset will have a new ID: BootID                    ***;
%macro jackmerge(jackdata,longdata,outdata);
proc sort data=&longdata;by id timeyrs;run;
proc sort data=&jackdata;by id;run;
data  &outdata;merge  &longdata &jackdata;by id;run;
data &outdata;set &outdata;where obsnum ne .;run;
%mend;


/*************** Calculate BCa interval **************/
proc sql
noprint;
select mean(e_ti) /* put mean of jackknifed values into macro variable */
into :meanjack
from jacksum;
quit;
data jacksum3;
set jacksum;
cubed=(&meanjack - e_ti)**3; /* create cubed value of difference */
squared=(&meanjack - e_ti)**2; /* create squared value of difference */
```

```
run;
proc means
data=jacksum3
noprint;
output out=jacksum4
sum(cubed)=sumcube /* find sum of cubed values */
sum(squared)=sumsquar; /* find sum of squared values */
run;
data accel;
set jacksum4;
accel=sumcube / (6 * (sumsquar**1.5)); /* plug values into equation for */
keep accel; /* the acceleration statistic */
run;


data ciends;
merge accel
bias;
part1=(bias + probit(0.025)) / (1 - (accel*(bias + probit(0.025))));
part2=(bias + probit(0.975)) / (1 - (accel*(bias + probit(0.975))));
alpha1=probnorm(bias + part1);
alpha2=probnorm(bias + part2);
n1=alpha1*5000; /*this depends on how many we run*/
n2=alpha2*5000; /*this depends on how many we run*/
call symput('n1', put(floor(n1), 5.)); /* Create macro variables with values */
call symput('n2', put(ceil(n2), 5.)); /* of N1 and N2 for later use */
run;

proc sort
data=test;
by e_ti;
run;

data ci_bca;
set test end=eof;
retain conf_lo conf_hi;
if _n_=&n1 then conf_lo=e_ti; /* select values for upper and lower */
if _n_=&n2 then conf_hi=e_ti; /* limits using N1 and N2 values */
if eof then output;
keep conf_lo conf_hi;
run;
```

B.3 Bias Corrected Accelerated (BCa) Bootstrap Confidence Intervals

The estimation of BCa confidence intervals was first put forward by Efron and Tibshirani (1994). Letting

$$\begin{aligned} \widehat{T} &= \text{The actual estimate from original data} \\ \widehat{T^*}(m) &= \text{The estimate from the } m^{th} \text{ bootstrap sample} \\ B &= \text{The total number of bootstrap samples,} \end{aligned}$$

then the bias correction can be expressed as

$$b = \Phi^{-1}\left(\frac{\sum_{m=1}^{B} I_{\widehat{T^*}(m)<\widehat{T}}}{B}\right).$$

Letting

$$\begin{aligned} \widehat{T_{(i)}} &= \text{Estimate calculated on the jacknifed sample with the } i^{th} \text{ observation removed} \\ N &= \text{The number of jacknifed samples} \\ \widehat{T_{(\bullet)}} &= \text{The mean of the n jackknife samples,} \end{aligned}$$

the acceleration parameter can be expressed as

$$a = \frac{\sum_{i=1}^{N}(\widehat{T_{(\bullet)}} - \widehat{T_{(i)}})^3}{6\left\{\sum_{i=1}^{N}(\widehat{T_{(\bullet)}} - \widehat{T_{(i)}})^2\right\}^{\frac{3}{2}}}.$$

The bias and acceleration parameters are then used to compute the endpoints $(\alpha_1 \times B, \alpha_2 \times B$ of the $100(1-2\alpha)$ confidence interval:

$$\alpha_1 = \tilde{\Phi}\left(b + \frac{b + \tilde{\Phi}^{-1}(\alpha)}{1 - a(b + \tilde{\Phi}^{-1}(\alpha))}\right)$$

$$\alpha_2 = \tilde{\Phi}\left(b + \frac{b + \tilde{\Phi}^{-1}(1-\alpha)}{1 - a(b + \tilde{\Phi}^{-1}(1-\alpha))}\right)$$

where $\tilde{\Phi}(x)$ is the standard normal cumulative distribution function.

Table B.1: *Possible combinations of outcomes which result in termination after 5, 6, and 7 visits*

E_1	E_2	E_3	E_4	E_5	E_6	E_7
			S_5			
0	**0**	0	1	1		
1	**1**	0	1	1		
0	**1**	0	1	1		
			S_6			
0	**0**	**0**	0	1	1	
1	**0**	**0**	0	1	1	
0	**1**	**0**	0	1	1	
0	**0**	**1**	0	1	1	
1	**0**	**1**	0	1	1	
			S_7			
0	**0**	0	0	0	1	1
1	**0**	0	0	0	1	1
0	**1**	0	0	0	1	1
0	**0**	1	0	0	1	1
1	**0**	1	0	0	1	1
0	**0**	0	1	0	1	1
1	**0**	0	1	0	1	1
0	**1**	0	1	0	1	1

Appendix C

Appendix: Chapter 6

C.1 SAS Program for Time to Threshold Estimation for a Left Censored Outcome

