

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Statistical Models for Time to Event Outcomes  
and Determination of Causal Relationships in  
an Observational HIV/TB Study

Lucy Campbell  
Department of Statistical Sciences  
University of Cape Town

Supervisors:  
Associate Professor Francesca Little  
and Associate Professor Andrew Boulle

5 December 2012

# Contents

|          |                                                                                                          |           |
|----------|----------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                                                                      | <b>1</b>  |
| 1.1      | International Epidemiological Databases to Evaluate AIDS (IeDEA) . . . . .                               | 2         |
| 1.1.1    | IeDEA Southern Africa . . . . .                                                                          | 2         |
| 1.2      | Background to research problem . . . . .                                                                 | 3         |
| 1.2.1    | South African ART guidelines . . . . .                                                                   | 4         |
| 1.2.2    | Observational studies showing improvements in survival resulting from ART initiation during TB treatment | 5         |
| 1.2.3    | Randomized controlled trials showing survival benefits of early ART initiation . . . . .                 | 6         |
| 1.3      | Objective and aims for thesis . . . . .                                                                  | 9         |
| 1.3.1    | Clinical objective and aims . . . . .                                                                    | 9         |
| 1.3.2    | Statistical Objective and Aims . . . . .                                                                 | 12        |
| 1.4      | Nature of cohort study and data available . . . . .                                                      | 12        |
| 1.5      | Statistical methods used . . . . .                                                                       | 19        |
| <b>I</b> | <b>Time to ART Analysis</b>                                                                              | <b>20</b> |
| <b>2</b> | <b>Introduction to Time to ART Analysis</b>                                                              | <b>21</b> |
| <b>3</b> | <b>Methodology</b>                                                                                       | <b>23</b> |
| 3.1      | Overview of time to event analysis . . . . .                                                             | 23        |
| 3.2      | Cox Proportional Hazards Model . . . . .                                                                 | 27        |
| 3.2.1    | Testing the proportional hazards assumption . . . . .                                                    | 28        |
| 3.3      | Parametric Survival Models . . . . .                                                                     | 29        |
| 3.3.1    | Proportional Hazards Models . . . . .                                                                    | 29        |
| 3.3.2    | Accelerated Failure Time Models . . . . .                                                                | 31        |
| 3.4      | Model checking . . . . .                                                                                 | 38        |
| 3.4.1    | Assessing Adequacy of Model Fit . . . . .                                                                | 38        |

|           |                                                                                              |            |
|-----------|----------------------------------------------------------------------------------------------|------------|
| 3.4.2     | Investigation of influential observations . . . . .                                          | 39         |
| 3.5       | Competing Risks Models . . . . .                                                             | 40         |
| 3.5.1     | Bias in the Kaplan-Meier estimates when there are<br>competing risks . . . . .               | 41         |
| 3.5.2     | Methods for assessing competing risks . . . . .                                              | 42         |
| <b>4</b>  | <b>Analysis of Time to ART Initiation</b>                                                    | <b>51</b>  |
| 4.1       | Summary statistics and Kaplan-Meier plots . . . . .                                          | 53         |
| 4.2       | Cox Proportional Hazards Model . . . . .                                                     | 59         |
| 4.2.1     | Cox Proportional Hazards Regressions . . . . .                                               | 59         |
| 4.2.2     | Graphical assessment of proportional hazards assumption                                      | 61         |
| 4.2.3     | Using interactions of time with covariates to model<br>non-proportional hazards . . . . .    | 64         |
| 4.3       | Parametric Survival Analysis Models . . . . .                                                | 65         |
| 4.3.1     | Models using the proportional hazards metric . . . . .                                       | 65         |
| 4.3.2     | Accelerated Failure Time Models . . . . .                                                    | 68         |
| 4.3.3     | Comparison of Parametric Models . . . . .                                                    | 77         |
| 4.3.4     | Analysis of deviance residuals for Models 8 & 9 . . . . .                                    | 79         |
| 4.3.5     | Sensitivity Analysis . . . . .                                                               | 81         |
| 4.4       | Competing Risks Models . . . . .                                                             | 84         |
| 4.4.1     | Comparison of patient characteristics by failure event . . . . .                             | 87         |
| 4.4.2     | Competing Risks Regressions . . . . .                                                        | 89         |
| 4.4.3     | Model checking for competing risks regressions . . . . .                                     | 100        |
| <b>5</b>  | <b>Conclusion for Time to ART Analysis</b>                                                   | <b>105</b> |
| <b>II</b> | <b>Outcome Analysis</b>                                                                      | <b>111</b> |
| <b>6</b>  | <b>Causal Models Methodology</b>                                                             | <b>112</b> |
| 6.1       | Introduction . . . . .                                                                       | 112        |
| 6.2       | Directed Acyclic Graphs (DAGs) . . . . .                                                     | 113        |
| 6.2.1     | DAG concepts and terminology . . . . .                                                       | 115        |
| 6.3       | Biases . . . . .                                                                             | 117        |
| 6.3.1     | Confounding and Confounders . . . . .                                                        | 117        |
| 6.3.2     | Selection Bias . . . . .                                                                     | 118        |
| 6.3.3     | Bias relating to time-varying confounding . . . . .                                          | 119        |
| 6.4       | Marginal Structural Models . . . . .                                                         | 121        |
| 6.4.1     | Counterfactual Outcomes . . . . .                                                            | 123        |
| 6.4.2     | Explanations and definitions of assumptions required<br>for valid causal inference . . . . . | 123        |

|          |                                                                                          |            |
|----------|------------------------------------------------------------------------------------------|------------|
| 6.4.3    | Inverse Probability of Treatment and Censoring Weights                                   | 126        |
| 6.5      | Estimating Inverse Probability Weights (IPW)                                             | 128        |
| 6.5.1    | Modelling the time-dependent intercept                                                   | 128        |
| 6.6      | Performing weighted Regression                                                           | 129        |
| 6.7      | Advantages and Disadvantages of MSMs                                                     | 130        |
| 6.8      | Multiple Imputation                                                                      | 130        |
| 6.8.1    | Motivation for the use of multiple imputation                                            | 130        |
| 6.8.2    | Missing Data Mechanisms                                                                  | 133        |
| 6.8.3    | Imputation Methods                                                                       | 135        |
| 6.8.4    | Imputation Model                                                                         | 141        |
| <b>7</b> | <b>Causal Models</b>                                                                     | <b>143</b> |
| 7.1      | Introduction                                                                             | 143        |
| 7.2      | Unweighted Models                                                                        | 148        |
| 7.3      | Model checking for unweighted models                                                     | 155        |
| 7.3.1    | Assessment of model fit                                                                  | 157        |
| 7.3.2    | Deviance Residuals                                                                       | 157        |
| 7.3.3    | Sensitivity of model estimates to influential patients                                   | 160        |
| 7.4      | Weighted Models                                                                          | 163        |
| 7.4.1    | Inverse probability of treatment weights                                                 | 164        |
| 7.4.2    | Inverse probability of censoring weights                                                 | 169        |
| 7.4.3    | Inverse Probability of Treatment and Censoring Weighted Models                           | 174        |
| 7.4.4    | Investigation of the distribution of the weights to check the weight model specification | 176        |
| 7.5      | Multiple Imputation                                                                      | 178        |
| 7.5.1    | Predictors of missingness                                                                | 180        |
| 7.5.2    | Missing data mechanisms                                                                  | 182        |
| 7.5.3    | Implementation of Imputation                                                             | 182        |
| 7.5.4    | Multiple Imputation Results                                                              | 185        |
| 7.5.5    | Model checking using imputed data-sets                                                   | 191        |
| 7.5.6    | Sensitivity of results to assumptions about missing values                               | 191        |
| 7.5.7    | Summary of estimated hazard ratios of ART under varying modelling assumptions            | 194        |
| 7.6      | Plausibility of assumptions required for valid causal inference                          | 195        |
| 7.6.1    | Conditional exchangeability                                                              | 196        |
| 7.6.2    | No structural non-positivity                                                             | 197        |
| 7.6.3    | Consistency                                                                              | 202        |
| 7.6.4    | No measurement error                                                                     | 202        |
| 7.6.5    | No model misspecification                                                                | 203        |

|          |                                                                                                                          |            |
|----------|--------------------------------------------------------------------------------------------------------------------------|------------|
| 7.7      | Extent to which potential disadvantages of MSMs may affect my analysis . . . . .                                         | 203        |
| <b>8</b> | <b>Conclusion for Outcome Analysis</b>                                                                                   | <b>205</b> |
| 8.1      | Problems arising from using an observational cohort for analysis                                                         | 207        |
| 8.2      | Ideal clinical trial . . . . .                                                                                           | 208        |
| <b>9</b> | <b>Conclusion</b>                                                                                                        | <b>210</b> |
| 9.1      | Suitability of selected methods to cope with the design and data                                                         | 210        |
| 9.2      | Other methods that could have been used . . . . .                                                                        | 213        |
| 9.3      | Clinical objective and aims . . . . .                                                                                    | 214        |
| 9.3.1    | Effect of predictors of earlier ART initiation on the time to ART initiation . . . . .                                   | 215        |
| 9.3.2    | Impact of early ART initiation on survival . . . . .                                                                     | 215        |
| <b>A</b> | <b>Abbreviations</b>                                                                                                     | <b>216</b> |
| <b>B</b> | <b>Papers Published</b>                                                                                                  | <b>218</b> |
| <b>C</b> | <b>Additional Analyses, Graphs and Tables</b>                                                                            | <b>219</b> |
| C.1      | Sensitivity Analyses . . . . .                                                                                           | 219        |
| C.1.1    | Consideration of inclusion of stage and type of TB in time to ART models . . . . .                                       | 219        |
| C.1.2    | Stratified Kaplan-Meier survival plots excluding patients who have no CD4 count near the start of TB treatment . . . . . | 224        |
| C.1.3    | Sensitivity of generalized Gamma models to assumptions about missing CD4 counts . . . . .                                | 224        |
| C.2      | Assessment of proportional hazards . . . . .                                                                             | 227        |
| C.3      | Log-logistic and log-Normal Models . . . . .                                                                             | 229        |
| <b>D</b> | <b>Code</b>                                                                                                              | <b>233</b> |
| D.1      | Time to ART Analysis for patients from all 3 clinic sites who do initiate ART . . . . .                                  | 233        |
| D.1.1    | Tables 4.3-4.6, B.5 and B.6 . . . . .                                                                                    | 233        |
| D.1.2    | Example of code to create plot of Cox-Snell residuals against the estimated cumulative hazard . . . . .                  | 235        |
| D.2      | Competing risks analysis . . . . .                                                                                       | 236        |
| D.2.1    | Figure 3.4 . . . . .                                                                                                     | 236        |
| D.2.2    | Figures 3.5-3.7 and 3.12 . . . . .                                                                                       | 238        |
| D.2.3    | Models for competing risks analysis, Table 4.10 . . . . .                                                                | 243        |
| D.2.4    | Figures 4.13 and 4.15 . . . . .                                                                                          | 244        |

|       |                                                                                                        |     |
|-------|--------------------------------------------------------------------------------------------------------|-----|
| D.2.5 | Figure 4.16 plot 1.a)                                                                                  | 244 |
| D.3   | Inverse Probability Weighting                                                                          | 247 |
| D.3.1 | Estimating Stabilized Inverse Probability of Treatment Weights (SIPTWs)                                | 247 |
| D.3.2 | Estimating the Stabilized Inverse Probability of Censoring Weights (SIPCWs)                            | 249 |
| D.3.3 | Figure 7.7: Median, interquartile range and 5th and 95th percentiles of stabilized IPTWs and IPCWs     | 251 |
| D.3.4 | Figure 7.5: Box plots of stabilized IPTWs by baseline CD4 count category                               | 252 |
| D.3.5 | Code for performing weighted Cox regressions in Stata when there are multiple observations per patient | 252 |
| D.4   | Marginal Structural Models                                                                             | 253 |
| D.4.1 | Table 7.11: Cox regressions using IPTCW to assess effect on hazard of death of being on ART            | 253 |
| D.5   | Multiple Imputation                                                                                    | 253 |
| D.5.1 | Multivariate Normal Data Augmentation (MVN DA)                                                         | 254 |
| D.5.2 | Multiple Imputation by Chained Equations (MICE)                                                        | 257 |

# List of Tables

|      |                                                                                                                                                      |     |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 1.1  | Summary statistics by patient characteristics, Part I . . . . .                                                                                      | 16  |
| 1.2  | Summary statistics by patient characteristics, Part I CR . . . .                                                                                     | 17  |
| 1.3  | Summary statistics by patient characteristics, Part II . . . . .                                                                                     | 18  |
| 4.1  | Characteristics of patients . . . . .                                                                                                                | 55  |
| 4.2  | Days from start of TB treatment to ART initiation by whether<br>patient had a CD4 count near the time of starting TB treat-<br>ment or not . . . . . | 58  |
| 4.3  | Cox PH 123 . . . . .                                                                                                                                 | 60  |
| 4.4  | Weibull PH 4 & 5 . . . . .                                                                                                                           | 67  |
| 4.5  | Weibull AFT 6 & 7 . . . . .                                                                                                                          | 71  |
| 4.6  | Generalized Gamma AFT, Models 8 & 9 . . . . .                                                                                                        | 73  |
| 4.7  | Comparison of model fit statistics for Parametric Models . . .                                                                                       | 79  |
| 4.8  | Sensitivity analysis for assumptions about patients having miss-<br>ing CD4 counts . . . . .                                                         | 83  |
| 4.9  | Summary statistics . . . . .                                                                                                                         | 88  |
| 4.10 | Competing Risks Regressions, Models 14-19 . . . . .                                                                                                  | 91  |
| 4.11 | Comparison of regressions on sub-distribution hazards and<br>cause-specific hazards for time to death pre-ART . . . . .                              | 93  |
| 4.12 | Comparison of regressions on sub-distribution hazards and<br>cause-specific hazards for ART initiation and death pre-ART .                           | 95  |
| 6.1  | Classification of missing-data mechanisms . . . . .                                                                                                  | 134 |
| 7.1  | Repeated values before ART initiation . . . . .                                                                                                      | 146 |
| 7.2  | Categories for time-updated variables . . . . .                                                                                                      | 148 |
| 7.3  | Unweighted Models for differing treatment functions . . . . .                                                                                        | 149 |
| 7.4  | Unweighted Cox regressions using treatment function ‘On ART’                                                                                         | 150 |
| 7.5  | Unweighted Cox regressions using treatment function ‘Weeks<br>not on ART’ . . . . .                                                                  | 151 |

|      |                                                                                                                                     |     |
|------|-------------------------------------------------------------------------------------------------------------------------------------|-----|
| 7.6  | Unweighted Cox regressions that include most recent categories for time-updated covariates . . . . .                                | 156 |
| 7.7  | PLR to estimate IPTW . . . . .                                                                                                      | 166 |
| 7.8  | MSM using IPTW: on ART . . . . .                                                                                                    | 169 |
| 7.9  | PLR to estimate IPCW . . . . .                                                                                                      | 171 |
| 7.10 | MSM using IPCW: on ART . . . . .                                                                                                    | 174 |
| 7.11 | MSM using IPTCW: on ART . . . . .                                                                                                   | 175 |
| 7.12 | MSM using IPTCW: Weeks not on ART . . . . .                                                                                         | 176 |
| 7.13 | Analysis of precision and variance trade-off by truncating weights                                                                  | 177 |
| 7.14 | Unweighted all versus complete cases . . . . .                                                                                      | 179 |
| 7.15 | Logistic regressions for indicators of missing baseline values . .                                                                  | 181 |
| 7.16 | Comparison of means before and after imputation . . . . .                                                                           | 186 |
| 7.17 | Comparison of proportions . . . . .                                                                                                 | 187 |
| 7.18 | Comparison of results for non-MI and MI data . . . . .                                                                              | 188 |
| 7.19 | Comparison of Complete Case versus MI . . . . .                                                                                     | 190 |
| 7.20 | Comparison of unweighted Cox Regressions using treatment function ‘On ART’ for different assumptions about missing values . . . . . | 192 |
| 7.21 | Comparison of unweighted and weighted Cox Regressions . . .                                                                         | 194 |
| 7.22 | Exploration of positivity by category . . . . .                                                                                     | 198 |
| 7.23 | Exploration of positivity by CD4 category and weeks in study                                                                        | 199 |
| 7.24 | Analysis of bias and variance trade-off in different weighting model specifications . . . . .                                       | 201 |
| C.1  | Cox PH sens . . . . .                                                                                                               | 222 |
| C.2  | Generalized Gamma AFT, Models 8s & 9s . . . . .                                                                                     | 223 |
| C.3  | Sensitivity analysis for assumptions about patients having missing CD4 counts ( $\hat{\sigma}$ ) . . . . .                          | 225 |
| C.4  | Sensitivity analysis for assumptions about patients having missing CD4 counts ( $\hat{\kappa}$ ) . . . . .                          | 226 |
| C.5  | Log-logistic AFT 10 & 11 . . . . .                                                                                                  | 231 |
| C.6  | Log-Normal AFT 12 & 13 . . . . .                                                                                                    | 232 |

# List of Figures

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |    |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Map of IeDEA-SA                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |    |
|     | Map obtained from the IeDEA-SA website (ISPM, 2010) . . .                                                                                                                                                                                                                                                                                                                                                                                                                                     | 2  |
| 1.2 | Time to ART Analysis . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 10 |
| 1.3 | Outcome Analysis . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 11 |
| 1.4 | Timing of CD4 counts . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 14 |
| 1.5 | Relationship between Time to ART and Outcome Analyses . .                                                                                                                                                                                                                                                                                                                                                                                                                                     | 19 |
| 3.1 | Shapes of baseline hazard functions for PH Models . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                   | 30 |
| 3.2 | Shapes of hazard curves for Generalized Gamma AFT Model .                                                                                                                                                                                                                                                                                                                                                                                                                                     | 37 |
| 3.3 | Bias in Kaplan-Meier (KM) survival estimates . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                        | 42 |
| 3.4 | Cox PH and cumulative incidence graphs . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                              | 48 |
| 3.5 | Cause-specific hazard and cumulative incidence functions . . .                                                                                                                                                                                                                                                                                                                                                                                                                                | 49 |
| 3.6 | Relative cause-specific hazard and cumulative incidence func-<br>tions . . . . .                                                                                                                                                                                                                                                                                                                                                                                                              | 49 |
| 3.7 | Stacked cause-specific hazard and cumulative incidence functions                                                                                                                                                                                                                                                                                                                                                                                                                              | 50 |
| 4.1 | Kaplan-Meier survival plots for time to ART initiation, N=1444                                                                                                                                                                                                                                                                                                                                                                                                                                | 56 |
| 4.2 | Smoothed hazard curves estimated using Cox PH Model<br>Stratified hazard curves by categorization of each predictor variable,<br>whilst keeping other predictor variables constant at covariate pattern<br>that yielded shortest time to ART initiation. This covariate pattern cor-<br>responded to enrolment clinic site 3, TB diagnosis in the calendar period<br>2006-2008, TB diagnosis ART clinic and CD4 count near the start of TB<br>treatment lower than 50 cells/ $\mu$ L. . . . . | 61 |
| 4.3 | Non-parametric smoothed hazard curves . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                               | 62 |
| 4.4 | Comparison of log-log survival curves by covariate<br>Estimates adjust to the baseline values of the other variables included in<br>Model 1 . . . . .                                                                                                                                                                                                                                                                                                                                         | 63 |
| 4.5 | Comparison of hazard functions for Weibull Models . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                   | 69 |
| 4.6 | Generalized Gamma AFT hazard and survival curves - base-<br>line: mode . . . . .                                                                                                                                                                                                                                                                                                                                                                                                              | 76 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |     |
|------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.7  | Comparison of model fit of Parametric Models . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 78  |
| 4.8  | Plots of deviance residuals for Models 8 and 9 . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 79  |
| 4.9  | Comparison of predicted hazard curves for Model 9 . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                              | 80  |
| 4.10 | Comparison of predicted survival curves for Model 9 . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                            | 80  |
| 4.11 | Comparison of generalized Gamma survival curves by CD4 count<br>Baseline covariates (solid blue line) correspond to the most frequent values observed for a covariate, which were enrolment at clinic site 1, TB diagnosis in the calendar period 2006-2008, TB diagnosis in TB clinic (before enrolment for HIV care) and CD4 count near the start of TB treatment lower than 50 cells/ $\mu$ L (LHS) or the grouped category of no CD4 count and CD4 count lower than 50 cells/ $\mu$ L (RHS). . . . . | 82  |
| 4.12 | Cause-specific hazard and cumulative incidence functions . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                         | 86  |
| 4.13 | Comparison of log-log survival curves by covariate<br>Estimates adjust to the baseline values of the other variables included in the Model . . . . .                                                                                                                                                                                                                                                                                                                                                     | 90  |
| 4.14 | Log-log plots for time to death . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 94  |
| 4.15 | Cause-specific hazard curves . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 96  |
| 4.16 | CIF Curves . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 98  |
| 4.17 | Cox-Snell residual plots for Models 17 to 19 . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 100 |
| 4.18 | Deviance residual plots for Models 17 to 19 . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 101 |
| 4.19 | Selected Delta-betas for Models 17 to 19 . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 102 |
| 4.20 | Likelihood displacement value plots for Models 17 to 19 . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                          | 103 |
| 5.1  | Component delays to ART initiation . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 109 |
| 6.1  | Convergence diagnostics of MCMC process: converged process                                                                                                                                                                                                                                                                                                                                                                                                                                               | 137 |
| 6.2  | Convergence diagnostics of MCMC process: non-converged process . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                 | 138 |
| 7.1  | Smoothed Hazard Curves . . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 153 |
| 7.2  | Cox-Snell residual plots for Models 1.a to 2.d<br>Model labels correspond to those used in Tables 7.3-7.6 . . . .                                                                                                                                                                                                                                                                                                                                                                                        | 158 |
| 7.3  | Deviance residual plots for Models 1.a to 2.d<br>Model labels correspond to those used in Tables 7.3-7.6 . . . .                                                                                                                                                                                                                                                                                                                                                                                         | 159 |
| 7.4  | Likelihood displacement value plots for Models 1.a to 2.d<br>Model labels correspond to those used in Tables 7.3-7.6 . . . .                                                                                                                                                                                                                                                                                                                                                                             | 161 |
| 7.5  | Box plots of stabilized IPTWs by baseline CD4 count category                                                                                                                                                                                                                                                                                                                                                                                                                                             | 167 |
| 7.6  | Box plots of stabilized IPCWs by baseline CD4 count category                                                                                                                                                                                                                                                                                                                                                                                                                                             | 172 |
| 7.7  | Median, interquartile range and 5th and 95th percentiles of stabilized IPTWs and IPCWs . . . . .                                                                                                                                                                                                                                                                                                                                                                                                         | 173 |

|     |                                                                                              |     |
|-----|----------------------------------------------------------------------------------------------|-----|
| 7.8 | Distribution of variables . . . . .                                                          | 183 |
| C.1 | Kaplan-Meier survival plots for time to ART initiation, N=1162224                            |     |
| C.2 | Comparison of KM vs Cox survival curves by covariate, N=1444227                              |     |
| C.3 | Comparison of KM vs Cox survival curves by CD4 count,<br>N=1444 . . . . .                    | 228 |
| C.4 | Possible shapes of hazard functions for Log-logistic and Log-<br>Normal AFT Models . . . . . | 230 |

# Acknowledgements

I am extremely grateful to Francesca Little for her dedicated support and advice and for always being available for me to ask her questions, as well as setting aside large amounts of time to read through my work and make helpful suggestions and comments. I am very thankful to Andrew Boulle for his help in understanding the dataset and how to go about starting to analyse it, as well as his expert opinion on clinical issues and astute comments on my work.

I enjoyed the opportunity to work with Steve Lawn and colleagues while doing analyses for two papers that we co-authored, during which time I was able to learn from their insight into analyses of HIV/TB co-infected patients. I was also privileged to be able to attend a number of courses and conferences, many of which contributed to my understanding of the statistical methods I used and possible alternatives, as well as helping me gain a broader perspective on the clinical research questions related to my data.

I would like to thank my colleagues from CIDER and from the Department of Statistical Sciences at the University of Cape Town (UCT), who have made the last three years interesting and enjoyable and offered encouragement, support and advice as needed. I am very grateful to my family and friends for their ongoing love, encouragement, support and prayers. I am very thankful to God, who created and sustains me and has often encouraged me and helped me persevere when I felt like I couldn't carry on.

I am grateful to the National Institutes of Health (NIH), the Harry Crossley Foundation and the Department of Statistical Sciences at UCT for the scholarships I received through them. I would like to thank the IeDEA-SA data centres for access to the data I used, as well as acknowledge the site investigators, site staff and patients, the provincial government of the Western Cape, Médecins Sans Frontières (MSF), the Desmond Tutu HIV Centre and those who die without access to ART.

## Abstract

In the Western Cape Province of South Africa, there is a high prevalence of patients infected with both the Human Immunodeficiency Virus (HIV) and tuberculosis (TB). Patients co-infected with HIV/TB have a higher risk of death than patients infected with only one of these diseases. Providing Antiretroviral Therapy (ART) to patients can help to restore their immune systems. There has been concern among clinicians, however, about the safety and efficacy of initiating ART while a patient is on TB treatment. Some of these concerns relate to the possible development of the Immune Reconstitution Inflammatory Syndrome (IRIS), drug interactions leading to sub-therapeutic doses, the high pill burden during the initial phase of TB treatment (which could lead to poor adherence) and also the difficulty of determining which drug is responsible for causing an adverse reaction if treatment for TB and ART are started in quick succession. A key question of interest is therefore the optimal time to initiate ART in patients who are on treatment for TB.

The clinical objective in this thesis is therefore to analyse time to ART initiation. The two main clinical aims are to explore the effect on the time to ART initiation of factors that have been found to be predictors of how soon a patient initiates ART and to determine the causal effect of ART initiation on the survival of patients.

I explore these questions using observational data on patients co-infected with HIV/TB that was collected by three clinics in the Western Cape. One of the challenges of using data from three different clinics is that these clinics did not always provide the same information on patients and sometimes followed different protocols about dates that were recorded, the timing of measurements and the patients for whom information was collected. A drawback of using observational data is that measurements on patients are not always recorded within time intervals of interest, nor are they always recorded regularly. A second shortcoming of observational data is that treatment of patients is not assigned at random. This means that it is probable that patients who have a worse prognosis a-priori will be observed to initiate ART sooner than those who have a better prognosis. This could bias the estimated effect of ART initiation on survival. I will therefore use methods that attempt to reduce potential biases, as well as explore the sensitivity of assumptions that I make where possible.

The statistical objective of the thesis is to explore the suitability of models for time to event, competing risks and causal analysis to address the clinical questions and cope with the nature of the data.

Both clinical questions involve the use of time to event data and I therefore use a variety of survival analysis methods. In order to explore the effects of predictors of ART initiation on the time to ART initiation, I use Cox regressions, parametric regressions and competing risks regressions. To determine the causal effect of ART initiation on survival, I use marginal structural models to account for potential time-dependent confounding. This involves constructing inverse probability of treatment and censoring weights and weighting the analysis model of interest by these weights. I also perform multiple imputation in order to test the sensitivity of using missing categories for the values of covariates for which measurements were not recorded within the time intervals of interest versus using imputed values.

For the exploration of the effect of factors that are predictive of ART initiation on the time to ART initiation, the generalized Gamma accelerated failure time model gives the best model fit. I find that patients who receive HIV care at clinic site 3, have very low CD4 counts, are diagnosed with TB in later years and are diagnosed with TB as a part of the screening process for ART initiation have the shortest mean time to ART initiation (30 days, 95% Confidence Interval (CI): 26 to 34 days). The longest mean time to ART initiation is estimated to be 231 days (95% CI: 214 to 247). Few patients experienced competing risks, so very similar information was obtained when focusing on the relative risk of ART at a given time versus the relative risk of ART initiation before another event.

Using inverse probability of treatment and censoring weighted Cox regression models, I find that patients who are on ART at a given time interval have a 47% reduced hazard of death (Hazard Ratio (HR): 0.53, robust 95% CI: 0.32 to 0.86), while patients for whom ART is delayed by an additional week have a 4% increased hazard of death (HR: 1.04, robust 95% CI: 1.01 to 1.07). The use of multiple imputation for unrecorded values of measurements rather than using missing categories does not substantially alter results.

# Chapter 1

## Introduction

This thesis presents an analysis of data from an observational cohort study of patients co-infected with tuberculosis (TB) and the Human Immunodeficiency Virus (HIV). The clinical objective is to analyse the time to Antiretroviral Therapy (ART) initiation and the impact of its timing on survival of patients who are on TB treatment. The thesis focuses on statistical methods for the analysis of time to event outcomes and statistical methods that are able to eliminate or reduce biases inherent in observational studies, including statistical methods for determining causal relationships.

The data and research questions were obtained from the Southern African regional data centre of the International Epidemiological Databases to Evaluate the AIDS (IeDEA, where AIDS stands for the Acquired Immune Deficiency Syndrome). I briefly explain the purpose of IeDEA and state its specific aims for the Southern African region. Thereafter I provide background information to explain the importance of investigating the implementation and timing of early ART initiation in patients co-infected with HIV and TB. This is followed by an overview of published papers that focused on the impact of the timing of ART initiation in patients co-infected with HIV and TB. I then introduce the clinical objective and aims of this thesis, followed by a description of the structure of the cohort study and the characteristics of the data, all of which motivated the statistical methods I explored.

## 1.1 International Epidemiological Databases to Evaluate AIDS (IeDEA)

IeDEA is an international research group funded by the National Institute of Allergy and Infectious Diseases (NIAID). It develops and implements methods to enable efficient collection and harmonization of data in order to answer many important HIV/AIDS research questions which require large data-sets. Regional centres have been established in Southern Africa, Central Africa, West Africa, Asia and Australia, North America, Caribbean and Central and South America (Institute of Social and Preventive Medicine, University of Bern (ISPM), 2010).

### 1.1.1 IeDEA Southern Africa

The aims of IeDEA-SA are to:

“Conduct clinical, epidemiological and health services research in order to inform HIV/AIDS and TB service delivery in the region, to increase the capacity for delivering ART, and, ultimately, to improve the prognosis of people living with HIV and AIDS in Southern Africa” (ISPM, 2010).

These aims are accomplished through collecting and analysing data from multiple cohorts in the region (see Figure 1.1) and publishing the results.

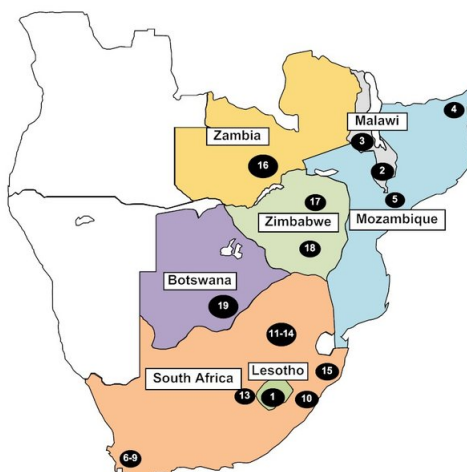


Figure 1.1: Map of IeDEA-SA

Map obtained from the IeDEA-SA website (ISPM, 2010)

## 1.2 Background to research problem

People infected with the HIV have a high risk of developing TB. Raviglione et al. (1997) reviewed the results from different cohort studies showing that for people infected with HIV, the probability of developing activated TB within one year was between 5 and 15% depending on their level of immune suppression. This is much higher than the estimated lifetime risk of 5-10% for people who are not infected with HIV (Aaron et al., 2004). The Western Cape province of South Africa has a particularly high burden of HIV and TB co-infection. Lawn et al. (2006) reported that 25% of patients enrolling for ART at the Gugulethu Community Health Service between 2002 and 2005 had prevalent TB. Boule et al. (2010b) reported that in Khayelitsha, the percentage of patients initiating ART while on treatment for tuberculosis increased from 22.7% in 2001/2002 to 42.4% in 2007.

This high burden of HIV/TB co-infection is a concern as there is a higher risk of death for patients who are co-infected with HIV and TB than those who either are only HIV-positive or only infected with TB or have neither disease. Whalen et al. (1995) found that for patients infected with HIV, the risk of death was twice as high in patients co-infected with TB than for who that did not have TB (odds ratio (OR) adjusted for age, intravenous drug use, CD4 count, previous opportunistic infections and ART was 2.17, 95% CI: 1.27-3.71).

The treatment of patients co-infected with HIV and TB requires careful management. Potential interactions between anti-tuberculosis and antiretroviral drugs may result in sub-therapeutic levels or toxicities (Aaron et al., 2004). If TB treatment and ART are started in quick succession, it can be complicated to determine which drug is responsible for any toxicities that develop (Burman and Jones, 2001). Clinicians are often also concerned about possible lack of adherence as a result of the high pill burden, particularly when ART is initiated during the first phase of treatment for TB (Dean et al., 2002; Dheda et al., 2004).

Additionally, patients receiving ART may experience Immune Reconstitution Inflammatory Syndrome (IRIS), a paradoxical worsening of TB symptoms that does not indicate TB treatment failure, but rather restoration of the patient's immune system. Lawn et al. (2007) found that the risk of developing TB IRIS was higher for patients who had lower CD4 counts and those who initiated ART sooner after starting TB treatment (than those who had

higher CD4 counts or initiated ART longer after starting TB treatment). Dheda et al. (2004) pointed out that patients who are at higher risk of death or a new AIDS defining illness during the early phase of TB treatment are also at high risk of developing paradoxical reactions and require close monitoring if TB treatment and ART are started at the same time (or close together). Nonetheless, Lawn et al. (2007) reported that the risk of death associated with TB IRIS was low and could probably be decreased with better strategies for diagnosing and managing TB IRIS.

These studies show that there is a trade-off between the advantages of early ART initiation and the potential dangers and complications. Data was therefore collected to enable analyses that would provide further insight into the relationship between the timing of ART initiation for patients on TB treatment and the morbidity and mortality of such patients. Further details of the objectives are given in the IeDEA TB concept sheet drawn up by Wood et al. (2007).

Although it would also be interesting to look at other outcomes such as the incidence of IRIS, AIDS Defining Illnesses (ADIs), toxicities, hospitalization or other adverse events, as well as assessing adherence and factors affecting it, there was very little data available to inform analyses surrounding these issues. I therefore focus on survival benefits for patients who initiate ART soon after the start of TB treatment both in my review of other studies below and in my analyses.

### 1.2.1 South African ART guidelines

I give a brief summary of the South African ART guidelines that were in place during the time that the data I have used was collected, as well as recent updates to these guidelines. These were both informed by some of the studies outlined below. The 2004 South African ART guidelines for patients co-infected with HIV/TB recommended initiating ART after the completion of two months of TB treatment for patients with a CD4 count  $< 200$  cells/ $\mu$ L and/or a history of World Health Organisation (WHO) stage IV disease. For patients with CD4 counts of  $< 50$  cells/ $\mu$ L or that had other serious HIV-related illnesses, the recommendation was that the patient should complete at least two weeks of TB treatment and be tolerating TB treatment before initiating ART. In 2010 these guidelines were updated such that all HIV/TB co-infected patients with a CD4 count  $\leq 350$  cells/ $\mu$ L and any who had MDR/XDR TB (irrespective of CD4 count) were eligible for ART. It was

advised that any patients co-infected with TB that had a CD4 count  $\leq 350$  cells/ $\mu$ L should complete two to a maximum of eight weeks of TB treatment before initiating ART, but that in general patients should commence ART as soon as they are tolerating the TB treatment (usually within two to four weeks).

### 1.2.2 Observational studies showing improvements in survival resulting from ART initiation during TB treatment

A number of observational studies have shown that patients who start receiving Highly Active Antiretroviral Therapy (HAART) during TB therapy have lower hazards of death and ADIs than those who do not (Girardi et al., 2001; Dean et al., 2002; Dheda et al., 2004; Manosuthi et al., 2006; Akksilp et al., 2007; Sanguanwongse et al., 2008; Maruza et al., 2012). The median CD4 counts for these studies range between about 50 and 170 cells/ $\mu$ L. Although these studies strengthen the evidence base indicating that it is preferable to initiate ART while patients are on TB treatment, they do not address the important question of the optimal timing for ART initiation during TB treatment.

Sungkanuparph et al. (2006) reported good TB treatment outcomes, few adverse events and only one paradoxical reaction and one death among 29 patients having median CD4 count of 74 (range: 23 to 189) cells/ $\mu$ L who initiated ART at a median of 8 weeks (range 4 to 12 weeks) after starting TB treatment. They therefore recommended early initiation of ART (within 4 to 12 weeks of starting TB treatment) in patients with advanced HIV disease (CD4 count  $< 200$  cells/ $\mu$ L). This was, however, a very small study.

Manosuthi et al. (2006) demonstrated higher survival probabilities among a group of 1003 patients (median CD4 count 53, IQR: 20 to 129 cells/ $\mu$ L for 197 patients for whom this was available) who started ART sooner after their TB diagnoses (within six, nine or twelve months versus later than six, twelve or nine months).

Tansuphasawadikul et al. (2007) reported that among 101 patients with CD4 counts lower than 100 cells/ $\mu$ L, those who die and those who experience opportunistic infections (OIs) have longer delays between TB diagnosis and ART initiation. This could indicate that earlier ART initiation is beneficial,

but it could also be a consequence of factors that result in poorer outcomes in patients also being the cause of longer delays between TB treatment and ART initiation.

Velasco et al. (2009) found that among 313 patients with a median CD4 count of 160 (IQR: 69 to 289) the hazard of death was lower for patients who started HAART within the first two months of TB treatment, but that this survival effect was greater in the short term (six months after the start of TB treatment) than in the long term (twelve months after the start of TB treatment). Nonetheless survival was still greater among patients who started HAART earlier.

Franke et al. (2011) performed a retrospective review and used marginal structural models and simulated two-year survival curves to calculate survival probabilities for patients starting ART 15, 30, 60 and 180 days after starting TB treatment or not at all (the results from this study were only published after I began the analysis for my thesis). They found that the predicted two-year survival probability for patients starting ART 15 days after the start of TB treatment was significantly greater than that of any other timing of ART initiation when the CD4 count was set at 50 or 100 cells/ $\mu\text{L}$ , but not for CD4 counts of 200 or 300 cells/ $\mu\text{L}$ . 322 patients (with a median first CD4 count of 113, range: 1 to 350) were included in the study.

The evidence from these studies seems to point toward starting ART within two months of starting TB treatment and possibly earlier, particularly for patients with lower CD4 counts.

### **1.2.3 Randomized controlled trials showing survival benefits of early ART initiation**

The results from the following randomized controlled trials were published shortly after I began the analysis of the data used for my thesis and address the question of when to initiate ART in patients on TB treatment.

Karim et al. (2010) conducted the “Starting Antiretroviral Therapy at Three Points in Tuberculosis (SAPIT)” trial in Durban, South Africa. They found that among patients with CD4 counts lower than 500 cells/ $\mu\text{L}$ , the rate of death was lower for patients assigned to start ART while on TB treatment than for those who were assigned to complete TB treatment before initiating

ART. This result held true in subsets of patients with CD4 counts above and below 200 cells/ $\mu$ L. Although the incidence of TB-associated IRIS was higher amongst patients assigned to initiate ART while on TB treatment, no deaths were attributable to TB-associated IRIS. No difference was found in the proportions of patients being diagnosed with grade 3 or 4 adverse events (other than IRIS).

Philpott and Schüklenk (2010); Wilson and Meintjes (2010) and Boulle et al. (2010a) critiqued the study by Karim et al. (2010), saying that patients with CD4 counts lower than 200 cells/ $\mu$ L should not have been randomized to receiving ART after the completion of TB treatment. They argued that it was predictable based on prior observational studies that patients with CD4 counts lower than 200 cells/ $\mu$ L would have higher mortality if they only started ART after TB treatment rather than during TB treatment. In addition, the 2004 South African ART guidelines (in force at the start of the SAPIT trial) recommended that patients having CD4 counts lower than 200 cells/ $\mu$ L that were diagnosed with TB should start ART after completing two months of TB treatment (or after completing two weeks and as soon as they are tolerating TB treatment if the patients CD4 counts were below 50 cells/ $\mu$ L). Cohen (2010), however, reports some clinicians and public health officials as saying that the criticism by Philpott and Schüklenk (2010) was too negative and that the results from this study would be helpful in assuring clinicians that starting ART during TB treatment does significantly reduce mortality.

The SAPIT study also included a comparison of patients starting ART within the first four weeks of TB treatment (early ART), versus within the first four weeks of the continuation phase of TB treatment (late ART). Karim et al. (2011) found an increased risk of IRIS (Incidence-Rate Ratio (IRR) 2.62, 95% CI: 1.48-4.82) and of switching antiretroviral drugs as a result of adverse events (10 versus 1,  $p = 0.006$ ) for patients receiving early ART compared to late ART, but no overall AIDS-free survival benefit (IRR 0.89, 95% CI: 0.44-1.79). The median CD4 count was 150 cells/ $\mu$ L and did not differ between groups. In a subset of patients with CD4 cell counts lower than 50 cells/ $\mu$ L however, Karim et al. (2011) found a reduction in the combined incidence of AIDS or death (IRR 0.32, 95% CI: 0.07-1.13) that they suggest outweighs the risks of IRIS (IRR 4.71, 95% CI: 1.48-19.64) and the need to switch antiretroviral drugs because of adverse events (3 versus 0 patients) for patients starting ART earlier versus later with respect to the time they started TB treatment.

Another randomized controlled trial comparing mortality among patients with differing times of starting ART after the start of TB treatment was performed by Blanc et al. (2011) among HIV positive Cambodian patients. This study only included patients with CD4 counts lower than 200 cells/ $\mu$ L that had a TB diagnosis confirmed by a positive smear for acid-fast bacilli. Blanc et al. (2011) found that patients initiating ART two weeks after starting TB treatment had a lower hazard of death than those starting ART eight weeks after the start of TB treatment (HR 0.62, 95% CI: 0.44-0.86). Since the median CD4 count for all patients was very low (25 cells/ $\mu$ L) this result is consistent with that of Karim et al. (2011).

A third study that investigated the optimal timing of initiating ART during TB treatment was conducted by Havlir et al. (2011). They enrolled patients from four continents who had confirmed or probable TB diagnoses and CD4 counts lower than 250 (median 77) cells/ $\mu$ L. They found a reduced proportion of patients dying or experiencing a previously undiagnosed AIDS-defining event in the subset of patients with CD4 counts lower than 50 cells/ $\mu$ L (15.5% versus 26.6%, 95% CI: 1.5-20.5) for those starting ART earlier (within 2 weeks of initiating treatment for TB) versus later (8-12 weeks after initiating treatment for TB). For patients with CD4 cell counts of 50 cells/ $\mu$ L or greater, the difference in proportions of patients dying and experiencing AIDS-defining events was not statistically significant ( $p = 0.67$ ) for those starting ART earlier versus later. Havlir et al. (2011) did find an increased proportion of patients being diagnosed with TB-associated IRIS in the earlier ART compared to the later ART group (11% versus 5%,  $p = 0.002$ ), but no deaths were attributed to TB-associated IRIS. They also did not find a significant difference in the occurrence of drug toxicity or laboratory abnormalities in patients who started ART earlier versus later.

Despite the greater risk of patients developing IRIS and needing to switch antiretroviral drugs as a result of toxicity, all three studies (Karim et al. 2011, Blanc et al. 2011 and Havlir et al. 2011) are in agreement that early ART initiation (two to four weeks after starting TB treatment) is associated with lower mortality for patients having very low CD4 counts (lower than 50 cells/ $\mu$ L or a median of 25 cells/ $\mu$ L).

## 1.3 Objective and aims for thesis

In this section, the clinical objective and aims for the thesis are explained, followed by a summary of the statistical objective and aims.

### 1.3.1 Clinical objective and aims

The clinical objective of this study is to look at the implementation and benefits of early ART initiation for patients co-infected with HIV and TB that are being treated for TB. The two main aims are:

- I. To determine the effect of predictors of earlier ART initiation on the time to ART initiation for patients who are being treated for TB
- II. To determine the impact of early ART initiation on survival of patients who have started TB treatment.

I address these two questions in the **Time to ART Analysis** and **Outcome Analysis** sections of my thesis respectively. The specific issues addressed are outlined in more detail under their respective headings below.

Both these questions have been considered in other observational studies (see for example Section 1.2.2). Results from randomized controlled trials addressing the second question were also published during the analysis period of this thesis (see Section 1.2.3 for further details). Nonetheless, consideration of these questions within the context of public sector health clinics in the Western Cape may be able to contribute valuable insight into the implementation and benefits of ART in this setting. Lawn and Wood (2007) advise that the optimal timing of ART initiation may not be the same in different settings. Franke et al. (2011) and Lawn et al. (2011d) also draw attention to the fact that randomized controlled trials are often carried out under conditions that are not necessarily generalizable to specific populations of interest. For example, randomized controlled trials often require patients to have smear-positive pulmonary TB for them to be included in the trial, but patients with lower CD4 counts often have smear-negative pulmonary or extra-pulmonary TB, which are associated with a lower probability of survival (Lawn et al., 2011d).

### Time to ART Analysis (Part I)

In preliminary analyses of the data used for this thesis (Lawn et al., 2011a,b, see Appendix B), the clinic site at which patients accessed HIV care, the calendar period in which they were diagnosed with TB, whether they were diagnosed with TB before or after enrolling for HIV care (in the TB or ART clinics respectively) and their CD4 counts close to the time of TB diagnosis were identified as factors that played a significant role in determining the length of time from the patient starting treatment for TB to initiating ART. I therefore chose to further investigate and compare the magnitude of the effects of these factors under differing modelling assumptions.

For this analysis, the baseline time was the start of TB treatment (which could occur after or before a patient enrolled for HIV care). The event of interest was ART initiation while on TB treatment, as illustrated in Figure 1.2.

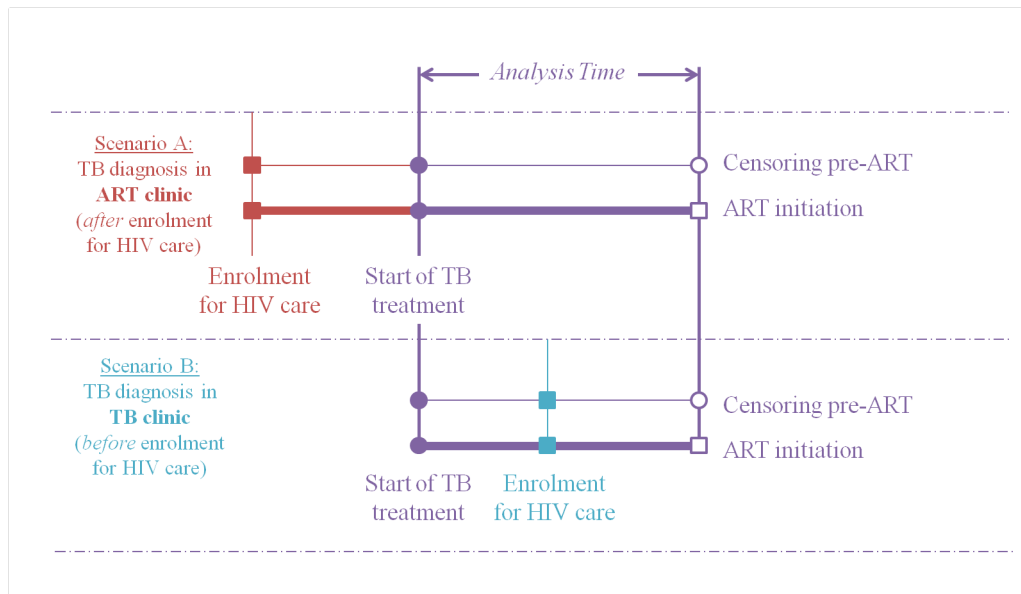


Figure 1.2: Time to ART Analysis

Some patients, however, were not observed to initiate ART while on TB treatment as a result of death pre-ART, permanent deferral for ART or the completion of TB treatment. These patients are denoted as censored.