```
/* Generate the censoring variable */
data vl;set paper3.wu_aids;logvl=lgcopy;if lgcopy eq 1.69897 then logvl=2;
 /*The dataset previously imputed half the LOD*/
censor=0;if logvl eq 2 then censor=1;time=day;
run;

/* Fit LMM without censoring - using imputed value ***/
data spl;
set vl;
k1 = 14;
k2 = 60;
if time <= k1 then timespl2 = 0;
if time > k1 then timespl2 = time - k1;
if time <= k2 then timespl3 = 0;
if time > k2 then timespl3 = time - k1- k2;
timemonth=time/30.5;
timeyear=time/365;
timespl2year=timespl2/365;
timespl3year=timespl3/365;
run;
proc sort data=spl;by id day;run;
/*Now try to include censoring */
ods output ParameterEstimates=p1;
```

```
proc nlmixed data=spl ;
parms a0=5.0709 t0=-46.7 t1=41.58 t2=4.8 sigsqe=0.098 sd_u1=0.52 sd_u3=6.6
sd_u4=10.5  g13=-0.3958 g14=1.0904 g34=-40;

/* Calculate pi, 3.1416...... */

pi=2*arsin(1);

/* Define the linear model with a random intercept (u1) and slopes (u3 and u4) */
mu = a0 + t0*timeyear + t1*timespl2year + t2*timespl3year + u1
+ u3*timespl2year + u4*timespl3year;
/* Define likelihood function for observations censored and not censored */
/* censor = 0 is just the PDF of the normal distribution for non-LOD values */
/* censor = 1 is for lower LOD censored values */
/* Note that mu is the predicted mean and logvl (log base 10 VL) is  observed data */

if censor=0 then L=(1/(sqrt(2*pi*sigsqe)))*exp(-(logvl-mu)**2/(2*sigsqe));
if censor=1 then L=probnorm((logvl-mu)/sqrt(sigsqe));

/* Define the log(likelihood), model, and random effects */
LL=log(L);
model logvl ~ general(LL);
random u1 u3 u4~ N([0, 0, 0], [sd_u1*sd_u1, g13, sd_u3*sd_u3, g14, g34, sd_u4*sd_u4])
subject=id;
predict  a0 + t0*timeyear + t1*timespl2year + t2*timespl3year + u1  + u3*timespl2year
+ u4*timespl3year  out=preds;
run;
ods output close;

data covp2 (keep=link residvar); set p1;
  if Parameter='sigsqe' then residvar=Estimate;
   if residvar eq . then delete;
   link=1;
run;
data preds;set preds;link=1;if pred<0 then delete;run; /*excludes implausible values****/
data predres;merge preds covp2;by link;run;

/***** Compute cumulative probabilities *******/
data prob;set predres;p=1-cdf('NORMAL',3,pred,sqrt(residvar));run;
/******** Need to calculate the Cn recursively but first manually do up to C5 *******/
/**** generate observation number**********/
proc sort data=prob;by id timemonth;run;
data probc2 ;
   set prob ;
   by id day;
   length visit c lagp lag2p q lagq lag2q lagc lag2c 8;
```