## Outcome Analysis (Part II)

Interest in this section is on the causal effect that the timing of ART initiation has on patients' survival. I explore this effect using two different ways of quantifying the timing of ART initiation:

1. Binary time-updated variable indicating whether the patient is on ART or not
2. Variable denoting the cumulative amount of time that a patient is not on ART from the baseline time, i.e. the delay in ART initiation.

For this analysis, patients were only considered to have entered the study if they had both enrolled for ART and started TB treatment, so the baseline time was defined as the later of the start of TB treatment and enrolment for ART, as shown in Figure 1.3.

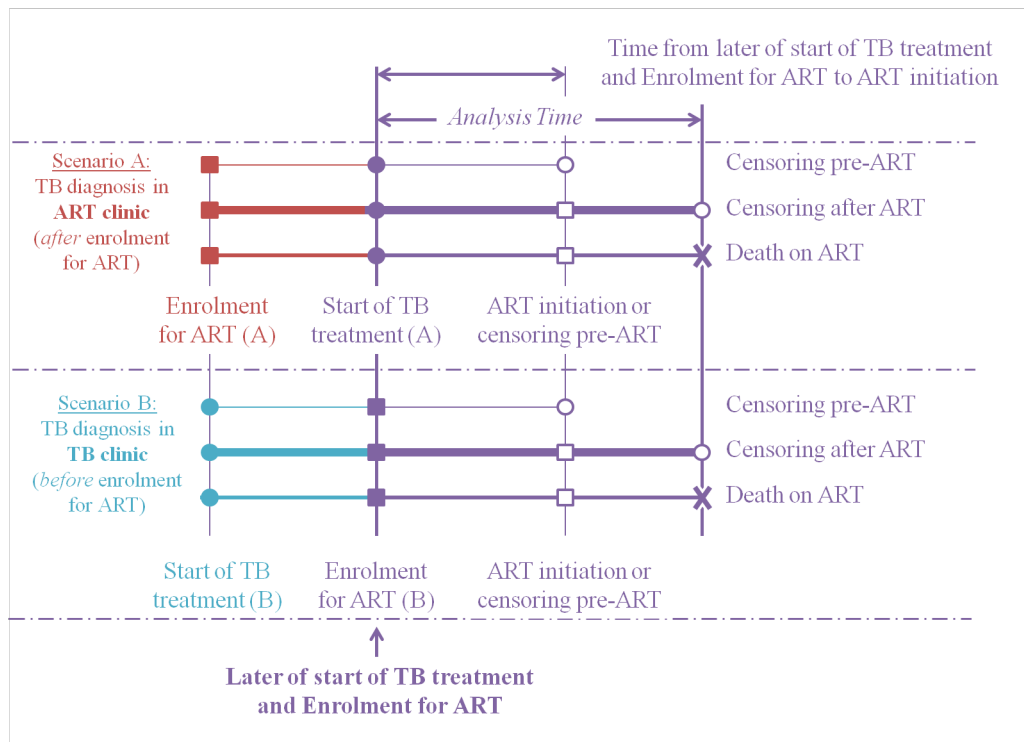


Figure 1.3: Outcome Analysis

The term 'Enrolment for ART' is used to denote the day on which a patient began the screening process for ART initiation. The event of interest was

death within 18 months of the later of the start of TB treatment and enrolment for ART. Patients who were transferred out or lost to follow-up, as well as those who were still in care at the end of the study were treated as censored.

### 1.3.2 Statistical Objective and Aims

The statistical objective of this thesis was to become familiar with the theory and application of statistical methods for dealing with analysis of observational data-sets, with particular reference to time to event analyses and determination of causal relationships. The aims were to:

1. Use appropriate models to explore the effect of different variables on the time to an event in the presence of non-proportional hazards
2. Assess the impact of competing risks on the relative times to event for patients with different covariate values
3. Use inverse probability of treatment and censoring weighted models to obtain unbiased causal estimates for the effect of a time-varying treatment on outcome
4. Compare the use of missing categories to the use of multiply imputed values

## 1.4 Nature of cohort study and data available

The data for this thesis is from a multi-centre study comprising patient information from ART and TB clinics located in three areas in the Western Cape during the period 2002-2008 (a subset of the centres marked 6-9 (bottom left-hand corner) on the IeDEA-SA map shown in Figure 1.1). All patients included in the study both enrolled for HIV care in an ART clinic and started TB treatment prior to ART initiation. The ordering of these two events, however, may have differed (see Figures 1.2 and 1.3). Some patients enrolled for HIV care and were then diagnosed with TB in the ART clinic before they started ART (scenario A). Others were diagnosed with TB in a TB clinic and subsequently enrolled for HIV care at the ART clinic (scenario B). For patients from site 1 and 3 the date at which a patient was considered to have enrolled for HIV care was the date that the screening process for ART initiation began (referred to as ‘enrolment for ART’ in analyses including only

patients from site 1, as in Figure 1.3).

There were 1772 patients from all three clinics, 1444 of whom initiated ART while on TB treatment. Where an end date for time of TB treatment was not recorded, TB treatment was assumed to end six months after it started for patients who had new TB infections and after eight months for patients being retreated for TB. Due to not having information about patients who did not initiate ART from two of the three clinics, the analysis of predictors of time to ART initiation (referred to as ‘Time to ART Analysis’) included only the subset of 1444 patients who did initiate ART while on TB treatment, thereby including no censored observations. A further analysis was performed including only patients from clinic site 1, which included all patients from this clinic who both started TB treatment and enrolled for ART, whether or not they were recorded to initiate ART. Patients from this clinic were also used for the Outcome Analysis. Eight patients had no values recorded for time-updated measurements, two patients had irreconcilable dates and a further 22 patients had no values recorded for time updated measurements beyond the baseline time. These patients were all excluded from the Outcome Analysis, leaving 861 patients of whom 133 (15%) died within 18 months of the later of being diagnosed with TB and enrolling for ART.

Tables 1.1 to 1.3 present an overview of the data aimed at showing patient characteristics for which data were available and that will be considered as potential predictors of ART initiation and/or death, as well as patient numbers for each of the analyses. While these tables present categorical versions for all variables, both categorical and continuous versions of age, the calendar period of TB diagnosis, CD4 count, haemoglobin level and weight were considered in subsequent models. I chose to categorize the calendar period of ART initiation as before 2006 versus 2006 and later, as an increase in provider-initiated HIV testing and counselling for TB patients began in 2005 in the Western Cape (Lawn et al., 2011c).

If clinic protocol for the management of HIV positive patients was followed and there were no complications, patients should have a first CD4 count two weeks after enrolling for ART. ART would then be initiated two weeks later, assuming the patient met the criteria for ART initiation. Patients at site 1 would then have CD4 count measurements taken every 16 weeks from the time of ART initiation, those at sites 2 and 3 would have CD4 counts taken every six months after ART initiation. As a result of the CIPRA-SA trial, some patients at site 3 may have had CD4 counts taken more often than at six-monthly intervals (Catherine Orrel, personal communication). Figure 1.4

shows that the CD4 counts cluster around 16 weeks after ART initiation for patients from site 1, six months (24 weeks) after ART initiation for site 2 and three months (12 weeks) for patients from site 3.

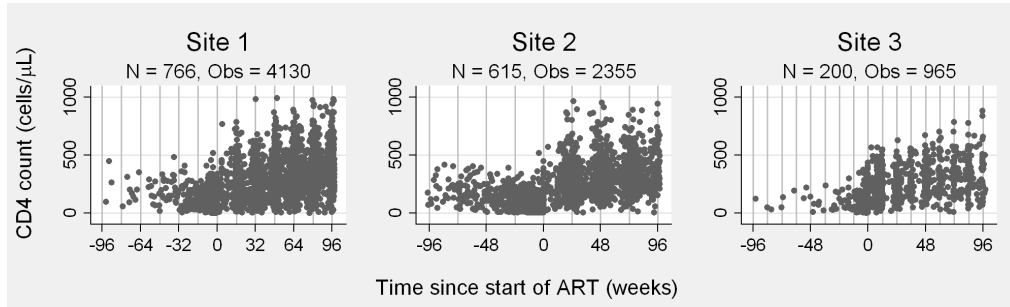


Figure 1.4: Timing of CD4 counts

The clinic site refers to the location of the clinic, so if a patient was enrolled for HIV care at ART clinic site 1, then they would also have received TB care at TB clinic site 1.

CD4 count, haemoglobin level (indicating whether patient has anaemia) and weight measurements were not always recorded sufficiently close to the baseline time of interest, thus resulting in some missing baseline values. I therefore chose to categorize these variables and include a missing category. This seemed a reasonable way of analysing these variables as clinicians often use cut-offs in decision-making. Anaemia classification and weight category were only considered as variables to be included in models for the Outcome Analysis.

For the Time to ART Analysis, which used the start of TB treatment as the baseline date, baseline CD4 counts were obtained by taking the closest CD4 count to the start of TB treatment that satisfied criteria of being taken not earlier than 183 days before or 91 days after the start of TB treatment and not more than seven days after ART initiation. For the Outcome Analysis, which used the later of the start of TB treatment and Enrolment for ART as the baseline date, the value taken closest to baseline time that satisfied criteria of being taken not earlier than 183 days before or 183 days after the baseline date and not more than seven days after ART initiation, was considered to be the baseline value for each of the measurements CD4 count, haemoglobin level and weight.

For the Outcome Analysis, I chose to categorize CD4 counts into five cate-

gories, 0-49, 50-99, 100-199,  $\geq 200$  and missing as these are categories that are commonly used in decision making by clinicians and including too many categories in this analysis may have lead to positivity problems. For the Time to ART Analysis, for which more patients were included, I chose to include an additional CD4 count category by splitting the category 100-199 into 100-149 and 150-199. Categories for anaemia classification and weight were chosen based on clinical rationale and such that sufficient numbers of patients were in each grouping (to prevent positivity problems). In Section 7.5, I compare the use of multiple imputation rather than missing categories to enable analyses using patients for whom baseline values were not recorded.

Table 1.1 shows the numbers and percentages of patients and the median and interquartile ranges (IQRs) for the time from the start of TB treatment to ART initiation (in months) by category for all patients who did initiate ART from all three clinic sites. There were 1444 patients with a median time to ART initiation of 2.63 (IQR: 1.54 to 4.16) months. Patients treated in clinic site 3, those treated in 2006 or after, those diagnosed with TB in the ART clinic and those with CD4 counts lower than 50 cells/ $\mu$ L had comparably shorter median times to ART initiation than other patients.

Table 1.2 shows the number and percentages by category for all patients from site 1, as well as the medians and interquartile ranges of the time from the start of TB treatment to ART initiation by category for patients who did initiate ART. The median time to ART initiation (3.12, IQR: 1.61 to 5.09) was longer than that of patients from all three clinic sites. The median times to ART initiation were comparably shorter for patients who were diagnosed with TB in 2006 or later, those who were diagnosed with TB in the ART clinic and those who had CD4 counts lower than 100 cells/ $\mu$ L.

Table 1.3 shows the numbers and percentages of all patients from site 1 included in the outcome analysis and the numbers and percentages and median and interquartile ranges by category of the time from the later of the start of TB treatment and enrolment for ART (baseline time) to death for patients who died within 18 months of the baseline time. The median time to death was 2.79 (IQR: 1.45 to 5.45) months. Patient groups having comparably shorter median times to death were those who were being treated for a new TB infection, those with WHO clinical stage IV of HIV, those with CD4 counts lower than 50 cells/ $\mu$ L, those with moderate to severe or unknown anaemia classification and those for whom no weight measurement was taken near the baseline time.

Table 1.1: Summary statistics for the time from the start of TB treatment to ART initiation for patients who did initiate ART from all three sites

| Characteristic                                   | All<br>N | (col %) | Months to ART<br>Median (IQR) |               |
|--------------------------------------------------|----------|---------|-------------------------------|---------------|
| Gender                                           |          |         |                               |               |
| Female                                           | 866      | (60%)   | 2.63                          | (1.61 - 4.17) |
| Male                                             | 578      | (40%)   | 2.63                          | (1.48 - 4.14) |
| Age in categories                                |          |         |                               |               |
| 18-24                                            | 149      | (10%)   | 2.63                          | (1.68 - 3.94) |
| 25-39                                            | 973      | (67%)   | 2.53                          | (1.48 - 4.11) |
| $\geq 40$                                        | 322      | (22%)   | 2.97                          | (1.81 - 4.37) |
| Clinic Location                                  |          |         |                               |               |
| Site 1                                           | 671      | (46%)   | 2.73                          | (1.41 - 4.34) |
| Site 2                                           | 594      | (41%)   | 2.83                          | (1.87 - 4.27) |
| Site 3                                           | 179      | (12%)   | 1.97                          | (1.12 - 3.09) |
| Calendar period of TB diagnosis                  |          |         |                               |               |
| 2006 or later                                    | 754      | (52%)   | 1.28                          | (2.17 - 3.45) |
| Before 2006                                      | 690      | (48%)   | 2.04                          | (3.32 - 4.76) |
| Clinic in which TB diagnosed                     |          |         |                               |               |
| ART Clinic                                       | 477      | (33%)   | 1.94                          | (0.92 - 2.96) |
| TB Clinic                                        | 967      | (67%)   | 3.19                          | (1.87 - 4.50) |
| TB classification <sup>a</sup>                   |          |         |                               |               |
| Smear-negative pulmonary                         | 416      | (29%)   | 2.63                          | (1.45 - 4.37) |
| Smear-positive pulmonary                         | 461      | (32%)   | 2.43                          | (1.51 - 3.94) |
| Extra-pulmonary                                  | 567      | (39%)   | 2.79                          | (1.64 - 4.30) |
| TB treatment classification                      |          |         |                               |               |
| Re-treatment (baseline)                          | 419      | (29%)   | 2.66                          | (1.51 - 4.60) |
| New TB infection                                 | 1025     | (71%)   | 2.63                          | (1.54 - 4.07) |
| WHO Clinical Stage of HIV                        |          |         |                               |               |
| I-III                                            | 621      | (43%)   | 2.56                          | (1.58 - 4.14) |
| IV                                               | 823      | (57%)   | 2.69                          | (1.54 - 4.21) |
| CD4 count category (cells/ $\mu$ L) <sup>b</sup> |          |         |                               |               |
| 0-49                                             | 449      | (31%)   | 1.64                          | (0.95 - 2.73) |
| 50-99                                            | 282      | (20%)   | 2.25                          | (1.64 - 3.48) |
| 100-149                                          | 210      | (15%)   | 2.53                          | (1.84 - 3.48) |
| 150-199                                          | 142      | (10%)   | 2.89                          | (2.00 - 4.04) |
| $\geq 200$                                       | 79       | (5%)    | 3.12                          | (2.23 - 4.50) |
| missing                                          | 282      | (20%)   | 4.88                          | (4.07 - 5.82) |
| All patients                                     | 1444     | (100%)  | 2.63                          | (1.54 - 4.16) |

<sup>a</sup> Patients with both pulmonary and extra-pulmonary TB were classified as having extra-pulmonary disease

<sup>b</sup> Measurement taken closest to baseline time that satisfied criteria of being taken not earlier than 183 days before or 91 days after baseline and not more than 7 days after start of ART

Table 1.2: Summary statistics for all patients from site 1 used in the competing risks analysis with time to ART initiation as event of interest

| Characteristic                                   | All |         | Initiate ART |         | Months to ART <sup>a</sup> |               |
|--------------------------------------------------|-----|---------|--------------|---------|----------------------------|---------------|
|                                                  | N   | (col %) | N            | (col %) | Median (IQR)               |               |
| Gender                                           |     |         |              |         |                            |               |
| Female                                           | 524 | (59%)   | 460          | (59%)   | 3.22                       | (1.61 - 5.36) |
| Male                                             | 369 | (41%)   | 316          | (41%)   | 3.12                       | (1.61 - 4.90) |
| Age in categories                                |     |         |              |         |                            |               |
| 18-24                                            | 91  | (10%)   | 81           | (10%)   | 3.45                       | (1.81 - 5.32) |
| 25-39                                            | 602 | (67%)   | 524          | (68%)   | 2.96                       | (1.53 - 5.06) |
| ≥ 40                                             | 200 | (22%)   | 171          | (22%)   | 3.48                       | (1.81 - 5.06) |
| Calendar period of TB diagnosis                  |     |         |              |         |                            |               |
| 2006 or later                                    | 481 | (54%)   | 411          | (53%)   | 2.40                       | (1.31 - 4.17) |
| Before 2006                                      | 412 | (46%)   | 365          | (47%)   | 4.11                       | (2.30 - 5.91) |
| Clinic in which TB diagnosed                     |     |         |              |         |                            |               |
| ART Clinic                                       | 211 | (24%)   | 189          | (24%)   | 1.38                       | (0.69 - 2.73) |
| TB Clinic                                        | 682 | (76%)   | 587          | (76%)   | 3.78                       | (2.17 - 5.52) |
| TB classification <sup>b</sup>                   |     |         |              |         |                            |               |
| Smear-negative pulmonary                         | 225 | (25%)   | 196          | (25%)   | 2.84                       | (1.4 - 4.9)   |
| Smear-positive pulmonary                         | 293 | (33%)   | 248          | (32%)   | 3.06                       | (1.63 - 5.08) |
| Extra-pulmonary                                  | 375 | (42%)   | 332          | (43%)   | 3.25                       | (1.77 - 5.32) |
| TB treatment classification                      |     |         |              |         |                            |               |
| Re-treatment (baseline)                          | 266 | (30%)   | 235          | (30%)   | 3.58                       | (1.74 - 5.72) |
| New TB infection                                 | 627 | (70%)   | 541          | (70%)   | 3.09                       | (1.61 - 4.86) |
| WHO Clinical Stage of HIV                        |     |         |              |         |                            |               |
| I-III                                            | 613 | (69%)   | 538          | (69%)   | 3.09                       | (1.61 - 5.13) |
| IV                                               | 280 | (31%)   | 238          | (31%)   | 3.25                       | (1.61 - 5.09) |
| CD4 count category (cells/ $\mu$ L) <sup>c</sup> |     |         |              |         |                            |               |
| 0-49                                             | 197 | (22%)   | 178          | (23%)   | 1.48                       | (0.85 - 2.73) |
| 50-99                                            | 135 | (15%)   | 125          | (16%)   | 1.87                       | (1.35 - 3.06) |
| 100-149                                          | 102 | (11%)   | 90           | (12%)   | 2.25                       | (1.45 - 3.22) |
| 150-199                                          | 70  | (8%)    | 65           | (8%)    | 2.73                       | (1.64 - 4.11) |
| ≥ 200                                            | 58  | (6%)    | 49           | (6%)    | 2.99                       | (2.23 - 4.14) |
| missing                                          | 331 | (37%)   | 269          | (35%)   | 5.45                       | (4.47 - 6.83) |
| All patients                                     | 893 | (100%)  | 776          | (100%)  | 3.12                       | (1.61 - 5.09) |

<sup>a</sup> Only the 776 patients who were observed to initiate ART were included

<sup>b</sup> Patients with both pulmonary and extra-pulmonary TB were classified as having extra-pulmonary disease

<sup>c</sup> Measurement taken closest to baseline time that satisfied criteria of being taken not earlier than 183 days before or 91 days after baseline and not more than 7 days after start of ART

Table 1.3: Summary statistics including time to death for patients from site 1 who were included in the outcome analysis

| Characteristic                                   | All        |               | Dead       |               | Months to death <sup>a</sup> |                      |
|--------------------------------------------------|------------|---------------|------------|---------------|------------------------------|----------------------|
|                                                  | N          | (col %)       | N          | (col %)       | Median (IQR)                 |                      |
| Gender                                           |            |               |            |               |                              |                      |
| Female                                           | 505        | (59%)         | 76         | (57%)         | 2.78                         | (1.33 - 7.34)        |
| Male                                             | 356        | (41%)         | 57         | (43%)         | 2.79                         | (1.48 - 5.26)        |
| Age in categories                                |            |               |            |               |                              |                      |
| 18-24                                            | 71         | (8%)          | 10         | (8%)          | 3.24                         | (1.68 - 8.77)        |
| 25-39                                            | 588        | (68%)         | 87         | (65%)         | 2.66                         | (1.18 - 4.96)        |
| $\geq 40$                                        | 202        | (23%)         | 36         | (27%)         | 3.22                         | (1.68 - 6.78)        |
| Calendar period of TB diagnosis                  |            |               |            |               |                              |                      |
| 2006 or later                                    | 462        | (54%)         | 70         | (53%)         | 2.81                         | (1.15 - 5.42)        |
| Before 2006                                      | 399        | (46%)         | 63         | (47%)         | 2.66                         | (1.68 - 6.41)        |
| Clinic in which TB diagnosed                     |            |               |            |               |                              |                      |
| TB Clinic                                        | 657        | (76%)         | 102        | (77%)         | 2.48                         | (1.22 - 4.96)        |
| ART Clinic                                       | 204        | (24%)         | 31         | (23%)         | 4.57                         | (1.64 - 11.04)       |
| TB classification <sup>b</sup>                   |            |               |            |               |                              |                      |
| SNP                                              | 213        | (25%)         | 37         | (28%)         | 2.73                         | (1.45 - 7.49)        |
| SPP                                              | 281        | (33%)         | 45         | (34%)         | 2.89                         | (1.22 - 4.93)        |
| Extra-pulmonary                                  | 367        | (43%)         | 51         | (38%)         | 2.66                         | (1.71 - 6.41)        |
| TB treatment classification                      |            |               |            |               |                              |                      |
| Re-treatment (baseline)                          | 255        | (30%)         | 48         | (36%)         | 3.48                         | (2.09 - 6.18)        |
| New TB infection                                 | 606        | (70%)         | 85         | (64%)         | 2.33                         | (1.18 - 5.45)        |
| WHO Clinical Stage of HIV                        |            |               |            |               |                              |                      |
| I-III                                            | 590        | (69%)         | 76         | (57%)         | 3.42                         | (1.66 - 7.44)        |
| IV                                               | 271        | (31%)         | 57         | (43%)         | 2.46                         | (1.22 - 4.27)        |
| CD4 count category (cells/ $\mu$ L) <sup>c</sup> |            |               |            |               |                              |                      |
| 0-49                                             | 274        | (32%)         | 61         | (46%)         | 2.10                         | (1.15 - 3.42)        |
| 50-99                                            | 205        | (24%)         | 23         | (17%)         | 4.21                         | (2.73 - 6.64)        |
| 100-199                                          | 260        | (30%)         | 23         | (17%)         | 4.93                         | (2.07 - 11.33)       |
| $\geq 200$                                       | 89         | (10%)         | 13         | (10%)         | 3.45                         | (1.77 - 11.10)       |
| missing                                          | 33         | (4%)          | 13         | (10%)         | 4.27                         | (1.71 - 4.76)        |
| Baseline Anaemia Classification <sup>c</sup>     |            |               |            |               |                              |                      |
| severe/moderate                                  | 365        | (42%)         | 65         | (49%)         | 2.50                         | (1.15 - 4.63)        |
| mild/none                                        | 412        | (48%)         | 41         | (31%)         | 4.14                         | (2.07 - 8.05)        |
| missing                                          | 84         | (10%)         | 27         | (20%)         | 2.37                         | (1.48 - 7.49)        |
| Baseline Weight <sup>c</sup>                     |            |               |            |               |                              |                      |
| (0; 50)                                          | 130        | (15%)         | 30         | (23%)         | 2.81                         | (2.10 - 6.93)        |
| [50; 60)                                         | 253        | (29%)         | 30         | (23%)         | 3.17                         | (1.91 - 5.42)        |
| [60; $\infty$ )                                  | 227        | (26%)         | 13         | (10%)         | 3.42                         | (1.15 - 4.93)        |
| missing                                          | 251        | (29%)         | 60         | (45%)         | 2.40                         | (1.07 - 5.11)        |
| <b>All patients</b>                              | <b>861</b> | <b>(100%)</b> | <b>133</b> | <b>(100%)</b> | <b>2.79</b>                  | <b>(1.45 - 5.45)</b> |

<sup>a</sup> Follow-up was administratively censored 18 months after the baseline time

<sup>b</sup> Patients with both pulmonary and extra-pulmonary TB were classified as having extra-pulmonary disease; SNP: Smear-negative pulmonary, SPP: Smear-positive pulmonary

<sup>c</sup> Measurement taken closest to baseline time that satisfied criteria of being taken not earlier than 183 days before or 183 days after baseline and not more than 7 days after start of ART

Figure 1.5 shows how the Time to ART and Outcome Analyses relate to each other. Predictors of time to ART initiation may also be predictors of time to death. The time to ART initiation is also a predictor of the time to death. I have not, however, performed a formal path analysis.

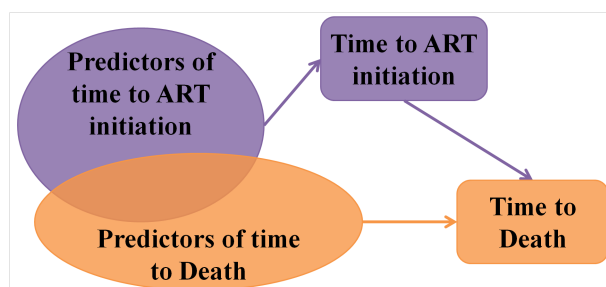


Figure 1.5: Relationship between Time to ART and Outcome Analyses

## 1.5 Statistical methods used

There are many different methods that could potentially be used to analyse these data in order to deal with complications arising as a result of the observational nature of this data set and to answer questions of interest. Since the outcomes for both the Time to ART and Outcome Analyses were time to event variables (time to ART initiation and time to death respectively), survival analysis techniques were used. These included Cox proportional hazards models, parametric models and competing risks models. In addition, I implemented the method of marginal structural models to attempt to obtain unbiased causal estimates of the effect of the timing of ART initiation on death (since CD4 count, haemoglobin measurements and weight were potential time-dependent confounders). Finally, I performed multiple imputation as an alternative to complete case analysis or using a missing category for patients who had no values recorded for a specific measurement (such as CD4 count, haemoglobin level or weight) near the baseline time.

The next section will introduce methodology for the Time to ART Analysis, which will be followed by a section describing the implementation of those methods to the data and the results obtained. Thereafter I will introduce further methodology used for the Outcome Analysis, followed by a section explaining the implementation and presenting the results of applying those methods to the data.

# Part I

## Time to ART Analysis

## Chapter 2

# Introduction to Time to ART Analysis

In this section of the thesis, I address the first aim: to determine the effect that different predictors of ART initiation have on the time to ART initiation. It thus focuses on the pre-ART phase of the study. Analyses of the time from the start of TB treatment to ART initiation are presented and discussed. The response is a *time to event* type of response with censoring (or competing risks) since not all subjects initiated ART while on TB treatment (some died pre-ART, some were recorded as permanently deferred for ART initiation and others completed their TB treatment without initiating ART).

The statistical methodology used is regression type models for time to event data, including:

- Cox Proportional Hazards (PH) models
- Parametric Survival Analysis Models
- Competing Risks models

The methodology that I present in Chapter 3 follows closely that of Cleves et al. (2008) and Collett (2003) for the Cox proportional hazards and parametric survival analysis sections. The Competing risks section incorporates the methodology presented in Fine and Gray (1999), Gichangi and Vach (2005) and Putter et al. (2007).

Cox proportional hazards regression is commonly used in the analysis of time to event data. It is a semi-parametric method that assumes that the

effect of a factor at one point in time is the same at a later point in time. It makes no assumptions other than this proportional hazards assumption and can thus be useful when the modeller does not want to impose a specific form on the shape of the hazard function on the data. Parametric survival analysis is an alternative that does impose restrictions on shape of the hazard functions. Parametric models can be fitted in either a proportional hazards or an accelerated failure time framework. Accelerated failure time models focus on the effect of variables on the time to an event. Competing Risks analysis is important when a patient may experience more than one event, that is, they may experience an event other than the event of interest (such as death), which may prevent (or substantially affect the probability of) the event of interest (for example, ART initiation) occurring.

In Chapter 4, I present the analysis of the time to ART initiation using the methods mentioned above. After introducing the data I used (using summary statistics and Kaplan-Meier plots) I start the formal analysis of the data using the Cox proportional hazards model. Next, I move to a parametric approach, where I fit models using both the proportional hazards and accelerated failure time metrics. The latter allow me to use time (to event) as the response, rather than the hazard function. Parameterizing the ancillary parameters enables me to model different underlying hazard functions for different strata of the variables. Lastly, I use a competing risks framework, which allows me to consider the impact on the event of interest (ART initiation while on TB treatment) of each of the competing risks (death, permanent deferral for ART and the completion of TB treatment), rather than grouping those who experience any of these competing events as censored observations.

# Chapter 3

## Methodology

### 3.1 Overview of time to event analysis

In time-to-event analyses, there are typically four functions of interest. The first is the *probability density* function of the time to event response,  $T$ ,

$$\begin{aligned} f(t) &= \Pr(T = t) \\ &= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}, \end{aligned} \tag{3.1}$$

where  $F(t) = \Pr(T < t)$  is the probability that the event of interest has occurred prior to time  $t$ . The second is the *survival* function that measures the probability of surviving to time  $t$  and beyond without experiencing the event of interest,

$$\begin{aligned} S(t) &= \Pr(T \geq t) \\ &= 1 - \int_0^t f(u) du \\ &= 1 - F(t). \end{aligned} \tag{3.2}$$

The third is the *hazard* function, which is the rate of occurrence of the event of interest up to time  $t$  or the ‘instantaneous’ probability of the event occurring

at time  $t$ , given survival up to that time,

$$\begin{aligned}
 h(t) &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \right\} \\
 &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{1}{\Delta t} \frac{\Pr(t \leq T < t + \Delta t)}{\Pr(T \geq t)} \right\} \\
 &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{1}{\Delta t} \frac{F(t + \Delta t) - F(t)}{S(t)} \right\} \\
 &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)} \\
 &= f(t)/S(t).
 \end{aligned} \tag{3.3}$$

Also, since  $S(t) = 1 - F(t)$ , it follows that  $\log(S(t)) = \log(1 - F(t))$  and using rules for taking derivatives of logs,

$$\frac{d}{dt} \log(1 - F(t)) = \frac{\frac{d}{dt}(1 - F(t))}{1 - F(t)} = -\frac{f(t)}{S(t)},$$

so we have

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \{\log S(t)\}. \tag{3.4}$$

The fourth function is the *integrated* or *cumulative hazard* function

$$\begin{aligned}
 H(t) &= \int_0^t h(u) du \\
 &= -\log S(t),
 \end{aligned} \tag{3.5}$$

or, equivalently,  $S(t) = \exp(-H(t))$ , giving an important link between the hazard and survival models that are used in regression models for time to event outcomes.

Empirical estimates of the survival function can be obtained using the Kaplan-Meier estimator,

$$\hat{S}_{KM}(t) = \prod_{j:t_j \leq t} \left\{ 1 - \frac{d_j}{n_j} \right\}, \tag{3.6}$$

where  $d_j$  is the number of individuals experiencing the event of interest at time  $t_j$  and  $n_j$  is the number of individuals at risk at time  $t_j$ .

An important ingredient of these formulae is the *risk set* in the denominator. This is the number of subjects still under observation who have not yet experienced the event of interest.

If a subject leaves the risk set before experiencing the event of interest, we refer to the subject as *censored*. A subject may be censored if, for example, the subject drops out of the study or the study period ends before the subject experiences the event of interest (known as right-censoring). By removing censored subjects from the risk set, we decrease the denominator and hence increase the estimated probability of a subject experiencing the event of interest (compared to the estimate that would have been obtained had the censored subject been considered to remain in the risk set). This implicitly assumes that censored subjects could still experience the event of interest some time in the unobserved future.

However, in the presence of competing events, this assumption is not valid. A competing event is an event which, if it occurred prior to the event of interest, would prevent the occurrence of the event of interest or substantially change its probability of occurrence (Gooley et al., 1999). For example, if the event of interest is ART initiation, then death pre-ART is a competing event, since subjects who die before starting ART will not be able to initiate ART.

When dealing with competing risks, focus is on specific causes (or an event of interest and its competing risks), rather than overall event-free survival. This leads to the introduction of four further functions of interest that model the joint distribution of the observed survival time  $T$  and the cause of failure  $D$ . The observed survival time  $T$  is the minimum of several potential latent survival times. The risk of the event of interest and the competing risks are all possible causes of failure and are indexed by  $k$ , where  $k = 1, 2, \dots, K$  and  $K$  is the total number of possible causes of failure. The *cause-specific hazard* function,

$$\lambda_k(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t}, \quad (3.7)$$

models the probability of failing from cause  $k$  within a small interval of time,  $\Delta t$ , given that a patient has survived up to time  $t$ . From this equation, we

can derive the *cumulative cause-specific hazard* as

$$\Lambda_k(t) = \int_0^t \lambda_k(s) ds, \quad (3.8)$$

which gives the number of times we would expect to observe failures (caused by event  $k$ ) over a given period (Cleves et al., 2008). The survival function  $S(t)$ , which denotes the probability of not having failed from any cause  $k$  by time  $t$  can then be obtained from the cumulative cause-specific hazards as

$$S(t) = \exp \left( - \sum_{k=1}^K \Lambda_k(t) \right). \quad (3.9)$$

The *cumulative incidence function* of cause  $k$ ,  $I_k(t) = \Pr(T \leq t, D = k)$ , is the probability that a subject fails from cause  $k$  before time  $t$ . It can be obtained from the survival and cause-specific hazard functions as follows:

$$I_k(t) = \int_0^t \lambda_k(s) S(s) ds. \quad (3.10)$$

The effect of covariates can either be modelled on the cause-specific hazards leading to an analogue of Cox proportional hazards regression or on the cumulative incidence function (Putter et al., 2007). Instead of the cause-specific hazard function given in Equation 3.7 above, Fine and Gray (1999) define the sub-distribution hazard

$$\begin{aligned} \bar{\lambda}_k(t) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr\{t \leq T \leq t + \Delta t, D = k | T \geq t \cup (T \leq t \cap D \neq k)\}}{\Delta t} \\ &= \frac{\frac{d}{dt} \bar{I}_k(t)}{1 - \bar{I}_k(t)} \\ &= -\frac{d}{dt} \log\{1 - \bar{I}_k(t)\}. \end{aligned} \quad (3.11)$$

This relationship between  $\bar{\lambda}_k(t)$  and  $\bar{I}_k(t)$  is similar to that between  $h(t)$  and  $F(t)$  shown in (3.4) (substituting  $S(t)$  with  $1 - F(t)$ ).

Putter et al. (2007) explain that the difference between the sub-distribution hazard  $\bar{\lambda}_k$  and the cause-specific hazard  $\lambda_k$  is that for the sub-distribution hazard, subjects that experience a competing event are not removed from the risk set, while for the cause-specific hazard, subjects are removed from the

risk set if they experience any failure event (event of interest or a competing event). This is identified by the union of the set  $(T \leq t \cap D \neq k)$  with the set  $(T \geq t)$  in the conditional component of the numerator for  $\bar{\lambda}_k$ . By keeping subjects who experience a competing event in the risk set, the sub-distribution hazard takes into account that those subjects cannot experience the event of interest (before the competing event).

## 3.2 Cox Proportional Hazards Model

Cox proportional hazards regression models the association between the time to event response and possible predictors by modelling the log of the relative hazard of the event as a linear combination of the predictor variables. The Cox proportional hazards model is based on the equation (Cox, 1972):

$$h_i(t) = h_0(t) \exp(\boldsymbol{\beta}' \mathbf{x}_i), \quad (3.12)$$

where:

$h_0(t)$  is the baseline hazard (hazard for subjects with  $\mathbf{x} = \mathbf{0}$ )

$\boldsymbol{\beta}' \mathbf{x}_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$

$\beta_1, \beta_2, \dots, \beta_k$  are unknown coefficients of the  $k$  predictor variables,  $X_1, X_2, \dots, X_k$

$x_{ji}$  is the value of the  $j^{th}$  predictor variable,  $X_j$ ,  $j = 1, 2, \dots, k$ , for the  $i^{th}$  individual,  $i = 1, 2, \dots, n$ .

The exponentiated regression coefficients  $e^{\beta_j}$ , known as *hazard ratios*, are estimates of the change in the relative hazard of the event associated with a one unit increase in a particular predictor variable (for continuous variables) or the presence of a certain factor compared to its absence (for binary predictor variables or indicator variables).

Hazard ratios greater than one indicate that the presence of a particular factor (if an indicator variable), or its increase (if a continuous variable) increases the hazard of experiencing the event of interest. Hazard ratios smaller than one indicate that the presence of a particular factor (if an indicator variable), or its increase (if a continuous variable) reduces the hazard of experiencing the event of interest.

Cox proportional hazards regression assumes that the impact of a predictor variable on the hazard of an event, as measured by the relative hazard or hazard ratio, stays constant over time, the so-called ‘proportional hazards’ assumption. It also assumes that any censoring is random (such that subjects that are censored would not have been more or less likely to experience the event of interest sooner than subjects that remained under observation). If these assumptions are not met, we need to consider alternative regression models.

### 3.2.1 Testing the proportional hazards assumption

I focus on the graphical assessment of the proportional hazards assumption. The validity of the assumption of proportional hazards can be assessed graphically by comparing *log-log* plots (plots of  $-\log[-\log \hat{S}(t)]$  versus the natural logarithm of time) for each value of a covariate (Cleves et al., 2008). These plots should be parallel if the assumption of proportional hazards (3.12) is to be accepted. This can be seen as follows (for a univariate model, in which I denote the dependency of the survival time on the value of covariate  $X$  as  $S(t|x_i)$ ):

$$\begin{aligned}
 S(t|x_i) &= \exp \left[ - \int_0^t h(s|x_i) ds \right] && \text{using (3.4)} \\
 &= \exp \left[ - \int_0^t h_0(t) \exp(\beta x_i) ds \right] && \text{using (3.12)} \\
 &= \left\{ \exp \left[ - \int_0^t h_0(t) ds \right] \right\}^{\exp(\beta x_i)} \\
 &= \{S_0(t)\}^{\exp(\beta x_i)}
 \end{aligned}$$

Then taking logs on both sides and multiplying by -1

$$-\log\{S(t|x_i)\} = -\exp(\beta x_i) \log\{S_0(t)\}$$

And again taking logs on both sides and multiplying by -1

$$-\log[-\log\{S(t|x_i)\}] = -(\beta x_i) - \log[-\log\{S_0(t)\}]$$

Finally, rearranging the equation, it can be seen that:

$$-(\beta x_i) = \log[-\log\{S_0(t)\}] - \log[-\log\{S(t|x_i)\}].$$

Thus, for a binary covariate  $X$ , the difference between the values of  $-\log[-\log \hat{S}(t|x_i)]$  for those with  $X_i = 1$  versus those with  $X_i = 0$  is  $-\hat{\beta}$ , which is constant with respect to time.

A second method described by Cleves et al. (2008) for assessing the assumption of proportional hazards graphically is to compare plots of survival curves  $S(t|x_i)$  versus time that use the Kaplan-Meier estimator  $\hat{S}_{KM}(t|x_i)$  (which is model agnostic) with those from the Cox regression  $\hat{S}_0(t)^{\exp(\hat{\beta}x_i)}$  (which assumes proportional hazards). If the assumption of proportional hazards is satisfied, these plots should be very similar.

In the absence of proportional hazards, the Cox model can be stratified, or the relationship of a covariate-specific-hazard and time can be explicitly included in the model through the inclusion of interaction terms between the covariate and (functions of) time.

### 3.3 Parametric Survival Models

The Cox proportional hazards model is a semi-parametric model, as it makes the assumption of proportional hazards, but does not impose a functional form for the hazard (Collett, 2003).

Non-parametric (distribution-free) and semi-parametric models compare subjects at the times that the event of interest occurs. By contrast, parametric models specify distributions that take into account all the information about a subject over the whole time that the subject is under observation (from study entry to exit via failure or being censored). Parametric models are thus more efficient than non-parametric models (Collett, 2003).

#### 3.3.1 Proportional Hazards Models

Parametric proportional hazards models have the same general form as Equation 3.12 given above for the Cox regression model. They differ, however, in that a functional form for the baseline hazard  $h_0(t)$  is specified. The following specifications are often used:

$$\begin{array}{ll} h_0(t) &= 1 \quad \text{exponential} \\ h_0(t) &= \exp(pt) \quad \text{Gompertz} \\ h_0(t) &= pt^{p-1} \quad \text{Weibull} \end{array}$$

where  $t$  is time and  $p$  is a shape parameter.

The functional form of the baseline hazard for the exponential model is a horizontal line crossing the y-axis at 1. The Gompertz and Weibull models are more flexible, allowing a variety of monotonic increasing or decreasing functions (see Figure 3.1). They also both include the exponential function as a special case (for  $p = 0$  in the Gompertz Model and  $p = 1$  in the Weibull Model).

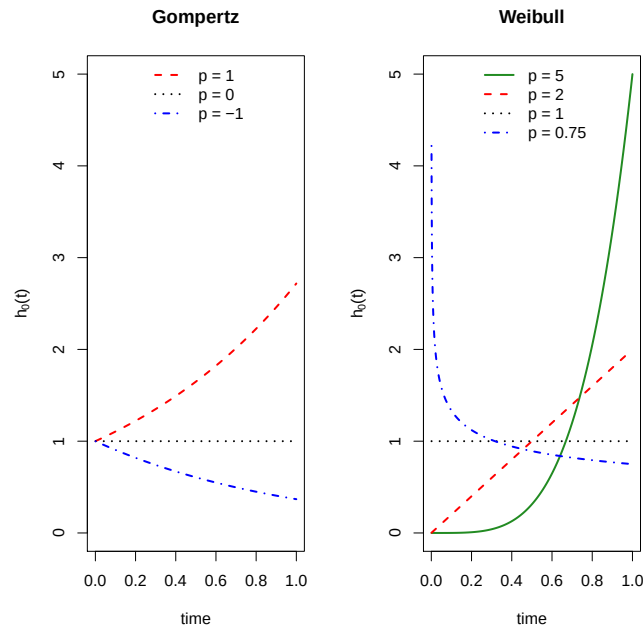


Figure 3.1: Some possible shapes of the baseline hazard functions for parametric survival models

Parametric models are useful if the modeller

- has an idea of what the shape of the baseline hazard should be
- wants to impose this functional form on the data in order to obtain the most efficient estimates for the coefficients  $\beta_j$  and
- wants to estimate the baseline hazard  $h_0(t)$  subject to these constraints.

If, however, the modeller does not have any idea a-priori of what the baseline hazard looks like and wants a baseline hazard that is very flexible, then Cox regression may be more appropriate (Cleves et al., 2008).

An advantage of some parametric proportional hazards models (such as the exponential and the Weibull) is that they can also be parameterized in the accelerated failure time (AFT) framework. AFT models focus on time rather than relative hazards. This makes it possible to estimate the mean time to event after fitting an AFT model using integration (calculating the area under the survival curve). The 95% confidence interval for the mean can also be calculated using the delta method.

### 3.3.2 Accelerated Failure Time Models

Collett (2003) gives the log-linear form of an AFT model as follows:

$$\log T_i = \mu + \boldsymbol{\alpha}'\mathbf{x}_i + \sigma\epsilon_i, \quad (3.13)$$

where:

$\boldsymbol{\alpha}'\mathbf{x}_i = \alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_k x_{ki}$  is the linear component of the model

$\alpha_1, \alpha_2, \dots, \alpha_k$  are unknown coefficients of the  $k$  predictor variables,  $X_1, X_2, \dots, X_k$

$x_{ji}$  is the value of the  $j^{th}$  predictor variable,  $X_j$ ,  $j = 1, 2, \dots, k$ , for the  $i^{th}$  individual,  $i = 1, 2, \dots, n$

$\mu$  is an intercept parameter corresponding to the value of  $\log T$  when all  $\mathbf{x}_i = \mathbf{0}$  and

$\epsilon_i$  is the difference between observed and modelled times on the log scale (scaled by  $\sigma$ ), and thus a random variable that can be assumed to have a particular underlying probability distribution.

The survivor function of  $T_i$ , the random variable associated with the survival time of the  $i^{th}$  individual, is then:

$$\begin{aligned} S_i(t) &= \Pr\{T_i \geq t\} \\ &= \Pr\{\exp(\mu + \boldsymbol{\alpha}'\mathbf{x}_i + \sigma\epsilon_i) \geq t\} \\ &= \Pr\{\exp(\mu + \sigma\epsilon_i) \geq t / \exp(\boldsymbol{\alpha}'\mathbf{x}_i)\} \\ &= S_0\{t / \exp(\boldsymbol{\alpha}'\mathbf{x}_i)\}. \end{aligned} \quad (3.14)$$

Using the relationship between the hazard and survivor function given in Equation 3.4 we can obtain the general hazard function for an AFT model from Equation 3.14 as follows:

$$\begin{aligned}
 S_i(t) &= S_0(t/e^{\alpha'x_i}) \\
 -\log[S_i(t)] &= -\log[S_0(t/e^{\alpha'x_i})] && \text{(taking logs and multiplying by -1)} \\
 -\frac{d}{dt}\{\log[S_i(t)]\} &= -\frac{d}{dt}\{\log[S_0(t/e^{\alpha'x_i})]\} && \text{(taking derivatives on both sides)} \\
 h_i(t) &= e^{-\alpha'x_i}h_0(t/e^{\alpha'x_i}) && \text{(from (3.4)).}
 \end{aligned} \tag{3.15}$$

### General forms of the PH and AFT Models

For proportional hazards (PH) models (based on the hazard for one group being proportional to the hazard of other group):

$$h_A(t) = \psi h_B(t)$$

and equivalently

$$S_A(t) = [S_B(t)]^\psi$$

with  $\psi = \exp(\beta'x)$

For accelerated failure time (AFT) models (based on the survival time of subjects in one group being a multiple of the survival time of subjects in the other group):

$$h_A(t) = \phi^{-1}h_B(t/\phi)$$

and equivalently

$$S_A(t) = S_B(t/\phi)$$

with  $\phi = \exp(\alpha'x)$  and:

$\phi < 1$  implies that the time to the event is accelerated (shorter) and

$\phi > 1$  implies that the time to event is decelerated (longer).

The next section looks at a commonly used model that can be parameterized in both the PH and AFT metrics, the Weibull model.

### Weibull Model

If survival times are assumed to have a Weibull distribution, the hazard function for the  $i^{th}$  individual in the PH framework can be written as:

$$h_i(t) = \lambda_i p t^{p-1},$$

where  $\lambda_i$  is parameterized as  $e^{\beta' \mathbf{x}_i}$ . The baseline hazard function would thus be  $h_0(t) = p t^{p-1}$  as given above. If the AFT framework is chosen instead, then Equation 3.15 can be used to write the hazard function for the  $i^{th}$  individual as:

$$\begin{aligned} h_i(t) &= e^{-\alpha' \mathbf{x}_i} p (t / e^{\alpha' \mathbf{x}_i})^{p-1} \\ &= e^{-p \alpha' \mathbf{x}_i} p t^{p-1}, \end{aligned}$$

such that  $\lambda_i$  is parameterized as  $e^{-p \alpha' \mathbf{x}_i}$ .

It can thus be seen that if the survival time follows a Weibull distribution and we estimate the coefficients  $\beta$  in the PH framework, we can obtain estimates for the coefficients  $\alpha$  of the AFT framework by dividing the coefficients  $\beta$  by the shape parameter of the Weibull distribution,  $p$  and multiplying by  $-1$ . This relationship only holds if the shape parameter  $p$  is held constant (i.e. it remains the same for subjects with any covariate pattern).

The exponential distribution can also be parameterized in both the PH and AFT frameworks, since it is a special form of the Weibull model obtained when  $p = 1$ .

The exponential, Gompertz (only in PH framework) and Weibull models are commonly used for modelling monotonic (increasing or decreasing) hazard functions, while the Log-logistic and Log-Normal models allow hazard functions that are non-monotonic to be modelled.

### Generalized Gamma Model

The Generalized Gamma distribution is useful for determining the appropriate functional form of a parametric model for survival data, since it includes as special cases the exponential, Log-normal and Weibull distributions (Collett, 2003).

The probability density of the Generalized Gamma distribution is

$$f(t) = \frac{\theta \lambda^{\rho\theta} t^{\rho\theta-1} \exp\{-(\lambda t)^\theta\}}{\Gamma(\rho)} \quad (3.16)$$

for  $0 \leq t < \infty$ , where  $\lambda > 0$  is a scale parameter and  $\rho > 0$  and  $\theta > 0$  are shape parameters and  $\Gamma(x)$  is the gamma function, defined by

$$\Gamma(x) = \int_0^\infty s^{x-1} e^{-s} ds.$$

The Generalized Gamma survivor function is given by

$$S(t) = 1 - I\{\rho, (\lambda t)^\theta\}, \quad (3.17)$$

where  $I\{\rho, (\lambda t)^\theta\}$  is the *incomplete gamma function*, given by

$$I\{\rho, (\lambda t)^\theta\} = \frac{1}{\Gamma(\rho)} \int_0^{(\lambda t)^\theta} u^{\rho-1} e^{-u} du. \quad (3.18)$$

The hazard function can then be obtained using Equation 3.3 such that we have

$$h(t) = \frac{[\theta \lambda^{\rho\theta} t^{\rho\theta-1} \exp\{-(\lambda t)^\theta\}]/\Gamma(\rho)}{1 - I\{\rho, (\lambda t)^\theta\}}. \quad (3.19)$$

For programming purposes, the above parameterization of the Generalized Gamma model is not that useful. Prentice (1974) explain that the Newton-Raphson method has not been found to work well and that solutions to the log-likelihood equations cannot always be found. He thus developed an alternative parameterization which is used in many software implementations of the generalized Gamma model. This parameterization is implemented in Stata and presented below using the notation of Cleves et al. (2008):

$$f(t) = \begin{cases} \frac{\gamma^\gamma}{\sigma t \sqrt{\gamma} \Gamma(\gamma)} \exp(z\sqrt{\gamma} - u) & \kappa \neq 0 \\ \frac{1}{\sigma t \sqrt{2\pi}} \exp(-z^2/2) & \kappa = 0 \end{cases}, \quad (3.20)$$

$$S(t) = \begin{cases} 1 - I\{\gamma, u\} & \kappa > 0 \\ 1 - \Phi(z) & \kappa = 0 \\ I\{\gamma, u\} & \kappa < 0 \end{cases} \quad (3.21)$$

and

$$h(t) = \begin{cases} \frac{\frac{\gamma^\gamma}{\sigma t \sqrt{\gamma} \Gamma(\gamma)} \exp(z\sqrt{\gamma}-u)}{1-I\{\gamma, u\}} & \kappa > 0 \\ \frac{\frac{1}{\sigma t \sqrt{2\pi}} \exp(-z^2/2)}{1-\Phi(z)} & \kappa = 0 \\ \frac{\frac{\gamma^\gamma}{\sigma t \sqrt{\gamma} \Gamma(\gamma)} \exp(z\sqrt{\gamma}-u)}{I\{\gamma, u\}} & \kappa < 0 \end{cases} , \quad (3.22)$$

where

$$\gamma = |\kappa|^{-2}$$

$$z = \frac{\text{sign}(\kappa)\{\log(t)-\mu\}}{\sigma}$$

$$u = \gamma e^{(|\kappa|z)}$$

$$I\{\gamma, u\} = \frac{1}{\Gamma(\gamma)} \int_0^u e^{-t} t^{\gamma-1} dt \quad \text{is the incomplete gamma distribution}$$

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \quad \text{is the standard normal cumulative distribution.}$$

The parameterization of the probability density function of the Generalized Gamma given by Collett (2003) (Equation 3.16) can be obtained from that used in Cleves et al. (2008) (Equation 3.20) by letting

$$\gamma = \rho$$

$$|\kappa| = \frac{1}{\sqrt{\rho}}$$

$$\mu = \frac{1}{\theta} \log(\rho) - \log(\lambda)$$

$$\sigma = \frac{1}{\theta \sqrt{\rho}}$$

It then follows that  $u = (\lambda t)^\theta$  as required for the equivalence of the parameterizations for the survivor function of the Generalized Gamma (Equations 3.17 and 3.21 for the Collett (2003) and Cleves et al. (2008) parameterizations respectively).

The Generalized Gamma model is implemented in Stata (StataCorp, 2009b) by parameterizing  $\mu_i$  as  $\alpha'x_i$  and treating  $\kappa$  and  $\sigma$  as *ancillary parameters* to be estimated from the data.

These ancillary parameters are the shape parameters and provide us with a tool for allowing the hazard function to take on different shapes for different values of the covariates. The Generalized Gamma hazard function is thus very flexible, allowing for a wide range of shapes, including as special cases the following distributions under the given parameter assumptions (StataCorp, 2009b; Collett, 2003):

| Distribution | Parameterization            |                               |
|--------------|-----------------------------|-------------------------------|
|              | Collett                     | Stata                         |
| exponential  | $\theta = 1$ and $\rho = 1$ | $\kappa = 1$ and $\sigma = 1$ |
| Gamma        | $\theta = 1$                | $\sigma =  \kappa $           |
| Log-normal   | $\rho \rightarrow \infty$   | $\kappa = 0$                  |
| Weibull      | $\rho = 1$                  | $\kappa = 1$                  |

Figure 3.2 shows some of the possible shapes for the Generalized Gamma hazard curve when each of the above criteria are satisfied. The value of  $\mu$  has been set to zero for each of these curves unless otherwise specified.

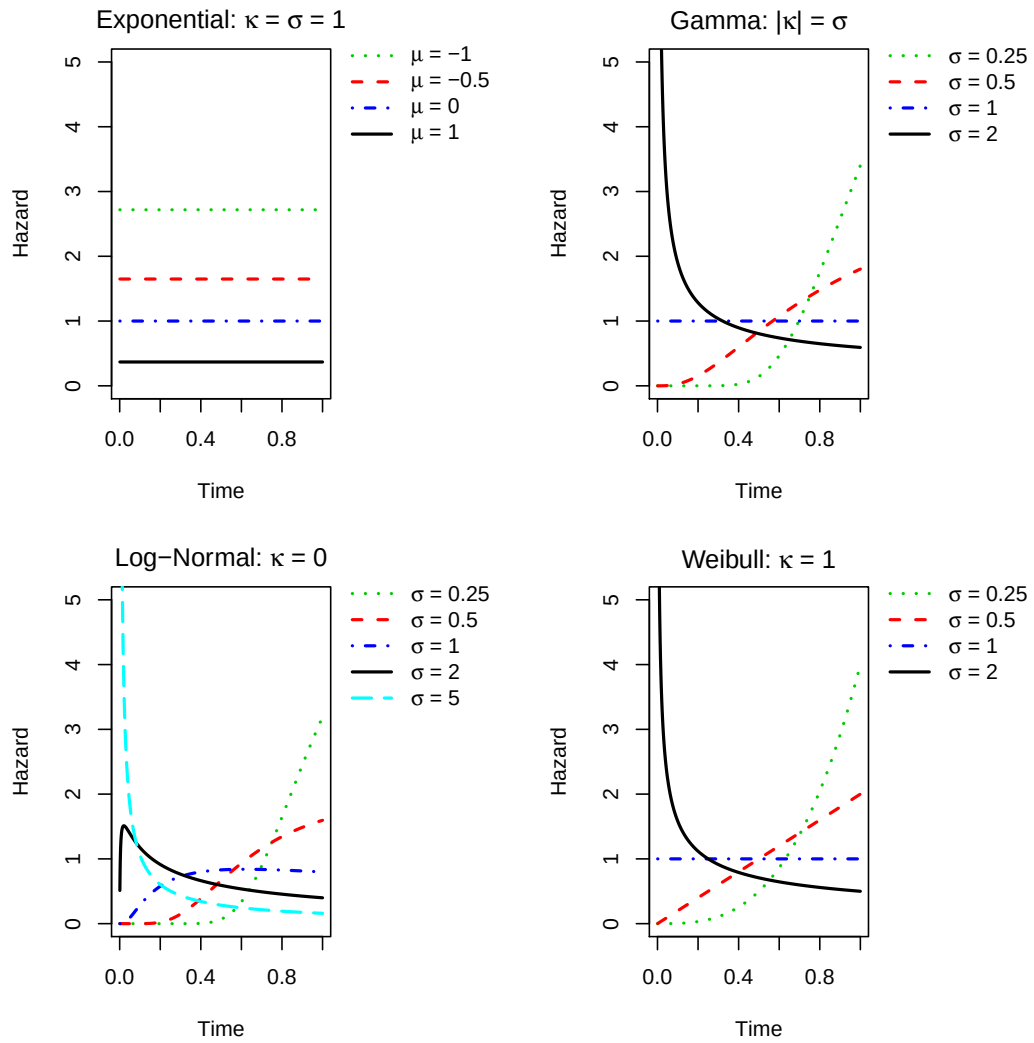


Figure 3.2: Some possible shapes of hazard curves for the Generalized Gamma AFT Model

## 3.4 Model checking

In addition to testing the assumption of proportional hazards for the Cox model, the validity of both Cox and parametric regression models can further be assessed by checking the overall model fit using Cox-Snell residuals, identifying outliers using deviance residuals and investigating the influence of individual observations on the overall model fit.

### 3.4.1 Assessing Adequacy of Model Fit

#### Cox-Snell residuals

Cox-Snell residuals are often used to assess model fit (Collett, 2003). The Cox-Snell residual for the  $i^{th}$  individual,  $i = 1, 2, \dots, n$ , is given by

$$\begin{aligned} r_{CSi} &= \exp(\hat{\beta}' \mathbf{x}_i) \hat{H}_0(t_i) \\ &= \hat{H}_i(t_i) \\ &= -\log(\hat{S}_i(t_i)) \end{aligned} \tag{3.23}$$

where  $\hat{H}_0(t_i)$  is the estimated baseline cumulative hazard function,  $\hat{H}_i(t_i)$  the estimated cumulative hazard function and  $\hat{S}_i(t_i)$  the estimated survivor function for the  $i^{th}$  individual observed at the survival time  $t_i$  of that individual.

The random variable  $Y_i = -\log(S(T_i))$ , where  $T_i$  is a random variable for the survival time of the  $i^{th}$  individual and  $S_i(t_i)$  is the corresponding survival function, has an exponential distribution with a mean of one (Collett, 2003). If a model fits the data well, then the properties of the estimates  $\hat{S}_i(t_i)$  should be similar to those of the values  $S_i(t_i)$ . It follows that the Cox-Snell residuals, which are a function of the estimates  $\hat{S}_i(t_i)$ , should also be exponentially distributed with a mean of one.

Since the exponential distribution is not symmetric about the mean and the survival times are correlated with the Cox-Snell residuals, plots of the residuals against the subject numbers or observed survival times or the rank order of the survival times are not helpful. A useful plot is that of the cumulative hazard of the Cox-Snell residuals against the residuals themselves. This is because the cumulative hazard at time  $t$  for an exponential distribution with a mean of one is simply  $t$ . Thus the values of the Cox-Snell residuals and their cumulative hazards will be equal if the model fits the data well,

resulting in a straight line through the origin with a gradient of one. If the plot of the Cox-Snell residuals versus their cumulative hazards shows systematic departure from a straight line through the origin, or does not have an intercept of zero or a gradient of one, this may suggest that alternative models need to be considered. Plots of the Cox-Snell residuals against the cumulative hazard may be more reliable for assessing the fit of parametric regression models, than for Cox regression models. (Collett, 2003)

### Deviance Residuals

Deviance residuals are useful for examining the fit of the model and determining whether there are outlying observations. This can be done by plotting the deviance residuals against the linear prediction of the model, or against the analysis time. Such a plot should show random scatter in the absence of censored patients. When there are censored patients, the deviance residuals for these patients will cluster around zero (Klein and Moeschberger, 1997). Any deviance residuals that are comparatively large in absolute value may indicate patients who are outliers and should be investigated further. Deviance residuals are calculated as

$$r_{Di} = \text{sign}(\delta_i - r_{CSi}) \sqrt{-2[\delta_i - r_{CSi} + \delta_i \log(r_{CSi})]}, \quad (3.24)$$

where  $\text{sign}(x) = -1$  if  $x < 0$  and  $\text{sign}(x) = 1$  if  $x \geq 0$ ,  $\delta_i$  is one for a patient who experiences the event of interest and zero otherwise and  $r_{CSi}$  is the Cox-Snell residual defined in 3.23 (Collett, 2003).

### 3.4.2 Investigation of influential observations

In all statistical modelling, there is an interest in identifying single observations that have a large effect on the estimated values of the model coefficients.

#### Delta-betas and likelihood displacement values

Ideally, one is interested in the values of the  $j^{th}$  coefficient,  $\hat{\beta}_j$ , when the  $i^{th}$  observation is both included and excluded from the estimation sample. Delta-betas are statistics that can be calculated from the estimated model based on all observations and that will estimate the change in the  $\hat{\beta}_j$  due to the deletion of the  $i^{th}$  observation. They are a product of the vector of score residuals,  $\mathbf{r}_{Si}$ , and the variance-covariance matrix of the estimated coefficient

vector  $\text{var}(\hat{\beta})$ . The score residuals are the contribution of the  $i^{\text{th}}$  observation to the partial derivative of the log-likelihood of the model with regard to the  $j^{\text{th}}$  coefficient. For parametric regression models, the delta-betas are calculated as  $\text{var}(\hat{\theta})\mathbf{r}_{Si}$ , where  $\theta$  represents both the shape and scale parameters.

The delta-betas can be aggregated over all  $j$  parameters to give a single measure of the influence of the  $i^{\text{th}}$  patient on all the parameters in the model using  $LD_i = \mathbf{r}'_{Si}\text{var}(\hat{\beta})\mathbf{r}_{Si}$ . This is an estimate of the likelihood displacement value, which is calculated as minus twice the difference between the maximum log-likelihoods of the model including all patients and the model excluding one patient. Collett (2003) suggest examining plots of the likelihood displacement values against the rank order of survival times or against the patient numbers. Any relatively large likelihood displacement values correspond to potentially influential patients.

### Comparison of Models

Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are often used to compare the fit of two models estimated using maximum likelihood. The model having the lowest AIC or BIC value is considered the better model. These values are calculated as follows:

$$\text{AIC} = -2\log(\text{likelihood}) + 2k \quad (3.25)$$

$$\text{BIC} = -2\log(\text{likelihood}) + k\log(N) \quad (3.26)$$

where  $k$  is the degrees of freedom of the model and  $N$  is the number of independent terms in the likelihood (StataCorp, 2009b).

## 3.5 Competing Risks Models

As outlined briefly in Section 3.1, in survival analysis, interest centres on the time to an event. Often, there is one main event of interest. Subjects who may still experience this event are part of the *risk set*. Any events which would preclude a subject from experiencing the event of interest or substantially change the probability of the event of interest occurring are known as *competing risks* (Gooley et al., 1999). Subjects may leave the risk set as a result of *failure* (experiencing the event of interest or a competing risk) or as a result of being censored (due to loss to follow-up or end of study).

All the models described in sections 3.2 and 3.3 assume that patients not experiencing the event of interest are censored and that this censoring is random. In a competing risks setting, the influence of competing risks on the hazard or cumulative incidence of the event of interest is considered explicitly.

For the data on which this thesis is based, in the analysis of time to ART initiation, subjects are considered to be part of the risk set if they are on TB treatment, still under observation and have not yet initiated ART. The main event of interest is thus ART initiation while on TB treatment. The competing events are permanent deferral for ART, death pre-ART, and the end of TB treatment. Permanent deferral for ART is treated as a competing risk, since it is unlikely that patients who are permanently deferred will initiate ART while on TB treatment and thus the probability of the event of interest is substantially changed for subjects that are permanently deferred.

### **3.5.1 Bias in the Kaplan-Meier estimates when there are competing risks**

In traditional survival analysis, Kaplan-Meier survival curves are often used to assess survival times visually. Kaplan-Meier survival estimates assume independence of the distributions for the event of interest and censoring. If the competing risks are not independent of the event of interest, then treating them as censored observations will result in bias in the Kaplan-Meier survival estimates (Putter et al., 2007).

For example, in Figure 3.3, the Kaplan-Meier survival curve for ART initiation (with competing risks treated as censoring events) and the Kaplan-Meier failure curve for the occurrence of any of the competing risks (with ART initiation treated as censoring) cross. This should not be, as patients cannot experience ART initiation and a competing event (they are disjoint sets of events) and thus the sum of probabilities of experiencing ART initiation or a competing event should be less than or equal to one (Putter et al., 2007).

This bias arises because the Kaplan-Meier method of estimation assumes that subjects that are treated as censored may still experience the event of interest. If, however, subjects actually left the risk set as a result of a competing event (and as such will never experience the event of interest), then this method of estimation overestimates the probability of failure (and underestimates the corresponding probability of survival) (Putter et al., 2007).

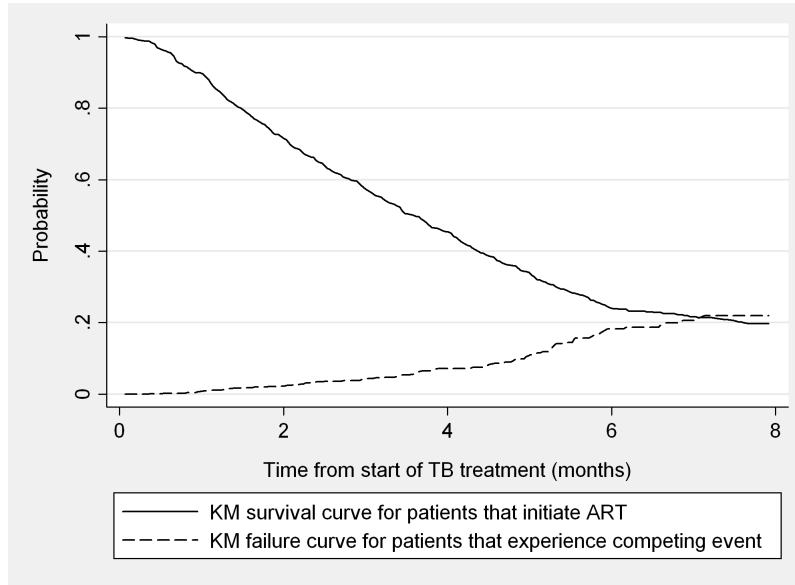


Figure 3.3: Bias in Kaplan-Meier (KM) survival estimates

### 3.5.2 Methods for assessing competing risks

Gichangi and Vach (2005) explain that there are two viewpoints from which we can consider competing risks. The first is from the perspective of what has happened up to a specific time point and the second focuses on what is likely to happen around a specific time point given that the subject has survived up to that time.

Starting with the second viewpoint, we examine the *cause-specific hazard*, which is the hazard of a subject failing as a result of an event of interest given the possibility of leaving the risk set as a result of experiencing a competing event (Putter et al., 2007).

The joint distribution of the survival time  $T$  and the cause of failure  $D$ , possibly given covariates  $\mathbf{X}$  must therefore be modelled. Equations for the cause-specific hazard, the cumulative cause-specific hazard and the cumulative incidence function for cause  $k$  are given in Section 3.1 (see equations 3.7 to 3.10).

The cumulative incidence function for cause  $k$ ,  $I_k(t) = \int_0^t \lambda_k(s)S(s)ds$  (3.10), is also known as the *sub-distribution function*, as it is always smaller than one and thus not a proper probability distribution (Putter et al., 2007). The

Kaplan-Meier estimator for failure, which treats causes other than  $k$  as censored, is calculated as

$$1 - S_k(t) = \int_0^t \lambda_k(s) S_k(s) ds \quad (3.27)$$

where  $S_k(t) = \exp(-\Lambda_k(t))$  describes the probability of survival from cause  $k$  in the absence of competing events if the event time distributions and the censoring distribution are independent (Putter et al., 2007).

Putter et al. (2007) comment that the bias in the Kaplan-Meier estimator can be shown by noting that  $S_k(t) \geq S(t)$ , which implies that  $1 - S_k(t) \geq I_k(t)$ , with equality at  $t$  only if no competing events have occurred, i.e.  $\sum_{j=1, j \neq k}^K \Lambda_j(t) = 0$ .

### Estimation of functions

I use the notation of Putter et al. (2007) to show how the cumulative incidence functions can be estimated.

For the ordered distinct times at which failure from any cause  $k$  occurs,  $0 < t_1 < t_2 < \dots < t_N$ , the Kaplan-Meier estimator of the survival time is given in Equation 3.6. A discretized version of the cause-specific hazard function from Equation 3.7,  $\lambda_k(t_j) = \Pr(T = t_j, D = k | T > t_{j-1})$  would be estimated by

$$\hat{\lambda}_k(t_j) = \frac{d_{kj}}{n_j}, \quad (3.28)$$

the proportion of subjects at risk at time  $t_j$  that fail from cause  $k$ .

The unconditional probability of failing from cause  $k$  at time  $t_j$ ,  $\Pr(T = t_j, D = k)$  is the product of the cause-specific hazard of failure from cause  $k$  and the probability of not yet having failed from any cause up to time  $t_j$  and is estimated as

$$\hat{p}_k(t_j) = \hat{\lambda}_k(t_j) \hat{S}(t_{j-1}). \quad (3.29)$$

The cumulative incidence of failure from cause  $k$  at time  $t$  is estimated as the sum of all estimated probabilities  $\hat{p}_k(t_j)$  of failing from cause  $k$  at time  $t_j$  up to time  $t$ :

$$\hat{I}_k(t) = \sum_{j: t_j \leq t} \hat{p}_k(t_j) \quad (3.30)$$

A competing risks analogue to Cox proportional hazards regression may be performed on the cause-specific hazards in order to estimate the effect of covariates on the cause-specific failure of interest using

$$\lambda_k(t|\mathbf{X}) = \lambda_{k,0}(t) \exp(\boldsymbol{\beta}'_k \mathbf{X}), \quad (3.31)$$

where  $\lambda_{k,0}(t)$  is the baseline cause-specific hazard for cause  $k$  and  $\boldsymbol{\beta}_k$  is the vector of coefficients for cause  $k$  corresponding to the vector of covariates  $\mathbf{X}$ .

The covariate effects  $\boldsymbol{\beta}_k$  are proportional for the cause-specific hazards. This simple effect of the covariates on the cause-specific hazards does not lead to a predictable effect of the covariates on the cumulative incidence functions, however, since the cumulative incidence function for cause  $k$  depends not only on the cause-specific hazard for cause  $k$ , but also on the cause-specific hazards for all the other causes through the overall survival function  $S(t)$  (Putter et al., 2007).

The cumulative incidence function for cause  $k$  depends on the effect of the covariate on all the causes (not only cause  $k$ ), as well as on the baseline cause-specific hazards for each cause (Putter et al., 2007). The cumulative incidence function for cause  $k$  given covariate vector  $\mathbf{X}$  expressed in terms of cause-specific hazard functions is as follows:

$$I_k(t|\mathbf{X}) = \int_0^t \lambda_k(s|\mathbf{X}) \exp\left(-\sum_{k=1}^K \int_0^s \lambda_k(u|\mathbf{X}) du\right) ds \quad (3.32)$$

Thus the above methodology is useful for looking at competing risks from the viewpoint of what happens around time point  $t$ , but it is less useful for analysing and interpreting competing risks from the viewpoint of what has happened up until time point  $t$ . This is because the modelling of  $I_k(t|\mathbf{X})$  happens through the modelling of the cause-specific hazard functions.

### Fine and Gray Methodology

Fine and Gray (1999) assert that in some situations it will be more important to have direct information on the probability of having ever experienced the event of interest up to a certain time than to have information on the cause-specific hazard of experiencing an event at a particular point in time. They thus propose a method by which the effect of covariates can be estimated directly with reference to the cumulative incidence function.

Fine and Gray (1999) assume some known and increasing function  $g(\cdot)$  such that

$$g\{\bar{I}_k(t|\mathbf{X})\} = h_0(t) + \boldsymbol{\beta}'\mathbf{X} \quad (3.33)$$

where “ $h_0(t)$  is a completely unspecified, invertible and monotone increasing function”;  $\mathbf{X}$  is a vector of  $p$  covariates and  $\boldsymbol{\beta}$  is the corresponding vector of regression coefficients.

This means that after transformation, the conditional cumulative incidence functions for two individuals with covariate vectors  $\mathbf{X}_1$  and  $\mathbf{X}_2$  will satisfy a vertical shift model, that is

$$g\{\bar{I}_k(t|\mathbf{X}_1)\} - g\{\bar{I}_k(t|\mathbf{X}_2)\} = \boldsymbol{\beta}'(\mathbf{X}_1 - \mathbf{X}_2) \quad (3.34)$$

for all  $t$ . The regression coefficients measure the effect of the difference in  $\mathbf{X}_1$  and  $\mathbf{X}_2$

Fine and Gray (1999) model the sub-distribution hazard defined in Equation 3.11 using a proportional hazards framework similar to the Cox proportional hazards model,

$$\bar{\lambda}_k(t|\mathbf{X}) = \bar{\lambda}_{k,0}(t) \exp\{\boldsymbol{\beta}'\mathbf{X}\}, \quad (3.35)$$

where  $\bar{\lambda}_{k,0}(t)$  is a completely unspecified, nonnegative function in  $t$ .

Using Equations 3.11 and 3.35,  $\bar{I}_k(t|\mathbf{X})$  can be written as

$$\bar{I}_k(t|\mathbf{X}) = 1 - \exp\left[-\int_0^t \bar{\lambda}_{k,0}(s) \exp\{\boldsymbol{\beta}'\mathbf{X}\} ds\right] \quad (3.36)$$

The transformation  $g(u) = \log\{-\log(1 - u)\}$  then gives

$$g\{\bar{I}_k(t|\mathbf{X})\} = \log\left[\int_0^t \bar{\lambda}_{k,0}(s) ds\right] + \boldsymbol{\beta}'\mathbf{X} \quad (3.37)$$

and comparing Equations 3.33 and 3.37 it can be seen that

$$h_0(t) = \log\left[\int_0^t \bar{\lambda}_{k,0}(s) ds\right]. \quad (3.38)$$

The regression coefficients and baseline hazard from the Cox transformation model for  $\bar{I}_k$  can thus be interpreted without reference to the sub-distribution hazard (Fine and Gray, 1999), because when taking the difference of the  $g$ -transformations of the cumulative incidence functions described by Fine and

Gray (1999) for subjects with differing covariates (for example  $\mathbf{X}_1$  and  $\mathbf{X}_2$ ), only the covariate vectors and regression coefficients will remain, as shown in Equation 3.34.

The exponentiated regression coefficients,  $e^{\beta}$ , can be obtained from the ratio of two sub-distribution hazards and are thus also termed *sub-distribution hazard ratios*, or *sub-hazard ratios*. These can be interpreted in a similar way as the hazard ratios in a Cox regression model.

The sub-hazard ratio can be thought of as the relative risk of the event of interest being experienced before any other event, whereas the cause-specific hazard ratio is simply the relative risk of the event of interest. This is because the risk set for calculating  $\bar{\lambda}_k$  includes all subjects that have not experienced the event of interest, while the risk set for calculating  $\lambda_k$  includes only subjects that have not experienced any event.

### Graphical Analysis

It is helpful to be able to examine the predicted hazard and cumulative incidence curves graphically. In order to plot hazard curves, it is necessary to smooth over the cause-specific hazards calculated at each failure time,  $\hat{\lambda}_k(t_j)$ . A simple method of smoothing over the hazards  $\hat{\lambda}_k(t)$  to obtain an estimate for the cause-specific hazard at time  $t$  is to use a kernel estimator.

A kernel estimator provides a weighted average of the estimates  $\hat{\lambda}_k(t)$  that are close to time  $t$  using a weighting function that is positive and integrates to one (Gichangi and Vach, 2005). I used the Epanechnikov kernel (StataCorp, 2009b), for which

$$K[t] = \begin{cases} \frac{3}{4}(1 - \frac{1}{5}t^2)/\sqrt{5} & \text{if } |t| < \sqrt{5} \\ 0 & \text{otherwise} \end{cases} \quad (3.39)$$

Kernel density estimates were obtained at each time  $t_j, j = 1, 2, \dots, J$  at which failure occurred (either as a result of observing the event of interest or a competing risk). They were calculated as follows

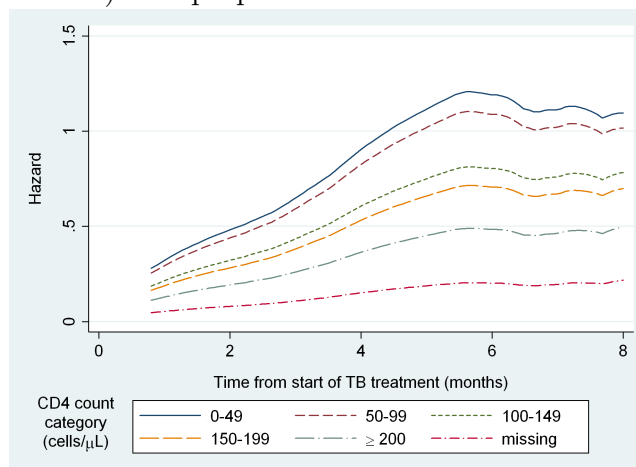
$$\tilde{\lambda}_k(t) = \frac{1}{h} \sum_{j=1}^J \hat{\lambda}_k(t_j) K\left(\frac{t - t_j}{h}\right) \quad (3.40)$$

The bandwidth  $h$  was determined using

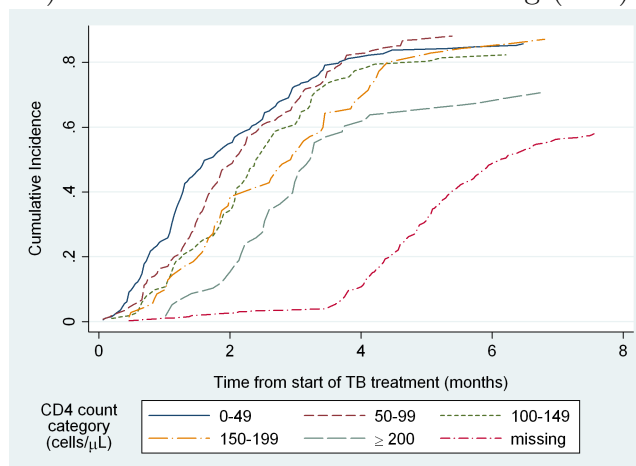
$$\begin{aligned} m &= \min \left( \sqrt{\text{Var}(t)}, \frac{\text{IQR}(t)}{1.349} \right) \\ h &= \frac{0.9m}{J^{1/5}} \end{aligned} \tag{3.41}$$

Figure 3.4 illustrates the smoothed hazard curves calculated using (3.39)-(3.41) with  $\hat{\lambda}_k(t|\mathbf{X})$  (a) and the cumulative incidence curves calculated according to Equations 3.32 (b) and 3.36 (c). This illustrates how an apparent simple relationship between and ordering of the categories of a covariate in the hazard function (a) is not necessarily seen in the corresponding cumulative incidence curves calculated according to (3.32) (b). The use of the Fine and Gray (1999) methodology, however, which focuses on being able to interpret relationships between variables for the cumulative incidence function, results in cumulative incidence curves for which the relationship between and ordering of the categories of a covariate do appear to be more simple, as seen in Figure 3.4.c.

a) Cox proportional hazards curves



b) Cumulative Incidence curves using (3.32)



c) Cumulative Incidence curves using (3.36)

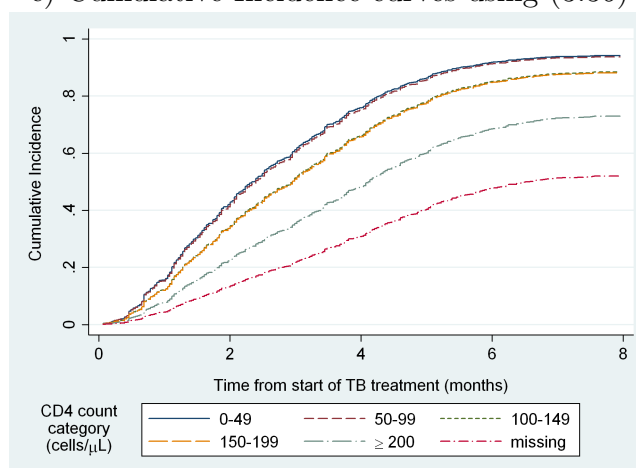


Figure 3.4: Cox proportional hazards graph and cumulative incidence graphs of time from start of TB treatment to ART initiation by CD4 count category near time of starting TB treatment

In addition to analyses of plots that stratify the hazard and cumulative incidence functions for the event of interest by covariate value, it is also informative to analyse plots that include cause-specific hazard functions and cause-specific cumulative incidence curves for each of the failure events. This gives an appreciation of the magnitude of the cause-specific hazards and cumulative incidences of failure events relative to each other. Three different types of graphs can contribute to this analysis (Gichangi and Vach, 2005):

i) A plot of the individual functions for each failure event on one graph:

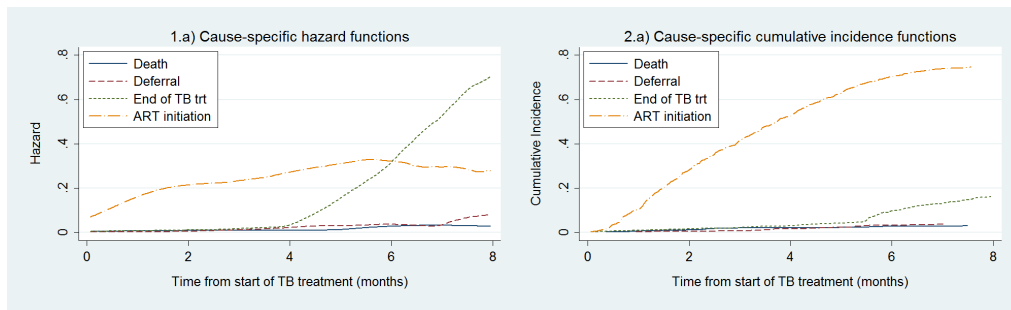


Figure 3.5: Cause-specific hazard and cumulative incidence functions

ii) A plot of the stacked functions, such that the bottom line (or area) corresponds to the cause-specific hazard (or cumulative incidence) function of one type of failure event having occurred (perhaps the event with the lowest hazard or cumulative incidence) and the top line corresponds to the hazard (or cumulative incidence) that any of the failure events have occurred:

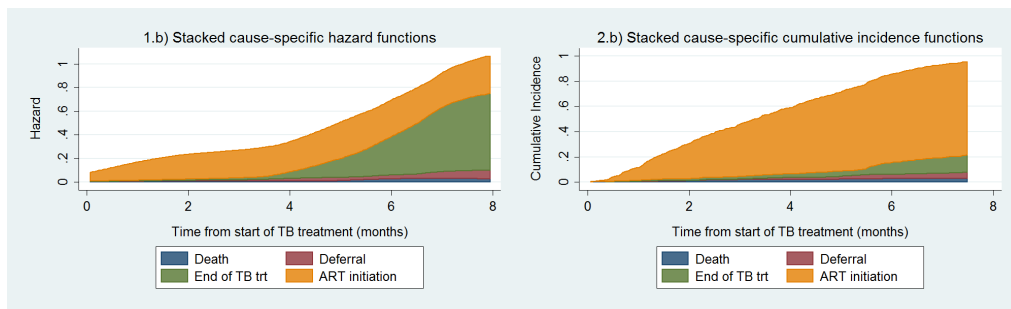


Figure 3.6: Relative cause-specific hazard and cumulative incidence functions

iii) A plot of the stacked relative functions, such that the total area of the graph is one and corresponds to any of the failure events having occurred

and the areas for each cause-specific hazard or cumulative incidence function are proportional to this:

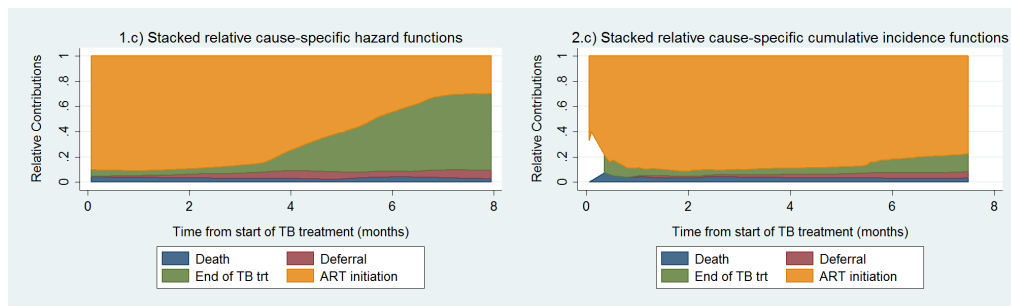


Figure 3.7: Stacked cause-specific hazard and cumulative incidence functions

## Chapter 4

# Analysis of Time to ART Initiation

The analysis presented in this chapter focuses on evaluating the effect of different baseline patient or programme characteristics on the time from the start of TB treatment to ART initiation. These effects will also be compared across different models. Data was collected from three different clinic sites for this analysis, but for clinic sites 2 and 3, no data was available on patients who did not start ART. For this reason, different subsets of data were used for different analyses depending on the aims of each analysis. In the Cox proportional hazards models and parametric survival analysis models of this chapter, a subset of the data including patients from all 3 clinics that initiated ART while on TB treatment was used (1444 patients). This allowed the data from all three clinics to be used and enabled an assessment of whether the time to ART initiation differs by clinic. In the final section (competing risks models), a subset of data including only patients from clinic site 1, some who did initiate ART and others who did not, was used (893 patients). For this section of the analysis, it would not have made sense to use data for which there was no information on patients who experienced competing risks.

The main response variable of interest was the time from the start of TB treatment to ART initiation. Results from preliminary analyses that I conducted to explore the factors affecting delays to ART initiation in the three clinic sites used for this thesis were published in Lawn et al. (2011a). I found that in crude analyses, patients who were younger, those receiving HIV care at clinic sites 1 and 3 (compared to site 2), those who started TB treatment in earlier years, those who had lower CD4 counts, those who were being treated

for TB for the first time and those who had smear positive pulmonary TB (compared to smear-negative pulmonary TB) had shorter delays to ART initiation. In the analysis with multiple covariates, only the clinic site, CD4 count and calendar period of starting TB treatment remained significantly associated with the time to ART initiation. A second paper (Lawn et al., 2011b), including only patients from clinic site 1, focused on delays in starting ART for patients who were diagnosed with TB in a TB clinic before enrolling for HIV care compared to delays for patients who were diagnosed with TB in an ART clinic as part of the screening process before ART initiation. The time from the start of TB treatment to ART initiation was much shorter for the latter group. The variables I have chosen to include in models for this section are therefore the three factors found to be significant predictors of the time to ART initiation in the analysis including multiple covariates of the first paper (Lawn et al., 2011a), as well as whether the patient was diagnosed with TB before or after enrolling for HIV care (which was seen to significantly affect the delay to ART initiation in the second paper (Lawn et al., 2011b)):

- The clinic site (3 sites). Since the clinic site refers to the location of the clinic, a patient was enrolled for HIV care at ART clinic site 1, would have received TB care at TB clinic site 1.
- The type of clinic in which the patients was diagnosed with TB (at a TB clinic, i.e. before enrolment for HIV care) or at an ART clinic (i.e. after enrolment for HIV care)
- The calendar period in which the patient was diagnosed with TB (before 2006, versus 2006 or later)
- The CD4 count taken closest to the time of starting TB treatment that satisfied the criteria of being taken not more than 6 months prior to or 3 months after the start of TB treatment and not more than 7 days after ART initiation. The measured CD4 counts were grouped into categories of 0-49, 50-99, 100-149, 150-199 and  $\geq 200$  cells/ $\mu$ L. An additional category labelled ‘missing’ for those who did not have a CD4 count taken near the start of TB treatment was also included. Sensitivity analysis for excluding these patients or assuming that they have very high or very low CD4 counts will be shown for the generalized Gamma accelerated failure time model (see Section 4.3.5).

Thought was also given as to whether the WHO clinical stage of HIV and the type of TB should be considered as factors to be included in the models (see discussion in Appendix C.1.1). The results from including these variables in

selected models are shown in Tables C.1 and C.2 in the Appendix. Stage and type of TB will not, however, be included in any of the models for the differing methods that are considered and compared in this section of analysis.

## 4.1 Summary statistics and Kaplan-Meier plots

Prior to formal modelling of the time to ART initiation, associations between the predictors and a binary simplification of the duration variable were explored by simple summary statistics and univariate and multiple logistic regression models. I chose to group patients into those who started ART within 12 weeks of TB treatment and those who did not. I chose 12 weeks as it was a convenient cut-off close to the median of the time from the start of TB treatment to ART initiation. Using a binary outcome makes comparison of proportions of patients easy and also allows for a formal comparison of these proportions using logistic regression.

Table 4.1 describes the patients who initiated ART within 12 weeks of starting TB treatment and those who started after 12 weeks with regard to the distribution of their baseline characteristics. Estimates of odds ratios (OR) and 95% confidence intervals (CI) from univariate and multiple logistic regressions (LR) are also shown.

The logistic regression analysis indicates that patients for whom the following are true have lower relative odds of initiating ART within 12 weeks of starting TB treatment than patients for whom they are not true:

- ART clinic at which patient enrolled was site 1 or 2 (rather than 3)
- TB was diagnosed before 2006 (rather than after)
- TB was diagnosed in TB clinic (i.e. before enrolment for HIV care) as opposed to in the ART clinic (i.e. after enrolment for HIV care)
- Patients with relatively high CD4 counts near the time of starting TB treatment, or no CD4 count taken near the time of starting TB treatment. In other words, increasing CD4 is associated with a greater delay in starting ART and it appears as if patients with no CD4 may either have less compromised immune systems or not be under observation (may enrol for HIV care a while after starting TB treatment).

I used Kaplan-Meier plots to visually illustrate the survival distributions in the different strata defined by the individual categorical predictor variables. These distributions show the proportion of patients who have not yet experienced the event of interest at successive time points and thus compares the strata with regard to the delay in ART initiation.

Table 4.1: Baseline characteristics of patients by time from start of TB treatment to ART initiation

| Characteristic<br>Clinic Location                | Time to ART initiation |        |            |        | All  |        | LR for ART initiation within 12 weeks |           |          |           |
|--------------------------------------------------|------------------------|--------|------------|--------|------|--------|---------------------------------------|-----------|----------|-----------|
|                                                  | 0-12 weeks             |        | > 12 weeks |        |      |        | Crude                                 |           | Adjusted |           |
|                                                  | n                      | (col%) | n          | (col%) | n    | (col%) | OR                                    | 95% CI    | OR       | 95% CI    |
| 1                                                | 340                    | (46%)  | 331        | (48%)  | 671  | (47%)  | 0.47                                  | 0.33-0.66 | 0.41     | 0.27-0.65 |
| 2                                                | 285                    | (38%)  | 309        | (44%)  | 594  | (41%)  | 0.42                                  | 0.29-0.60 | 0.16     | 0.10-0.25 |
| 3 (baseline)                                     | 123                    | (16%)  | 56         | (8%)   | 179  | (12%)  | 1.00                                  | -         | 1.00     | -         |
| Year in which TB diagnosed                       |                        |        |            |        |      |        |                                       |           |          |           |
| 2002-2005                                        | 273                    | (36%)  | 417        | (60%)  | 690  | (48%)  | 0.38                                  | 0.31-0.48 | 0.40     | 0.31-0.51 |
| 2006-2008 (baseline)                             | 475                    | (64%)  | 279        | (40%)  | 754  | (52%)  | 1.00                                  | -         | 1.00     | -         |
| Clinic in which TB diagnosed                     |                        |        |            |        |      |        |                                       |           |          |           |
| TB Clinic                                        | 409                    | (55%)  | 558        | (80%)  | 967  | (67%)  | 0.30                                  | 0.24-0.38 | 0.23     | 0.17-0.31 |
| ART Clinic (baseline)                            | 339                    | (45%)  | 138        | (20%)  | 477  | (37%)  | 1.00                                  | -         | 1.00     | -         |
| CD4 count category (cells/ $\mu$ L) <sup>a</sup> |                        |        |            |        |      |        |                                       |           |          |           |
| 0 (baseline)                                     | 338                    | (45%)  | 111        | (16%)  | 449  | (31%)  | 1.00                                  | -         | 1.00     | -         |
| 50                                               | 169                    | (23%)  | 113        | (16%)  | 282  | (20%)  | 0.49                                  | 0.36-0.68 | 0.41     | 0.29-0.58 |
| 100                                              | 115                    | (15%)  | 95         | (14%)  | 210  | (14%)  | 0.40                                  | 0.28-0.56 | 0.30     | 0.21-0.45 |
| 150                                              | 66                     | (9%)   | 76         | (11%)  | 142  | (10%)  | 0.29                                  | 0.19-0.42 | 0.22     | 0.14-0.33 |
| 200                                              | 29                     | (4%)   | 50         | (7%)   | 79   | (5%)   | 0.19                                  | 0.11-0.32 | 0.13     | 0.07-0.22 |
| missing                                          | 31                     | (4%)   | 251        | (36%)  | 282  | (20%)  | 0.04                                  | 0.03-0.06 | 0.03     | 0.02-0.05 |
| All patients <sup>b</sup>                        | 748                    | (52%)  | 696        | (48%)  | 1444 | (100%) |                                       |           |          |           |

LR: Logistic Regression, OR: Odds Ratio, CI: Confidence Interval, col%: column % (within covariate)

<sup>a</sup> CD4 count taken closest start of TB treatment that satisfied criteria of being taken not earlier than 183 days before or 91 days after start of TB treatment and not more than 7 days after start of ART

<sup>b</sup> Row % given for this row

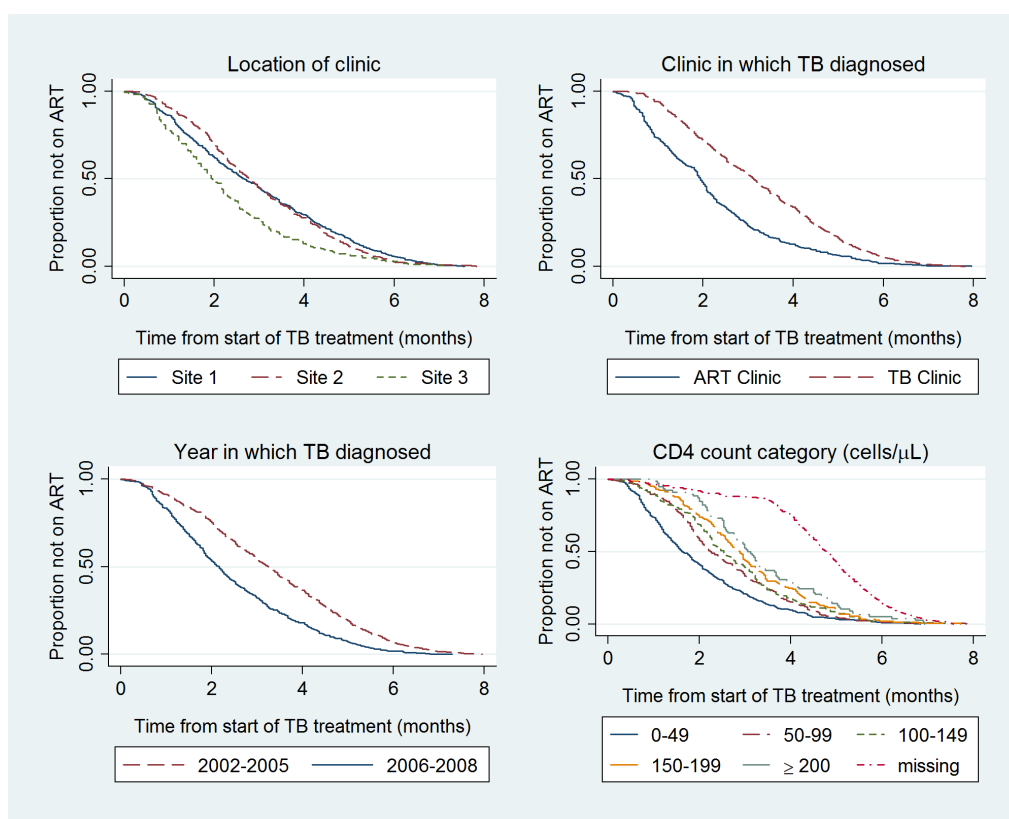


Figure 4.1: Kaplan-Meier survival plots for time from start of TB treatment to ART initiation,  $N=1444$

The Kaplan-Meier survival plots presented in Figure 4.1 show the faster ART initiation for patients enrolled at clinic site 3, those diagnosed with TB in an ART clinic, those diagnosed with TB in 2006 or later and those with lower CD4 cell counts near the time of starting TB treatment.