```
   if first.id then call missing(of visit lag:);
   visit+1;
   q=1-p;
   select (visit);
     when (1) c=.;
     when (2) c=lagp*p + lagq*p + lagp*q ;
     when (3) c=lag2p*lagp*p + lag2q*lagp*p + lag2p*lagp*q + lag2q*lagp*q + lag2p*lagq*p;
     otherwise c=lag2c*lagp*q + lagc*p ;
   end;
   output;
   lag2p=lagp;
   lagp=p;
   lag2q=lagq;
   lagq=q;
   lag2c=lagc;
   lagc=c;
   retain lag: ;
run;

/******** Compute stopping probabilities Sij *************/
data forsij;set probc2;by id; lag3c=lag3(c);lag3p=lag3(p);lag3q=lag3(q);
if visit eq 2 then sij=(lagq*q);
if visit eq 3 then sij=(lag2p*lagq*q);
if visit eq 4 then sij=(lag3p*lag2p*lagq*q) + (lag3q*lag2p*lagq*q);
if visit gt 4 then  sij=lag3c*lag2p*lagq*q;
t_ij_sij= timemonth*sij;if t_ij_sij eq . then t_ij_sij=0;
run;
/******** Compute expected time ***********************/
data e_ti ;
  set forsij ;
  by id;
  if first.id then
   do;
    e_ti = 0;
   end;
  e_ti + t_ij_sij;
run;
```

Appendix D

Appendix: Chapter 7

D.1 SAS Program for Time to Threshold Estimation for Ordinal Outcomes

```
proc glimmix data=pans  pconv=1e-6 method=quad(qpoints=10);
class treat;
model cgi_sev =  treat*xtime
/ dist=multinomial link=cumlogit;
random intercept xtime / subject=id type=un;
output out=modelpred pred=xb;
run;

data prob;set modelpred;p=1-(exp(xb) / (1+exp(xb)));where _LEVEL_ eq 3;run;
data probc2 ;
   set prob ;
   by id xtime;
   length visit c lagp lag2p q lagq lag2q lagc lag2c 8;
   if first.id then call missing(of visit lag:);
   visit+1;
   q=1-p;
   select (visit);
     when (1) c=.;
     when (2) c=lagp*p + lagq*p + lagp*q ;
     when (3) c=lag2p*lagp*p + lag2q*lagp*p + lag2p*lagp*q + lag2q*lagp*q + lag2p*lagq*p;
     otherwise c=lag2c*lagp*q + lagc*p ;
   end;
   output;
```

```
   lag2p=lagp;
   lagp=p;
   lag2q=lagq;
   lagq=q;
   lag2c=lagc;
   lagc=c;
   retain lag: ;
run;

/******** Compute stopping probabilities Sij *************/
data forsij;set probc2;by id; lag3c=lag3(c);lag3p=lag3(p);lag3q=lag3(q);
if visit eq 2 then sij=(lagq*q);
if visit eq 3 then sij=(lag2p*lagq*q);
if visit eq 4 then sij=(lag3p*lag2p*lagq*q) + (lag3q*lag2p*lagq*q);
if visit gt 4 then  sij=lag3c*lag2p*lagq*q;
t_ij_sij= xtime*sij;if t_ij_sij eq . then t_ij_sij=0;
run;
/******** Compute expected time ***********************/
data e_ti ;
  set forsij ;
  by id;
  IF first.id THEN
   DO;
    e_ti = 0;
   END;
  e_ti + t_ij_sij;
run;
```

Appendix E

Appendix: Chapter 8

E.1 SAS Program for MNAR Imputation