The survival curve for patients who did not have a CD4 count near the time of starting TB treatment appears quite different to that of patients who do have a CD4 count recorded (see bottom right plot of Figure 4.1). These patients appear to have a much longer time to ART initiation from the start of TB treatment than patients for whom a CD4 count was measured.

The sensitivity of the results to the inclusion of the missing CD4 category was examined by repeating this analysis for patients who did have a CD4 count near the time of starting TB treatment. The results were very similar to those that include patients who do not have a CD4 count (see Figure C.1 in the Appendix C.1.2).

Table 4.2 further compares patients with and without CD4 counts near the start of TB treatment with regard to other important predictors for ART initiation (from Table 4.1). It shows that the proportion of patients not having CD4 counts near the time of starting TB treatment are lower for patients who enrolled clinic site 2, those diagnosed with TB in later years and those who were diagnosed with TB in the ART clinic.

The median time to ART initiation for patients not having a CD4 count near the time of starting TB treatment is much longer than that of patients who did have a CD4 count for each category of the predictors shown.

The relative durations (within categories of a predictor) are similar among those with and without CD4 cell counts. Table 4.2 thus shows that the results given previously (that patients attending clinics at Site 3, those who were diagnosed with TB in the calendar period 2006-2008 and those who were diagnosed with TB in the ART clinic had a much shorter time from the start of TB treatment to ART initiation) are true both for those having a CD4 count and those not having a CD4 count. I have thus chosen to include patients not having a CD4 count near the start of TB treatment (by including them in a separate CD4 count category denoted ‘missing’) for the remainder of the Time to ART analysis.

Formal modelling of the time to ART initiation was done using Cox proportional hazards models, parametric survival analysis models and competing risks models. These models are presented and compared in the following sections.

Table 4.2: Days from start of TB treatment to ART initiation by whether patient had a CD4 count near the time of starting TB treatment or not

|                              | Has CD4 |       |              | No CD4 |       |               | All  |        |              |
|------------------------------|---------|-------|--------------|--------|-------|---------------|------|--------|--------------|
|                              | N       | (%)   | Median (IQR) | N      | (%)   | Median (IQR)  | N    | (%)    | Median (IQR) |
| Clinic Location              |         |       |              |        |       |               |      |        |              |
| Site 1                       | 471     | (70%) | 60 (35-91)   | 200    | (30%) | 151 (128-177) | 671  | (100%) | 83 (43-132)  |
| Site 2                       | 551     | (93%) | 84 (56-124)  | 43     | (7%)  | 151 (74-182)  | 594  | (100%) | 86 (57-130)  |
| Site 3                       | 140     | (78%) | 53 (32-76)   | 39     | (22%) | 133 (105-171) | 179  | (100%) | 60 (34-94)   |
| Year in which TB diagnosed   |         |       |              |        |       |               |      |        |              |
| 2002-2005                    | 504     | (73%) | 82 (55-114)  | 186    | (27%) | 154 (186-154) | 690  | (100%) | 101 (62-145) |
| 2006-2008                    | 658     | (87%) | 60 (36-93)   | 96     | (13%) | 140 (112-162) | 754  | (100%) | 66 (39-105)  |
| Clinic in which TB diagnosed |         |       |              |        |       |               |      |        |              |
| TB Clinic                    | 718     | (74%) | 77 (50-108)  | 249    | (26%) | 152 (129-178) | 967  | (100%) | 97 (57-137)  |
| ART Clinic                   | 444     | (93%) | 59 (28-89)   | 33     | (7%)  | 65 (28-130)   | 477  | (100%) | 59 (28-90)   |
| All                          | 1162    | (80%) | 69 (42-102)  | 282    | (20%) | 149 (124-177) | 1444 | (100%) | 80 (47-127)  |

## 4.2 Cox Proportional Hazards Model

I fit the Cox proportional hazards models to estimate the joint impact of the four baseline predictor variables on the relative hazard of ART initiation. The analysis was for a subset of patients who all initiated ART and hence included no censored observations. The advantage of the Cox model over the logistic regression model is that it did not require a categorization of the time to ART initiation and through the use of the equation  $S(t) = \exp(-H(t))$  could indirectly measure the impact of the baseline characteristics on the delay to ART initiation, as measured by the survival function.

### 4.2.1 Cox Proportional Hazards Regressions

Models of the following form were fit:

$$\begin{aligned}
 \log \left[ \frac{h_i(t)}{h_0(t)} \right] = & \underbrace{\beta_1 X_{1i} + \beta_2 X_{2i}}_{\text{clinic site 1 \& 2 vs 3}} \\
 & + \underbrace{\beta_3 X_{3i}}_{\substack{\text{TB diagnosed} \\ \text{2002--2005} \\ \text{vs 2006--2008}}} + \underbrace{\beta_4 X_{4i}}_{\substack{\text{TB diagnosed} \\ \text{before enrolment} \\ \text{for HIV care}}} \\
 & + \underbrace{\beta_5 X_{5i} + \dots + \beta_9 X_{9i}}_{\text{CD4 categories vs } < 50 \text{ cells}/\mu\text{L}} .
 \end{aligned} \tag{4.1}$$

The results from this model are presented as Model 1 in Table 4.3. These indicate that patients who enrolled at clinic site 2, those diagnosed with TB before 2006, those diagnosed with TB before enrolment for HIV care and those with higher (or missing) CD4 counts near the time of starting TB treatment had a lower hazard of initiating ART. These results are also portrayed graphically in Figure 4.2. They may be biased, however, since the proportional hazards assumption is not satisfied (See Section 4.2.2).

Table 4.3: Cox Regressions, Models 1-3

| Characteristic               | Model 1             |             | Model 2               |             | Model 3               |             |
|------------------------------|---------------------|-------------|-----------------------|-------------|-----------------------|-------------|
|                              | $e^{\hat{\beta}_j}$ | 95% CI      | $e^{\hat{\beta}_j^*}$ | 95% CI      | $e^{\hat{\beta}_j^*}$ | 95% CI      |
| Location of Clinic           |                     |             |                       |             |                       |             |
| Site 1                       | 0.67                | 0.57 - 0.79 | 0.62                  | 0.46 - 0.84 | 0.66                  | 0.53 - 0.82 |
| Site 2                       | 0.35                | 0.29 - 0.42 | 0.29                  | 0.21 - 0.40 | 0.30                  | 0.24 - 0.38 |
| Site 3 (baseline)            | 1.00                | -           | 1.00                  | -           | 1.00                  | -           |
| Year in which TB diagnosed   |                     |             |                       |             |                       |             |
| 2002-2005                    | 0.63                | 0.56 - 0.70 | 0.49                  | 0.39 - 0.61 | 0.54                  | 0.46 - 0.64 |
| 2006-2008 (baseline)         | 1.00                | -           | 1.00                  | -           | 1.00                  | -           |
| Clinic in which TB diagnosed |                     |             |                       |             |                       |             |
| TB Clinic                    | 0.57                | 0.51 - 0.65 | 0.24                  | 0.19 - 0.31 | 0.30                  | 0.25 - 0.36 |
| ART Clinic (baseline)        | 1.00                | -           | 1.00                  | -           | 1.00                  | -           |
| CD4 count <sup>a</sup>       |                     |             |                       |             |                       |             |
| 0-49 (baseline)              | 1.00                | -           | 1.00                  | -           | 1.00                  | -           |
| 50-99                        | 0.72                | 0.62 - 0.83 | 0.35                  | 0.26 - 0.48 | 0.50                  | 0.40 - 0.62 |
| 100-149                      | 0.61                | 0.52 - 0.72 | 0.28                  | 0.20 - 0.40 | 0.39                  | 0.30 - 0.50 |
| 150-199                      | 0.51                | 0.42 - 0.62 | 0.22                  | 0.15 - 0.33 | 0.28                  | 0.20 - 0.39 |
| $\geq 200$                   | 0.40                | 0.31 - 0.51 | 0.14                  | 0.08 - 0.25 | 0.17                  | 0.10 - 0.29 |
| missing                      | 0.19                | 0.16 - 0.22 | 0.03                  | 0.02 - 0.05 | 0.04                  | 0.03 - 0.07 |

CI, Confidence Interval

$e^{\hat{\beta}_j}$  denotes the estimated hazard ratio of ART initiation averaged over the entire analysis period for a patient having covariate  $X_j$  compared to the baseline patient

$e^{\hat{\beta}_j^*}$  denotes the estimated hazard ratio of ART initiation at the start of TB treatment for a patient having covariate  $X_j$  compared to the baseline patient

$e^{\hat{\beta}_j^{**}}$  denotes the estimated time-varying hazard ratio of ART initiation for a patient having covariate  $X_j$  compared to the baseline patient

<sup>a</sup> CD4 count taken closest start of TB treatment that satisfied criteria of being taken not earlier than 183 days before or 91 days after start of TB treatment and not more than 7 days after start of ART

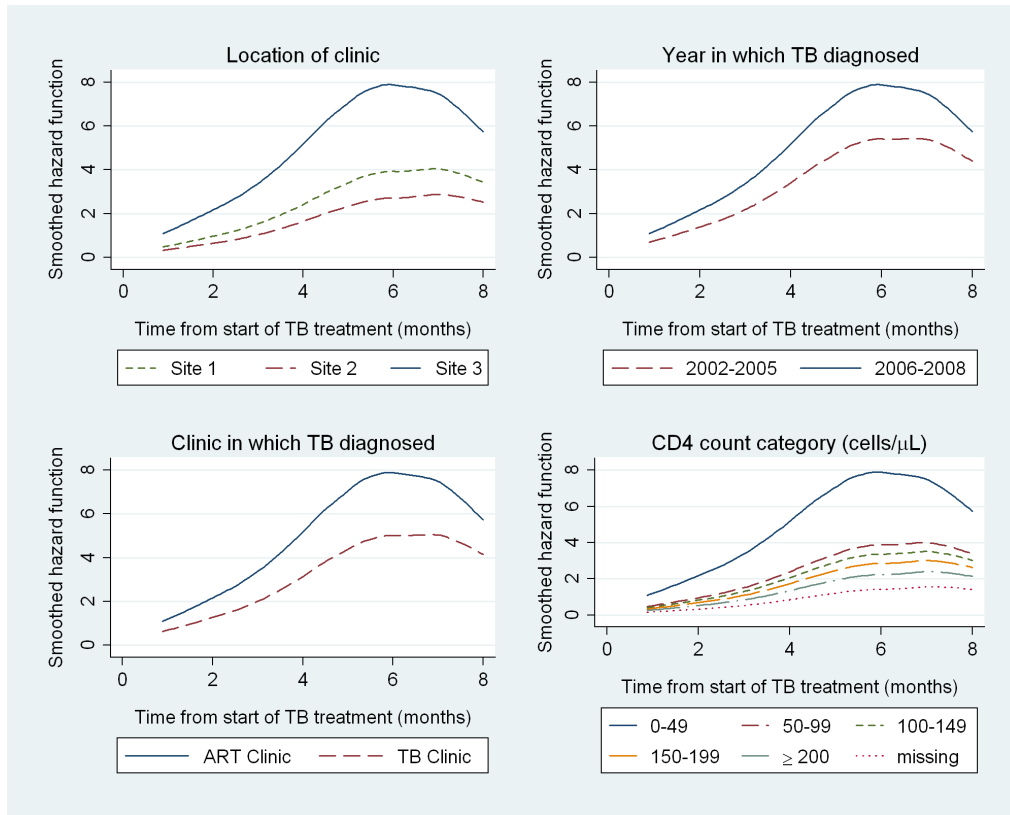


Figure 4.2: Smoothed hazard curves estimated using Cox PH Model

Stratified hazard curves by categorization of each predictor variable, whilst keeping other predictor variables constant at covariate pattern that yielded shortest time to ART initiation. This covariate pattern corresponded to enrolment clinic site 3, TB diagnosis in the calendar period 2006-2008, TB diagnosis ART clinic and CD4 count near the start of TB treatment lower than 50 cells/ $\mu$ L.

#### 4.2.2 Graphical assessment of proportional hazards assumption

Figure 4.3 shows the smoothed hazard curves stratified by variable category for each predictor variable from Model 1. None of the hazard curves stratified by predictor category look to be proportional over time, but the curved nature of these graphs and the fact that they are smoothed functions make the assumption of proportional hazards difficult to assess satisfactorily using only these plots.

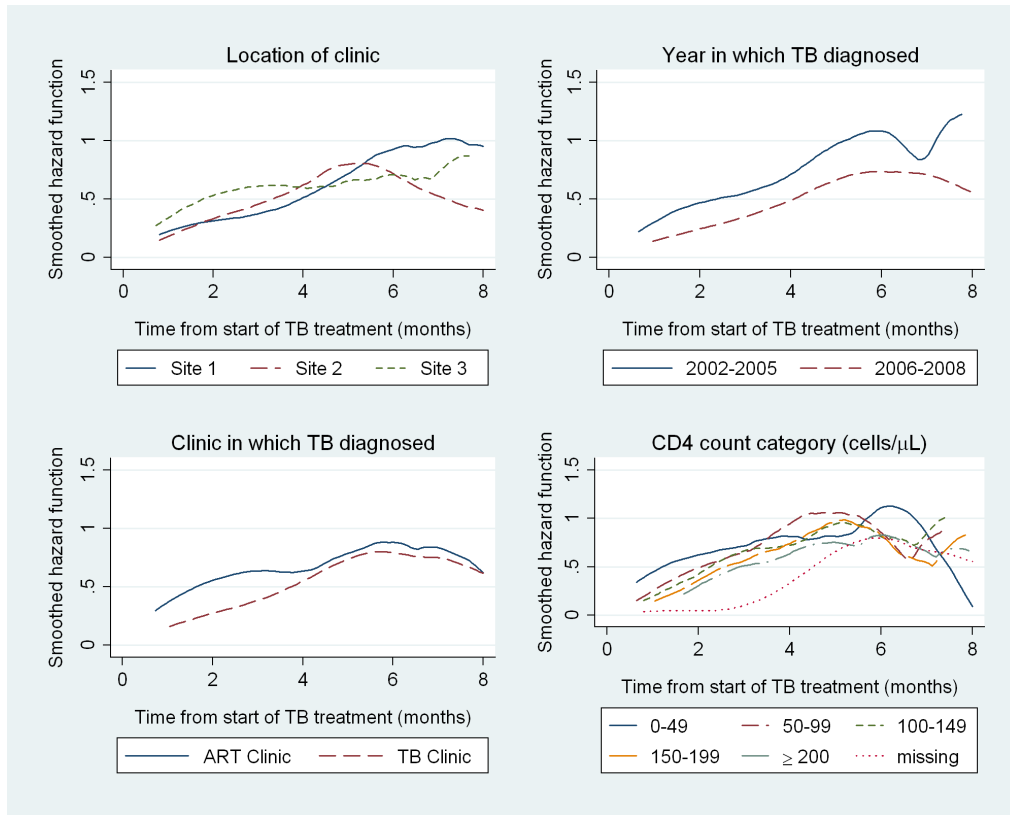


Figure 4.3: Non-parametric smoothed hazard curves

Plots of Kaplan-Meier estimates (observed) versus Cox proportional hazards regression estimates (predicted) are shown in Figures C.2 and C.3 in Appendix C.2. There is a slight deviation of the predicted (Cox) survival curve from the observed (Kaplan-Meier) survival curve for the curves grouped by the clinic in which TB was diagnosed. The deviations are much smaller for the survival curves grouped by calendar period of TB diagnosis and whether a patient is being treated for a new TB infection or is being retreated for TB. The deviations of the predicted values from the observed values can be more clearly seen in Figure C.3, which groups the survival curves by CD4 count category, particularly for those who did not have CD4 counts near the start of TB treatment.

Log-log survival curves are shown in Figure 4.4. The log-log survival curves appear almost parallel for most of the analysis time when stratified by clinic site and year of TB diagnosis. They are clearly not parallel, however, when stratified by the clinic in which TB was diagnosed and the CD4 count category. The main offenders of the proportional hazards assumption are thus

the clinic in which TB was diagnosed and the CD4 count taken near the time of TB diagnosis.

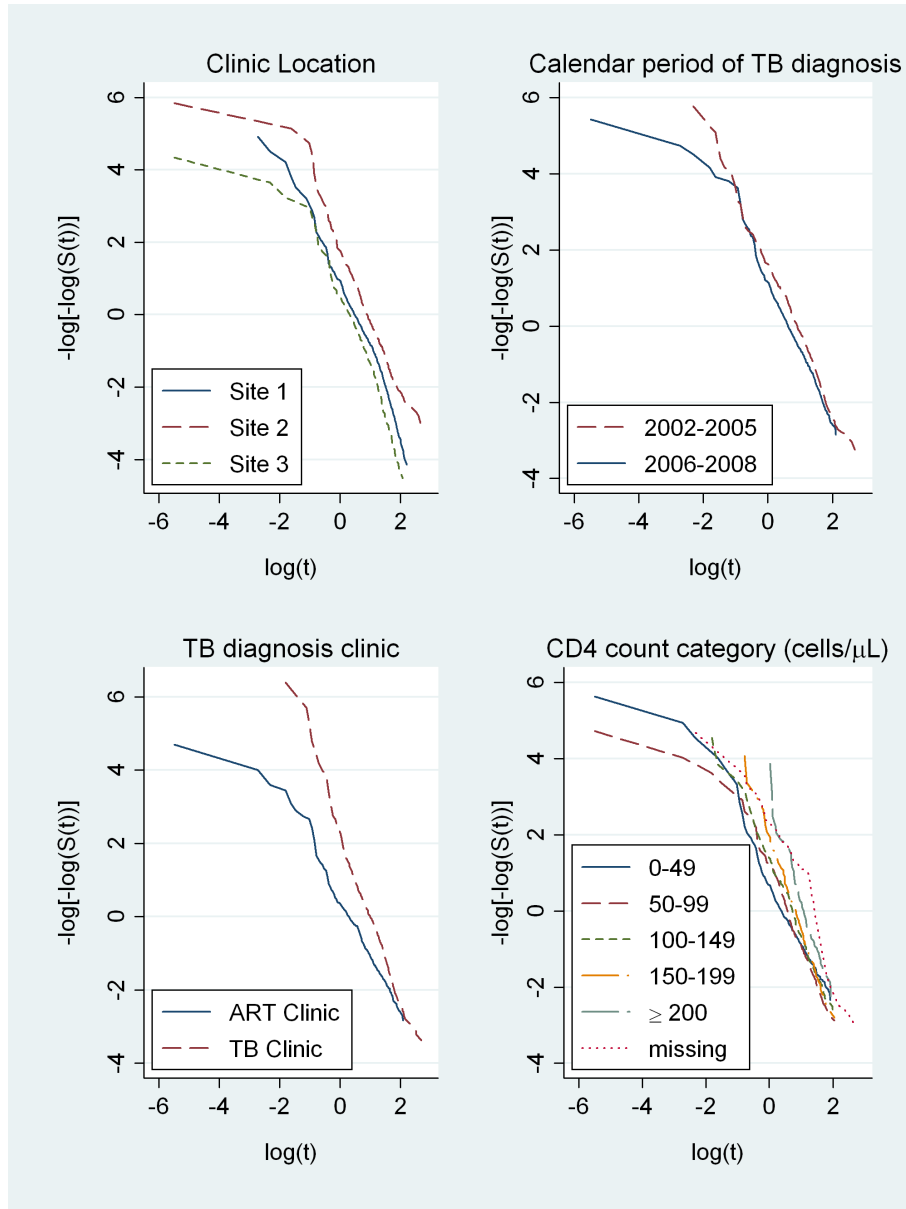


Figure 4.4: Comparison of log-log survival curves by covariate  
Estimates adjust to the baseline values of the other variables included in Model 1

### 4.2.3 Using interactions of time with covariates to model non-proportional hazards

One way of overcoming the problem of non-proportional hazards is to include an interaction of each variable with time (or a function of time). Equation 4.1 then changes to

$$\begin{aligned} \log \left[ \frac{h_i(t)}{h_0(t)} \right] &= \beta_1(t)X_{1i} + \beta_2(t)X_{2i} + \dots + \beta_9(t)X_{9i} \\ &= (\beta_1^* + \beta_1^{**} \times g(t))X_{1i} \\ &\quad + (\beta_2^* + \beta_2^{**} \times g(t))X_{2i} + \dots \\ &\quad + (\beta_9^* + \beta_9^{**} \times g(t))X_{9i} \end{aligned} \tag{4.2}$$

where  $X_{1i} \dots X_{9i}$  are as in 4.1 and  $g(t) = t$  in Model 2 and  $g(t) = \log(t)$  in Model 3. The estimated hazard ratios for these two models are in columns two and three of Table 4.3.

The estimated hazard ratios relating to the main effects of the covariates in models 2 and 3 (denoted  $\exp(\hat{\beta}_j^*)$  in Table 4.3) now refer to the effect of these covariates at time zero, i.e. at or immediately after the start of TB treatment. These effects are now greater (the estimated hazard ratios are lower) than the effects from model 1 (denoted  $\exp(\hat{\beta}_j)$ ) that were averaged over the entire time period. The estimated hazard ratios for the interaction terms (denoted  $\exp(\hat{\beta}_j^{**})$ ) measure the time-dependent effect modification. These are all greater than one, thus implying that the reduction in the time to ART initiation associated with having a particular value for a covariate at the start of TB treatment is not as great at later time points.

The reduction in the hazard ratios corresponding to the main effects of covariates for which an interaction with a function of time was included in the model were smaller for Model 3 (which used the natural logarithm of time as the function of time) as compared to Model 2 (which used time itself as the function). The hazard ratios corresponding to the interactions of the function of time with the covariates were larger for Model 3 than Model 2. Using  $g(t) = t$  allows a more rapid change in the hazard ratio over time, while using  $g(t) = \log(t)$  allows a more gradual change as time progresses.

To conclude, all three models indicate faster ART initiation in site 3, for patients diagnosed with TB from 2006 onward, for those diagnosed with TB after enrolment for HIV care and for those with lower CD4 counts. Models 2

and 3 indicate a diminishing effect of these baseline predictors as time progresses.

In the section that follows, we consider an alternative method of coping with non-proportional hazards: the use of parametric models that include ancillary parameters. The ancillary parameters allow for the modelling of survival functions that are not necessarily proportional over time for each value of a covariate.

### 4.3 Parametric Survival Analysis Models

As explained above (Section 3.3.2) the estimates from the Weibull Model can be presented in both the proportional hazards and accelerated failure time frameworks. I begin with the proportional hazards framework.

#### 4.3.1 Models using the proportional hazards metric

##### Weibull Model

Models of the following form were fit:

$$h_i(t) = \lambda_i p_i t^{p_i - 1} \quad (4.3)$$

where

$$\lambda_i = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_9 X_{9i}),$$

and the shape parameter  $p_i$  can be parameterized as

$$p_i = \exp(\beta_{p0} + \beta_{p1} X_{1i} + \beta_{p2} X_{2i} + \dots + \beta_{p9} X_{9i})$$

and  $X_{1i} \dots X_{9i}$  are as in (4.1)

The results are shown in Table 4.4. Patients enrolled at clinic site 2, those diagnosed with TB before 2006, those diagnosed with TB prior to enrolment for HIV care and those having higher CD4 counts or not having a CD4 count taken near the time of starting TB treatment had lower hazards of ART initiation. The hazard ratios of ART initiation ( $\exp(\hat{\beta}_j)$ ,  $j = 1, 2, \dots, 9$ ) for Model 5, for which the shape parameter  $p_i$  was allowed to vary for patients having different covariate patterns, were lower than those of Model 4 for

which the coefficients  $\hat{\beta}_{p1}, \hat{\beta}_{p2}, \dots, \hat{\beta}_{p9}$  of the shape parameter  $p_i$  were constrained to be one (such that the shape parameter  $p_i$  was constant for all covariate patterns). The coefficients for the shape parameter were not significantly different from zero for clinic location and calendar period of TB diagnosis, indicating that the shape parameter need not have been allowed to vary by level of these predictors. Holding the shape parameter constant for all levels of these predictors (not shown) does not significantly alter results for this model.

Table 4.4: Weibull proportional hazards regressions, Models 4 &amp; 5

| Characteristic               | Model 4 <sup>(1)</sup> |             | Model 5 <sup>(2)</sup> |             |                           |
|------------------------------|------------------------|-------------|------------------------|-------------|---------------------------|
|                              | $e^{\hat{\beta}_j}$    | 95% CI      | $e^{\hat{\beta}_j}$    | 95% CI      | $\hat{\beta}_{pj}$ 95% CI |
| Location of Clinic           |                        |             |                        |             |                           |
| Site 1                       | 0.68                   | 0.57 - 0.80 | 0.62                   | 0.47 - 0.82 | 0.020 -0.067 - 0.107      |
| Site 2                       | 0.35                   | 0.29 - 0.42 | 0.33                   | 0.25 - 0.45 | -0.002 -0.093 - 0.088     |
| Site 3 (baseline)            | 1.00                   | -           | 1.00                   | -           | 0.000 -                   |
| Year in which TB diagnosed   |                        |             |                        |             |                           |
| 2002-2005                    | 0.63                   | 0.56 - 0.70 | 0.53                   | 0.43 - 0.66 | 0.053 -0.010 - 0.116      |
| 2006-2008 (baseline)         | 1.00                   | -           | 1.00                   | -           | 0.000 -                   |
| Clinic in which TB diagnosed |                        |             |                        |             |                           |
| TB Clinic                    | 0.57                   | 0.51 - 0.65 | 0.23                   | 0.19 - 0.30 | 0.336 0.255 - 0.418       |
| ART Clinic (baseline)        | 1.00                   | -           | 1.00                   | -           | 0.000 -                   |
| CD4 count                    |                        |             |                        |             |                           |
| 0-49 (baseline)              | 1.00                   | -           | 1.00                   | -           | 0.000 -                   |
| 50-99                        | 0.72                   | 0.62 - 0.83 | 0.40                   | 0.30 - 0.53 | 0.246 0.146 - 0.346       |
| 100-149                      | 0.62                   | 0.52 - 0.73 | 0.31                   | 0.22 - 0.43 | 0.275 0.167 - 0.384       |
| 150-199                      | 0.52                   | 0.43 - 0.63 | 0.25                   | 0.16 - 0.38 | 0.263 0.139 - 0.388       |
| $\geq 200$                   | 0.41                   | 0.32 - 0.52 | 0.14                   | 0.08 - 0.26 | 0.364 0.209 - 0.519       |
| missing                      | 0.20                   | 0.17 - 0.24 | 0.06                   | 0.04 - 0.10 | 0.303 0.187 - 0.420       |

$e^{\hat{\beta}_j}$  is the hazard ratio of ART initiation for a patient having covariate  $X_j$  compared to the baseline patient.

- (1) For Model 4, the shape parameter  $p_i$  does not vary by covariate pattern, i.e.  $\hat{\beta}_{pj} = 0$  for all  $j \neq 0$ , such that  $p_i = \exp(\hat{\beta}_{p0}) = \exp(0.71) = 2.04$  for all patients (95% CI for  $p$ : 1.96-2.12).
- (2) For Model 5,  $p_i = \exp(\hat{\beta}_{p0} + \hat{\beta}_{p1}X_{1i} + \hat{\beta}_{p2}X_{2i} + \dots + \hat{\beta}_{p9}X_{9i})$ , with  $\hat{\beta}_{p0} = 0.33$ , 95% CI: 0.23-0.43 and  $\hat{\beta}_{pj}$  as given above.

### 4.3.2 Accelerated Failure Time Models

Two advantages of accelerated failure time (AFT) models over proportional hazards models are (i) that they focus on time rather than relative hazards and (ii) that they do not require the assumption of proportional hazards. First, I present the Weibull Model in the accelerated failure time framework, and thereafter the generalized Gamma model.

#### Weibull Model

Assuming a random variable  $\tau_i = \exp(-x_i\beta_x)t_i$  and  $\tau_i \sim \text{Weibull}(\alpha_0, p)$ , models of the following form were fit:

$$\log t_i = \boldsymbol{\alpha}'\mathbf{x}_i + z_i, \quad (4.4)$$

where

$$\boldsymbol{\alpha}'\mathbf{x}_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \dots + \alpha_9 X_{9i}$$

$X_{1i} \dots X_{9i}$  are as in (4.1)

$z_i$  has an extreme-value distribution with shape parameter  $\sigma_i = 1/p_i$ , which can be parameterized as

$$p_i = \exp(\alpha_{p0} + \alpha_{p1} X_{1i} + \alpha_{p2} X_{2i} + \dots + \alpha_{p9} X_{9i}).$$

Table 4.5 shows results for Model 6, which constrains the coefficients of the shape parameter  $p_i$  ( $\alpha_{pj}, j = 1, 2, \dots, 9$ ) to be zero (such that  $p_i$  does not vary by covariate value) and Model 7, which does allow the shape parameter  $p_i$  to vary by covariate value. The results for the main effects are shown in terms of  $\hat{\phi}_j = \exp(\hat{\alpha}_j)$ , which measures the extent of deceleration (or slowing down) in time to the event of interest. An alternative presentation in terms of  $\exp(-\hat{\alpha}_j)$  would have presented results in terms of acceleration in time to the event of interest.

The estimated coefficients  $\hat{\alpha}_j, j = 1, 2, \dots, 9$  from Model 6 are equal to  $-\hat{\beta}_j/\hat{p}$  from Model 4 (where  $\hat{p}$  is the shape parameter that does not vary by covariate value). This same relationship does not hold, however, for Models 5 and 7, for which the shape parameter  $p_i$  is allowed to vary by covariate value.

Both Models 6 and 7 show longer time to ART for patients attending a clinic in location 2, those diagnosed with TB before 2006, those diagnosed with TB before enrolment for HIV care and those having higher or missing

CD4 counts near the time of starting TB treatment. Model 7 shows significant modification in shape parameter by calendar period of TB diagnosis, timing of enrolment for HIV care relative to timing of TB diagnosis and CD4 count category.

The shortest predicted mean time to ART initiation is 35 days (95% CI: 32-38) for Model 6 and 33 days (95% CI: 30-37) for Model 7 (for patients enrolled at clinic site 3, diagnosed with TB in the calendar period 2006-2008 in the ART clinic and having a CD4 count lower than 50 cells/ $\mu$ L). The longest predicted mean times to ART initiation are 211 days (95% CI: 195-228) for Model 6 and 226 days (95% CI: 212-241) for Model 7 (for patients enrolled at clinic site 2, diagnosed with TB before enrolment for HIV care and not having a CD4 count measurement taken near the time of starting TB treatment).

Figure 4.5 shows the predicted hazard curves for patients who were diagnosed with TB in 2006 or later, had CD4 counts between 0 and 49 cells/ $\mu$ L near the time of starting TB treatment and received HIV care from clinic site 1 by whether they were diagnosed with TB in the TB clinic (before enrolment for ART) or in the ART clinic (after enrolment for ART as a part of the screening process that occurs before ART initiation) for each of the three Models 4, 5 and 7 (the predicted hazard curve for Model 6 is the same as that for Model 4).

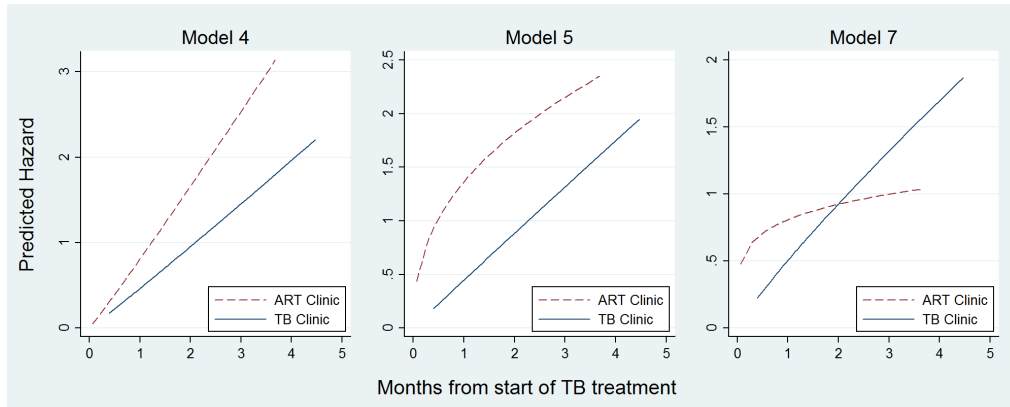


Figure 4.5: Comparison of predicted hazard curves (by clinic in which TB was diagnosed) from Weibull models in PH and AFT framework that do (Models 5 and 7) and do not (Model 4) allow the shape parameter to vary by covariate value

For Model 4, which does not allow the shape parameter to vary by covariate value, the two predicted hazard curves have the same shape. For Model 5, which is estimated using the proportional hazards framework and allows the shape parameter to vary by covariate value, the hazard curves have different shapes. For Model 7, which is estimated using the accelerated failure time framework and allows the shape parameter to vary by covariate value, the predicted hazard curves have a different shape and cross each other.

Table 4.5: Weibull AFT regressions, Models 6 &amp; 7

| Characteristic               | Model 6 <sup>(1)</sup>        |             | Model 7 <sup>(2)</sup>        |             |
|------------------------------|-------------------------------|-------------|-------------------------------|-------------|
|                              | $\hat{\phi}_j$ <sup>(3)</sup> | 95% CI      | $\hat{\phi}_j$ <sup>(3)</sup> | 95% CI      |
| Location of Clinic           |                               |             |                               |             |
| Site 1                       | 1.21                          | 1.12 - 1.31 | 1.19                          | 1.10 - 1.28 |
| Site 2                       | 1.67                          | 1.53 - 1.81 | 1.64                          | 1.52 - 1.77 |
| Site 3 (baseline)            | 1.00                          | -           | 1.00                          | -           |
| Year in which TB diagnosed   |                               |             |                               |             |
| 2002-2005                    | 1.26                          | 1.19 - 1.33 | 1.23                          | 1.17 - 1.29 |
| 2006-2008 (baseline)         | 1.00                          | -           | 1.00                          | -           |
| Clinic in which TB diagnosed |                               |             |                               |             |
| TB Clinic                    | 1.31                          | 1.24 - 1.39 | 1.45                          | 1.36 - 1.55 |
| ART Clinic (baseline)        | 1.00                          | -           | 1.00                          | -           |
| CD4 count                    |                               |             |                               |             |
| 0-49 (baseline)              | 1.00                          | -           | 1.00                          | -           |
| 50-99                        | 1.18                          | 1.09 - 1.27 | 1.27                          | 1.17 - 1.37 |
| 100-149                      | 1.27                          | 1.17 - 1.37 | 1.36                          | 1.26 - 1.48 |
| 150-199                      | 1.38                          | 1.26 - 1.51 | 1.54                          | 1.41 - 1.68 |
| $\geq 200$                   | 1.55                          | 1.38 - 1.74 | 1.68                          | 1.52 - 1.85 |
| missing                      | 2.19                          | 2.02 - 2.37 | 2.42                          | 2.25 - 2.60 |

<sup>(1)</sup> For Model 6, the shape parameter  $p$  does not vary by covariate pattern, i.e.  $\hat{\beta}_{pj} = 0$  for all  $j \neq 0$ , such that  $\hat{p}_i = \exp(\hat{\alpha}_{p0}) = \exp(0.71) = 2.04$  for all patients (95% CI for  $\hat{p}$ : 1.96-2.12).

<sup>(2)</sup> For Model 7,  $p_i = \exp(\hat{\alpha}_{p0} + \hat{\alpha}_{p1}X_{1i} + \hat{\alpha}_{p2}X_{2i} + \dots + \hat{\alpha}_{p9}X_{9i})$ , with  $\hat{\alpha}_{p0} = 0.18$  (95% CI: 0.03-0.33) and  $\hat{\alpha}_{pj}$  as given above.

<sup>(3)</sup>  $\hat{\phi}_j = \exp(\hat{\alpha}_j)$  is the decelerator, which indicates the multiplicative increase in the mean time to ART initiation for patients who have covariate  $X_j$  compared to the baseline patients.

### Generalized Gamma Model

The hazard ratio of the generalized Gamma was modelled using the relationship  $h(t) = \frac{f(t)}{S(t)}$ , where  $f(t)$  is given by (3.20) and  $S(t)$  is given by (3.21). The scale parameter is modelled as

$$\mu_i = \beta_{\mu 0} + \beta_{\mu 1}X_{1i} + \beta_{\mu 2}X_{2i} + \dots + \beta_{\mu 9}X_{9i}$$

and the shape parameters are modelled as

$$\kappa_i = \beta_{\kappa 0} + \beta_{\kappa 1}X_{1i} + \beta_{\kappa 2}X_{2i} + \dots + \beta_{\kappa 9}X_{9i}$$

and

$$\log(\sigma_i) = \beta_{\sigma 0} + \beta_{\sigma 1}X_{1i} + \beta_{\sigma 2}X_{2i} + \dots + \beta_{\sigma 9}X_{9i},$$

where  $X_{1i}, X_{2i}, \dots, X_{9i}$  are as in (4.1).

The generalized Gamma AFT model is a generalization of many of the other possible forms of AFT models. Table 4.6 shows the results of the estimation of two generalized Gamma AFT models (Models 8 and 9). Model 8 did not allow the estimated values of  $\hat{\kappa}_i$  and  $\hat{\sigma}_i$  to vary by covariate value (i.e. only  $\hat{\beta}_{\kappa 0}$  and  $\hat{\beta}_{\sigma 0}$  were estimated from the data, the remaining  $\hat{\beta}_{\kappa j}$  and  $\hat{\beta}_{\sigma j}$ ,  $j = 1, 2, \dots, 9$ , were constrained to be zero). The 95% confidence intervals for  $\hat{\sigma}_i = \exp(\hat{\beta}_{\sigma 0})$  and  $\hat{\kappa}_i = \hat{\beta}_{\kappa 0}$  from Model 8 are mutually exclusive and do not include either 0 or 1. The null hypotheses that the exponential ( $\kappa_i = 1$  and  $\sigma_i = 1$ ), Weibull ( $\kappa_i = 1$ ), Log-normal ( $\kappa_i = 0$ ) or Gamma ( $\sigma_i = |\kappa_i|$ ) models (which are special cases of the Generalized Gamma) may be sufficient to model the data are thus all rejected. This suggests that the Generalized Gamma, which has two ancillary parameters that influence the shape of the hazard, may be needed to model this data (rather than one of the simpler models that only have one ancillary parameter).

For Model 8, which did not allow the ancillary parameters to vary by covariate value, the baseline patients had the shortest estimated mean time to ART initiation (32 days, 95% CI: 28-35). The decelerators shown ( $e^{\hat{\beta}_{\mu j}}$ ) indicate the multiplicative increase in the mean time to ART initiation for a patient having the shown covariate value (compared to a patient with baseline values of the covariates).

Table 4.6: Generalized Gamma AFT, Models 8 &amp; 9

| Characteristic               | Model 8 <sup>(1)</sup>    |             |                           | Model 9 <sup>(2)</sup> |                          |               |
|------------------------------|---------------------------|-------------|---------------------------|------------------------|--------------------------|---------------|
|                              | $e^{\hat{\beta}_{\mu j}}$ | 95% CI      | $e^{\hat{\beta}_{\mu j}}$ | 95% CI                 | $\hat{\beta}_{\sigma j}$ | 95% CI        |
| Location of Clinic           |                           |             |                           |                        |                          |               |
| Site 1                       | 1.20                      | 1.10 - 1.31 | 1.05                      | 0.94 - 1.16            | 0.10                     | -0.05 - 0.25  |
| Site 2                       | 1.67                      | 1.53 - 1.82 | 1.50                      | 1.34 - 1.68            | 0.04                     | -0.11 - 0.19  |
| Site 3 (baseline)            | 1.00                      | -           | 1.00                      | -                      | 0.00                     | -             |
| Year in which TB diagnosed   |                           |             |                           |                        |                          |               |
| 2002-2005                    | 1.27                      | 1.20 - 1.34 | 1.20                      | 1.12 - 1.28            | -0.04                    | -0.13 - 0.04  |
| 2006-2008 (baseline)         | 1.00                      | -           | 1.00                      | -                      | 0.00                     | -             |
| Clinic in which TB diagnosed |                           |             |                           |                        |                          |               |
| TB Clinic                    | 1.42                      | 1.33 - 1.51 | 1.40                      | 1.27 - 1.54            | -0.40                    | -0.50 - -0.31 |
| ART Clinic (baseline)        | 1.00                      | -           | 1.00                      | -                      | 0.00                     | -             |
| CD4 count                    |                           |             |                           |                        |                          |               |
| 0-49 (baseline)              | 1.00                      | -           | 1.00                      | -                      | 0.00                     | -             |
| 50-99                        | 1.25                      | 1.16 - 1.35 | 1.39                      | 1.23 - 1.56            | -0.33                    | -0.45 - -0.21 |
| 100-149                      | 1.35                      | 1.24 - 1.48 | 1.56                      | 1.37 - 1.76            | -0.37                    | -0.51 - -0.24 |
| 150-199                      | 1.48                      | 1.34 - 1.64 | 1.51                      | 1.31 - 1.75            | -0.40                    | -0.54 - -0.26 |
| $\geq 200$                   | 1.68                      | 1.48 - 1.91 | 1.76                      | 1.51 - 2.05            | -0.51                    | -0.69 - -0.32 |
| missing                      | 2.34                      | 2.15 - 2.55 | 2.84                      | 2.52 - 3.19            | -0.54                    | -0.67 - -0.41 |
|                              |                           |             |                           |                        | 0.43                     | 0.08 - 0.79   |

$e^{\hat{\beta}_{\mu j}}$  is the decelerator, which indicates the multiplicative increase in the mean time to ART initiation for patients who have covariate  $X_j$  compared to the baseline patients

- (1) For Model 8, the ancillary parameters  $\sigma$  and  $\kappa$  are assumed to be the same for all covariate patterns, i.e.  $\hat{\beta}_{\sigma j} = \hat{\beta}_{\kappa j} = 0$  for all  $j \neq 0$ , such that  $\hat{\sigma} = \exp(\hat{\beta}_{\sigma 0}) = \exp(-0.67) = 0.51$  for all patients (95% CI for  $\hat{\sigma}$ : 0.49 to 0.53) and  $\hat{\kappa} = \hat{\beta}_{\kappa 0} = 0.67$  for all patients (95% CI: 0.57 to 0.78).
- (2) For Model 9,  $\log(\hat{\sigma}) = \hat{\beta}_{\sigma 0} + \hat{\beta}_{\sigma 1}X_{1i} + \hat{\beta}_{\sigma 2}X_{2i} + \dots + \hat{\beta}_{\sigma 9}X_{9i}$ , where  $\hat{\beta}_{\sigma 0} = -0.24$  (95% CI: -0.41 to -0.07) and  $\hat{\kappa} = \hat{\beta}_{\kappa 0} + \hat{\beta}_{\kappa 1}X_{1i} + \hat{\beta}_{\kappa 2}X_{2i} + \dots + \hat{\beta}_{\kappa 9}X_{9i}$ , where  $\hat{\beta}_{\kappa 0} = 1.17$  (95% CI: 0.71 to 1.63).

So, for example, using the mean time given above and the values of the decelerators for Model 8 shown in Table 4.6, we can determine the following: for a patient who enrolled for HIV care at clinic site 3, who was diagnosed with TB in 2006-2008 and who was diagnosed with TB in the ART clinic (all baseline covariate values), but who had a CD4 count in the category 50-99 cells/ $\mu$ L the estimated mean time to ART initiation would be  $32 \times 1.25 = 40$  days (95% CI: 36-44). As the CD4 category values increase (further away from the baseline category), this multiplicative increase in time to ART increases as well. For patients with no recorded CD4 count the time to ART is 2.34 times the time to ART for patients with a CD4 count recorded between the values of 0 and 49. For patients who enrolled at clinic sites 1 and 2, the time to ART initiation was 20% and 67% greater respectively compared to patients who enrolled at clinic site 3. Patients who were diagnosed with TB in earlier years (2002 to 2005 compared to 2006 and later) had a 26% increased delay to ART initiation. Patients who were diagnosed with TB in the TB clinic (before enrolment for HIV care) had a 31% increased time to ART initiation compared to patients diagnosed with TB in the ART clinic.

These relationships are not as simple when ancillary parameters are allowed to vary by covariate pattern, as in Model 9. I have thus chosen to describe the effects of estimates for each of the coefficients  $e^{\hat{\beta}_{\mu j}}$ ,  $e^{\hat{\beta}_{\sigma j}}$  and  $e^{\hat{\beta}_{\kappa j}}$  by referring to their effect on the shape of the hazard curve.

Decelerators  $e^{\hat{\beta}_{\mu j}}$  greater than one result in hazard curves that are less steep and have lower intercepts than the hazard curve corresponding to baseline values of covariates. All values for covariates other than the baseline values had decelerators greater than one in both Model 8 and Model 9. The estimated values for  $e^{\hat{\beta}_{\mu j}}$  are not that different for Model 9 compared to Model 8 (though they are lower in Model 9 for all covariates other than the CD4 count categories, for which they are higher).

Values of  $\hat{\beta}_{\sigma j}$  greater than zero also result in a hazard curves that are less steep, but do not affect the intercepts of the hazard curves. Only clinic sites 1 and 2 have estimated values  $\hat{\beta}_{\sigma j}$  greater than zero. For the other covariates, the effect of the decrease in slope resulting from decelerators above one is offset to some extent by the values for  $\hat{\beta}_{\sigma j}$  that are below zero. Values for  $\hat{\beta}_{\kappa j}$  that are greater than zero result in hazard curves that are more curved and slightly steeper. Fairly large changes ( $\hat{\beta}_{\kappa j} > 1$  or 2) are required for there to be a very noticeable effect on the hazard curve, however and the estimates for this data were all less than 0.6 in absolute value.

Figure 4.6 shows the hazard and survival curves by different levels of each covariate displayed in Model 9 of Table 4.6. I chose to use the most frequent value for each covariate (patients enrolled at clinic site 1, diagnosed with TB in 2006-2008 before enrolment for HIV care and having a CD4 count of lower than 50 cells/ $\mu$ L) as the baseline hazard and survival curves (solid blue lines). Note that these are different to the baseline covariates used in the table, which corresponded to those with the shortest time to ART initiation.

The most marked differences with regard to shape can be seen in the hazard and survival curves stratified by CD4 count category, where both the lowest CD4 category and the missing CD4 category have very different shapes from the survival curves for the other CD4 count categories. These different shapes highlight the much faster or earlier ART initiation for the  $CD4 \leq 50$  group and the very long time from the start of TB treatment to ART initiation for those not having CD4 counts taken near the time of starting TB treatment. The hazard curves stratified by clinic also have quite different shapes, with the hazard curve for clinic site 3 being much steeper (patients initiating ART sooner) than that of clinic site 2. The curves for clinic sites 1 and 3 diverge after about one month of TB treatment, with the curve for clinic site 1 becoming less steep.

In model 9, which allowed the estimated values of  $\hat{\kappa}_i$  and  $\hat{\sigma}_i$  to vary by covariate value, the shortest estimated mean time to ART initiation of 30 days (95% CI: 26-34) was for patients enrolled at clinic site 3, who were diagnosed with TB after enrolment for HIV care and in the calendar period 2006-2008 and who had a CD4 count lower than 50 cells/ $\mu$ L near the time of starting TB treatment. The longest estimated mean time to ART initiation of 230 days (95% CI: 214-247) was for patients who were enrolled at clinic site 2 after their TB diagnosis and who were diagnosed with TB before 2006 and did not have a CD4 count near the time of starting TB treatment. This was longer than the estimated mean time to ART initiation from Model 8, of 223 days (95% CI: 204-241). Allowing the ancillary parameters to vary by covariate pattern resulted in a slightly larger range of estimates of mean time to ART initiation, with the shortest mean time being shorter and the longest mean time being longer, for Model 9 compared to Model 8.

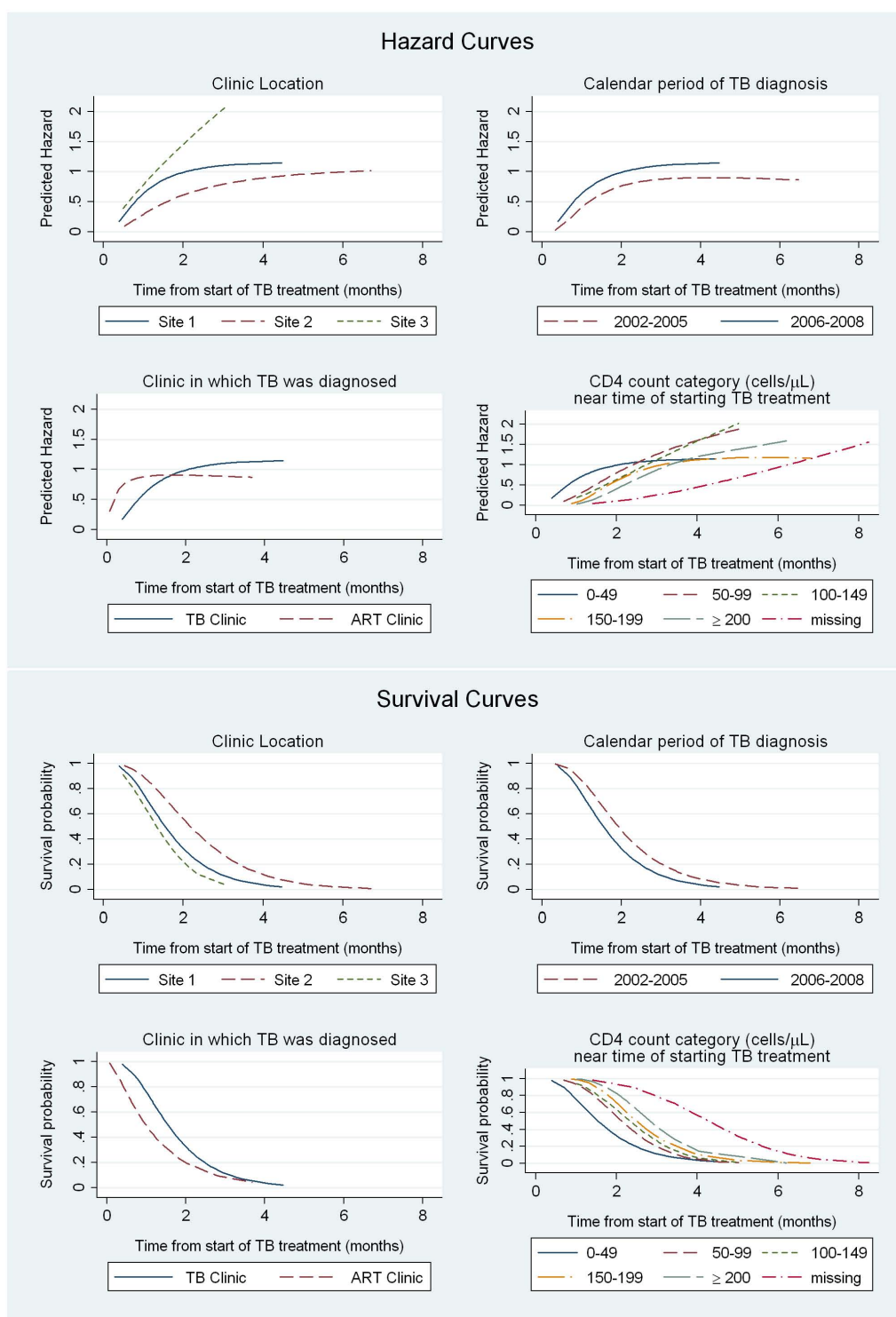


Figure 4.6: Generalized Gamma AFT hazard and survival curves by covariate (using mode of covariates as baseline)

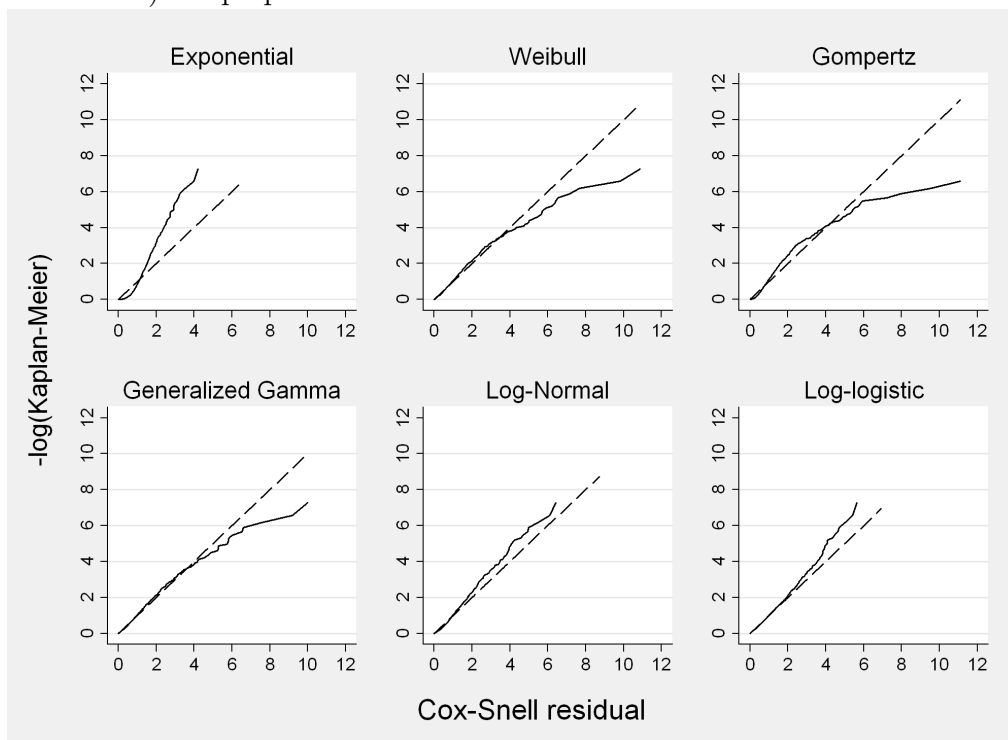
### 4.3.3 Comparison of Parametric Models

AFT Models using the log-logistic and log-Normal distributions were also fitted (see Tables C.5 and C.6 in Appendix C.3). Table 4.7 shows the log-likelihood (ll) of the null model and the estimated model, as well as the degrees of freedom and the values for the Akaike and Bayesian Information Criteria (AIC and BIC) for the models fitted. All models used the covariates clinic location, year of TB diagnosis, clinic in which TB was diagnosed and CD4 count category.

For the models that did not allow the shape parameter to vary by covariate value, the model with the highest log-likelihood and the lowest AIC and BIC was the log-logistic model. Out of all the models, however, the model with the highest log-likelihood and lowest AIC was the generalized Gamma model that allowed the values of the shape parameters to vary by covariate value (corresponding to Model 9 above). The model with the lowest BIC was the Weibull model that allowed the shape parameter to vary by covariate value (corresponding to Model 7 above).

Figure 4.7 shows plots of the cumulative hazard of the Cox-Snell residuals. If the models fit the data perfectly, the graphs would be a straight line through the origin (shown by the dotted lines). In panel A, which shows the graphs for the models that constrained the shape parameters to be constant for all values of covariates used in the models, the Log-Normal and Log-logistic models appear to fit the data better than the Weibull and generalized Gamma Models, especially for higher values of the Cox-Snell residuals. The generalized Gamma model that allows the shape parameters to vary by covariate value, however, appears to have the best fit to the data out of all the models shown.

A) Shape parameters constant for all values of covariates



B) Shape parameters allowed to vary by covariate value

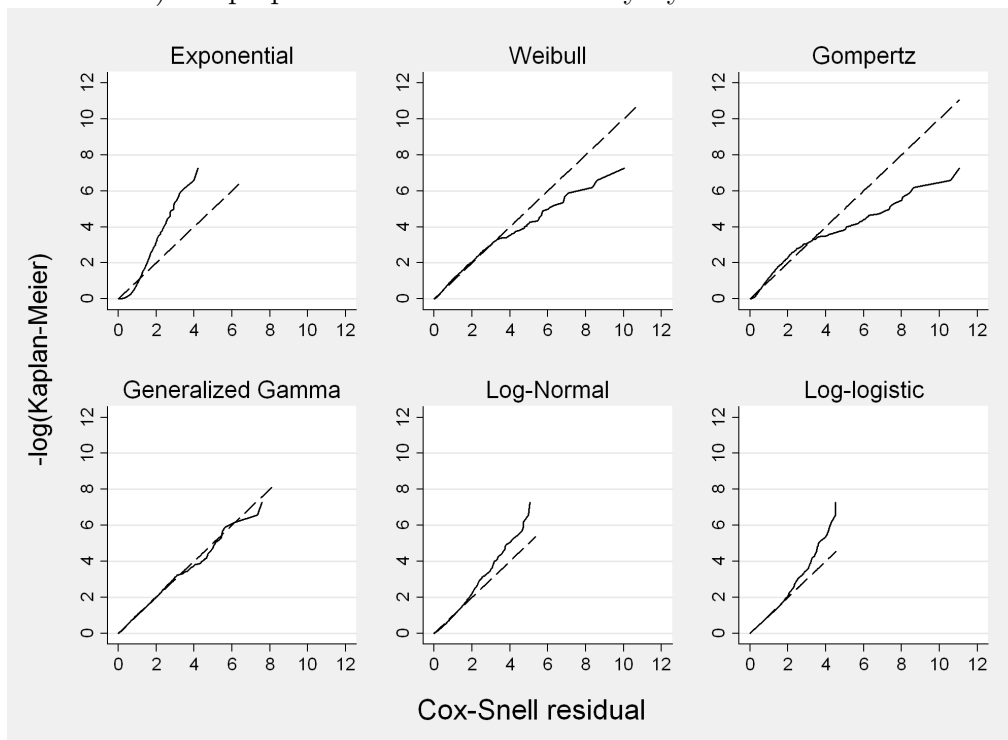


Figure 4.7: Comparison of model fit of Parametric Models

Table 4.7: Comparison of model fit statistics for Parametric Models

| Model              | ll(null)  | ll(model) | df | AIC      | BIC      |
|--------------------|-----------|-----------|----|----------|----------|
| exponential        | -1779.170 | -1653.608 | 10 | 3327.215 | 3379.967 |
| Gompertz           | -1634.584 | -1331.160 | 11 | 2684.321 | 2742.348 |
| Log-logistic       | -1619.364 | -1184.949 | 11 | 2391.898 | 2449.925 |
| Log-normal         | -1672.300 | -1296.943 | 11 | 2615.887 | 2673.913 |
| Weibull            | -1543.420 | -1201.745 | 11 | 2425.491 | 2483.517 |
| Generalized Gamma  | -1541.495 | -1186.132 | 12 | 2396.264 | 2459.566 |
| Gompertz*          | -1413.615 | -1218.585 | 20 | 2477.170 | 2582.674 |
| Log-logistic*      | -1523.400 | -1090.811 | 20 | 2221.621 | 2327.125 |
| Log-normal*        | -1535.431 | -1148.590 | 20 | 2337.180 | 2442.683 |
| Weibull*           | -1474.814 | -1082.332 | 20 | 2204.665 | 2310.168 |
| Generalized Gamma* | -1280.426 | -1052.295 | 30 | 2164.590 | 2322.846 |

\*Shape parameters allowed to vary by covariate value

#### 4.3.4 Analysis of deviance residuals for Models 8 & 9

I present results from model checking for Models 8 and 9 only, since Model 9 is the best model based on the comparison of models given above and Model 8 is a simplified form of Model 9 (same distribution, but ancillary parameters not allowed to vary by covariate value). Figure 4.8 shows plots of the deviance residuals for Models 8 and 9 against the linear predictors.

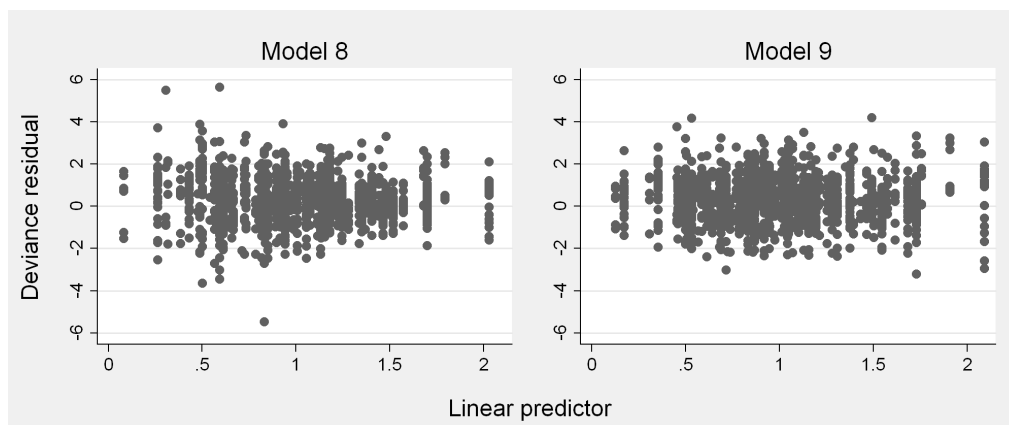


Figure 4.8: Plots of deviance residuals for Models 8 and 9

The plot for Model 9 looks like random scatter, indicating reasonable model fit, though there are some deviance residuals greater than two in absolute value. For Model 8, there appear to be some large deviance residuals corresponding to low to medium values of the linear predictor (the three patients with the largest deviance residuals had very short or very long times to ART initiation).

Re-estimating the models after excluding those patients for whom the absolute values of the deviance residuals were greater than two (88 patients for Model 8 and 94 patients for Model 9) did result in changes in the parameter estimates, however the same general conclusions held. Figure 4.9 shows a comparison of the predicted hazard curves for Model 9 when all patients are included versus when patients with deviance residuals greater than two in absolute value are excluded.

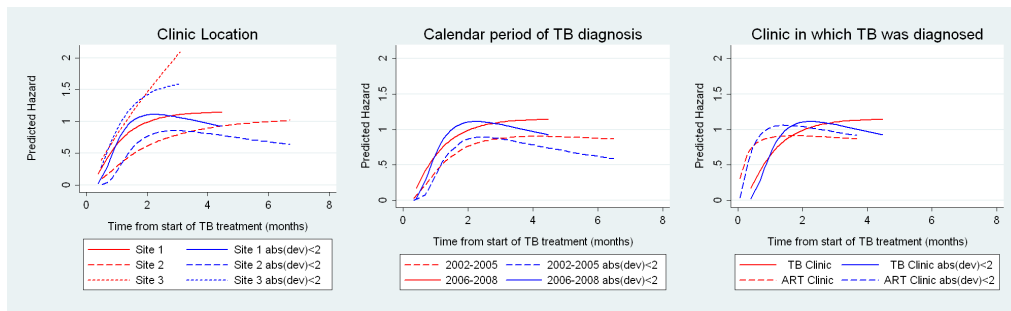


Figure 4.9: Comparison of predicted hazard curves for Model 9 when patients with large deviance residuals are included (red) versus excluded (blue)

The predicted hazards of ART initiation generally start slightly lower, incline faster and are higher at one to two months after the start of TB treatment and lower thereafter (for the model excluding patients who have deviance residuals greater than two in absolute value compared to the model including all patients). The changes are much less apparent for the predicted survival curves, which show slightly higher survival up until about two months and thereafter very similar or slightly lower predicted survival probabilities (see Figure 4.10).

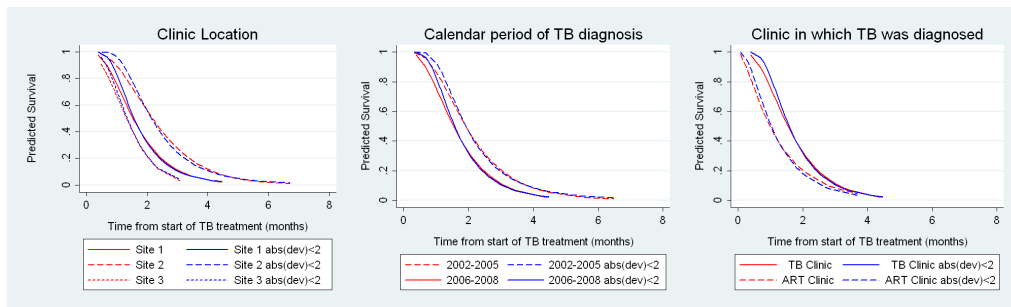


Figure 4.10: Comparison of predicted survival curves for Model 9 when patients with large deviance residuals are included (red) versus excluded (blue)

### 4.3.5 Sensitivity Analysis

#### Treatment of missing CD4 counts

Throughout this analysis, I have included patients not having a CD4 count near the start of TB treatment using a separate CD4 count category denoted ‘missing’. To test the sensitivity of the results to different treatment of missing CD4 counts, four additional models were fitted as described below:

- GG1 the chosen generalized Gamma AFT model
- GG2 a model that excludes patients who do not have a CD4 count near the start of TB treatment
- GG3 a model that assumes that the CD4 count was  $\geq 200$  cells/ $\mu\text{L}$  for all patients who did not have a recorded CD4 count near the time of starting TB treatment
- GG4 a model that assumes that the CD4 count was  $< 50$  cells/ $\mu\text{L}$  for all patients who did not have a recorded CD4 count near the time of starting TB treatment.

For Models GG2 and GG3, the estimated decelerators ( $e^{\hat{\beta}_{\mu j}}$ ) (see Table 4.8) and values for  $\hat{\beta}_{\sigma j}$  and  $\hat{\beta}_{\kappa j}$  (see Tables C.3 and C.4 in Appendix C.1.3) were quite similar to those of Model GG1 (aside from the decelerator for CD4  $\geq 200$  cells/ $\mu\text{L}$  in Model GG3, which was much larger than that of Model GG1, as a result of the inclusion of the patients with missing CD4 counts in that category). For Model GG4, however, the decelerators were greater (and further from one) for covariates not relating to CD4 count. The decelerators for those having CD4 counts between 50 and 199 cells/ $\mu\text{L}$  were lower than one, indicating a faster time to ART initiation for these patients as compared to the combined grouping of patients not having a CD4 count near the time of starting TB treatment or having a CD4 count lower than 50 cells/ $\mu\text{L}$  (but this effect was only significant for those with CD4 counts in the category 50-99 cells/ $\mu\text{L}$ ).

It could possibly be contended that patients who had no CD4 count recorded near the time of starting TB treatment had longer delays to ART initiation simply because they had no CD4 count recorded. The fact that they survived the longer time to ART initiation, however, may indicate that their CD4 counts near the time of starting TB treatment were relatively high. Furthermore, patients who appear sicker at the time of starting TB treatment are more likely to be urged (and feel the need) to attend an ART clinic

(sooner), where they would have their CD4 count taken and enrol for ART if they meet eligibility criteria. Such patients are also likely to have lower CD4 counts since a low CD4 count is an indicator of a weakened immune system. It thus seems unlikely that many patients not having a CD4 count recorded near the start of TB treatment have a very low CD4 count. The scenario in Model GG3, where patients not having a CD4 count are assumed to have a CD4 count  $\geq 200$  cells/ $\mu$ L, is more likely and this is confirmed by the results presented here. Patients not having a CD4 count recorded, however, may experience greater delays in ART initiation than those having a CD4 count, hence the choice to model those with a missing CD4 count in their own category.

Figure 4.11 shows the survival curves by CD4 count category for Model GG1 and Model GG4. The effect of grouping those not having a CD4 count near the time of starting TB treatment with those who had a CD4 count lower than 50 cells/ $\mu$ L is that the survival curve for those patients becomes less steep than that of only those who have a CD4 count lower than 50 cells/ $\mu$ L (from Model GG1) and more steep than that of only those not having a CD4 count. This combined survival curve also crosses the survival curves for all the other categories of CD4 counts. This once again illustrates the implausibility of assuming that missing CD4 counts correspond to values of 0-49.

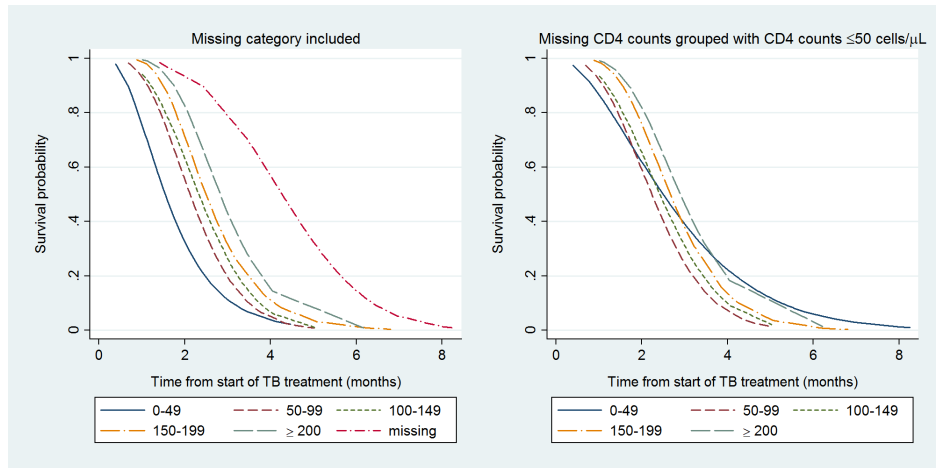


Figure 4.11: Comparison of generalized Gamma survival curves by CD4 count Baseline covariates (solid blue line) correspond to the most frequent values observed for a covariate, which were enrolment at clinic site 1, TB diagnosis in the calendar period 2006-2008, TB diagnosis in TB clinic (before enrolment for HIV care) and CD4 count near the start of TB treatment lower than 50 cells/ $\mu$ L (LHS) or the grouped category of no CD4 count and CD4 count lower than 50 cells/ $\mu$ L (RHS).

Table 4.8: Comparison of decelerators for models making different assumptions about the appropriate treatment of patients with no CD4 count near the start of TB treatment

| Characteristic               | Model GG1                                     |             | Model GG2                                     |             | Model GG3                                       |             | Model GG4                                   |             |
|------------------------------|-----------------------------------------------|-------------|-----------------------------------------------|-------------|-------------------------------------------------|-------------|---------------------------------------------|-------------|
|                              | missing category<br>$e^{\hat{\beta}_{\mu,j}}$ | 95% CI      | missing excluded<br>$e^{\hat{\beta}_{\mu,j}}$ | 95% CI      | missing $\geq 200$<br>$e^{\hat{\beta}_{\mu,j}}$ | 95% CI      | missing $< 50$<br>$e^{\hat{\beta}_{\mu,j}}$ | 95% CI      |
| Location of Clinic           |                                               |             |                                               |             |                                                 |             |                                             |             |
| Site 1                       | 1.05                                          | 0.94 - 1.16 | 1.10                                          | 0.96 - 1.26 | 1.08                                            | 0.96 - 1.21 | 1.29                                        | 1.12 - 1.48 |
| Site 2                       | 1.50                                          | 1.34 - 1.68 | 1.58                                          | 1.39 - 1.79 | 1.44                                            | 1.28 - 1.62 | 1.53                                        | 1.34 - 1.75 |
| Site 3 (baseline)            | 1.00                                          | -           | 1.00                                          | -           | 1.00                                            | -           | 1.00                                        | -           |
| Year in which TB diagnosed   |                                               |             |                                               |             |                                                 |             |                                             |             |
| 2002-2005                    | 1.20                                          | 1.12 - 1.28 | 1.26                                          | 1.16 - 1.37 | 1.23                                            | 1.15 - 1.32 | 1.38                                        | 1.27 - 1.51 |
| 2006-2008 (baseline)         | 1.00                                          | -           | 1.00                                          | -           | 1.00                                            | -           | 1.00                                        | -           |
| Clinic in which TB diagnosed |                                               |             |                                               |             |                                                 |             |                                             |             |
| TB Clinic                    | 1.40                                          | 1.27 - 1.54 | 1.39                                          | 1.26 - 1.53 | 1.45                                            | 1.32 - 1.60 | 1.48                                        | 1.33 - 1.64 |
| ART Clinic (baseline)        | 1.00                                          | -           | 1.00                                          | -           | 1.00                                            | -           | 1.00                                        | -           |
| CD4 count                    |                                               |             |                                               |             |                                                 |             |                                             |             |
| 0-49 (baseline)              | 1.00                                          | -           | 1.00                                          | -           | 1.00                                            | -           | 1.00                                        | -           |
| 50-99                        | 1.39                                          | 1.23 - 1.56 | 1.37                                          | 1.22 - 1.54 | 1.39                                            | 1.23 - 1.56 | 0.83                                        | 0.73 - 0.93 |
| 100-149                      | 1.56                                          | 1.37 - 1.76 | 1.51                                          | 1.34 - 1.71 | 1.55                                            | 1.37 - 1.75 | 0.91                                        | 0.80 - 1.02 |
| 150-199                      | 1.51                                          | 1.31 - 1.75 | 1.48                                          | 1.29 - 1.70 | 1.55                                            | 1.34 - 1.79 | 0.92                                        | 0.81 - 1.06 |
| $\geq 200$                   | 1.76                                          | 1.51 - 2.05 | 1.75                                          | 1.50 - 2.04 | 2.53                                            | 2.26 - 2.83 | 1.00                                        | 0.84 - 1.19 |
| missing                      | 2.84                                          | 2.52 - 3.19 |                                               |             |                                                 |             |                                             |             |

$e^{\hat{\beta}_{\mu,j}}$  is the decelerator, which indicates the multiplicative increase in the mean time to ART initiation for patients who have covariate  $X_j$  compared to the baseline patients.

## 4.4 Competing Risks Models

Up until this point in the analysis of time to ART initiation, I have only included patients who did initiate ART. This may have introduced bias, in that estimates of the time to ART initiation may have been underestimates as a result of excluding patients who did not initiate ART. In this section I explore the effect of competing events on the time to and hazard of ART initiation while on TB treatment. The analyses in this section include only data from clinic site 1. For clinic sites 2 and 3, no data was available on patients who did not initiate ART, so no comparison can be made for these sites between patients who did and did not initiate ART.

The main event of interest is ART initiation while on TB treatment and the competing events are death, permanent deferral and the end of TB treatment. In the absence of a given date for the end of TB treatment, it was assumed to have been completed after 6 months for patients who were being treated for a new TB infection (3 patients) or after 8 months for patients who were being retreated for TB (4 patients). In describing the ART programme at clinic site 1, Lawn et al. (2005) states that patients may be permanently deferred for ART if they elect to obtain treatment elsewhere, repeatedly miss scheduled follow-up appointments or relocate out of the area. They may also be permanently deferred for psychosocial reasons, such as not accepting that they are infected with HIV.

Death is a competing event, since patients who die before starting ART will not be observed to initiate ART. Although it is possible that patients classified as permanently deferred could initiate ART in the future, being classified as permanently deferred does substantially affect the probability of initiating ART while on TB treatment and I have thus considered permanent deferral to be a competing event. The end of TB treatment is considered a competing event, since the event of interest is ART initiation *while on TB treatment*.

Of 893 patients enrolled at clinic site 1, only 28 (3%) were observed to die, 34 (4%) were ‘permanently deferred’ for ART prior to the end of TB treatment and 157 (18%) had an end of TB date prior to ART initiation, death or permanent deferral. These low incidences mean that we do not expect to see a big impact of these competing events on ART initiation.

### Graphical analysis of hazard functions and cumulative incidence functions for competing risks

The graphs presented here are similar to those presented in Figure 2 of Gichangi and Vach (2005), but have been constructed using my data. They are also the same as those shown in Section 3.5.2, but have been presented here again for ease of reference. Figure 4.12 (1.a) shows the individual cause-specific hazards of death, deferral, end of TB treatment and ART initiation. The hazards of death and deferral were very low for all times from the start of TB treatment until eight months thereafter. The hazard for death did increase slightly from about five months after the start of TB treatment, while the hazard of deferral was slightly greater from about four months and increased quite quickly from seven months after the start of TB treatment. The hazard of TB treatment ending was very low initially, but increased quickly from about four months after the start of TB treatment and continued to rise up to eight months after the start of TB treatment.

The hazard of ART initiation was fairly low initially (though substantially higher than the hazards of the competing risks) but continued to increase until about five months after the start of TB treatment and then declined.

Figure 4.12 (1.b) shows the stacked cause-specific hazard functions, such that the top of the area corresponding to the hazard for ART initiation is equivalent to the cumulative incidence of failure from any cause. These hazard functions were drawn (from bottom to top) using equations (3.39)-(3.41) and:

- i) treating death as the event of interest and the other events as censoring events
- ii) treating the combined outcome of death or deferral as the event of interest and end of TB treatment or ART initiation as censoring events
- iii) treating the combined outcome of death, deferral or the end of TB treatment as the event of interest and ART initiation as a censoring event
- iv) assuming all subjects fail and none are censored.

Figure 4.12 (1.c) shows the relative contributions of the different causes (or hazards). From both Figures 1.b and 1.c, it can clearly be seen that ART initiation is the greatest cause of failure for the first four months of the analysis. Thereafter, the hazard of TB treatment ending quickly rises, such that

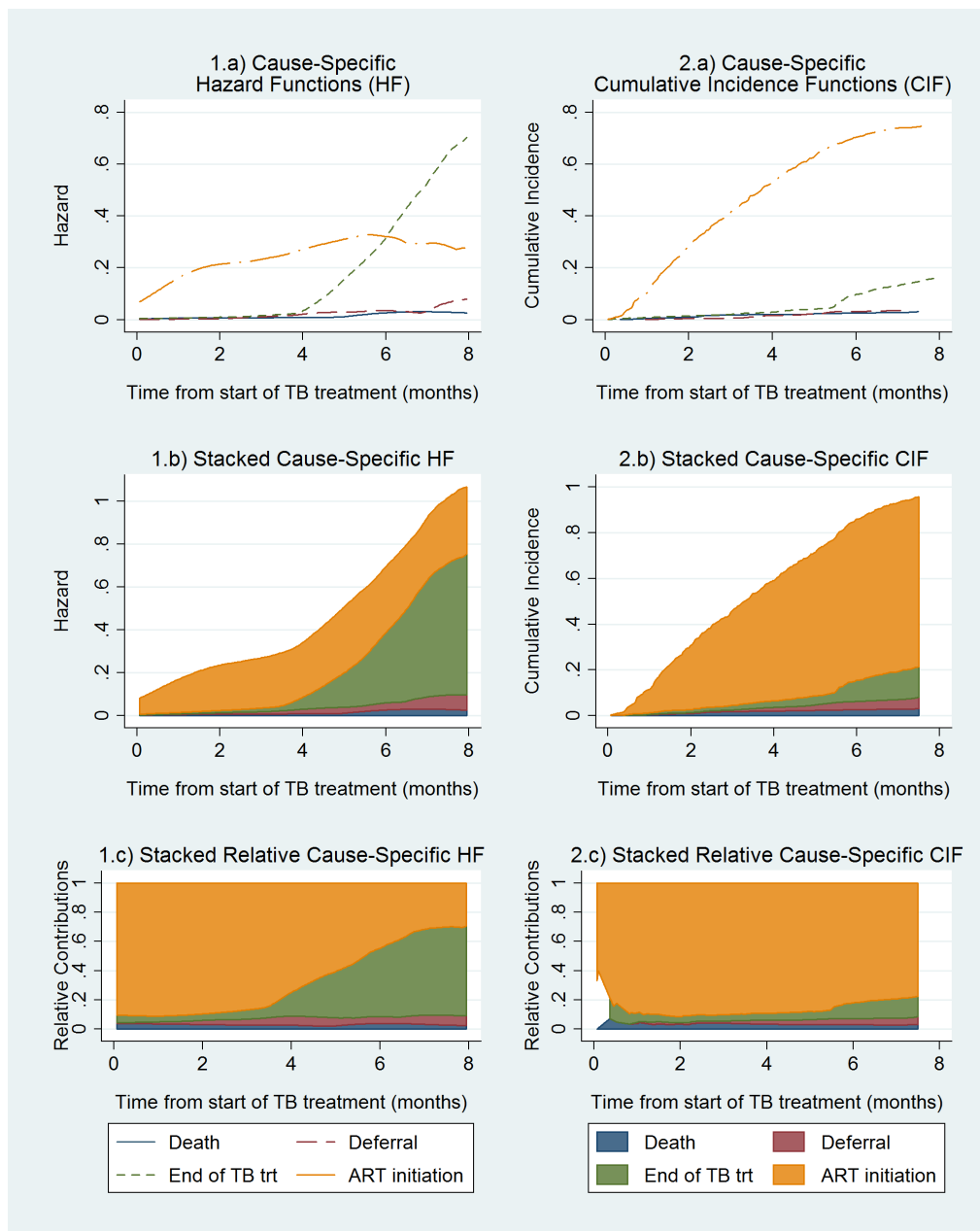


Figure 4.12: Cause-specific hazard and cumulative incidence functions

the end of TB treatment becomes the greatest cause of failure from about six months after the start of TB treatment (which is the usual period of TB treatment for patients who are being treated for a new TB infection that do not default or have other complications). Failures from death and deferral form a small proportion of the number of failures at any time.

The cumulative incidence functions shown in Figure 4.12 (2.a-c) tell a similar story. They make it clear, however, that even though the hazard of ending TB treatment becomes large by eight months after the start of TB treatment (surpassing that of ART initiation), the cumulative incidence of ART initiation remains much larger than that of TB treatment ending.

#### **4.4.1 Comparison of patient characteristics by failure event**

A failure event is an event which results in a patient leaving the risk set and can either be the event of interest (ART initiation while on TB treatment) or a competing event (death pre-ART, permanent deferral for ART or the end of TB treatment). Table 4.9 shows the median and interquartile ranges of CD4 counts (near start of TB treatment) and of days from the start of TB treatment to failure stratified by the failure event experienced by the patient and either the clinic in which TB was diagnosed or the calendar period of TB diagnosis.

The overall median CD4 count was similar for patients diagnosed in the ART or TB clinics and in earlier or later calendar periods, but did vary within those strata by failure event.

The median number of days from the start of TB treatment to outcome was shorter amongst those diagnosed with TB in the ART Clinic and in later calendar periods across all failure events.

Overall, median CD4 counts were lowest for patients who died pre-ART, followed by those who initiated ART and were highest for those who were permanently deferred for ART. The median days from the start of TB treatment to death were similar to the median number of days from the start of TB treatment to ART initiation and were much shorter than the median number of days to permanent deferral or the end of TB treatment. The median number of days from the start of TB treatment to the end of TB treatment

Table 4.9: Summary statistics by clinic in which TB was diagnosed and calendar period of TB diagnosis

|                                                                     | CD4 count (cells/ $\mu$ L)<br>near start of TB treatment (trt) |        |           |                      | Days from start of TB<br>treatment (trt) to failure |        |            |                      |
|---------------------------------------------------------------------|----------------------------------------------------------------|--------|-----------|----------------------|-----------------------------------------------------|--------|------------|----------------------|
|                                                                     | N                                                              | (%)    | Median    | (IQR) <sup>(1)</sup> | N                                                   | (%)    | Median     | (IQR) <sup>(2)</sup> |
| <b>Clinic in which TB diagnosed</b>                                 |                                                                |        |           |                      |                                                     |        |            |                      |
| Percentages are out of all patients in shown failure event category |                                                                |        |           |                      |                                                     |        |            |                      |
| <b>ART Clinic</b>                                                   |                                                                |        |           |                      |                                                     |        |            |                      |
| Death                                                               | 5                                                              | (33%)  | 27        | (11-41)              | 7                                                   | (25%)  | 42         | (16-55)              |
| ART initiation                                                      | 164                                                            | (35%)  | 75        | (28-136)             | 176                                                 | (26%)  | 39         | (21-69)              |
| Deferral                                                            | 5                                                              | (31%)  | 152       | (102-177)            | 5                                                   | (15%)  | 98         | (83-147)             |
| End of TB trt                                                       | 22                                                             | (39%)  | 124       | (59-203)             | 23                                                  | (15%)  | 168        | (49-216)             |
| All                                                                 | 196                                                            | (35%)  | 80        | (34-141)             | 211                                                 | (24%)  | 43         | (21-83)              |
| <b>TB Clinic</b>                                                    |                                                                |        |           |                      |                                                     |        |            |                      |
| Death                                                               | 10                                                             | (67%)  | 38        | (19-76)              | 21                                                  | (75%)  | 117        | (71-176)             |
| ART initiation                                                      | 310                                                            | (65%)  | 74        | (33-141)             | 498                                                 | (74%)  | 103        | (58-144)             |
| Deferral                                                            | 11                                                             | (69%)  | 107       | (90-164)             | 29                                                  | (85%)  | 146        | (110-169)            |
| End of TB trt                                                       | 35                                                             | (61%)  | 92        | (36-193)             | 134                                                 | (85%)  | 180        | (167-214)            |
| All                                                                 | 366                                                            | (65%)  | 76        | (33-143)             | 682                                                 | (76%)  | 121        | (69-168)             |
| <b>Calendar period of TB diagnosis</b>                              |                                                                |        |           |                      |                                                     |        |            |                      |
| Percentages are out of all patients in shown failure event category |                                                                |        |           |                      |                                                     |        |            |                      |
| <b>2002-2005</b>                                                    |                                                                |        |           |                      |                                                     |        |            |                      |
| Death                                                               | 5                                                              | (33%)  | 27        | (11-122)             | 12                                                  | (43%)  | 108        | (44-177)             |
| ART initiation                                                      | 171                                                            | (36%)  | 71        | (24-132)             | 305                                                 | (45%)  | 110        | (61-153)             |
| Deferral                                                            | 4                                                              | (25%)  | 130       | (99-217)             | 12                                                  | (35%)  | 143        | (94-205)             |
| End of TB trt                                                       | 24                                                             | (42%)  | 101       | (45-203)             | 83                                                  | (53%)  | 184        | (167-217)            |
| All                                                                 | 204                                                            | (36%)  | 76        | (26-144)             | 412                                                 | (46%)  | 126        | (69-176)             |
| <b>2006-2008</b>                                                    |                                                                |        |           |                      |                                                     |        |            |                      |
| Death                                                               | 10                                                             | (67%)  | 38        | (19-68)              | 16                                                  | (57%)  | 74         | (57-135)             |
| ART initiation                                                      | 303                                                            | (64%)  | 75        | (37-140)             | 369                                                 | (55%)  | 64         | (38-106)             |
| Deferral                                                            | 12                                                             | (75%)  | 119       | (88-171)             | 22                                                  | (65%)  | 142        | (110-162)            |
| End of TB trt                                                       | 33                                                             | (58%)  | 118       | (48-193)             | 74                                                  | (47%)  | 171        | (149-213)            |
| All outcomes                                                        | 358                                                            | (64%)  | 77        | (38-141)             | 481                                                 | (54%)  | 77         | (43-139)             |
| <b>All patients</b>                                                 |                                                                |        |           |                      |                                                     |        |            |                      |
| Percentages are out of all patients for all failure events          |                                                                |        |           |                      |                                                     |        |            |                      |
| Death                                                               | 15                                                             | (3%)   | 37        | (11-76)              | 28                                                  | (3%)   | 75         | (50-161)             |
| ART initiation                                                      | 474                                                            | (84%)  | 74        | (31-140)             | 674                                                 | (75%)  | 83         | (43-132)             |
| Deferral                                                            | 16                                                             | (3%)   | 120       | (94-171)             | 34                                                  | (4%)   | 143        | (105-165)            |
| End of TB trt                                                       | 57                                                             | (10%)  | 112       | (48-199)             | 157                                                 | (18%)  | 177        | (162-214)            |
| All                                                                 | 562                                                            | (100%) | <b>77</b> | <b>(33-143)</b>      | 893                                                 | (100%) | <b>100</b> | <b>(50-160)</b>      |

<sup>(1)</sup> The median (IQR) CD4 counts were calculated based on 562 (62% of) patients who had a CD4 count available near the time of starting TB treatment (within the interval of 6 months prior to or 3 months after starting TB treatment and not more than 7 days after initiating ART).

<sup>(2)</sup> The median (IQR) times to failure event were calculated on all 893 patients.

was longer than the median number of days from the start of TB treatment to any of the other failure events across all stratifications and overall.

#### 4.4.2 Competing Risks Regressions

Models of the following form were fit:

$$\log \left[ \frac{\bar{\lambda}_{ki}(t)}{\bar{\lambda}_0(t)} \right] = \beta_0 + \beta_3 X_{3i} + \beta_4 X_{4i} + \dots + \beta_9 X_{9i} \quad (4.5)$$

where  $\bar{\lambda}_{ki}(t)$  is the sub-distribution hazard of patient  $i$  for cause  $k$  at time  $t$  (defined in 3.11),  $\bar{\lambda}_0(t)$  is the baseline sub-distribution hazard (which is common for all causes  $k$  and all individuals  $i$ ) and  $X_{3i} \dots X_{9i}$  are as in (4.1), and  $X_{1i}$  and  $X_{2i}$  that compared clinic sites 1 and to to 3 are not applicable since only patients from clinic site 1 are included in the analysis.

These will be compared to cause-specific hazard models of the form (4.1) that treat the competing events as censored observations.

#### Time to ART initiation

Table 4.10 shows the (sub)hazard ratios ( $e^{\hat{\beta}_j}$ ) and their corresponding 95% confidence intervals (CI's) assuming proportionality on the sub-distribution hazard (Model 14) or on the cause-specific hazard (Model 17) using all 893 patients from Site 1 (whether or not they initiated ART).

The log-log plots shown in Figure 4.13 suggest that the assumption of proportional hazards is not too badly violated for Model 17 (although there is some crossing of lines). I nevertheless considered the inclusion of time-varying covariates.

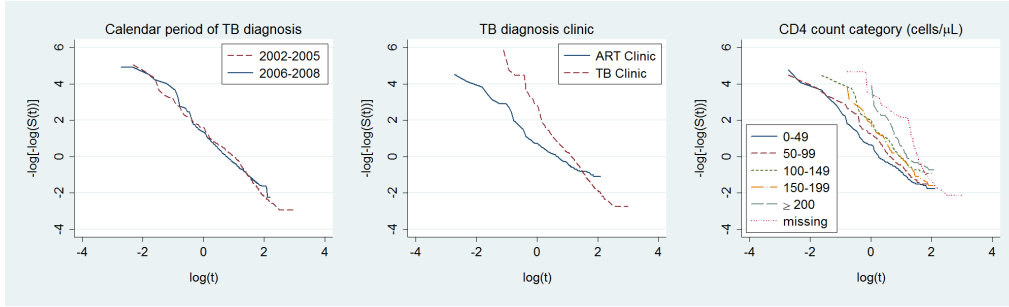


Figure 4.13: Comparison of log-log survival curves by covariate  
Estimates adjust to the baseline values of the other variables included in the Model

Using  $f(t) = t$  as the function of time by which to multiply the covariates to model the time-varying effects (Models 15 and 18), the effects at baseline ( $e^{\hat{\beta}_j^*}$ ) were a lot larger (hazard ratios further from one) than the corresponding effects averaged over all time ( $e^{\hat{\beta}_j}$  from Models 14 and 17) and the hazard ratios corresponding to the time-varying covariates ( $e^{\hat{\beta}_j^{**}}$ ) were all greater than one. This implies that the estimated effects of the covariates is greater at the baseline time than averaged over all time and that these effect decrease beyond the baseline time.

Using  $f(t) = \log(t)$  as the function of time by which to multiply the covariates to model the time-varying effects (Models 16 and 19), the estimated effects at baseline ( $e^{\hat{\beta}_j^*}$ ) were larger (further from one) than those for the effects averaged over all analysis time, but smaller than the estimated baseline effects using  $f(t) = t$  as the function of time. The estimated (sub)hazard ratios corresponding to the time-varying covariates ( $e^{\hat{\beta}_j^{**}}$ ) were all greater than one (and were greater than those corresponding to the function of time  $f(t) = t$ ).

All the models show that patients diagnosed in earlier years, those diagnosed with TB in the TB clinic and those with higher CD4 counts (or not having a CD4 count near the time of starting TB treatment) had lower (sub)hazards of ART initiation. For the models that do not include time-varying effects, the hazard ratios are slightly lower in the cause-specific hazards regression (Model 17) than the corresponding sub-distribution hazard ratios from the regression on the sub-distribution hazards (Model 14).