```
/************ Scenario 2 *******************************/
proc sort data=paper2.aaasens2;by bmicat id;run;
proc mi data=paper2.aaasens2 seed=486048  out=aaami_shift1
nimpute=10;
title âĂŹShift multiple imputationâĂŹ;
class bmicat;
var    age bmicat diam1 diam2 diam3 diam4 diam5 diam6;
monotone reg ;
mnar adjust (diam2 / shift=1);
mnar adjust (diam3 / shift=1);
mnar adjust (diam4 / shift=1);
mnar adjust (diam5 / shift=1);
mnar adjust (diam6 / shift=1);
run;



data aaami_shift1(drop=age bmi bmicat);set aaami_shift1;run;
proc transpose data=aaami_shift1 out=aaalong1;
   by _Imputation_ id;
run;
data aaalong1;
  set aaalong1 (rename=(col1=diam));
  timep=input(substr(_name_, 5), 5.);
  drop _name_;
```

```
run;

/********* Merge with full data that has windows of prediction *****/
data aaa;set paper2.aaa;
age=leeftijd08;
diam=dmaxabdominaalmm;
id=patient;
timep=timepoint;
if bmi lt 25 & bmi ne . then bmicat=1;
if bmi ge 25 & bmi & bmi lt 30  then bmicat=2;
if bmi ge 30 & bmi ne . then bmicat=3;
run;
proc sort data=aaa;by id timep ;run;
data aaa_drop(drop=diam);set aaa;run;

%macro aaa(num);
   %do i = 1 %to &num;
data aaa_&i;set aaa_drop;_Imputation_=&i;run;
  %end;
data alldata;set aaa_1-aaa_&num;run;
%mend;
%aaa(10)
proc sort data=alldata;by _Imputation_ id timep;run;
proc sort data=aaalong1;by _Imputation_ id timep;run;
data impute_window;merge alldata aaalong1;by _Imputation_ id timep;run;



/********** Stage 1 : Fit LMM *************************/
proc sort data=impute_window;by _Imputation_ id timeyears;run;
ods output CovParms=covp;
proc mixed data=impute_window method=reml covtest asycov ;
class _Imputation_ id bmicat;
model diam=  bmicat bmicat*timeyears age/ noint outp=predlin solution covb;
random intercept timeyears/  type=un  subject=id;
repeated / subject=id r;
by _Imputation_;
ods output SolutionF=mixparms CovB=mixcovb AsyCov=covv;;
run;
ods output close;

data mixparms2;
set mixparms;
if bmicat=1 and effect='bmicat' then effect='bmicat1';
if bmicat=2 and effect='bmicat' then effect='bmicat2';
if bmicat=3 and effect='bmicat' then effect='bmicat3';
```

```
if bmicat=1 and effect='timeyears*bmicat' then effect='timeyrs*bmicat1';
if bmicat=2 and effect='timeyears*bmicat' then effect='timeyrs*bmicat2';
if bmicat=3 and effect='timeyears*bmicat' then effect='timeyrs*bmicat3';
run;
data mixparms2;
set mixparms2 (drop=bmicat);
run;
data mixcovb2;
set mixcovb;
if bmicat=1 and effect='bmicat' then effect='bmicat1';
if bmicat=2 and effect='bmicat' then effect='bmicat2';
if bmicat=3 and effect='bmicat' then effect='bmicat3';
if bmicat=1 and effect='timeyears*bmicat' then effect='timeyrs*bmicat1';
if bmicat=2 and effect='timeyears*bmicat' then effect='timeyrs*bmicat2';
if bmicat=3 and effect='timeyears*bmicat' then effect='timeyrs*bmicat3';
run;

data mixcovb2;
set mixcovb2 (drop=bmicat);
run;
 proc mianalyze parms=mixparms2
                  covb(effectvar=rowcol)=mixcovb2;
      modeleffects age bmicat1 bmicat2 bmicat3 timeyrs*bmicat1 timeyrs*bmicat2 timeyrs*bmicat3;
run;
data covpx;set covp;if CovParm='UN(1,1)' then effect='d11';
if CovParm='UN(2,1)' then effect='d12';
if CovParm='UN(2,2)' then effect='d22';
if CovParm='Residual' then effect='sigma2';
run;
data covvx;set covv;if CovParm='UN(1,1)' then effect='d11';
if CovParm='UN(2,1)' then effect='d12';
if CovParm='UN(2,2)' then effect='d22';
if CovParm='Residual' then effect='sigma2';
Col1=CovP1;
Col2=CovP2;
Col3=CovP3;
Col4=CovP4;
run;


proc mianalyze parms=covpx covb(effectvar=rowcol)=covvx ;
      modeleffects d11 d12 d22 sig;
run;