Table 4.10: Competing Risks Regressions, Models 14-19

| Models for proportional hazards regression on the sub-distribution hazards |                         |             |                       |                         |                          |             |                         |             |                          |                |
|----------------------------------------------------------------------------|-------------------------|-------------|-----------------------|-------------------------|--------------------------|-------------|-------------------------|-------------|--------------------------|----------------|
| Characteristic                                                             | Model 14 <sup>(1)</sup> |             |                       | Model 15 <sup>(2)</sup> |                          |             | Model 16 <sup>(3)</sup> |             |                          | 95% CI         |
|                                                                            | $e^{\hat{\beta}_j}$     | 95% CI      | $e^{\hat{\beta}_j^*}$ | 95% CI                  | $e^{\hat{\beta}_j^{**}}$ | 95% CI      | $e^{\hat{\beta}_j^*}$   | 95% CI      | $e^{\hat{\beta}_j^{**}}$ |                |
| TB diagnosed before 2006                                                   | 0.92                    | 0.77 - 1.10 | 0.48                  | 0.35 - 0.67             | 1.23                     | 1.12 - 1.34 | 0.64                    | 0.49 - 0.83 | 1.48                     | 1.18 - 1.86    |
| Referred from TB Clinic                                                    | 0.70                    | 0.55 - 0.90 | 0.17                  | 0.11 - 0.26             | 1.83                     | 1.51 - 2.21 | 0.30                    | 0.23 - 0.39 | 3.78                     | 2.76 - 5.17    |
| CD4 count                                                                  |                         |             |                       |                         |                          |             |                         |             |                          |                |
| 0-49 (baseline)                                                            | 1.00                    | -           | 1.00                  | -                       | 1.00                     | -           | 1.00                    | -           | 1.00                     | -              |
| 50-99                                                                      | 0.98                    | 0.72 - 1.33 | 0.55                  | 0.34 - 0.87             | 1.29                     | 1.05 - 1.58 | 0.76                    | 0.57 - 1.02 | 1.49                     | 1.04 - 2.14    |
| 100-149                                                                    | 0.73                    | 0.52 - 1.02 | 0.33                  | 0.20 - 0.55             | 1.43                     | 1.15 - 1.77 | 0.49                    | 0.34 - 0.69 | 2.03                     | 1.37 - 3.00    |
| 150-199                                                                    | 0.74                    | 0.53 - 1.04 | 0.19                  | 0.10 - 0.37             | 1.74                     | 1.38 - 2.20 | 0.36                    | 0.22 - 0.58 | 2.78                     | 1.72 - 4.50    |
| $\geq 200$                                                                 | 0.45                    | 0.30 - 0.67 | 0.13                  | 0.06 - 0.25             | 1.74                     | 1.36 - 2.22 | 0.19                    | 0.11 - 0.34 | 3.43                     | 2.01 - 5.86    |
| missing                                                                    | 0.29                    | 0.23 - 0.37 | 0.00                  | 0.00 - 0.01             | 3.79                     | 2.75 - 5.22 | 0.00                    | 0.00 - 0.01 | 58.19                    | 15.80 - 214.30 |

| Models for proportional hazards regression on the cause-specific hazards |                         |             |                       |                         |                          |             |                         |             |                          |               |
|--------------------------------------------------------------------------|-------------------------|-------------|-----------------------|-------------------------|--------------------------|-------------|-------------------------|-------------|--------------------------|---------------|
| Characteristic                                                           | Model 17 <sup>(1)</sup> |             |                       | Model 18 <sup>(2)</sup> |                          |             | Model 19 <sup>(3)</sup> |             |                          | 95% CI        |
|                                                                          | $e^{\hat{\beta}_j}$     | 95% CI      | $e^{\hat{\beta}_j^*}$ | 95% CI                  | $e^{\hat{\beta}_j^{**}}$ | 95% CI      | $e^{\hat{\beta}_j^*}$   | 95% CI      | $e^{\hat{\beta}_j^{**}}$ |               |
| TB diagnosed before 2006                                                 | 0.87                    | 0.74 - 1.01 | 0.55                  | 0.40 - 0.75             | 1.15                     | 1.05 - 1.26 | 0.66                    | 0.51 - 0.84 | 1.32                     | 1.07 - 1.63   |
| Referred from TB Clinic                                                  | 0.63                    | 0.52 - 0.76 | 0.17                  | 0.11 - 0.25             | 1.76                     | 1.49 - 2.08 | 0.28                    | 0.22 - 0.37 | 3.56                     | 2.60 - 4.88   |
| CD4 count                                                                |                         |             |                       |                         |                          |             |                         |             |                          |               |
| 0-49 (baseline)                                                          | 1.00                    | -           | 1.00                  | -                       | 1.00                     | -           | 1.00                    | -           | 1.00                     | -             |
| 50-99                                                                    | 0.88                    | 0.69 - 1.11 | 0.53                  | 0.34 - 0.84             | 1.26                     | 1.03 - 1.54 | 0.72                    | 0.54 - 0.96 | 1.46                     | 1.05 - 2.05   |
| 100-149                                                                  | 0.59                    | 0.45 - 0.77 | 0.33                  | 0.19 - 0.56             | 1.37                     | 1.10 - 1.70 | 0.45                    | 0.32 - 0.65 | 1.88                     | 1.27 - 2.80   |
| 150-199                                                                  | 0.56                    | 0.42 - 0.76 | 0.22                  | 0.12 - 0.42             | 1.49                     | 1.18 - 1.88 | 0.35                    | 0.22 - 0.55 | 2.13                     | 1.34 - 3.39   |
| $\geq 200$                                                               | 0.38                    | 0.27 - 0.54 | 0.16                  | 0.08 - 0.34             | 1.52                     | 1.18 - 1.96 | 0.21                    | 0.11 - 0.40 | 2.65                     | 1.47 - 4.77   |
| missing                                                                  | 0.19                    | 0.15 - 0.23 | 0.00                  | 0.00 - 0.01             | 2.89                     | 2.24 - 3.75 | 0.00                    | 0.00 - 0.01 | 22.61                    | 10.31 - 49.58 |

<sup>(1)</sup> Models 14 and 17 assume that the effect of the predictor variables are constant over time.<sup>(2)</sup> Models 15 and 18 include interactions of the predictor variables with the time to ART initiation.<sup>(3)</sup> Models 16 and 19 include interactions of the predictor variables with the natural logarithm of time.  $e^{\hat{\beta}_j}$ ,  $e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 4.3, with hazard ratios replaced by sub-hazard ratios for Models 14-16.

### Time to death pre-ART

Although the main interest in this section is time to ART initiation, time to death pre-ART is also of interest. It may be helpful for physicians or policy-makers to know if there are certain patient or programme characteristics that are associated with a greater hazard of death pre-ART.

I have chosen to allow different variables to be included in the time to death model. Although predictors affecting the time to death may be expected to influence whether a patient is fast-tracked for ART, it may not be that exactly the same set of predictors is appropriate for both models. In particular, the calendar period of TB diagnosis does appear to influence the time to ART initiation, but does not appear to be associated with time to death. This could be because policy changes have affected the delay between the start of TB treatment and ART initiation (but have not influenced the time to death pre-ART for patients who are possibly very sick and are thus unable to initiate ART pre-death). The predictors related to the WHO clinical stage of a patient and whether they were being treated for a new TB infection or being retreated for TB did not appear to be strongly associated with the time to ART initiation (perhaps because the CD4 count of a patient influences whether they are fast-tracked for ART more strongly than stage and whether they are being retreated for TB or not), but were associated with the time to death pre-ART.

I also chose to categorize CD4 counts into those above versus below 50 cells/ $\mu$ L and missing for the death pre-ART models, since few patients who died pre-ART had very high CD4 counts and the small number of patients dying pre-ART did not support having many CD4 count categories (whereas for the time to ART initiation models, there are more patients and a greater range of CD4 counts and thus having more CD4 count categories seemed appropriate).

Table 4.11 summarizes the results from the multiple regressions on the sub-distribution hazards and cause-specific hazards.

Table 4.11: Comparison of regressions on sub-distribution hazards (SDH) and cause-specific hazards (CSH) for time to death pre-ART

| Characteristic          | Proportional hazards regression on |             |                       |             |
|-------------------------|------------------------------------|-------------|-----------------------|-------------|
|                         | sub-distribution hazard            |             | cause-specific hazard |             |
|                         | SHR                                | 95% CI      | HR                    | 95% CI      |
| CD4 count               |                                    |             |                       |             |
| 0-49                    | 0.26                               | 0.09 - 0.77 | 0.19                  | 0.07 - 0.57 |
| $\geq 50$ (baseline)    | 1.00                               | -           | 1.00                  | -           |
| missing                 | 0.89                               | 0.34 - 2.32 | 0.28                  | 0.11 - 0.72 |
| WHO Stage III of HIV    | 0.50                               | 0.24 - 1.05 | 0.53                  | 0.25 - 1.13 |
| Referred from TB Clinic | 0.73                               | 0.26 - 2.06 | 0.40                  | 0.15 - 1.08 |
| Retreatment for TB      | 0.51                               | 0.19 - 1.36 | 0.34                  | 0.12 - 0.93 |

SHR, SubHazard Ratio; CI, Confidence Interval; HR, Hazard Ratio

In both analyses, patients who had lower or missing CD4 counts (near the time of starting TB treatment), those who were categorized as being in WHO stage III or below at the time of enrolment for ART, those diagnosed with TB in the TB clinic (before enrolment for ART) and those who were being retreated for TB had lower (sub)hazards of death pre-ART. The effects were larger and more statistically significant in the regression on the cause-specific hazards for all covariates except WHO clinical stage of HIV disease. This is because the risk set for the sub-distribution hazard includes patients who have already failed from a competing event and thus is larger than the risk set for the cause-specific hazard, which excludes such patients. This means that the proportion of patients failing out of those in the risk set would be smaller, resulting in less of an effect when calculating the sub-distribution hazard compared to the cause-specific hazard. The assumption of proportional hazards is reasonable (see Figure 4.14).

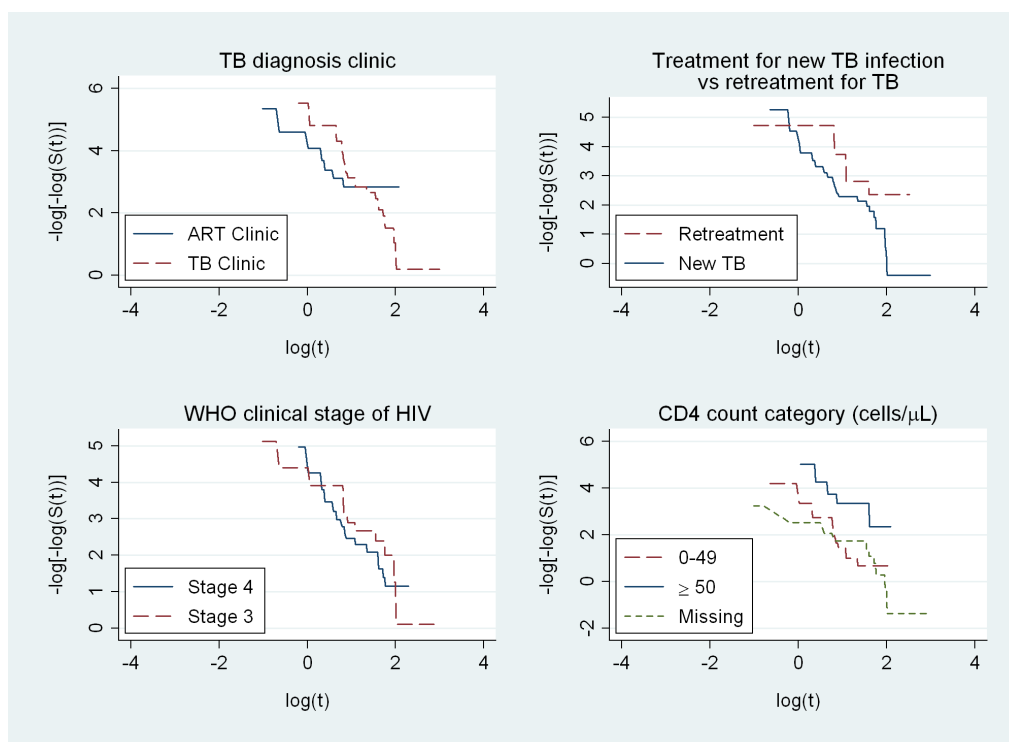


Figure 4.14: Log-log survival plots for time to death pre-ART with ART initiation, permanent deferral for ART and end of TB treatment as censoring events,  $N=893$

### Combined analysis of time to ART initiation and time to death pre-ART

Table 4.12 shows a comparison of the estimated covariate effects obtained when assuming that patients who experience a competing event are censored (cause-specific hazards regression) versus keeping them in the risk set (regression on sub-distribution hazards). The sub-distribution hazard of death is reduced by 26% for patients diagnosed with TB in earlier versus later years. This means that for the group of patients diagnosed with TB in the earlier years, a smaller proportion of patients die (compared to the group diagnosed with TB in later years) and thus a greater proportion remain at risk for ART initiation. The estimated reduction in ART initiation in earlier years compared to later years may thus be underestimated in the regression on the sub-distribution hazards. This is because patients who died would have been considered to still be in the risk set, thus contributing to the denominator, but not the numerator (since they cannot actually initiate ART). This effect is smaller (the hazard ratio is closer to one) than the corresponding effect es-

timated in the cause-specific hazards regression (which may overestimate the reduction by censoring patients who died, thus removing them from the risk set, which is equivalent to assuming that they are represented by patients who remain in the risk set and may thus still initiate ART in the future).

Table 4.12: Comparison of regressions on sub-distribution hazards and cause-specific hazards for ART initiation and death pre-ART

**Regression on sub-distribution hazards**

| Characteristic                      | ART Initiation      |             | Death pre-ART       |             |
|-------------------------------------|---------------------|-------------|---------------------|-------------|
|                                     | $e^{\hat{\beta}_j}$ | 95% CI      | $e^{\hat{\beta}_j}$ | 95% CI      |
| Calendar period of TB diagnosis     |                     |             |                     |             |
| 2006-2008 (baseline)                | 1.00                | -           | 1.00                | -           |
| 2002-2005                           | 0.92                | 0.77 - 1.09 | 0.74                | 0.34 - 1.59 |
| Clinic in which TB diagnosed        |                     |             |                     |             |
| ART Clinic (baseline)               | 1.00                | -           | 1.00                | -           |
| TB Clinic                           | 0.72                | 0.57 - 0.93 | 0.71                | 0.26 - 1.98 |
| CD4 count category (cells/ $\mu$ L) |                     |             |                     |             |
| $\leq 50$ (baseline)                | 1.00                | -           | 1.00                | -           |
| $> 50$                              | 0.76                | 0.59 - 0.98 | 0.26                | 0.09 - 0.78 |
| missing                             | 0.29                | 0.23 - 0.37 | 0.92                | 0.36 - 2.35 |

**Regression on cause-specific hazards**

| Characteristic                      | ART Initiation      |             | Death pre-ART       |             |
|-------------------------------------|---------------------|-------------|---------------------|-------------|
|                                     | $e^{\hat{\beta}_j}$ | 95% CI      | $e^{\hat{\beta}_j}$ | 95% CI      |
| Calendar period of TB diagnosis     |                     |             |                     |             |
| 2006-2008 (baseline)                | 1.00                | -           | 1.00                | -           |
| 2002-2005                           | 0.87                | 0.75 - 1.02 | 0.65                | 0.30 - 1.41 |
| Clinic in which TB diagnosed        |                     |             |                     |             |
| ART Clinic (baseline)               | 1.00                | -           | 1.00                | -           |
| TB Clinic                           | 0.64                | 0.53 - 0.77 | 0.43                | 0.16 - 1.12 |
| CD4 count category (cells/ $\mu$ L) |                     |             |                     |             |
| $\leq 50$ (baseline)                | 1.00                | -           | 1.00                | -           |
| $> 50$                              | 0.63                | 0.52 - 0.76 | 0.18                | 0.06 - 0.54 |
| missing                             | 0.19                | 0.15 - 0.24 | 0.29                | 0.11 - 0.74 |

Similar reasoning can be given for each covariate, using either ART initiation or death pre-ART as the event of interest and thus explaining why the effects for the regression on the cause-specific hazards are greater (hazard ratios further from one) than the corresponding effects for the regression on the sub-distribution function.

Figure 4.15 shows smoothed hazard curves for each of the causes by covariate.

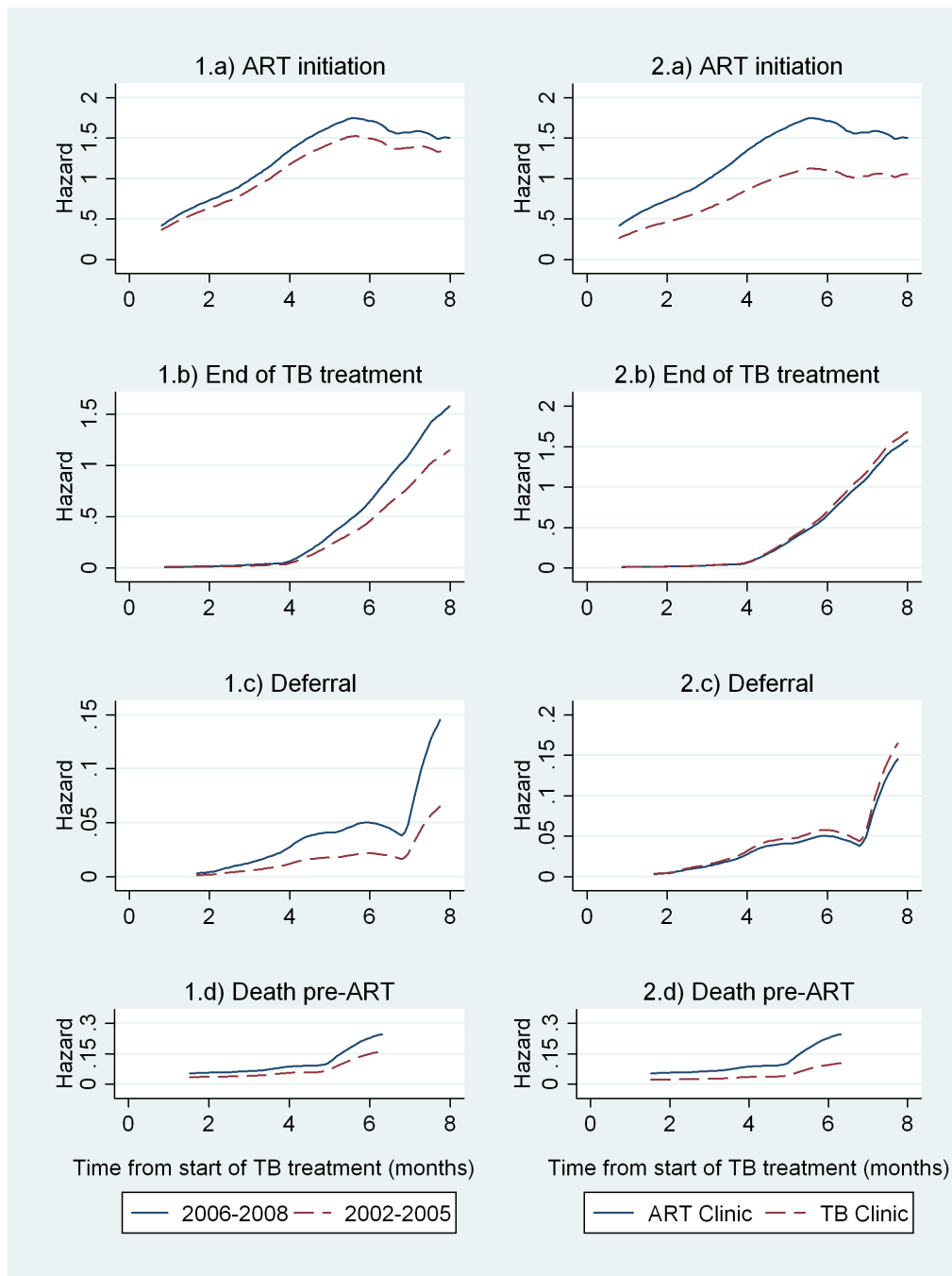


Figure 4.15: Smoothed cause-specific hazard curves by calendar period (1) and clinic (2) of TB diagnosis for each competing risk

These curves are calculated using estimates from multiple regressions on the cause-specific hazards and use TB diagnosis in 2006 or later, TB diagnosis in an ART clinic and CD4 count of lower than 50 cells/ $\mu$ L as the baseline covariate values.

The cause-specific hazards of ART initiation, end of TB treatment, deferral and death were all greater for patients diagnosed with TB in 2006 or later. The relative difference in the cause-specific hazards for patients diagnosed in later versus earlier years was particularly large for the cause-specific hazard of deferral.

For patients diagnosed with TB in the ART clinic, the cause-specific hazards of ART initiation and death pre-ART were greater (than for those diagnosed with TB in the TB clinic), but the cause-specific hazards of the end of TB treatment and deferral did not appear to differ by the clinic in which TB was diagnosed (or were slightly lower for patients diagnosed with TB in the ART clinic).

Figure 4.16 shows a comparison of cumulative incidence curves (similar to that presented by StataCorp (2009b), but using my data) for the outcomes of ART initiation (a) and death pre-ART (b) by calendar period of ART initiation (1) and clinic of TB diagnosis (2). The curves shown correspond to univariate analyses. The solid lines for the plots numbered 1.a) and 1.b) correspond to patients diagnosed with TB in 2006 or later, while the dashed lines correspond to patients diagnosed with TB before 2006. For the plots labelled 2.a) and 2.b) the solid lines correspond to patients diagnosed with TB in the ART clinic (as part of the screening process before ART initiation), while the dashed and dotted lines correspond to patients diagnosed with TB in the TB clinic (before enrolment for ART).

The black lines correspond to a cumulative incidence function estimated by stratifying the patients by group and therefore not making any assumption about proportionality of effects for covariates on either cause-specific hazards or sub-distribution hazards. Examining these curves suggests that the cumulative incidence functions for ART initiation and those for death pre-ART are not proportional for patients diagnosed with TB in earlier versus later years, nor for those diagnosed in the ART versus TB clinic.

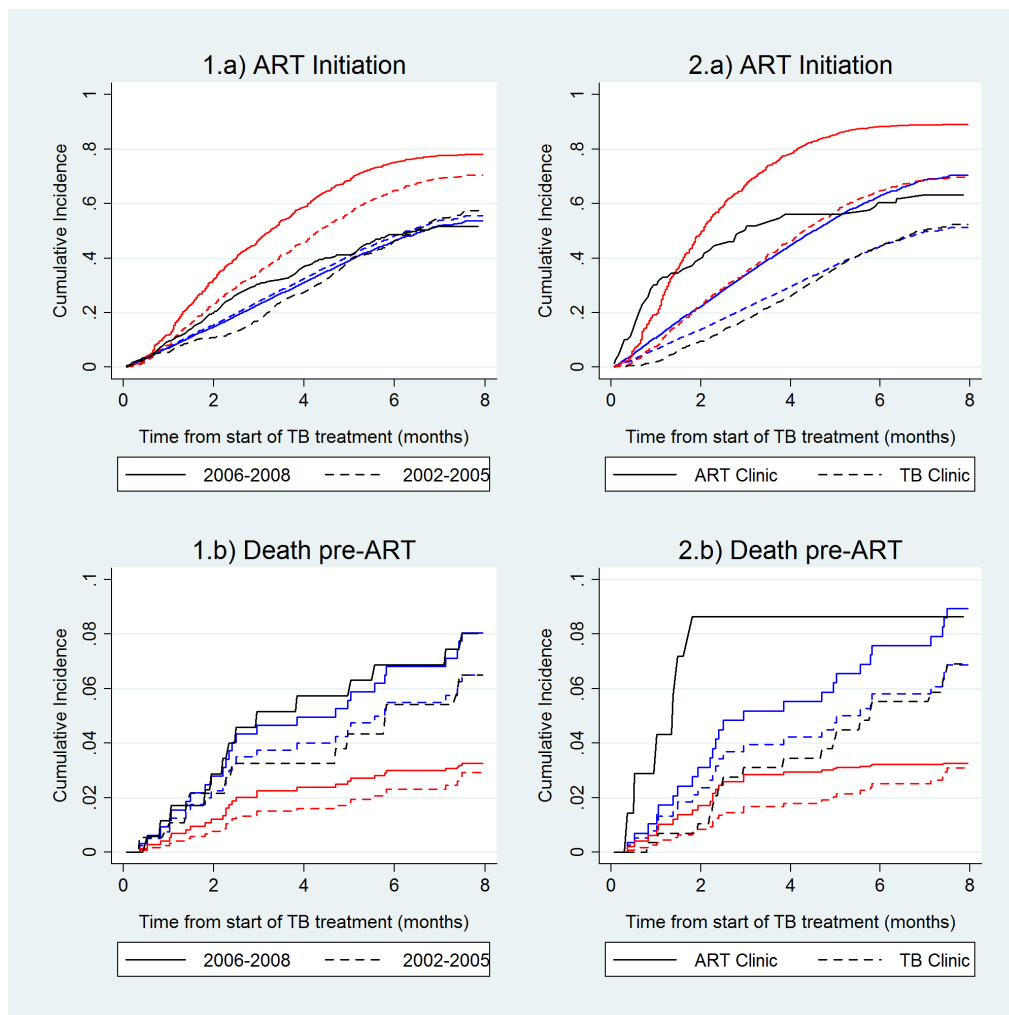


Figure 4.16: Comparison of cumulative incidence functions for ART initiation (a) and death pre-ART (b) by calendar period (1) and clinic (2) of TB diagnosis under different assumptions (i-iii):

- i) No assumptions about effect of covariates
- ii) Effect of covariates proportional for cause-specific hazards of all events
- iii) Effect of covariates proportional on sub-distribution hazard of event of interest

Comparisons adjusted for my data from those of StataCorp (2009b) pages 207-211 and 226-227 and Putter et al. (2007).

The red lines correspond to cumulative incidence functions that are estimated under the assumption that the effects of covariates are proportional for the cause-specific hazards of all events. The cumulative incidence curves estimated in this way were higher for the event of ART initiation, but lower for the event of death pre-ART.

The blue lines correspond to cumulative incidence functions estimated under the assumption that the effect of the covariates on the sub-distribution hazard of the event of interest is proportional. The cumulative incidence curves estimated in this way were generally between the curves estimated under no proportionality assumptions (the curve corresponding to the value of the covariate for which the incidence of the event of interest was higher (solid lines) was slightly lower for the curve calculated assuming proportionality of covariate effects on the sub-distribution hazard (blue curve) than the corresponding solid curve estimated without using proportionality assumptions (black curve), while for the curves corresponding to the value of the covariate for which the cumulative incidence of the event of interest was lower (dashed lines) this ordering was reversed).

In conclusion, these graphs indicate that the assumption of proportional hazards is not satisfied. Models such as Model 15, 16, 18 and 19 in Table 4.10, which allow for interactions of the predictor variables with a function of time, are thus to be preferred over models which assume constant effects of the predictor variables over time (such as Models 14 and 17).

### 4.4.3 Model checking for competing risks regressions

I have chosen to show results from model checking only for the time to ART analysis (and not for the time to death analysis), as this was the main analysis of interest. For the regressions on the sub-distribution hazards, Cox-Snell residuals and deviance residuals could not be obtained and likelihood displacement values were also not easily obtainable using Stata/IC 11.

#### Assessment of model fit for cause-specific hazard models

The plots of the Cox-Snell residuals against the estimated cumulative hazard functions (see Figure 4.17) suggest adequate model fit for models 18 and 19 (which included time-varying covariates), but suggest poor fit for model 17 (for which the proportional hazards assumption might not hold).

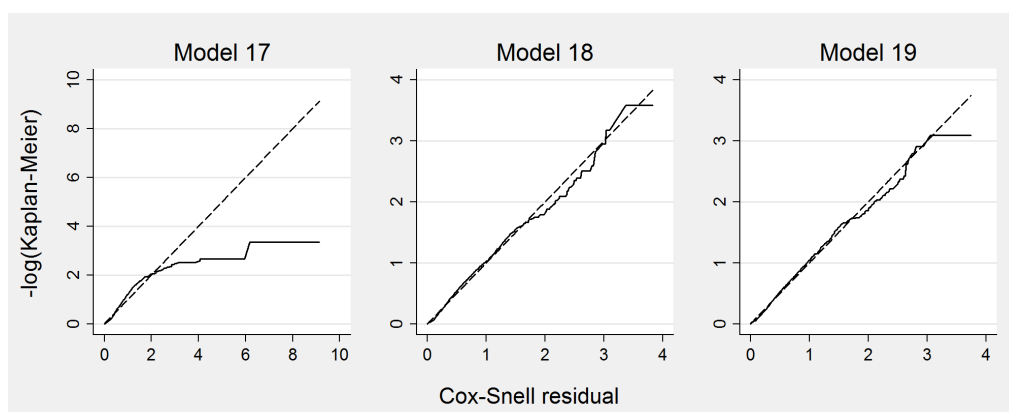


Figure 4.17: Cox-Snell residual plots for Models 17 to 19

#### Examination of outliers for cause-specific hazard models

The plots of the deviance residuals against the analysis time show a downward trend (see Figure 4.18). This indicates that the predicted times to ART were often longer than the observed times for patients who were observed to initiate ART soon after the start of TB treatment (large positive residuals at early times), but shorter than the observed times for patients who were observed to initiate ART long after the start of TB treatment (large negative residuals at later times).

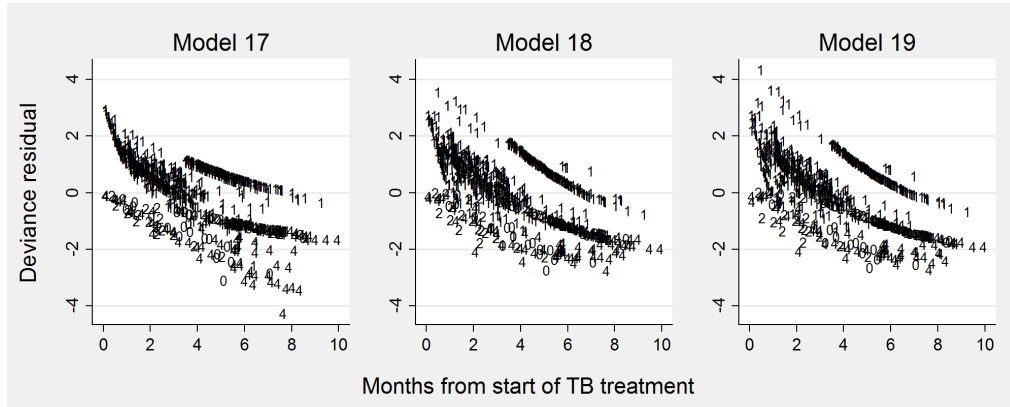


Figure 4.18: Deviance residual plots for Models 17 to 19

The largest negative deviance residual corresponded to the same patient in each of models 17 to 19. This patient was diagnosed with TB in 2006 or later as part of the screening process for ART and had a CD4 count lower than 50 cells/ $\mu$ L, but did not initiate ART before completing TB treatment. The patient with the largest positive residual for model 17 was also diagnosed with TB in 2006 or later as part of the screening process for ART, but had a CD4 count of 50-100 cells/ $\mu$ L and started ART two days after starting TB treatment. The patient with the largest positive deviance residual for both models 18 and 19 was diagnosed with TB before 2006 as part of the screening process for ART initiation and did not have a CD4 count recorded near the time of starting TB treatment, but started ART two weeks after starting TB treatment.

### Sensitivity of model estimates to influential patients

For Models 14 to 16, based on the sub-distribution hazard, I examined plots of the delta-betas for each covariate against the survival time and identified relatively large values. I then compared the results of excluding patients who had relatively large delta-betas for one of the covariates from the model to the original model for each of the three models.

Figure 4.19 shows plots of the delta-betas against the rank of survival time for the covariates for which the largest delta-beta values were obtained for each of Models 14 to 16.

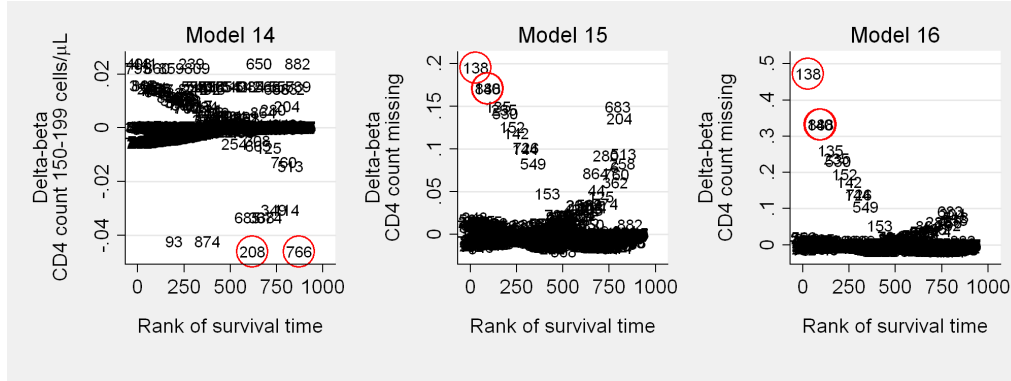


Figure 4.19: Selected Delta-betas for Models 17 to 19

For Model 14, there were few patients who had large delta-beta values for any of the covariates. The largest two absolute values of delta-betas were for those of the CD4 count category 150-199 cells/ $\mu$ L (both were -0.046). These patients both had CD4 counts in the range 150-199 cells/ $\mu$ L and were diagnosed with TB in 2006 or later as part of the screening process for ART initiation. One was permanently deferred for ART 147 days after TB diagnosis and the other ended TB treatment without initiating ART 244 days after starting TB treatment, so they were at the extreme end of the range for follow-up time. Excluding these two patients resulted in very small decreases in the sub-distribution hazard ratio of ART initiation for patients diagnosed with TB before 2006 and those diagnosed with TB before enrolling for ART and a smaller effect (sub-hazard ratio closer to one) for patients having CD4 counts in the category 150-199 cells/ $\mu$ L.

The largest three delta-beta values for Models 15 (0.17, 0.17 and 0.20) and 16 (0.33, 0.33 and 0.47) correspond to the same three patients and are all for the coefficient associated with the missing CD4 count category. These patients were all diagnosed with TB before 2006 as part of the screening process for ART initiation (in the ART clinic), had no recorded CD4 count near the time of TB diagnosis and started ART within one month of starting TB treatment (one within two weeks), so were at the short end of the follow-up range. Excluding these patients from the models resulted in small changes on the coefficient scale, except for the baseline and time-varying coefficients corresponding to a missing CD4 count. The baseline coefficients became smaller (more negative, corresponding to lower sub-distribution hazard ratios) and the time-varying coefficients became greater (corresponding to sub-distribution hazard ratios greater than one), indicating a low sub-hazard of ART initiation at baseline for patients not having a measured CD4

count that quickly increased over time.

For Models 17 to 19, based on the cause-specific hazard, I used the likelihood displacement value rather than delta-beta values to identify influential observations. Although these different measures do not provide exactly the same information, the exclusion of patients with large likelihood displacement values from their respective models should give a reasonable idea of the sensitivity of the estimated parameters from the models to influential observations. Use of the likelihood displacement was much simpler, since it requires examining only one plot per model to determine influential observations.

Plots of the likelihood displacement values for each patient against the rank of the survival times are shown for Models 17 to 19 in Figure 4.20.

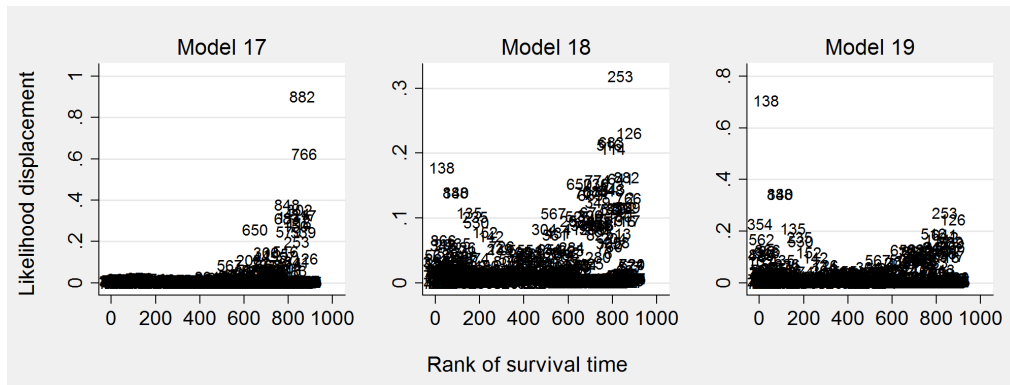


Figure 4.20: Likelihood displacement value plots for Models 17 to 19

To assess the impact of patients who had large likelihood displacement values, I refitted the models excluding these patients.

The patient who had the highest likelihood displacement value for Model 17 was diagnosed with TB in 2006 or later as part of the screening process for ART and had a CD4 count  $< 50$  cells/ $\mu$ L, but did not initiate ART before the end of TB treatment. Excluding this patient resulted in lower hazard ratios of ART initiation for each of the variables in the model compared to the baseline values.

The patient who had the highest likelihood displacement value for Model 18 was diagnosed with TB before 2006 in a TB clinic (before enrolling for ART), had a CD4 count of 50-99 cells/ $\mu$ L and was permanently deferred for

ART seven months after starting TB treatment. The changes to parameter estimates for this model were small, with a slight decrease (increase) in the baseline (time-varying) coefficient for a CD4 count less than 50 cells/ $\mu$ L and increase (decrease) in the baseline (time-varying) coefficients for CD4 counts greater than 200 or missing.

The patient who had the highest likelihood displacement value for Model 19 is the same patient who had the highest delta-beta values for Models 15 and 16. This patient was diagnosed with TB before 2006 as part of the screening process for ART initiation and did not have a CD4 count near the time of starting TB treatment, but initiated ART two weeks after starting TB treatment. The main changes when refitting this model were a smaller (more negative) baseline coefficient and a larger time-varying coefficient for patients not having a CD4 count.

To summarize, the regressions on the cause-specific hazards that allowed time-varying coefficients (Models 18 and 19) showed reasonable fit using the Cox-Snell residual plots. The plots of the deviance residuals against analysis time for Models 17, 18 and 19 showed downward trends indicating over-prediction of the time to ART at earlier times and under-prediction at later times. Excluding patients who had large values for delta-betas or likelihood displacement values from the models for which these were calculated did cause some changes in effect estimates, but the overall message of the results remained consistent.

## Chapter 5

# Conclusion for Time to ART Analysis

I estimated the effects of predictors of ART initiation using a wide range of models, some simple and others more complex. The same pattern of results was found for all models:

- Patients diagnosed with TB in later years had a shorter times to ART initiation
- Patients diagnosed with TB in the TB clinic (i.e. before enrolment for HIV care) had longer times to ART initiation
- Patients who had lower CD4 counts near the time of starting TB treatment had shorter times to ART initiation; patients who did not have a CD4 counts near the time of starting TB treatment had longer times to ART initiation
- Patients enrolled at clinic site 2 had longer times to ART initiation and those enrolled at clinic site 3 had shorter times to ART initiation than those enrolled at clinic site 1.

Allowing for time-varying-coefficients in the Cox regression analysis using patients from all 3 clinics showed (not surprisingly) that the effect of these baseline factors are greatest at baseline and have less of an impact as time since start of TB treatment increases.

### Parametric Analysis

Hazard ratios for the Cox proportional hazards model (not allowing time-varying coefficients) and the Weibull model in the proportional hazards framework that did not allow the ancillary parameter to vary by covariate were very similar. For the AFT models that did not allow ancillary parameters to vary by covariate, the decelerators for the different models were comparable, but in general they were smaller and closer to one for the Weibull model and larger and further from one for the log-Normal and log-logistic models than for the generalized Gamma model. Comparison of the estimated effects from AFT models that allow for ancillary parameters to vary by covariate is not straight forward, since the shapes of the hazard functions for those models differ.

The generalized Gamma model allowing for ancillary parameters to vary by covariate appeared to have the best model fit. For this model, the shortest predicted mean time to ART initiation was 30 days (95% CI: 26 to 34 days) and the longest mean time to ART initiation was 231 days (95% CI: 214 to 247). Interpreting the effects of models that allow varying ancillary parameters requires caution, since these change the shape of the hazard and survival functions. Graphical analysis is thus helpful in addition to examination of estimated hazard ratios and values for ancillary parameters and predictions of mean time to event. If the goal is exploratory analysis, however, non-parametric graphical analysis should perhaps be used instead (smoothing functions if necessary), since it does not make any assumptions about the form of the model.

### Competing Risks Analysis

In the competing risks analysis, I found (as expected) that for the time to ART analysis, for which the cause-specific hazard of interest (ART initiation while on TB treatment) is large and the cause-specific hazards of the competing risks (death pre-ART, permanent deferral and end of TB treatment) are relatively small, the adjusted sub-hazard ratios (which are calculated using the sub-distribution hazard for which subjects who experience a competing event are not removed from the risk set) were similar to (though estimating a slightly smaller effect than) the adjusted hazard ratios (calculated using the cause-specific hazards for which subjects that experience a competing event are removed from the risk set).

In the time to death analysis, however, for which the cause-specific hazard of the event of interest (death pre-ART) is small and the cause-specific hazards of the competing risks (particularly ART initiation and also end of TB treatment at later times) are relatively large, far fewer adjusted sub-hazard ratios were significantly associated with the time to death than adjusted cause-specific hazard ratios.

Although it can be informative to compare the estimates obtained for regressions on the cause-specific versus sub-distribution hazards, the appropriate model should be chosen based on the analysis question of interest. The cause-specific hazard models should be used if interest is on the relative risk of an event around a specific time, since the risk set for calculating the cause-specific hazard includes only patients who have not experienced any failure event, such that the cause-specific hazard ratio is the relative risk of the event of interest ignoring the competing risks. The sub-distribution hazard models should be used if interest is on the relative risk of an event up to a given time, since the risk set for calculating the sub-distribution hazard includes all subjects who have not yet experienced the event of interest (even if they have experienced another event and will never experience the event of interest), such that the sub-distribution hazard ratio estimates the relative risk of the event of interest being experienced first.

The most important aspect of the competing risks analysis is the graphical analysis, which allows easy comparison of the magnitude of the different competing risks and their cumulative incidences, facilitating explicit focus on and understanding of the interplay between the different competing risks and the factors that affect them. I presented graphs of the cause-specific hazards stratified by covariates for each of the competing events (Figure 4.15), which showed that diagnosis of TB in the ART clinic was associated with a much higher hazard of ART initiation (than diagnosis of TB in the TB clinic), while diagnosis of TB in 2006 or later (compared to before 2006) was associated with a much higher hazard of deferral. Both these factors were associated with a higher hazard of death pre-ART. A comparison of cumulative incidence curves estimated under different proportionality assumptions and stratified by covariate for both time to ART initiation and time to death (Figure 4.16) showed that the assumptions of proportionality did not always hold, but that the assumption of covariate effect being proportional on the sub-distribution hazard for the event of interest often gave a closer approximation to the non-parametric estimate (which made no proportionality assumptions) than the assumption that the cause-specific hazards of all events were proportional.

### Proportional Hazards Assumption

The proportional hazards assumption did not hold for many of the models fitted in this analysis. For this reason, parametric models such as the Weibull and Generalized Gamma were considered as an alternative to the Cox proportional hazards model. These models allow different underlying hazard functions for different strata of the predictor variables through the parameterization of ancillary parameters. For competing risks analysis, however, the methodology has been developed around simpler semi-parametric models that do assume proportional hazards. One way of overcoming this assumption is including interactions of a function of time with the predictor variables. An alternative may be to consider the development of an extension to the Fine and Gray methodology that would allow the use of non-proportional models.

### Potential Sources of Bias

For the analysis of the time to ART initiation using all three clinics, only patients who did initiate ART were included. These estimates should therefore be understood as descriptive for this subset of patients and not be used to predict the time to ART initiation for a patient who has just started TB treatment, since they do not take into account patients who do not initiate ART (within the study period).

The expected time from enrolment for HIV care until starting ART was longer for patients enrolled for HIV care at clinic site 2 compared to sites 1 and 3, since the date of enrolment for HIV care corresponded to the first time the patient visited the clinic for site 2 (which may have been a while before the patient became eligible for ART and started the screening process for ART initiation), whereas patients from sites 1 and 3 were only recorded as enrolling for HIV care when they started the screening process for ART initiation. This may have affected the estimates of the effect that being diagnosed with TB in the ART (versus TB clinic) had on the time to ART initiation (the effect for patients from sites 1 and 3 only would have been greater).

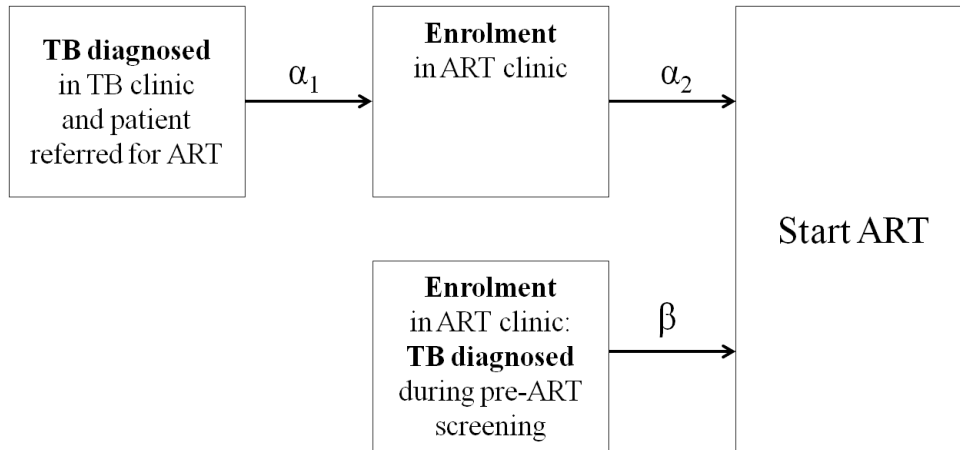
The timing of CD4 count measurements also differed between clinics. Patients from site 2 generally had CD4 counts taken closer to the time of starting TB treatment and thus fewer patients from this site had no CD4 count taken near the time of starting TB treatment. This may have been because

a greater proportion of patients were enrolled for HIV care before being diagnosed with TB, as well as integration of the ART and TB clinics, which began in 2004 (Garone et al., 2011), but was only considered as complete in 2009 (Kerschberger et al., 2012). I chose to use CD4 counts as a categorical variable and include a ‘missing’ category for patients who had no recorded CD4 count within the interval of 183 days before or 91 days after the start of TB treatment and not more than 7 days after initiating ART. Results were not very sensitive to excluding patients who had no CD4 count near the time of starting TB treatment. I explore the use of multiple imputation rather than using missing categories for the Outcome Analysis in Section 7.5.

This section presented models for the overall delay between the start of TB treatment and ART initiation. In an earlier analysis of data from clinic site 1, we focused attention on the component delays from the start of TB treatment to ART initiation (Lawn et al., 2011b).

### Component delays from start of TB treatment to ART initiation

Figure 5.1 shows the component delays. We used simple descriptive statistics and regressions on sub-distribution hazards to explore these component delays.



Conceptual diagram showing the component time periods ( $\alpha_1$ ,  $\alpha_2$  and  $\beta$ ) between tuberculosis (TB) diagnosis and starting antiretroviral therapy (ART) for patients with TB diagnosed either in a TB clinic or in the ART clinic (published in Lawn et al. (2011b)).

Figure 5.1: Component delays to ART initiation

We found that for patients diagnosed with TB in the TB clinic, the time

from TB diagnosis to enrolment in the ART clinic ( $\alpha_1$ ) was almost twice as long as the time from enrolment in the ART clinic to starting ART ( $\alpha_2$ ), but that this latter time was similar to the time from the start of TB treatment to ART initiation ( $\beta$ ) for patients diagnosed with TB during pre-ART screening. The time from the start of TB treatment to ART initiation was therefore almost three times longer for patients diagnosed in the TB clinic than for patients diagnosed with TB as part of pre-ART screening.

We found that delays between TB diagnosis and enrolment in the ART clinic ( $\alpha_1$ ) were shorter in more recent calendar periods and for patients who had smear-positive TB (probably as a result of changes in treatment guidelines), while delays between enrolment in the ART clinic (for patients already on TB treatment) and ART initiation ( $\alpha_2$ ) were shorter for patients who had lower CD4 counts. For patients who were diagnosed with TB as part of the pre-ART screening process, those who had lower CD4 counts, were diagnosed with TB in earlier years and had smear-positive pulmonary TB had shorter times to ART initiation.

# Part II

## Outcome Analysis

# Chapter 6

## Causal Models Methodology

### 6.1 Introduction

In this section, I seek to analyse the causal effect of ART initiation (and its timing) on the survival of HIV infected patients who are on treatment for TB.

It is important to understand that just because two variables are associated with each other does not imply that one causes the other. For example, if all people with TB wear a mask and people that have TB have a higher mortality rate than the general population, then I might observe that wearing a mask is associated with higher mortality. This does not imply, however, that wearing a mask causes death.

Another important concept is that of *counterfactual outcomes*. The key idea is that although I cannot know the outcome of a specific patient had that patient received a treatment other than the treatment that they actually received, if I have groups of similar patients who receive different treatments, then I can compare the outcomes of those patients and so estimate the causal effect of the treatment. This will be explained in more detail in Section 6.4.1.

When analyses are conducted on observational data, rather than using ideal randomized controlled trials, there is the potential for biased estimation of causal effects. One potential source of bias is *confounding*. An example of a situation in which there is confounding is when patients with a particular value for a variable are more likely both to experience a specific treatment and the outcome of interest. For example, patients with a very low CD4

count are more likely both to start ART sooner after the start of TB treatment and to die. If I ignored the CD4 counts of patients, I might come to the conclusion that starting ART sooner after the start of TB treatment is harmful to patients. When treatment is only assigned at one time, this bias can be overcome by adjusting for the confounders (variables that cause confounding, such as CD4 count in the example above) using stratification.

When the treatment received by a patient may change over time and the value of a variable that is predictive of the outcome may change over time and its current value predicts future treatment, then that variable is known as a *time-dependent confounder*. If past treatment values also affect future values of the time-dependent confounder, then attempting to estimate the effect of the treatment on the outcome using traditional methods such as stratification is likely to provide biased results (Hernán et al., 2000).

Another potential bias is *selection bias*, which occurs when patients who are analysed are not representative of all patients for whom inferences are being made. Patient drop-out can result in selection bias.

*Marginal structural models* (MSMs, which are regression models weighted using inverse probability of treatment and censoring weights) can be used to eliminate these biases and provide valid causal estimates under assumptions that will be introduced and explained in Section 6.4.

In order to illustrate the concepts of association, causation, confounding and selection bias more clearly, it is helpful to introduce *directed acyclic graphs* (DAGs). DAGs are a useful tool to demonstrate and better understand the causal structure of a hypothesis or scenario and to make explicit the assumptions that are to be used in analysis.

## 6.2 Directed Acyclic Graphs (DAGs)

This section is a summarised version of that presented by Hernán and Robins (2012) adapted to my own analysis as far as possible.

A directed acyclic graph DAG is made up of nodes (representing variables) and arrows (representing the relationships between variables). These arrows are directed. A directed graph is acyclic if it is not possible to start at one variable and trace a route following the direction of the arrows that returns

to that variable. An example of a simple DAG is as follows:

$$\text{CD4}_0 \longrightarrow \text{ART} \longrightarrow \text{Death} \quad (6.1)$$

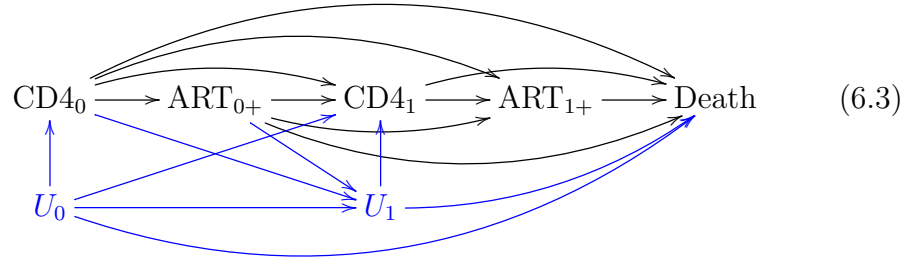
This DAG indicates that patients' baseline CD4 counts ( $\text{CD4}_0$ ) affect whether they receive the treatment (ART) and whether they die (within the study period). Whether patients receive the treatment (ART) also affects whether or not they are observed to die by the end of the study period.

If patients who do not initiate ART shortly after the baseline time may have a subsequent CD4 count taken (at say time 1) and may then initiate ART, a more complex DAG can be constructed as follows:

$$\text{CD4}_0 \longrightarrow \text{ART}_{0+} \longrightarrow \text{CD4}_1 \longrightarrow \text{ART}_{1+} \longrightarrow \text{Death} \quad (6.2)$$

This DAG indicates that patients' baseline and updated CD4 counts affect future ART, future CD4 and the eventual outcome (death). In addition, ART soon after baseline and at later times affects future CD4, future ART and the eventual outcome.

Let  $U$  denote the underlying health status of a patient, which cannot be measured directly, but can be estimated using variables such as CD4 count, weight or haemoglobin, which are indicators of (or proxies for) how healthy a patient is (or is not). Including the time-updated variables  $U_t$  in the above DAG results in DAG 6.3 (assuming that the underlying health statuses of patients affect their current and future CD4 counts and whether they die, but not whether they initiate ART). For simplicity, I show only CD4 count as a potential proxy for a patient's underlying health status and as a predictor of ART initiation and death (other proxies such as weight and haemoglobin level could also have been included). The resulting DAG has the same structure (but different variable labels) to Figure 1b of Robins et al. (2000).



This DAG indicates (in addition to the comments made for DAG 6.2; corresponding additional nodes and arrows shown in blue) that the underlying health statuses of patients at time  $U_0$  affect their baseline and future CD4 counts, the future underlying health statuses of the patients and whether they die within the study period. The baseline CD4 counts and whether they initiate ART just after baseline also affect the future underlying health statuses of patients, which in turn affects the future CD4 counts of patients and whether they die within the study period.

### 6.2.1 DAG concepts and terminology

An arrow pointing from a variable  $A$  to a variable  $B$ , implies that  $A$  *causes*  $B$  and  $B$  is an *effect* of  $A$ .  $A$  is then referred to as the *parent* of  $B$  and  $B$  as the *child* of  $A$ . So, for example, in DAG 6.3, the parent  $CD4_0$  causes the child  $ART_{0+}$ . If a variable  $C$  can be reached by following the direction of the arrows from a variable  $A$ , then  $A$  is said to be an *ancestor* of  $C$  and  $C$  is a *descendant* of  $A$ .  $CD4_0$  is an ancestor of  $ART_{0+}$ ,  $U_1$ ,  $CD4_1$ ,  $ART_{1+}$  and  $Death$ .  $CD4_1$  is a descendant of  $U_0$ ,  $CD4_0$ ,  $ART_{0+}$  and  $U_1$ .

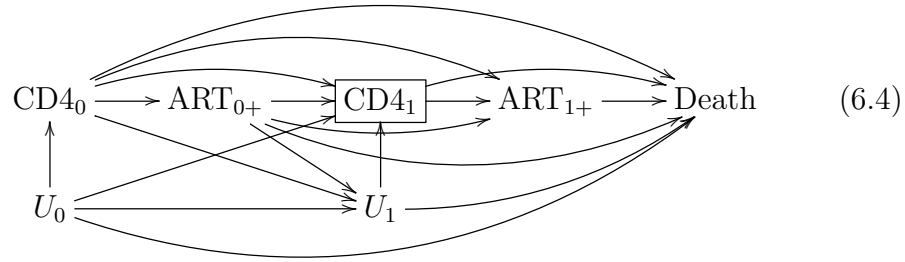
A DAG is said to be a *causal DAG* if all common causes of any pair of variables is included in the DAG. So, for example, if DAG 6.3 is a true representation of the underlying data, then DAG 6.2 is not a causal DAG, since (for example)  $U_0$  is a common cause of both the baseline and subsequent CD4 counts and is not included in DAG 6.2.

The lack of an arrow from a variable  $A$  to a variable  $B$  represents the assumption that  $A$  does not cause  $B$ . So, for example, I have not drawn an arrow from  $U_0$  to  $ART_{0+}$  since I am making the assumption that patients are not initiated on ART based **directly** on their unmeasurable underlying health status. I do nonetheless assume that the unmeasurable underlying health status ( $U_0$ ) affects the initiation of ART ( $ART_{0+}$ ) **indirectly**, through its effect on the measurable CD4 count ( $CD4_0$ ), since the latter can be used by clinicians for decision-making.

A *path* is formed by a series of nodes connected by arrows (for example,  $CD4_0 \rightarrow ART_{0+} \rightarrow CD4_1 \rightarrow ART_{1+} \rightarrow \text{Death}$ ).

A *collider* is a variable that is a common effect of two other variables in the DAG, such as the CD4 counts of patients at time 1 ( $CD4_1$ ) in DAG 6.3, which is a common effect of (amongst other things) whether the patients are on ART before time 1 ( $ART_{0+}$ ) and their underlying health status at time 1 ( $U_1$ ).

If patients are only observed conditional on the value of a certain variable, this can be represented by drawing a box around that variable. So, for example, if I only have information on patients who had CD4 counts lower than 200 cells/ $\mu\text{L}$  at time 1, then the corresponding DAG would be represented as follows:



The variable  $CD4_1$  is said to be *conditioned on* in DAG 6.4. Conditioning on a variable can also occur through stratification or adjustment for that variable (such as including it as a covariate in a regression equation).

A path is said to be *open* if there are no colliders along the path that have not been conditioned on (or have conditioned on descendants) and no non-colliders along the path that have been conditioned on. A path that is not open is *blocked*. So, for example, in DAG 6.4, the path

$$ART_{0+} \rightarrow CD4_1 \rightarrow ART_{1+} \rightarrow \text{Death}$$

is blocked (since  $CD4_1$  is not a collider in this path and is conditioned on), whereas in DAG 6.3 this path is open. The path

$$ART_{0+} \rightarrow CD4_1 \leftarrow U_0$$

is open in DAG 6.4 (since  $CD4_1$  is a collider in this path and has been conditioned on), but is not open in DAG 6.3. The path

$$ART_{0+} \rightarrow U_1 \leftarrow CD4_0$$

is open in DAG 6.4 (since  $CD4_1$  is a descendant of the collider  $U_1$  in this path and has been conditioned on), but is not open in DAG 6.3.

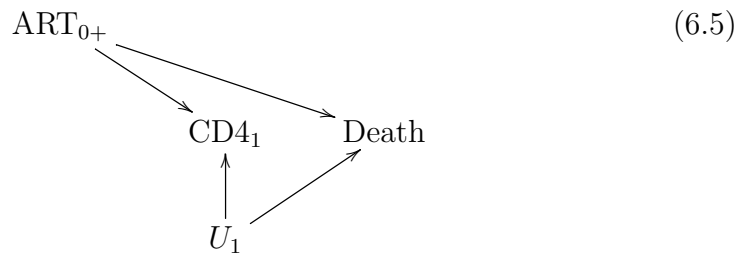
Two variables are said to be *associated* if there is an open path between them (not necessarily in the direction of the arrows). If there is an open path between two variables that does follow the direction of the arrows, then there is *causation*. I am particularly interested in showing whether the use of ART has a causal effect on the time to death.

When there is a variable that is a common cause of both treatment and outcome (such as  $U_0$ ,  $CD4_0$ ,  $U_1$  and  $CD4_1$ ) then there is said to be a *backdoor* path. If such a backdoor path cannot be blocked, it will be a source of *confounding bias*. Confounding bias is the bias that results in the estimated effect of treatment on outcome incorporating a non-causal associational effect of treatment on outcome (rather than only the causal effect of treatment on outcome).

## 6.3 Biases

### 6.3.1 Confounding and Confounders

Traditionally, a *confounder* has been described as a variable that “is associated with the treatment and with the outcome conditional on the treatment (or in the untreated) and is not along the causal pathway between treatment and outcome” (Hernán and Robins, 2012). Standard practise has been to adjust for such variables through stratification. In many situations, this is the appropriate course of action. In some situations, however, adjusting for a variable defined in this way can introduce bias (rather than remove bias). DAG 6.5 shown below does not show all associations that would be expected between variables based on clinical knowledge, but has purposefully been made simple in order to demonstrate the concept of inducing selection bias by adjusting for a variable that is not a confounder.



In DAG 6.5,  $CD4_1$  is associated with  $ART_{0+}$  and with Death conditional on  $ART_{0+}$  (through the path  $CD4_1 \leftarrow U_1 \rightarrow \text{Death}$ ) and is not along the causal pathway between  $ART_{0+}$  and Death.  $CD4_1$  would thus traditionally be called a confounder (assuming DAG 6.5 was the correct causal DAG for the analysis of interest). There is no bias for the effect of  $ART_{0+}$  on death in this DAG, however, as the only association between  $ART_{0+}$  and Death is the causal association of  $ART_{0+}$  on Death. The path

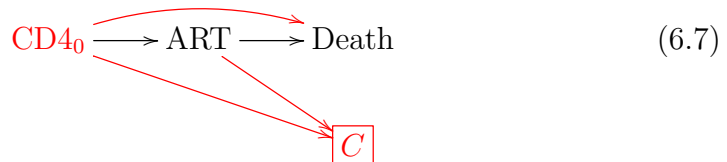
$$ART_{0+} \rightarrow CD4_1 \leftarrow U_1 \rightarrow \text{Death} \quad (6.6)$$

is blocked by  $CD4_1$ , which is a common cause of  $ART_{0+}$  and  $U_1$ . If, however, I adjust or stratify for  $CD4_1$ , then I open the previously blocked pathway (6.6), introducing a non-causal association between ART and Death and thus introducing bias. Hernán and Robins (2012) therefore recommend defining a confounder as a variable that causes confounding bias, rather than using the traditional definition given above.

When treatment and covariates change over time, a time-varying covariate is a time-dependent confounder if post-baseline values of the time-varying covariate predict both future treatment and outcome (Robins and Hernán, 2009).

### 6.3.2 Selection Bias

In addition to confounding bias, there may also be selection bias. When patients who are included (or remain) in a study are not representative of the population for which inferences are being made (they are in some way different to patients who are not included in or do not remain in the study), then there is said to be selection bias. If a variable that is common effect of the treatment and outcome (or variables with which they are associated) has been conditioned on (or adjusted for), then there will be selection bias.



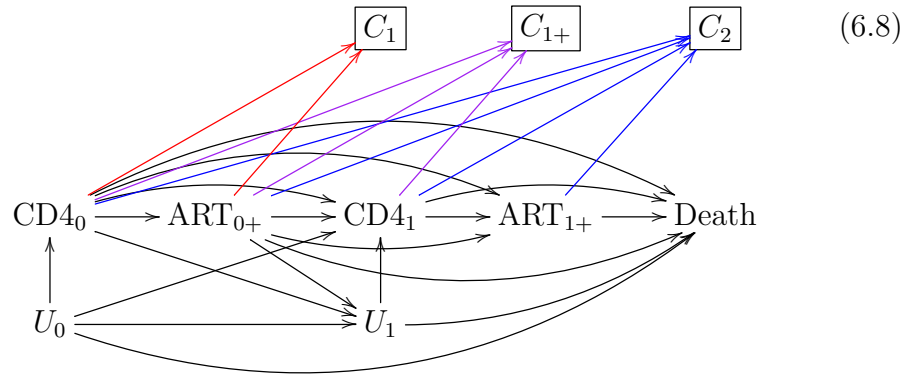
In DAG 6.7,  $C$  represents censoring. Since I only have information on the outcome for patients who are not censored,  $C$  is drawn with a box around it

to indicate that it is conditioned on. It is easy to see that there is selection bias in the population represented by this DAG, since  $C$  is conditioned on and is a common effect of a patient's CD4 count (associated with the outcome, death) and ART (the treatment). The non-causal path of association

$$\text{ART} \rightarrow C \leftarrow \text{CD4}_0 \rightarrow \text{Death}$$

between ART and Death (shown in red in DAG 6.7) is open when there is censoring. For example, if patients who have a high (or low) CD4 count are more likely to be censored, then the association between ART and death among patients who are not censored is likely to be different to the association between ART and death for patients who are censored.

DAG 6.8 shows selection bias in the context of a time-varying treatment and time-dependent confounder, where censoring may occur at multiple times in the study. The additional paths of association opened by censoring occurring between the assignment of ART just after baseline and the CD4 count measured at time 1 [ $C_1$ ], or between the CD4 count measured at time 1 and the assignment of treatment just after time 1 [ $C_{1+}$ ], or after the assignment of treatment occurring just after time 1 [ $C_2$ ] are shown in red, purple and blue respectively.



### 6.3.3 Bias relating to time-varying confounding

Robins et al. (2000) state that when there is a time dependent confounder for which the future values are affected by past treatment history, then the standard approach for measuring the effect of the time-varying treatment on the outcome (modelling the probability of the outcome as a function of the past treatment and past confounder histories using techniques such as stratification, regression or matching) may be biased. I will explain this using DAG 6.3.

In order to estimate the joint causal effect of  $ART_{0+}$  and  $ART_{1+}$  on Death, I need to adjust for the confounding effect of  $CD4_1$  on  $ART_{1+}$ . If I don't, then there is an association between  $ART_{1+}$  and Death that is not a causal effect (through the pathway  $ART_{1+} \leftarrow CD4_1 \rightarrow \text{Death}$ ). For example, among patients who did not initiate ART just after baseline, those who have lower CD4 counts at time 1 are more likely both to initiate ART just after time 1 and to die sooner (than patients with higher CD4 counts at time 1). Note: Since I assume that patients remain on ART once they have initiated it, there is no direct causal effect of  $CD4_1$  on  $ART_{1+}$  for patients who initiated ART just after baseline (and hence no open backdoor path through  $CD4_1$ ).

If I adjust for  $CD4_1$  by stratification, regression or matching, however, then I open up the paths of association

$$\begin{aligned} ART_{0+} &\rightarrow CD4_1 \leftarrow U_0 \rightarrow \text{Death} \\ ART_{0+} &\rightarrow CD4_1 \leftarrow U_0 \rightarrow U_1 \rightarrow \text{Death and} \\ ART_{0+} &\rightarrow CD4_1 \leftarrow U_1 \rightarrow \text{Death} \end{aligned}$$

between  $ART_{0+}$  and Death that were previously blocked by the collider  $CD4_1$ . Selection bias is induced, since  $CD4_1$  is a common effect of the treatment ( $ART_{0+}$ ) and variables associated with the outcome ( $U_0$  and  $U_1$ ). Even if there were no causal effect of ART on death,  $ART_{0+}$  and Death would still be associated if I conditioned on  $CD4_1$ . For example, within the strata of patients who have high CD4 counts, knowing that a patient did not initiate ART just after baseline would make it more likely that the patient had a better prognosis as indicated by the value of the unmeasurable variable  $U_0$ , which in turn would make it less likely that they died within the study period.

In addition, if I adjust for  $CD4_1$  using stratification, regression or matching, I would also block the indirect causal path  $ART_{0+} \rightarrow CD4_1 \rightarrow \text{Death}$ . The overall effect of ART initiation on death would therefore be underestimated, since the effect of ART initiation just after baseline on death that is mediated through the change in CD4 count would not be included. When adjusting for confounding, I therefore need to avoid adjusting for any variable that is intermediate on the pathway between the treatment and outcome (Hernán et al., 2002).

I will thus obtain bias in effect estimates if I do not adjust for  $CD4_1$  and if I do adjust for it using methods such as regression, stratification or matching. Alternative methods therefore need to be used to adjust for the updated

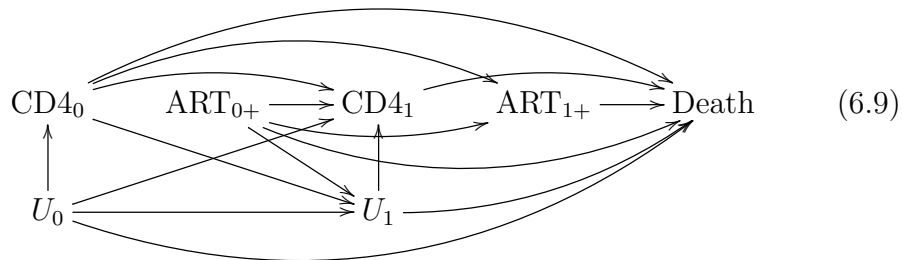
measures of time-dependent confounders (such as CD4 count) that will not induce selection bias by creating an association between the baseline treatment and the outcome, nor eliminate the component of the causal effect that is mediated by the time-updated confounders.

## 6.4 Marginal Structural Models

One method for appropriately estimating the effect of a time-varying treatment (in the presence of time-dependent confounding and selection bias) is the use of marginal structural models. These models are so called because causal models are often referred to as structural models and these models aim to estimate the causal effect of treatment on outcome (a marginal quantity) rather than the associational effect (where the outcome is conditional on the treatment received) (Hernán et al., 2000; Hernán and Robins, 2012). The method of fitting marginal structural models involves two steps:

1. Calculate the probability of a patient receiving the treatment (and censoring) history he/she actually received (weighting models are usually used to estimate weights based on the probabilities of treatment and censoring)
2. Perform a weighted regression analysis using the weights calculated in Step 1 (this weighted regression model is the marginal structural model).

The effect of weighting the study population using inverse probability of **treatment** weights (IPTWs) is that a pseudo-population is created in which covariate values no longer predict treatment. This can be shown in a DAG by removing the arrows that point from the covariate(s) into the treatment(s). For example, if I assume that DAG 6.3 is the appropriate DAG for my population, then using inverse probability of treatment weights would result in a population for which the DAG could be given by DAG 6.9 in which the arrows from covariates to treatments have been removed.



Selection bias can be adjusted for using inverse probability of **censoring** weights (IPCWs). The effect of weighting by IPCWs for a DAG is that the pseudo-population created can be represented by a DAG for which the arrows from the treatment and measured covariate(s) into the conditioned variable that is a common effect of treatment and outcome (or variables associated with them) (variables  $C$  and  $C_1, C_{1+}, C_2$  in DAG 6.7 and 6.8 respectively) can be removed. In the case of DAG 6.7 and 6.8, these variables are no longer connected to any other variables in the DAG and can thus be removed, resulting in DAG 6.1 and 6.3 shown previously.

The IPTWs can then be multiplied with the IPCWs resulting in combined inverse probability of treatment and censoring weights (IPTCWs) that can be used to weight the study population in order to create a pseudo-population in which past covariate history does not predict either treatment or censoring and past treatment history does not predict censoring.

For valid causal inference from a marginal structural model, a number of assumptions are required (Hernán and Robins, 2012):

1. Conditional exchangeability
2. Positivity
3. Well-defined interventions (or consistency)
4. No model misspecification
5. No measurement error

Assumptions 1-3 are known as the *identifiability* conditions.

Before defining and explaining these conditions, I introduce some notation and the concept of counterfactual outcomes. I will use the convention that uppercase letters denote random variables and lowercase letters denote realizations of these random variables. Let  $A_i(t)$  denote the treatment received by patient  $i$  at time  $t$  and  $\bar{A}_i(t) = A_i(0), A_i(1), \dots, A_i(t)$  the treatment history of patient  $i$  up to time  $t$  and  $\bar{A}_i$  the treatment history of patient  $i$  for all time.  $A_i(t) = 1$  if the patient receives treatment (say ART) at time  $t$  and 0 otherwise. I assume that random vectors for each patient are drawn independently from a distribution common to all patients such that the  $i$  subscript denoting an individual can be suppressed (Fewell et al., 2001).

### 6.4.1 Counterfactual Outcomes

Typically, the outcome (such as survival time  $T$ ) of a patient can only be observed under one treatment regime  $\bar{a}$  (for example initiating ART at time 1 and remaining on it thereafter,  $\bar{a} = 0, 1, 1, 1, \dots$ ). Outcomes under other treatment regimes (such as initiating ART at time 2,  $\bar{a}' = 0, 0, 1, 1, \dots$ ) cannot be observed for the same patient and are thus known as counterfactual (contrary to the fact) outcomes. Counterfactual outcomes will be denoted as  $T^{\bar{a}}$ . For ease of reference, the actual outcome of the patient, say  $T^{\bar{a}}$  is also termed a counterfactual outcome. We can therefore say that if a patient received treatment regime  $\bar{a}$  then the observed outcome  $T$  is equivalent to the counterfactual outcome  $T^{\bar{a}}$ . If  $\bar{a} \neq \bar{a}'$ , then the observed outcome  $T$  for a patient who received regime  $\bar{a}$  is not expected to be equal to the counterfactual outcome  $T^{\bar{a}'}$ . Although we cannot know the outcome that an individual would have had, had they received a different treatment regime, if we can assume that individuals in a population are similar, then we can compare the average outcomes under different treatment regimes and this can allow us to determine the causal effects of treatments in populations (under the assumptions explained below).

### 6.4.2 Explanations and definitions of assumptions required for valid causal inference

*Conditional exchangeability* means that conditional on measured covariates, the average observed outcome had treated patients not been treated is the same as the average outcome among untreated patients (and vice versa). This assumption implies no *unmeasured confounding* and can be encoded on a DAG by omitting any arrows from unmeasured variables to the treatment variables as in DAG 6.3. If there had been unmeasured confounding, DAG 6.3 would have included arrows from  $U_0$  to  $\text{ART}_{0+}$  and from  $U_0$  and  $U_{0+}$  to  $\text{ART}_{1+}$ . The generalized definition of conditional exchangeability (Robins and Hernán, 2009) can be expressed mathematically as

$$T^{\bar{a}} \perp\!\!\!\perp A(t) | \bar{A}(t-1) = \bar{a}(t), \bar{L}(t) = \bar{l}(t) \text{ for all regimes } \bar{a} \text{ and all } \bar{l}(t),$$

where

$\bar{L}(t)$  denotes the history of time-updated variables CD4 count, haemoglobin level and weight value through time  $t$  (baseline values of other variables such as gender, age and WHO clinical stage of HIV disease are also included in  $L(0)$ )

$T^{\bar{a}}$  denotes the survival time of patients who received ART according to the regime  $\bar{a}$  and

the symbol  $\perp$  is used to denote independence.

In words, the counterfactual outcome for a patient under the regime  $\bar{a}$  is independent of the treatment received by that patient at time  $t$  conditional on the treatment regime received by that patient up to time  $t - 1$  and the covariate history of that patient up to time  $t$ .

In ideal randomized controlled trials, (conditional) exchangeability of patients is assumed to hold if there is no informative censoring, since treatment is assigned randomly (or randomly within prespecified subsets of the population defined by measured covariates) and not based on the prognosis of the patient or any other factors that may affect outcome other than through the effect of the treatment. In observational studies, the lack of randomization means that it cannot be assumed that patients who are treated and patients who are untreated are exchangeable. Nonetheless, assuming no unmeasured confounding, conditional exchangeability can be obtained in a pseudo-population created by weighting the original population using inverse probability of treatment weights (see Section 6.4.3).

*Positivity* is said to hold when for every treatment and covariate combination of interest there is a non-zero probability of patients with a specific set of covariates receiving each level of the treatment. The generalized definition of positivity (Robins and Hernán, 2009) can be expressed mathematically as follows:

$$\begin{aligned} &\text{If } f_{\bar{A}(t-1), \bar{L}(t)}\{\bar{a}(t-1), \bar{l}(t)\} \neq 0, \\ &\text{then } f_{A(t)|\bar{A}(t-1), \bar{L}(t)}\{a(t)|\bar{a}(t-1), \bar{l}(t)\} > 0 \text{ for all } a(t). \end{aligned}$$

This condition states that if there is a non-zero probability of a patient having received treatment regime  $\bar{a}(t-1)$  up to time  $t-1$  and having covariate history  $\bar{l}(t)$  up to time  $t$ , then the probability of the patient receiving treatment  $a(t)$  at time  $t$  conditional on having received regime  $\bar{a}(t-1)$  up to time  $t-1$  and having covariate history  $\bar{l}(t)$  up to time  $t$  must be greater than zero for all possible treatment values  $a(t)$  at time  $t$ .

Note that it is possible that this positivity condition does hold in an observational study even if there is no patient observed to have treatment  $a(t)$  and past treatment and covariate histories  $\bar{a}(t-1)$  and  $\bar{l}(t)$  as long as it is not impossible for structural/programmatic reasons that a patient with a specific

covariate and treatment history receives that treatment value (Hernán and Robins, 2012). Nonetheless, if there are too many random zeroes (combinations of covariate values for which all or no patients receive treatment), then effect estimates may be biased. Hernán and Robins (2012) explain that these effect estimates are calculated by using parametric models to smooth over the zeroes by borrowing information from strata for which some patients initiate ART to estimate the probability of ART initiation in strata in which there are random zeroes. Cole and Hernán (2008) caution that weighted estimates are more sensitive to random zeroes than usual regression estimates.

An intervention is *well-defined* if the method of administering the treatment can be clearly described. This will lead to *consistency*, which can be expressed mathematically (in its generalized form) as follows (Robins and Hernán, 2009):

If  $\bar{A} = \bar{a}$  for a given patient, then  $T^{\bar{a}} = T$  for that patient.

This means that if a patient is observed to have a particular treatment history, then the counterfactual outcome under that treatment history must equal the observed outcome for that patient.

*No model misspecification* means that the model(s) for determining the weights and the analysis model are correctly specified. Examples of model misspecification would be using a function of treatment that does not fully capture the actual effect of treatment on outcome, or not including an interaction term between the treatment variable and a covariate that is required. Hernán and Robins (2012) emphasize the importance of thinking carefully about whether models are correctly specified, as well as evaluating the sensitivity of parameter estimates to the specification of the model. They point out that increasing the number of parameters used in the model would decrease the chances of bias from model misspecification, but would increase the variability of effect estimates and so result in wider confidence intervals. Bang and Robins (2005) describe doubly robust estimates which could also be used to increase the robustness of marginal structural models to model misspecification, as they require only that either the model for estimating the inverse probability of treatment weights or the weighted analysis model be correctly specified for valid causal estimates.

*Measurement error* occurs when either the treatment or outcome (or their common causes or conditioned on common effects) are inaccurately measured (or recorded).

### 6.4.3 Inverse Probability of Treatment and Censoring Weights

As has been referred to above, conditional exchangeability is required in order to obtain valid causal estimates. For observational studies, this can be achieved by using inverse probability of treatment and censoring weights to create a pseudo-population in which the effect of ART on survival is the same as in the original population, but the probability of being on ART is dependent only on whether the patient was on ART previously (not on any time-dependent confounders, such as CD4 count) and likewise the probability of being censored is not associated with either the past treatment history or the past confounder history (Cole et al., 2003). The stabilized inverse probability of treatment and stabilized inverse probability of censoring weights can be estimated using the following formulae (Hernán et al., 2001):

$$SW^A(t) = \prod_{k=0}^t \frac{f\{A(k)|\bar{A}(k-1), L(0), C(k) = 0\}}{f\{A(k)|\bar{A}(k-1), \bar{L}(k), C(k) = 0\}}, \quad (6.10)$$

$$SW^C(t) = \prod_{k=0}^t \frac{\Pr[C(k) = 0|\bar{C}(k-1) = 0, \bar{A}(k-1), L(0), T > k]}{\Pr[C(k) = 0|\bar{C}(k-1) = 0, \bar{A}(k-1), \bar{L}(k), T > k]}, \quad (6.11)$$

where  $\bar{A}(-1)$  and  $\bar{C}(-1)$  are defined to be zero.

Informally, the numerator of (6.10) is the probability that a patient receives the treatment that he/she did receive at time  $k$  given his/her past treatment history up to time  $k-1$ , baseline covariates  $L(0)$  and that he/she remained uncensored up to time  $k$ , while the denominator is the probability that a patient received the treatment that he/she did receive given his/her past treatment history up to time  $k$  and his/her past covariate history through time  $k$  and that he/she remained uncensored up to time  $k$ .

Similarly, the numerator of (6.11) can be described as the probability that a patient remained uncensored at time  $k$  given that he/she was uncensored up to time  $k-1$  and his/her past treatment history up to time  $k-1$  and baseline covariates and that he/she was alive at time  $k$ , while the denominator can be described as the probability that the patient remained uncensored up to time  $k$  given that he/she was uncensored up to time  $k-1$  and his/her past treatment history up to time  $k-1$  and their past covariate history up to time  $k$  and that he/she was alive at time  $k$ .

In each case, the difference in numerator and denominator is that the nu-

erator only uses baseline covariates, while the denominator uses updated covariate values.

The mean of the estimated stabilized inverse probability weights should be one at each time point (Cole and Hernán, 2008). The stabilized inverse probability of treatment and censoring weights  $SW^{A,C}(t) = SW^A(t) \times SW^C(t)$  are the product of the stabilized inverse probability of treatment and stabilized inverse probability of censoring weights.

Unstabilized weights (numerators in above replaced with one) could also be used in order to create a pseudo-population in which the average outcome  $E[T^{\bar{a}}]$  for patients receiving treatment regime  $\bar{a}$  is the same as in the original population, but neither the probabilities of being on ART nor being censored are associated with the time-dependent confounders. The probability of treatment (where treatment is dichotomous) will be 50% in the pseudo-population created using unstabilized weights, but in the pseudo-population created using stabilized weights, the probability of treatment will be the same as in the actual population. The disadvantage of unstabilized weights is they can often vary a lot, resulting in effect estimates that are very variable (have large standard errors) and are not normally distributed (Robins et al., 2000).

In point treatment studies (where treatment is not time-dependent), inverse probability of treatment weighting is similar to standardization using the g-formula, which calculates the expected outcome among those who received treatment  $a$  as

$$E[T^a] = \sum_l E[T|A = a, L = l] \Pr[L = l] \quad (6.12)$$

for discrete  $L$ . Standardization effectively simulates the joint distribution of the treatment, outcome and confounders had all patients in the population been treated, while inverse probability weighting simulates the data that would have been obtained had all subjects been unconditionally randomized. When the treatment is time-varying, however, the g-formula suffers from the *null paradox*. Even if the sharp null hypothesis of no causal effect of treatment on outcome for all treatment regimes holds, a treatment effect will be found when the g-formula is used to obtain causal effect estimates in large studies, even if they were sequentially randomized studies. (Robins and Hernán, 2009)

## 6.5 Estimating Inverse Probability Weights (IPW)

For observational studies, the values for the stabilized inverse probability of treatment and censoring weights need to be estimated using the data. This can be done using pooled logistic regression models. In my analysis, patients who have started antiretroviral therapy are assumed to remain on it, thus the number of regimens is only  $K + 1$ . Patients may be assigned antiretroviral therapy at any weekly interval after the start of TB treatment from 1 week after the start of TB treatment until the end of follow-up, which is 18 months (or 78 weeks) after the start of TB treatment. There are therefore 79 regimens and over 800 patients, so it may be possible to estimate a saturated model (although the inclusion of covariates in the model will make this less feasible).

In other applications (such as a dichotomous regimen that a patient may stop and start repeatedly), there may be many more possible regimes than the number of patients in the study (Robins and Hernán, 2009). In these instances, it will be necessary to use unsaturated models to estimate the inverse probability weights. An example of a potential unsaturated model would be to use a function of the cumulative number of intervals on treatment (rather than modelling whether a patient was on or off treatment at each time interval in the study). Using an unsaturated (parametric) model opens the door to model misspecification. Although a number of hypothesized models could be tested, it cannot be known if the correct model has been tested and thus the possibility of model misspecification cannot be excluded for unsaturated models.

### 6.5.1 Modelling the time-dependent intercept

In order to obtain a consistent estimate of the effect of the treatment (ART) on the outcome of interest (time to death) in the regression analysis weighted using the inverse probability of treatment (and censoring) weights, the denominator of these weights must be consistently estimated so that the estimates of these weights do not vary considerably. This can be achieved by reducing the number of free parameters in the model by not fitting a separate intercept at each interval. This means that when calculating the intercept for each interval, information from patients shortly before and after that interval should also be used. This can be achieved by assuming the intercepts are

constant for windows of say 12 week periods, or by assuming that the intercepts are a smooth function of the time since the start of analysis (start of TB treatment) and using smoothing techniques (such as regression splines, smoothing splines or kernel regression) to estimate them. (Hernán et al., 2000, 2001)

A spline is a continuous function, made up of piecewise linear or polynomial functions that are joined at points known as knots. A restricted cubic spline is linear before the first and after the final knot and piecewise cubic polynomial between adjacent knots (StataCorp, 2009b).

Using restricted cubic splines, the original time variable,  $\mathcal{V}$  is replaced by  $n - 1$  new variables,

$$\begin{aligned} V_1 &= \mathcal{V} \\ V_{i+1} &= \frac{(\mathcal{V} - k_i)_+^3 - (k_n - k_{n-1})^{-1} \{ (\mathcal{V} - k_{n-1})_+^3 (k_n - k_i) - (\mathcal{V} - k_n)_+^3 (k_{n-1} - k_i) \}}{(k_n - k_1)^2} \\ &\quad i = 1, \dots, n - 2 \end{aligned} \tag{6.13}$$

where

$$(u)_+ = \begin{cases} u & \text{if } u > 0 \\ 0 & \text{if } u \leq 0 \end{cases}$$

and  $k_i$  are the knot values.

Harrell Jr. (2001) recommends using the percentiles (5, 27.5, 50, 72.5 and 95) of the data as knot points when creating a spline with five knot points.

## 6.6 Performing weighted Regression

Hernán et al. (2000) suggest the use of a pooled logistic regression to estimate the Cox proportional hazards regression as standard statistical software often does not allow the use of time-varying weights. D'Agostino et al. (1990) prove that the cross-sectional pooled logistic and Cox time dependent covariate regression methods are asymptotically equivalent when there are short time intervals between measurements, small probabilities of an event within any interval and a common intercept for the pooled logistic regression across all intervals. However, it is possible to treat the total follow-up time for a patient as several independent time intervals, thereby allowing the use of time-varying weights in the fitting of survival models. This does not affect

the estimates of the model coefficients or their standard errors. The Stata code is presented in Appendix D.3.5. I will therefore not make use of the pooled logistic approximation.

I have chosen to use the robust variance estimator method to obtain 95% confidence intervals. This method provides valid, but conservative confidence intervals (the true population parameter would be included in the 95% interval more than 95% of the time).

## 6.7 Advantages and Disadvantages of MSMs

Two advantages of using MSMs over other structural models is that they are simple to understand (as they correspond closely to standard regression techniques) and are also simple to implement in standard statistical software packages (Hernán et al., 2000). Some disadvantages of MSMs are as follows:

- Models for the conditional probability of treatment need to be correctly specified at each time interval (Hernán et al., 2000).
- The product in the denominator of the inverse probability weights may be very small for some patients resulting in them receiving very large weights if there are many time periods, which can lead to bias and imprecision (Robins and Hernán, 2009).
- The distribution of the stabilized IPTWs could be very variable and skewed if there are strong interactions between the treatment and co-variates. This may result in 95% confidence intervals for the parameters of the MSM that are very wide and do not cover the true parameters at least 95% of the time (unless the number of patients is very large) (Hernán et al., 2000).

## 6.8 Multiple Imputation

### 6.8.1 Motivation for the use of multiple imputation

While marginal structural models are able to adjust for potential biases due to censoring or time-dependent confounding, a further source of bias is missing data. The models used in the analysis need to be adjusted for markers of a patient's health (or lack thereof) that are taken close to the baseline time.

Three such markers that are available in the data set are the CD4 count measurements of patients, their haemoglobin levels and their weights. The values of these markers change over time. In a randomized controlled trial, I would have wanted patients to have measurements taken for each of these values either on the day of starting TB treatment, or within a short interval thereof.

In the dataset I am using, however, the closest measurement to the baseline time was sometimes months before or after that date. The range for the closest CD4 count near to the baseline time is from 139 days before to 364 days after baseline, with a median of 14 days after the baseline time (IQR: 0 to 16). Twelve (1.39%) of 861 patients had no CD4 count recorded. For haemoglobin level measurements, the range was from 138 days before to 364 days after the baseline time, with a median of 14 days (IQR: 0 to 18 days). Twenty-one (2.44%) patients had no haemoglobin measurement recorded. The closest weight measurement to the baseline time ranged from 722 days before to 619 days after the baseline time, with a median of 0 days after the baseline time (IQR: 0 to 0 days). 244 (28.34%) patients had no recorded weight measurement.

I decided to use intervals of six months before or after the baseline time within which the closest measurement to the baseline time would be considered the patient's baseline measurement (provided it was not taken more than a week after initiating ART). There were some patients who had no value for one or more of these indicators falling within those intervals. Specifically, 33 (3.83%) patients had no baseline CD4 count, 84 (9.76%) patients had no baseline haemoglobin measurement and 251 (29.15%) patients had no baseline weight measurement. 6 (0.70%) patients have no baseline value recorded for any of the three indicators, while 542 (62.95%) patients have baseline values recorded for all three of these variables. There are three possible ways of dealing with this:

1. Only include patients who have values recorded for each health marker variable within the specified intervals
2. Categorize the values for the markers and include a 'missing' category
3. Impute values for these markers at the baseline time

The problem with the first method is that results can only be applied to patients for whom values for all three markers are available near the baseline time, unless it can be assumed that these patients are similar to those patients for whom values were not available (which is an untestable assumption). This method also throws away some of the available data, in that

patients for whom I had values for only two of the markers, will not be included in the analysis and hence the information that I did have on those patients is not used.

The drawback of the second method is that it may not be optimal to categorize the data, although if categories are defined in such a way that they are meaningful clinically, then such an analysis may be helpful in that it uses the data that was available to clinicians at the time of making decisions about when to start the patient on ART. It also assumes that patients with missing values are a homogenous group, which is not necessarily true.

The third method may be the best option available if the data can be assumed to be Missing At Random (MAR; see Section 6.8.2) and care is taken that the inclusion of the imputed values does not change the relationship between the response variable and the variable for which values were imputed. Imputing only one value for each missing value and then proceeding as if the data is complete would be misleading (Schafer, 1997). This is because data that is completed using a single imputation based on a regression equation from a single random sample from the population of interest will have values that have a much lower variability than the variability that would have existed had real data been available for all the values (Graham, 2009). In addition, analyses based on a single imputation do not take into account the uncertainty surrounding the imputed values. Multiple imputation should thus be performed. This means that a number of different completed data sets are imputed independently. Each of these imputed data sets is then analysed separately and their results combined. This is done using Rubin's rules (1976) such that appropriate estimates for the required statistical quantities, such as the variance and confidence intervals for measurements of quantities of interest, can be obtained. These estimates incorporate both the variance between imputed data sets (that accounts for the uncertainty due to missingness) and the variance within each imputed data set (White et al., 2011).

The reason that multiple imputation is only performed for the baseline values is that it is only the baseline values that are included in the analysis model. Time-updated values are only used in estimation of the inverse probability of treatment and censoring weights (which are used to weight the regression model). When there is no value of a variable at a specific time, it seems to be common practise to carry forward the most recent value for that patient (Cole et al., 2003, 2005, 2007, 2010; Sterne et al., 2005; Gsponer et al., 2012). This is because the most recent value reflects the information that would have been available to the clinician, based on which the decision to

treat or not to treat would have been made. Cole et al. (2003) reported that imputing baseline values or excluding patients who did not have baseline values gave similar results. Fairall et al. (2008) compared multiply imputing missing values (time-updated and baseline) versus carrying forward missing values (excluding patients with missing baseline values) and found very little difference in the results. For estimation of the inverse probability of treatment and censoring weights, I therefore chose to use the method of carrying forward the most recently observed value for patients who had missing values. I used a missing category to allow inclusion of patients who did not have baseline values in these models for the weights.

### 6.8.2 Missing Data Mechanisms

There are three different missing data mechanisms usually referred to in the literature. These are Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR). In order to define these mechanisms precisely, some notation is required. The notation described here is that of Little and Rubin (2002). Let  $\mathbf{Y} = (y_{ij})$  be a  $n \times k$  matrix that contains the values of each variable  $j$  for each patient  $i$ . Let the complete data matrix  $\mathbf{Y}$  be partitioned into a matrix  $\mathbf{Y}_{\text{obs}}$  containing the values of variables that were observed for each patient and a matrix  $\mathbf{Y}_{\text{mis}}$  representing the unobserved values for each patient. Let  $\mathbf{M} = (m_{ij})$  be a second  $n \times k$  matrix encoding whether a specific value is missing (one) or not missing (zero). Let  $\boldsymbol{\phi}$  be a third matrix representing a set of unmeasured variables. The missing data mechanism can then be modelled as the conditional distribution  $f(\mathbf{M}|\mathbf{Y}, \boldsymbol{\phi})$ .

The criteria for classifying the mechanism of missingness in a dataset is summarised in Table 6.1 taken from Little and Rubin (2002). The missingness mechanism is known as MCAR if the probability that the value of a variable  $y_{ij}$  is missing does not depend on either the observed or unobserved values of the complete data matrix, so that we can write  $f(\mathbf{M}|\mathbf{Y}, \boldsymbol{\phi}) = f(\mathbf{M}|\boldsymbol{\phi})$ . If the probability of the value of a variable  $y_{ij}$  being missing depends on the values of the observed data matrix  $\mathbf{Y}_{\text{obs}}$ , but not on the unobserved values of the missing data matrix  $\mathbf{Y}_{\text{mis}}$ , such that we can write  $f(\mathbf{M}|\mathbf{Y}, \boldsymbol{\phi}) = f(\mathbf{M}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\phi})$ , then the missingness mechanism is known as MAR. If, however, the probability of the value of a variable  $y_{ij}$  being missing does depend on the unobserved values of the complete data matrix (as well as on the unmeasured and observed values), such that the conditional distribution  $f(\mathbf{M}|\mathbf{Y}, \boldsymbol{\phi})$  cannot be simplified, then that missingness mechanism is known as MNAR.

Table 6.1: Classification of missing-data mechanisms using criteria of Little and Rubin (2002)

| Depends on                               | MCAR | MAR | MNAR |
|------------------------------------------|------|-----|------|
| unmeasured variables, $\phi$             | Yes  | Yes | Yes  |
| observed data, $\mathbf{Y}_{\text{obs}}$ | No   | Yes | Yes  |
| missing data, $\mathbf{Y}_{\text{mis}}$  | No   | No  | Yes  |

Collins et al. (2001) describe MCAR to be the scenario when there is no statistical relationship between the process through which the missing values arise and what the underlying values of the variables actually are. This means that we assume that those with missing values for a particular variable are not systematically different from those for whom values were observed on that variable. Excluding those with a missing value on that variable would therefore not bias analyses, as those for whom values were observed could represent those for whom values were not observed. A complete case analysis could therefore be conducted (assuming there is sufficient power for the desired analyses using only complete cases).