/***  Create dataset of predictions and residual variance******/
```

```
data covp2 (drop=CovParm estimate subject); set covp;
  if CovParm='Residual' then residvar=estimate;
   if residvar eq . then delete;
   link=1;
run;
data preds;set predlin;link=1;run;
data predres;merge preds covp2;by _Imputation_ link;run;



/*************************************************/
/******** Calculate cumulative probabilities and Cij and Sij *****/

data prob;set predres;p=cdf('NORMAL',50,pred,sqrt(residvar));
proc sort data=prob;by _Imputation_ id timeyears;run;
data probc2 ;
   set prob ;
   by _Imputation_ id timep;
   length visit c lagp lag2p q lagq lag2q lagc lag2c 8;
   if first.id then call missing(of visit lag:);
   visit+1;
   q=1-p;
   select (visit);
     when (1) c=.;
     when (2) c=lagp*p + lagq*p + lagp*q ;
     when (3) c=lag2p*lagp*p + lag2q*lagp*p + lag2p*lagp*q + lag2q*lagp*q + lag2p*lagq*p;
     otherwise c=lag2c*lagp*q + lagc*p ;
   end;
   output;
   lag2p=lagp;
   lagp=p;
   lag2q=lagq;
   lagq=q;
   lag2c=lagc;
   lagc=c;
   retain lag: ;
run;

/******** Compute stopping probabilities Sij *************/
data forsij;set probc2;by _Imputation_ id; lag3c=lag3(c);lag3p=lag3(p);lag3q=lag3(q);
if visit eq 2 then sij=(lagq*q);
if visit eq 3 then sij=(lag2p*lagq*q);
if visit eq 4 then sij=(lag3p*lag2p*lagq*q) + (lag3q*lag2p*lagq*q);
if visit gt 4 then  sij=lag3c*lag2p*lagq*q;
t_ij_sij= timeyears*sij;if t_ij_sij eq . then t_ij_sij=0;
run;
```

```
/******** Compute expected time ***********************/
data e_ti ;
  set forsij ;
  by _Imputation_ id;
  if first.id then
   do;
    e_ti = 0;
   end;
  e_ti + t_ij_sij;
run;

data   e_ti2;set e_ti;by _Imputation_ id;
if last.id then lastrec = 1;run;
data e_ti3;set e_ti2;where lastrec=1; run;
```

Samenvatting

Biomerkers spelen een cruciale rol in het medisch onderzoek en helpen de arts om een diagnose te stellen en in het bijzonder om te bepalen wanneer het tijd is om te interveniëren. Jammer genoeg zijn nogal wat biomerkers onderhevig aan een grote mate van schommelingen en belangrijke meetfouten. Uiteraard leidt dit tot bezorgdheid over de geldigheid van medische beslissingen die op één enkele meting van een dergelijke merker gestoeld zijn. Verscheidene studies hebben aangetoond dat beslissingen, genomen op basis van verscheidene opeenvolgende metingen van eenzelfde merker, veel betrouwbaarder zijn. In dat verband stellen onderzoekers zogenaamde persistentie-criteria voor, waarmee verwezen wordt naar het twee (of meer) keer voorkomen van een verhoogde (of verlaagde) waarde van dezelfde merker; hierbij is de grenswaarde natuurlijk van groot belang (Amornkul et al., 2013; Zhang, 2015).

In Hoofdstuk 2 beschrijven we vier studies waar de tijd tot het overschrijden van een bepaalde grenswaarde van groot belang is. De studies komen uit de volgende gebieden: HIV/AIDS, cardiologie en psychiatrie. In Hoofdstuk 3 schetsen we de basisbeginselen van de analyse van longitudinale gegevens, met bijzondere aandacht voor het random effect model. Deze concepten vormen de bouwstenen voor de voorgestelde aanpak voor de schatting van de tijd tot grenswaarde, die we uitwerken in Hoofdstuk 4.

In Hoofdstuk 4 stellen we dus een methode voor om de tijd tot het overschrijden van een grenswaarde te schatten, op basis van persistentie-criteria. We maken daarbij gebruik van een tweetraps methode. In eerste instantie passen we een lineair gemengd model aan

de longitudinale metingen aan, waaruit dan voor de patiënt specifieke waarden volgen. Die zijn een functie van zogenaamde vaste effecten en empirische Bayes schatters. Op basis hiervan wordt de kans bepaald om twee opeenvolgende waarnemingen te hebben die onder (of boven) een bepaalde grenswaarde liggen. Door dit te doen voor elk van de geplande meetmomenten, kunnen we de verwachte overschrijdingstijd berekenen. Door het afleiden van een recursieve relatie van de zogenaamde continueringskansen op elk moment, kunnen we aantonen dat de berekening van de verwachte tijd eenvoudig en efficiënt is, en makkelijk geïmplementeerd kan worden met behulp van bestaande software. We passen deze methodologie toe op twee studies en voeren een sensitiviteitsanalyse uit om te achterhalen of ze robuust is tegen afwijkingen van de gemaakte veronderstellingen. Een mogelijke tekortkoming is dat de methode, gegeven random effecten, onafhankelijkheid veronderstelt van de residuen, de zogenaamde conditionele onafhankelijkheidsassumptie.

In Hoofdstuk 5 breiden we de methodologie, voorgesteld in Hoofstuk 4, zodanig uit dat ook seriële correlatie kan meegenomen worden. Veronderstellend dat de zogenaamde Markov eigenschap geldt, en met behulp van de kettingregel voor kansen, laten we zien dat de continueringskans op elk moment kan uitgedrukt worden als het product van conditionele kansen. We passen de methode toe op een cohort van HIV positieve personen, waarbij we de tijd tot aan een CD4 grenswaarde schatten. Om de impact na te gaan van het over het hoofd zien van seriële correlatie, vergelijken we de aanpak van vorig hoofdstuk met de hier voorgestelde. We stellen vast dat het verkeerd modelleren van de correlatiestructuur tot substantiële overschatting van de tijd tot grenswaarde kan leiden.

In Hoofdstuk 6 beschouwen we biomerkers die onderhevig zijn aan detectielimieten en/of aan censurering. Onze methodologie wordt aan dergelijke situaties aangepast. We stellen vast dat, mits het opnemen van censurering in de likelihood functie, de methode uit Hoofdstuk 4 gewoon kan gebruikt worden. We passen ze dan toe op metingen van virusdruk, gemeten in patiënten uit de ACTG 315 studie. Meer bepaald schatten we de tijd tot behandelsucces.

In Hoofdstukken 4, 5 en 6 gingen we uit van continue biomerkers. Uiteraard vormt dit een beperking. Immers, de gezondheidstoestand kan bijvoorbeeld ook gemeten worden op een ordinale schaal, zoals we die vaak tegenkomen in de psychiatrie. In Hoofdstuk 7 stellen we daarom een variant voor van de methode uit Hoofdstuk 4, dus voor het geval

van een ordinale merker. We passen de methode toe op gegevens van patiënten die behandeld worden voor schizofrenie, waarbij onze interesse uitgaat naar tijd tot remissie. De methodologie voor tijd tot grenswaarde, voorgesteld in Hoofdstukken 4 t.e.m. 7, gaat uit van de veronderstelling dat de ontbrekende gegevens ignorable zijn. Deze veronderstelling wordt onderworpen aan een sensitiviteitsanalyse in Hoofdstuk 8. De thesis sluit af met een samenvatting van de bijdragen, geleverd in de diverse hoofdstukken. Routes voor verder onderzoek worden tot slot aangegeven.