When the observed values of some variable(s) can completely predict whether the values of the variable of interest will be missing or not, then the missingness mechanism is MAR. The missing-data mechanism is MNAR when it is believed that the relationship between the value of a variable and whether or not it is missing cannot be completely accounted for by variables that have been observed (Collins et al., 2001).

### Ignorability

Multiple imputation is only appropriate if the missingness mechanism is *ignorable*. Harel and Zhou (2007) explain the two conditions for ignorability with reference to the Bayesian joint model for the complete data and the missingness,

$$P(\mathbf{Y}, \mathbf{M}, \boldsymbol{\theta}, \phi) = P(\mathbf{Y}|\boldsymbol{\theta})P(\mathbf{M}|\mathbf{Y}, \phi)P(\boldsymbol{\theta}, \phi),$$

where  $\boldsymbol{\theta}$  and  $\phi$  are unknown parameters of the conditional distributions for the complete data and missingness indicator matrices respectively. The two conditions are that:

1. The missing data is missing at random,  $P(\mathbf{M}|\mathbf{Y}, \phi) = P(\mathbf{M}|\mathbf{Y}_{\text{obs}}, \phi)$ .

2.  $\boldsymbol{\theta}$  and  $\boldsymbol{\phi}$  are *distinct*, i.e. their joint prior distribution factorizes into independent prior distributions,  $P(\boldsymbol{\theta}, \boldsymbol{\phi}) = P(\boldsymbol{\theta})P(\boldsymbol{\phi})$ , such that changing one of these parameters will not affect the other parameter.

Methods that assume that the missingness mechanism is ignorable do not require that there is no relationship between the reason a value is missing and what that value is, but that data that is observed is able to explain that relationship reasonably well (Schafer, 1997).

In order to impute data, an imputation method and an imputation model are required.

### 6.8.3 Imputation Methods

I considered two methods of imputation: data augmentation discussed by Schafer (1997) and Little and Rubin (2002) and chained equations (or regression switching) discussed by Van Buuren et al. (1999); Royston (2004, 2005); Royston and White (2011) and White et al. (2011) among others. The variables that required imputation in this data set were all continuous. After appropriate transformations (see Section 7.5.3 and Figure 7.8) I assumed approximate multivariate normality of the completed data (observed and unobserved values),  $\mathbf{Y} = (\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}})$  for both methods.

#### Data Augmentation

This method has Bayesian roots and is similar to the well-known Expectation Maximization (EM) algorithm (Schafer, 1997). Instead of the ‘Expectation’ and ‘Maximization’ steps of the EM algorithm, it uses ‘Imputation’ and ‘Posterior’ steps, which are explained below. The aim is to obtain random values for the missing observations,  $\mathbf{Y}_{\text{mis}}$  from the probability distribution of the missing values given the observed values,  $\Pr(\mathbf{Y}_{\text{mis}}|\mathbf{Y}_{\text{obs}})$  (StataCorp, 2009b). These values, however, depend on the distribution of the parameter values for the completed data matrix given the observed values,  $\Pr(\boldsymbol{\theta}|\mathbf{Y}_{\text{obs}})$ . This cyclical problem can be solved in the following way (Schafer, 1997; Little and Rubin, 2002):

1. Obtain initial parameter values  $\boldsymbol{\theta}^{(0)}$  (perhaps by expectation maximization methods)

2. Given a parameter value  $\boldsymbol{\theta}^{(t)}$  successively obtain draws of
  - $\mathbf{Y}_{\text{mis}}^{(t+1)}$  from  $\Pr(\mathbf{Y}_{\text{mis}}|\mathbf{Y}_{\text{obs}}, \boldsymbol{\theta}^{(t)})$  - ‘imputation step’ and
  - $\boldsymbol{\theta}^{(t+1)}$  from  $\Pr(\boldsymbol{\theta}|\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}^{(t+1)})$  - ‘posterior step’

These values form a Monte-Carlo Markov Chain (MCMC) sequence  $\{(\mathbf{Y}_{\text{mis}}^{(t)}, \boldsymbol{\theta}^{(t+1)}) : t = 1, 2, \dots\}$  which converges to the stationary distribution  $\Pr(\boldsymbol{\theta}, \mathbf{Y}_{\text{mis}}|\mathbf{Y}_{\text{obs}})$ .

In other words, whereas for the EM algorithm, the expectation of the complete-data sufficient statistics are calculated (Expectation step), for data augmentation, a random draw is taken from the complete-data sufficient statistics (Imputation step). Then instead of maximizing the complete-data likelihood (as in the Maximization step of EM), a draw is taken from the complete-data posterior (Posterior step of data augmentation).

It is helpful to determine the approximate number of iterations required,  $k$ , such that  $\boldsymbol{\theta}^{(t)}$  and  $\boldsymbol{\theta}^{(t+k)}$  are likely to be independent (Schafer, 1997). Then if the starting estimates are reasonable, a ‘burn-in’ period of length  $k$  should be long enough for stationarity to be reached by the  $k^{\text{th}}$  iteration. After the burn-in period, taking every  $k^{\text{th}}$  iterate of  $\boldsymbol{\theta}^{(t)}$  and  $\mathbf{Y}_{\text{mis}}^{(t)}$ ,  $t \in k, \dots, mk$  should result in  $m$  independent imputations.

In order to determine  $k$ , Schafer (1997) advises examining particular functions or components of  $\boldsymbol{\theta}$  for which convergence is expected to be slowest. When the observed-data posterior is nearly normal, he suggests a worst linear function,  $\xi(\boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1})^T(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_k)$  based on the expectation maximization procedure, where  $\hat{\boldsymbol{\theta}}_k$  is the convergent value of  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\theta}}_{k-1}$  is the estimate of  $\boldsymbol{\theta}$  one step prior to convergence. StataCorp (2009b) suggest using the number of iterations taken for the EM procedure (used to determine initial values) to converge as an approximation for the length of the burn-in period.

Time series and sample autocorrelation plots using the worst linear function can be inspected. After a sufficient number of iterations, there should be no visible trend in the plot of the worst linear function against the number of iterations. Serial dependence should also be seen to diminish, with correlations tending towards zero. Figure 6.1 shows the time-series and autocorrelation plots generated using the imputation equation I used to impute baseline CD4 counts, haemoglobin levels and weight measurements for the data I am using (see Section 7.5.3), but varying  $k$ , the length of the burn-in period. The EM process converged at iteration 29. For  $k = 100$ , the time

series plot for the worst linear function is possibly stationary, though there is one peak and a possible downward trend thereafter, whereas for  $k = 500$  there is clearly no trend in the time series plot for the worst linear function. The serial correlation does decrease quickly for  $k = 100$ , but the autocorrelations are fairly variable and not that close to zero, whereas for  $k = 500$ , the serial correlation decreases quickly and the autocorrelations are close to zero from about the 5th lagged autocorrelation onwards. Independence between successive draws of the imputed values might be obtained using 100 iterations (or a number between 100 and 500), but I will use the more conservative value of 500 to ensure independence. Figure 6.2 shows a MCMC process where convergence is not achieved (at least not within any reasonable number of iterations for the purpose of imputation). Although the EM procedure converges after 303 iterations, using a value for  $k$  as large as 10000 still results in a time-series plot that does not appear to have stabilized and autocorrelations that decrease very slowly.

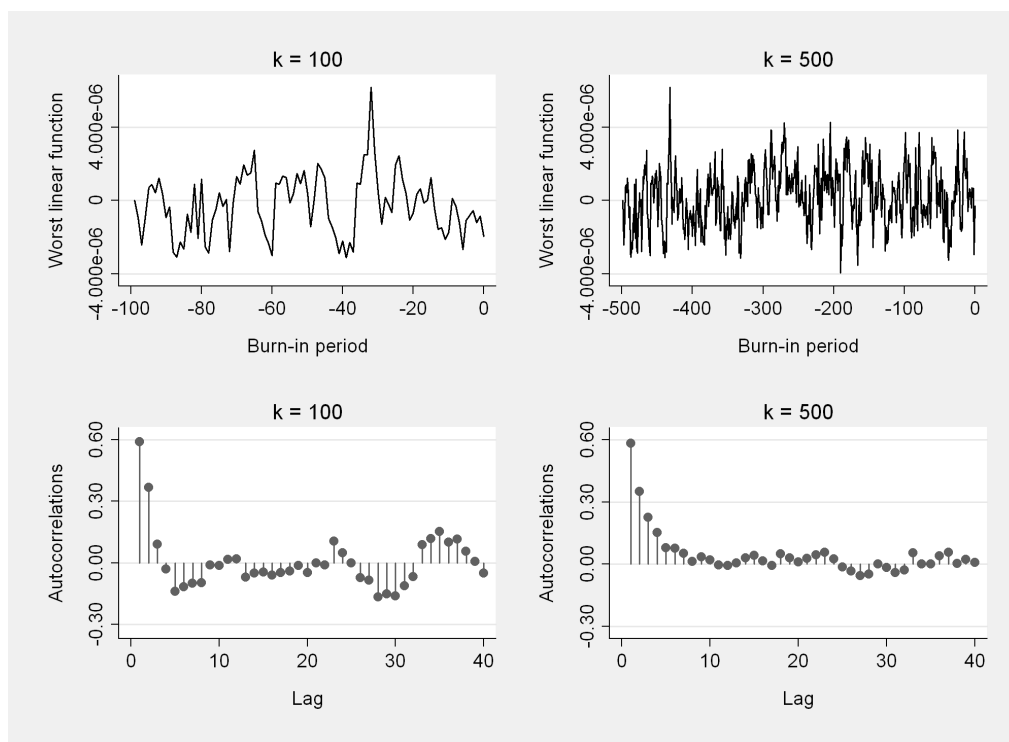


Figure 6.1: Convergence diagnostics of MCMC process: converged process

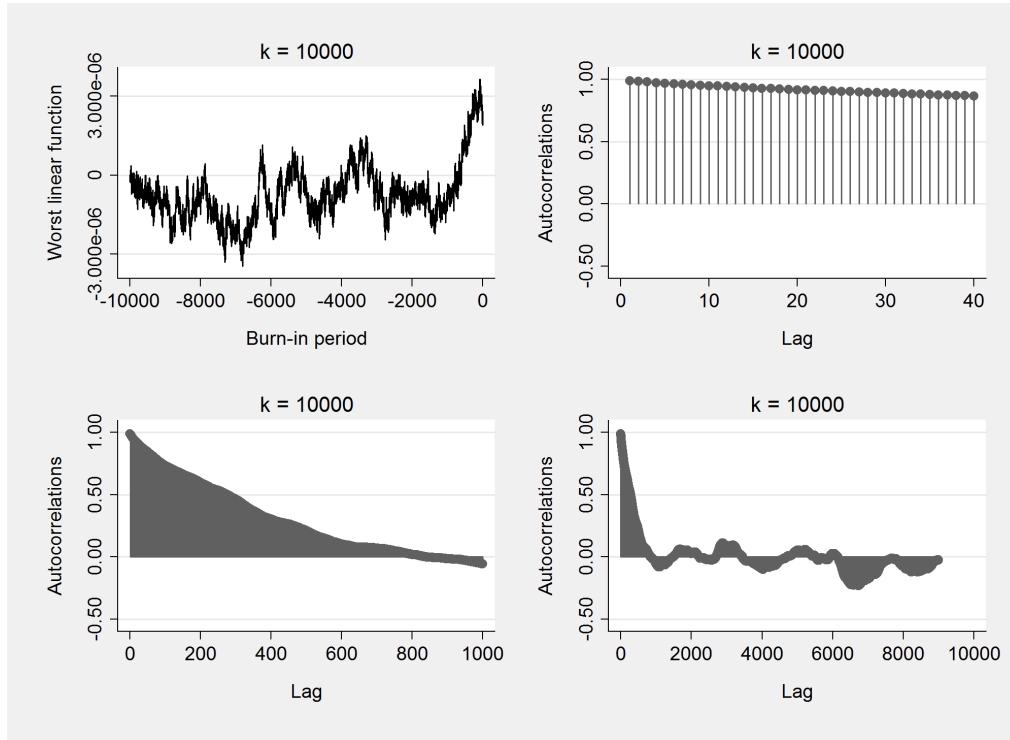


Figure 6.2: Convergence diagnostics of MCMC process: non-converged process

### Multiple Imputation by Chained Equations (MICE)

This method uses separate imputation equations for each variable that contains missing values. Where there is more than one variable that requires imputation (is incomplete) and the values of these variables are correlated with each other, this results in the need for an iterative procedure. This is because imputing the values for one of the variables requires that there be values for all observations of all the other variables in the imputation equation. In order to perform the imputation for the first variable with missing values, therefore, initial (or place-holder) values are required for the observations for which no values were available for other incomplete variables included in the imputation equation. Unobserved values for the first incomplete variable are then imputed using the place-holder values for the unobserved values of other incomplete variables. For each of the subsequent variables requiring imputation, the previously imputed values for other variables are used. The first variable is then imputed again using the updated imputed values of all the variables. This process continues until there is little or no change in the updated values being imputed (usually after about 10 cycles through the

data, but possibly more if the variables requiring imputation are strongly associated (White et al., 2011)).

The posterior predictive density is given by Van Buuren et al. (1999) as

$$p(\mathbf{Y}_{\text{mis}}|\mathbf{X}, \mathbf{R}) = \int p(\mathbf{Y}_{\text{mis}}|\mathbf{X}, \mathbf{R}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{R})d\boldsymbol{\theta},$$

where  $\mathbf{Y}_{\text{mis}}$  refers to missing observations,  $\mathbf{X}$  refers to covariate values used to predict  $\mathbf{Y}$ ,  $\mathbf{R}$  is a response indicator matrix, with  $R_{ij} = 1$  if  $Y_{ij}$  is observed and 0 otherwise,  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \log \boldsymbol{\sigma})$ ,  $\boldsymbol{\beta}$  is a regression weight and  $\boldsymbol{\sigma}$  is its standard deviation.

Multiple imputation by chained equations can be summarised as

1. Drawing a value of  $\boldsymbol{\theta}^*$  from  $p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{R})$  and then
2. Drawing a value  $\mathbf{Y}_{\text{mis}}^*$  from  $p(\mathbf{Y}_{\text{mis}}|\mathbf{X}, \mathbf{R}, \boldsymbol{\theta} = \boldsymbol{\theta}^*)$ , the conditional posterior distribution of  $\mathbf{Y}_{\text{mis}}^*$  given  $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ .

To begin with, values for missing observations are filled in at random for each variable to be imputed (such as by simple random sampling with replacement from the observed values of that variable) (Royston and White, 2011; White et al., 2011).

When a variable that has missing values to be imputed ( $\mathbf{Y}$ ) is modelled by a linear regression and the conventional uninformative prior is used, then an adapted version of the algorithm for obtaining  $m$  multiple imputations presented by Van Buuren et al. (1999) is as follows:

Regress the observed values (of a variable that has missing values that are to be imputed,  $\mathbf{Y}_{\text{obs}}$ ) on the corresponding observations of covariates that are predictors for that variable ( $\mathbf{X}_{\text{obs}}$ ) to obtain a vector of regression coefficients ( $\hat{\boldsymbol{\beta}}$ ). Incorporate uncertainty by replacing  $\hat{\boldsymbol{\beta}}$  with  $\hat{\boldsymbol{\beta}}_*$  from the posterior distribution of  $\hat{\boldsymbol{\beta}}$  given the covariate values and taking into account the missing indicator matrix using

$$\begin{aligned}\boldsymbol{\sigma}_*^2 &= (\mathbf{Y}_{\text{obs}} - \hat{\mathbf{Y}}_{\text{obs}})'(\mathbf{Y}_{\text{obs}} - \hat{\mathbf{Y}}_{\text{obs}})/g \text{ and} \\ \hat{\boldsymbol{\beta}}_* &= \hat{\boldsymbol{\beta}} + \boldsymbol{\sigma}_* \mathbf{W}^{\frac{1}{2}} \mathbf{D},\end{aligned}$$

where  $g$  is a random draw from the chi-squared distribution with degrees of freedom  $n_{\text{obs}} - r$ ;  $n_{\text{obs}}$  is the number of observed  $Y$ -values;  $r$  is the number of covariates used to impute  $\mathbf{Y}$ ;  $\mathbf{D}$  is an  $r$ -dimensional random

vector drawn from the multivariate standard normal distribution;  $\mathbf{W}^{\frac{1}{2}}$  is the triangular square root of  $\mathbf{W}$  obtained by Cholesky decomposition and  $\mathbf{W} = (\mathbf{X}'_{\text{obs}} \mathbf{X}_{\text{obs}})^{-1}$ .

Predict the missing values of  $\mathbf{Y}$ ,  $\hat{\mathbf{Y}}_{\text{mis}}$  using the corresponding values of the covariates ( $\mathbf{X}_{\text{mis}}$ ) and  $\hat{\beta}_*$  calculated above:  $\hat{\mathbf{Y}}_{\text{mis}} = \mathbf{X}_{\text{mis}} \hat{\beta}_*$ . Then draw an imputed value ( $Y_{\text{mis},i}^*$ ) from the posterior distribution  $p(\mathbf{Y}_{\text{mis}} | \mathbf{X}, \mathbf{R}, \theta = \theta^*)$  by either:

(a) **Using predicted values and adding a random error term**

$Y_{\text{mis},i}^* = \hat{Y}_{\text{mis},i} + z\sigma_i^*$ , where  $z$  is a random draw from the standard normal distribution,

or

(b) **Using prediction mean matching**

For each missing value  $Y_{\text{mis},i}$ ,  $i = 1, \dots, n_{\text{mis}}$  the patient whose  $\hat{Y}_{\text{obs},i}$  is closest to  $\hat{Y}_{\text{mis},i}$  is found and the value  $Y_{\text{obs},i}$  of that patient is then taken as the imputed value  $Y_{\text{mis},i}^*$ .

This approach is only recommended for continuous variables that do not (even approximately) satisfy the assumption of normality or for which the relationship with the covariates is non-linear (White et al., 2011). It should not be used for small sample sizes or where imputed values may be expected to be outside the range of the observed values.

As explained above, these procedures are performed iteratively for all variables requiring imputation until there is little or no change in the updated values. This forms one imputation set. The entire process is then repeated  $m$  times to create  $m$  complete data sets. White et al. (2011) suggest as a rule of thumb that the number of imputations should be greater than or equal to the proportion of incomplete cases, but caution that this may not be appropriate in all situations.

Since imputations obtained in this way are already statistically independent (Van Buuren et al., 1999), it is not necessary to ‘burn’ iterations between successive draws as was required in the data augmentation procedure explained above, which uses MCMC methods. Van Buuren et al. (1999) also remark that the choice of functional form used for the imputation model is not of critical importance unless there is very little uncertainty for the values of the missing observations.

A theoretical weakness of MICE is that there will not necessarily be a joint

probability distribution that is compatible with all the conditional distributions, in which case convergence to a stationary distribution for the imputed values may not be achieved (Gelman and Speed, 1993; Raghunathan et al., 2001; Van Buuren et al., 2006; Horton and Kleinman, 2007). Van Buuren et al. (2006) investigated the practical implications of this issue by performing a number of simulations and found that incompatible conditional distributions did not result in problematic imputation results, but acknowledged that further research is required. Azur et al. (2011) suggest that when there are few variables that need imputation and these variables are continuous and approximately normally distributed, it may be preferable to use a joint model such as the multivariate normal.

### 6.8.4 Imputation Model

Collins et al. (2001) suggest using a model that is inclusive, rather than restrictive. They recommend using all variables that could be predictive of whether the value for a variable is missing or not, as well as those to be included in subsequent analysis models - including dependent or outcome variable(s). This is because the correlation between the variable being imputed and any other variable will be suppressed toward zero if that variable is not included in the model. For this reason, any plausible interactions should also be included in the imputation model. Variables that may predict the value of the missing observation should also be included.

If necessary, variables should be transformed so as to make their distributions closer approximations to the normal distribution (Royston, 2009) or to constrain the variables to be within reasonable limits. In theory, the conditional distributions of the variables should be normally distributed, but in practise if the marginal distributions are approximately normally distributed this is usually good enough (Royston and White, 2011).

White and Royston (2009) found that associations between covariates and the outcome were biased toward the null when only the event indicator and the logarithm of the survival time were included in the imputation model, but that there was lower bias in results when the cumulative baseline hazard and event indicator were included in the imputation model. They suggest using the Nelson-Aalen estimator of the cumulative hazard, or estimating the cumulative baseline hazard using Cox regression to approximate the cumulative baseline hazard when it is unknown.

In section 7.5.4 I will present the results of using both methods described above to impute the missing baseline covariate values. The marginal structural models will then be refitted on the imputed data.

# Chapter 7

## Causal Models

### 7.1 Introduction

In this section, I introduce the rationale for using marginal structural models and describe some of the decisions and assumptions made to enable analysis. Thereafter, I show results from unweighted analysis models, followed by an explanation of the models used to estimate inverse probability of treatment and censoring weights and presentation of the results obtained from weighting analysis models using these weights. I then describe the process of performing multiple imputation on my data and present a comparison of results using multiple imputation versus complete cases or including missing categories. This is followed by a brief sensitivity analysis. Thereafter, I present some of the results from model checking. This is followed by a discussion of the assumptions required for valid causal analysis and whether they are likely to have been met for my analysis.

The objective of this analysis was to determine the impact of the delay between start of TB treatment and ART initiation on mortality. However, using the start of TB treatment as the baseline time was problematic as patients could not initiate ART without first enrolling for ART (starting the screening process for ART initiation) and furthermore, there was very little information at the time of starting TB treatment for patients who had not yet enrolled for ART (since measurements such as CD4 count and haemoglobin level, as well as the WHO clinical stage of HIV disease are usually recorded when a patient enrolls for ART). I thus chose to use the later of enrolling for ART or starting TB treatment as my baseline time. The median time

from enrolment for ART to the start of TB treatment was 5.86 weeks (IQR: 0.14 to 14.57 and range -122.57 to 71.29). I included a variable for the time between the start of TB treatment and enrolment for ART in my models in order to adjust for the differing timing of starting TB treatment relative to enrolment for ART for different patients.

861 patients from site 1 were used for analyses in this chapter, as explained in Section 1.4. I administratively censored patients at 18 months after the later of enrolment for ART or starting TB treatment. Since the outcome of interest was the time to death (if this occurred within 18 months of the later of starting TB treatment or enrolling for ART), time to event methods are appropriate to analyse the impact of ART and other variables on mortality. I will focus on Cox regression models. The effect on the hazard ratio for death of the following two time-updated measures of the timing of ART initiation was considered:

1. The effect of a patient being on ART or not at each time point.
2. The length of time since the later of the start of TB treatment and enrolment for ART that a patient has not been on ART at a given time interval.

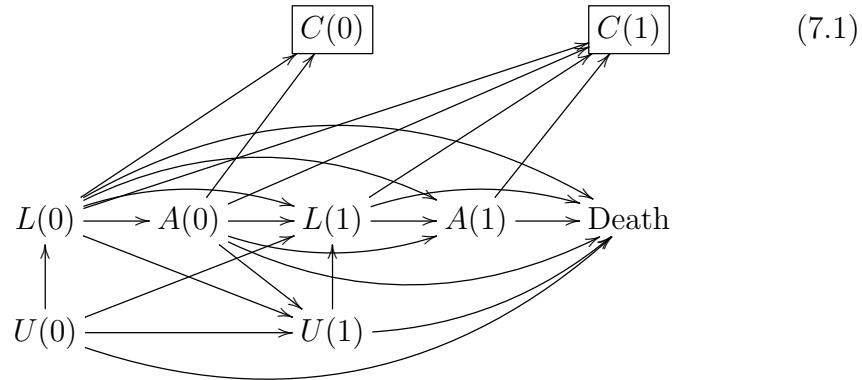
The CD4 count, level of haemoglobin and weight of the patient are time-updated measures that may be predictive both of whether (or how soon) a patient initiates ART and whether (or how soon) they die (within the period of analysis). These measures are also affected by whether a patient is on ART or not. They are thus potential time-dependent confounders and therefore cannot be adjusted for using standard modelling techniques (i.e. by including baseline and time-updated values in an analysis model). Marginal structural models provide an appropriate method for adjusting for these time-dependent confounders by adjusting for the probability of a patient starting ART given their current and baseline values using inverse-probability-of-treatment-weights (IPTW). For calculation of the IPTWs, values of CD4 counts, haemoglobin levels and weight measurements are carried forward until a new value is recorded. This means that the data used in the weight calculations is the same as that which was available to the clinicians.

The use of MSMs requires that the dataset be discretized into time intervals of appropriate length so that inverse probability weights can be calculated for each time interval. Although CD4 counts are taken at roughly 16-week intervals for patients who have initiated ART, the timing of CD4 count, haemoglobin and weight measurements prior to ART initiation is irregular,

especially when evaluated from the later of the time of starting TB treatment or enrolment for ART. In order to allow for accurate assignment of events such as ART initiation or death to a particular interval, smaller intervals are preferable (although too many intervals could result in positivity violations). I have chosen to work with one-week intervals.

Weighting the analysis model using these weights adjusts for the pre-treatment time-updated values without introducing selection bias (as would occur if these time-updated values were included as covariates in the analysis model) and without removing part of the overall effect of ART (since part of the way ART prolongs a patient's survival is mediated through improvement in these measures). Since the aim is to determine an unbiased causal estimate of the effect of ART on mortality, there is no need to take into account values of these measurements after the patient has initiated ART, as patients are assumed to remain on ART after initiating it (and so post-ART values do not affect the probability of treatment). Baseline values for these variables and other potential confounders are also included in the analysis model.

I show the causal structure that I assume underlies the (up to the end of the first time interval after baseline) using DAG 7.1 (see Section 6.2 for an explanation of DAGs).



$L(t)$  is a vector containing the values of the time-updated variables (CD4 count, haemoglobin level and weight) at time  $t$ .  $L(0)$  also includes values for the other baseline variables (such as gender, age and WHO clinical stage of HIV).  $U(t)$  represents the status of the patients' health at time  $t$ , which cannot be known directly, but can be approximated using measured variables, such as those in  $L(t)$ .  $A(t)$  indicates the treatment received by patients at time  $t$  ( $A(t) = 1$  if a patient receives ART at time  $t$  and 0 otherwise).  $C(t)$  indicates whether the patient has been censored by time  $t$  ( $C(t) = 1$  if censored

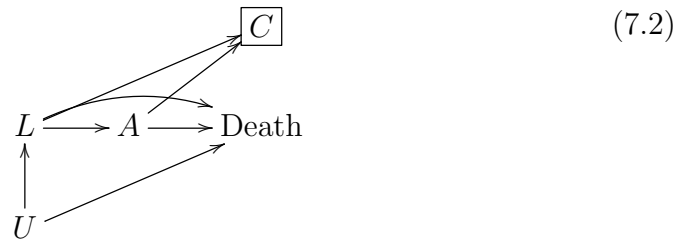
by time  $t$  and 0 otherwise). I assume that treatment assignment  $A(t)$  occurs after the measurements of the variables comprising  $L(t)$  and that censoring  $C(t)$  occurs after treatment assignment. I further assume that whether a patient initiates ART at time  $t$  is affected only by the past values of treatment and the values of measured covariates up to the time just prior to treatment assignment (i.e. I assume that treatment assignment is not affected by the unmeasured health status of a patient other than through the measured variables that approximate the patient's health status). This is the assumption of no unmeasured confounding. I also assume that censoring is affected only by past treatment and covariate values.

The extent (or lack thereof) of repeated measurements for each time-updated variable before the start of ART is shown in Table 7.1.

Table 7.1: Number (% out of 861) of patients having  $n$  values recorded after the later of the start of TB treatment and enrolment for ART up to the week of ART initiation

| $n$ | CD4 count |       | Haemoglobin |       | Weight |       |
|-----|-----------|-------|-------------|-------|--------|-------|
| 1   | 587       | (68%) | 572         | (66%) | 575    | (67%) |
| 2   | 48        | (6%)  | 48          | (6%)  | 380    | (44%) |
| 3   | 7         | (1%)  | 7           | (1%)  | 47     | (5%)  |

The very small number of patients having a second CD4 count or haemoglobin level measurement means that there may not be time-dependent confounding by these variables. Weight has more repeated measurements, but is perhaps the variable least expected to predict whether or not a patient initiates ART. It is thus unlikely that (measured) time-dependent confounding is a major bias in this dataset. For more than 50% of patients, the causal DAG could actually correspond to a point treatment study, where treatment is assigned based only on baseline variable values (DAG 7.2).



The lack of frequently updated measured variables may mean that these measured variables fail to accurately reflect the health status of the patient at the time that treatment decisions are made. If clinic visits are more frequent than the recording of time-updated CD4 count, haemoglobin and weight, then it is possible that the assumption of unmeasured confounding is not that plausible and that unrecorded factors are taken into account by clinicians when making treatment decisions. It may also be unlikely that whether a patient drops out of the study (is permanently deferred, transferred out or last to follow-up) is affected only by past values of measured variables and past treatment, in which case there may be selection bias caused by unmeasured variables.

Though CD4 count, haemoglobin level and weight are continuous variables, I decided to use them as categorical variables in order to allow the inclusion of patients for whom these values were missing by creating ‘missing’ categories. Using categories can also allow patients to be grouped according to clinical criteria that may be used to decide whether a patient is treated or not. Table 7.2 shows the categories used to group the time-updated variables.

The numbers and percentages shown refer to baseline measurements. The hazard ratios (95% CI) for death shown are for crude models including only one baseline categorical variable. Patients who had higher baseline values for the covariates shown had lower estimated hazards of death, while patients who had no recorded baseline value for a covariate had greater estimated hazards of death than patients for whom the covariates were in the lowest category (though the hazard ratio for the missing category was only significantly greater than one for patients who had a missing anaemia classification).

I initially grouped patients into five groups based on their haemoglobin level (g/dL):

Females (0; 8), [8; 10); [10; 12); [12;  $\infty$ ) and missing

Males (0; 8), [8; 11), [11; 13), [13;  $\infty$ ) and missing.

These categories were labelled ‘severe anaemia, moderate anaemia, mild anaemia, no anaemia and missing’. This resulted in too many groupings and some categories with very few patients (< 7% of patients were classified as severely anaemic at baseline). I therefore grouped patients with severe and moderate anaemia and labelled them as ‘anaemic’. Patients classified as having mild anaemia and no anaemia were grouped and termed ‘not anaemic’. These terms were used for ease of reference, since the main interest was

Table 7.2: Numbers and percentages of patients having a baseline value in the shown categories and the crude hazard ratio (HR) for death within 18 months of the later of starting TB treatment or enrolling for ART

| Variable                            | N   | (%)   | HR   | 95% CI    |
|-------------------------------------|-----|-------|------|-----------|
| Baseline CD4 count (cells/ $\mu$ L) |     |       |      |           |
| 0-49                                | 274 | (32%) | 1.00 | -         |
| 50-99                               | 205 | (24%) | 0.43 | 0.27-0.70 |
| 100-199                             | 260 | (30%) | 0.35 | 0.22-0.57 |
| $\geq 200$                          | 89  | (10%) | 0.63 | 0.35-1.15 |
| missing                             | 33  | (4%)  | 1.81 | 0.99-3.29 |
| Anaemic at baseline*                |     |       |      |           |
| yes                                 | 365 | (42%) | 1.00 | -         |
| no                                  | 412 | (48%) | 0.53 | 0.36-0.78 |
| missing                             | 84  | (10%) | 1.84 | 1.18-2.89 |
| Baseline Weight (kg)                |     |       |      |           |
| (0; 50)                             | 130 | (15%) | 1.00 | -         |
| [50; 60)                            | 253 | (29%) | 0.48 | 0.29-0.80 |
| [60; $\infty$ )                     | 227 | (26%) | 0.22 | 0.12-0.43 |
| missing                             | 251 | (29%) | 1.05 | 0.68-1.63 |

\*Patients were classified as anaemic if they had a haemoglobin level of  $< 10$  g/dL (females) or  $< 11$  g/dL (males)

whether patients who were severely (or moderately) anaemic had a higher hazard of death and possibly a different hazard of ART initiation (compared to patients who were not anaemic or only had mild anaemia).

## 7.2 Unweighted Models

Table 7.3 shows results from Cox proportional hazards models that include only the function of treatment of interest.

Table 7.3: Unweighted Models for differing treatment functions

| Model | Treatment Function | HR   | 95% CI      |
|-------|--------------------|------|-------------|
| 1.a)  | On ART             | 0.52 | 0.33 - 0.81 |
| 2.a)  | Weeks not on ART   | 1.04 | 1.02 - 1.07 |

Being on ART at a given interval (Model 1.a) resulted in a 48% reduction in the hazard of death. Model 2.a shows a 4% increase in the hazard of death for every additional week that a patient delayed initiation of ART (after they had both started TB treatment and enrolled for ART). These effects were both significant at the 5% level. The proportional hazards assumption was also satisfied for both models ( $p$ -values for test based on Schoenfeld residuals were  $> 0.1$ ; see Grambsch and Therneau (1994) or Cleves et al. (2008) for an explanation of the statistical test).

Next, I consider models that include other covariates. All models will include a function of the timing of ART (main exposure), the variable denoting the time between the start of TB treatment and enrolment for ART and the baseline values for the three potential time-dependent confounders: CD4 count category, anaemia classification and weight category. In addition I consider the inclusion of the following ‘optional’ baseline variables that may be risk factors for death and will only be selected for the model if they are significant at the 10% level: gender, age, WHO clinical stage of HIV, whether patient is being retreated for TB (versus being treated for a new TB infection), whether patient has smear-positive pulmonary or extra-pulmonary TB (compared to smear-negative pulmonary TB), whether patient was diagnosed with TB before 2006.

Models 1.b of Table 7.4 and 2.b of Table 7.5 show the results for models where the only optional baseline variables included were the age of the patient and whether they were recorded as having WHO clinical stage IV of HIV disease.

Table 7.4: Unweighted Cox regressions using treatment function ‘On ART’

|                                                              | Model 1.b           |             | Model 1.c             |             |                          |             |
|--------------------------------------------------------------|---------------------|-------------|-----------------------|-------------|--------------------------|-------------|
|                                                              | $e^{\hat{\beta}_j}$ | 95% CI      | $e^{\hat{\beta}_j^*}$ | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI      |
| On ART                                                       | 0.50                | 0.32 - 0.79 | 0.48                  | 0.31 - 0.75 | 1.00                     | -           |
| Months from start<br>of TB treatment to<br>Enrolment for ART | 1.03                | 0.96 - 1.10 | 1.12                  | 1.02 - 1.23 | 0.98                     | 0.97 - 0.99 |
| Age (years)                                                  | 1.02                | 1.00 - 1.04 | 1.02                  | 1.00 - 1.04 | 1.00                     | -           |
| WHO Clinical Stage                                           |                     |             |                       |             |                          |             |
| I-III                                                        | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| IV                                                           | 1.44                | 1.01 - 2.04 | 1.43                  | 1.01 - 2.04 | 1.00                     | -           |
| Baseline CD4 Count                                           |                     |             |                       |             |                          |             |
| 0-49                                                         | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| 50-99                                                        | 0.42                | 0.26 - 0.67 | 0.23                  | 0.11 - 0.49 | 1.16                     | 1.01 - 1.33 |
| 100-199                                                      | 0.34                | 0.21 - 0.55 | 0.14                  | 0.06 - 0.31 | 1.22                     | 1.07 - 1.40 |
| $\geq 200$                                                   | 0.53                | 0.29 - 0.98 | 0.23                  | 0.09 - 0.60 | 1.21                     | 1.04 - 1.41 |
| missing                                                      | 0.83                | 0.38 - 1.81 | 0.49                  | 0.15 - 1.60 | 1.14                     | 0.93 - 1.41 |
| Anaemic at baseline                                          |                     |             |                       |             |                          |             |
| yes                                                          | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| no                                                           | 0.66                | 0.44 - 0.99 | 0.36                  | 0.19 - 0.67 | 1.15                     | 1.03 - 1.28 |
| missing                                                      | 1.93                | 1.09 - 3.43 | 1.43                  | 0.62 - 3.31 | 1.08                     | 0.93 - 1.26 |
| Baseline Weight                                              |                     |             |                       |             |                          |             |
| (0; 50)                                                      | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| [50; 60)                                                     | 0.55                | 0.33 - 0.91 | 0.54                  | 0.32 - 0.90 | 1.00                     | -           |
| [60; $\infty$ )                                              | 0.27                | 0.14 - 0.52 | 0.27                  | 0.14 - 0.52 | 1.00                     | -           |
| missing                                                      | 1.21                | 0.78 - 1.88 | 1.19                  | 0.76 - 1.85 | 1.00                     | -           |

$e^{\hat{\beta}_j}$  denotes the estimated hazard ratio of death averaged over the entire analysis period for a patient having covariate  $X_j$  compared to the baseline patient

$e^{\hat{\beta}_j^*}$  denotes the estimated hazard ratio of death at the baseline time for a patient having covariate  $X_j$  compared to the baseline patient

$e^{\hat{\beta}_j^{**}}$  denotes the estimated time-varying hazard ratio of death for a patient having covariate  $X_j$  compared to the baseline patient

Table 7.5: Unweighted Cox regressions using treatment function ‘Weeks not on ART’

|                                                        | Model 2.b           |             | Model 2.c             |             |                          |             |
|--------------------------------------------------------|---------------------|-------------|-----------------------|-------------|--------------------------|-------------|
|                                                        | $e^{\hat{\beta}_j}$ | 95% CI      | $e^{\hat{\beta}_j^*}$ | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI      |
| Weeks not on ART                                       | 1.05                | 1.03 - 1.07 | 1.05                  | 1.02 - 1.07 | 1.00                     | -           |
| Months from start of TB treatment to Enrolment for ART | 1.03                | 0.96 - 1.10 | 1.12                  | 1.02 - 1.24 | 0.98                     | 0.97 - 1.00 |
| Age (years)                                            | 1.02                | 1.00 - 1.04 | 1.02                  | 1.00 - 1.04 | 1.00                     | -           |
| WHO Clinical Stage                                     |                     |             |                       |             |                          |             |
| I-III                                                  | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| IV                                                     | 1.47                | 1.03 - 2.09 | 1.47                  | 1.03 - 2.08 | 1.00                     | -           |
| Baseline CD4 count                                     |                     |             |                       |             |                          |             |
| 0-49                                                   | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| 50-99                                                  | 0.41                | 0.25 - 0.66 | 0.25                  | 0.12 - 0.52 | 1.14                     | 1.00 - 1.31 |
| 100-199                                                | 0.33                | 0.21 - 0.54 | 0.15                  | 0.07 - 0.33 | 1.21                     | 1.06 - 1.38 |
| $\geq 200$                                             | 0.47                | 0.25 - 0.89 | 0.30                  | 0.12 - 0.77 | 1.12                     | 0.96 - 1.31 |
| missing                                                | 0.78                | 0.35 - 1.70 | 0.59                  | 0.18 - 1.91 | 1.09                     | 0.87 - 1.35 |
| Anaemic at baseline                                    |                     |             |                       |             |                          |             |
| yes                                                    | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| no                                                     | 0.66                | 0.44 - 0.99 | 0.35                  | 0.19 - 0.66 | 1.15                     | 1.03 - 1.28 |
| missing                                                | 1.97                | 1.11 - 3.51 | 1.39                  | 0.60 - 3.24 | 1.09                     | 0.93 - 1.28 |
| Baseline Weight                                        |                     |             |                       |             |                          |             |
| (0; 50)                                                | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| [50; 60)                                               | 0.55                | 0.33 - 0.91 | 0.52                  | 0.31 - 0.87 | 1.00                     | -           |
| [60; $\infty$ )                                        | 0.27                | 0.14 - 0.52 | 0.26                  | 0.13 - 0.50 | 1.00                     | -           |
| missing                                                | 1.21                | 0.77 - 1.88 | 1.16                  | 0.75 - 1.81 | 1.00                     | -           |

$e^{\hat{\beta}_j}$ ,  $e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 7.4.

The equation for Model 1.b can be written as:

$$\begin{aligned}
 \log \left[ \frac{h_i(t)}{h_0(t)} \right] = & \underbrace{\beta_1 X_{1i}}_{\text{On ART}} + \underbrace{\beta_2 X_{2i}}_{\text{Months from start of TB treatment to enrolment for ART}} + \underbrace{\beta_3 X_{3i}}_{\text{Age}} + \underbrace{\beta_4 X_{4i}}_{\text{WHO Clinical Stage IV of HIV vs Stages I-III}} \\
 & + \underbrace{\beta_5 X_{5i} + \dots + \beta_8 X_{8i}}_{\text{CD4 categories vs } < 50 \text{ cells}/\mu\text{L (baseline)}} + \underbrace{\beta_9 X_{9i} + \beta_{10} X_{10i}}_{\text{not anaemic \& missing vs anaemic (baseline)}} \\
 & + \underbrace{\beta_{11} X_{11i} + \dots + \beta_{13} X_{13i}}_{\text{weight categories vs } < 50 \text{ kg (baseline)}}.
 \end{aligned} \tag{7.3}$$

For Model 2.b, the equation is the same as above, but with the label ‘On ART’ replaced by ‘Weeks not on ART’.

The treatment effect is slightly greater in these models than in the crude models (1.a and 2.a of Table 7.3). Patients who had a greater delay between starting TB treatment and enrolment for ART and those who were older had slightly increased hazards of death. Patients in WHO stage IV of HIV disease had a greater hazard of death than those in WHO stages I-III, which was expected as stage IV is an indicator of more advanced immune deterioration in HIV-positive patients. Patients with higher baseline CD4 counts and weight values and those who were not anaemic also had lower hazards of death than those with very low CD4 count ( $< 50$  cells/ $\mu\text{L}$ ) or very low weight ( $< 50$  kg) or that were anaemic at baseline. Patients with CD4 count  $\geq 200$  were probably quite sick at baseline (despite this not being reflected in their CD4 count measurements) otherwise they would not have met the criteria to start the screening process for ART.

Although those with missing baseline CD4 count categories had a lower hazard of death than those with baseline values  $< 50$  cells/ $\mu\text{L}$ , this difference was not statistically significant, suggesting that the prognosis for those with missing CD4 count categories is not much better than that of those having very low values for these measurements (only 4% of patients had no baseline CD4 count). Patients not having a baseline weight value and those not having a haemoglobin measurement (missing anaemia classification) had a higher hazard of death. For weight this effect was not significant, while for haemoglobin it was significant (but only 10% of patients had no baseline haemoglobin measurement).

These poor predicted outcomes for patients not having recorded baseline measurements is probably a combined result of such patients being quite sick at baseline (average unrecorded measurement values possibly similar to those in the lowest categories), as well as many of them experiencing a delay in ART initiation (as seen in Table 4.10 of Section 4.4.2). Not all the patients used for the analysis in Section 4.4.2 were included in this one, however, and the baseline time was the start of TB treatment, rather than the later of the start of TB treatment and enrolment for ART as in this analysis and hence the baseline CD4 count and the patients for whom this was missing may also have differed. In particular, for patients who were diagnosed with TB before enrolling for ART, the CD4 count of the patient at the time of starting TB treatment may have been high and may have declined by the time the patient enrolled for ART, especially if the delay between these two events was

substantial. Patients who enrolled for ART long after starting TB treatment were those who were more likely not to have a CD4 count near the time of starting TB treatment.

Note also that the sensitivity analysis for the treatment of missing CD4 counts in Section 4.3.5, for which it was thought the missing values for CD4 counts could be high was conducted only on patients who did initiate ART and included patients from all three clinic sites unlike this analysis, which included patients both who did and did not initiate ART and only included patients from site 1. These two analyses also had different baseline times and thus would have had different baseline CD4 counts and patients for whom these were missing.

The tests of proportional hazards based on Schoenfeld residuals were not satisfied for the time from the start of TB treatment to enrolment for ART, nor for the baseline CD4 count categories, nor for the baseline anaemia classification (for models using either of the two functions of treatment). The lack of proportionality of the hazards for the CD4 count categories as well as the lack of proportionality between the hazards of death for those who are and are not anaemic can be seen in Figure 7.1.

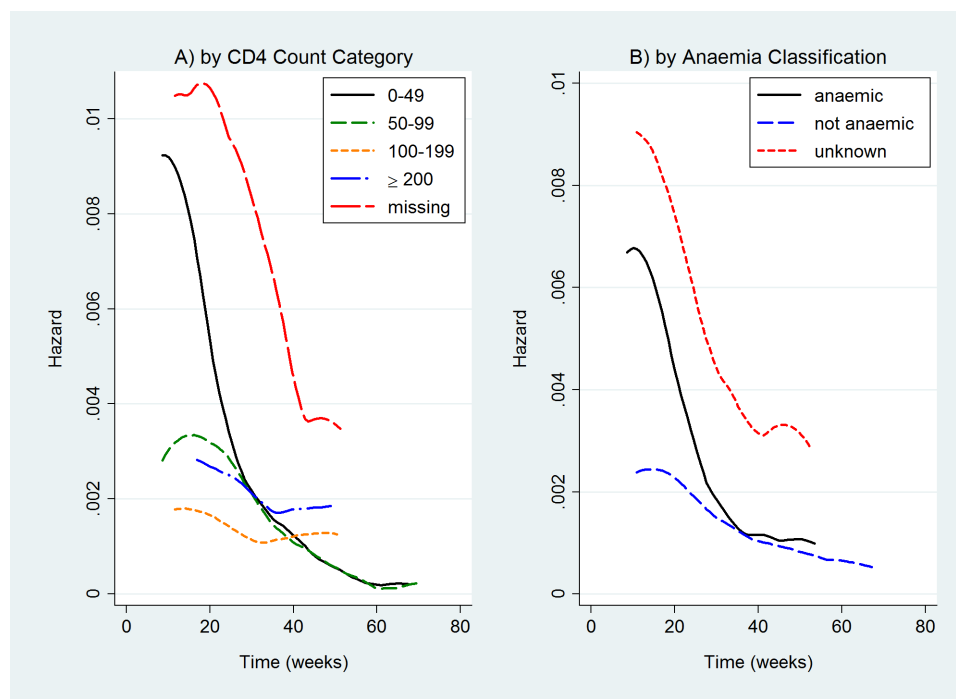


Figure 7.1: Smoothed Hazard Curves

In order to account for this lack of proportionality, I included interactions of time with each of the variables not satisfying the assumption of proportional hazards (Models 1.c and 2.c of Tables 7.4 and 7.5). An alternative to including time interactions is to stratify by variables that violate the proportional hazards assumption. This was not a good option for my analysis, however, as I wanted to estimate and interpret the effect on the hazard of death for these variables. In addition, it does not make sense to stratify on the continuous variable ‘time from start of TB treatment to enrolment for ART’ and there were too many categories of variables to stratify on both CD4 count category and anaemia classification.

The equation for Model 1.c shown in Table 7.4 can be written as

$$\begin{aligned} \log \left[ \frac{h_i(t)}{h_0(t)} \right] &= \beta_1(t)X_{1i} + \beta_2(t)X_{2i} + \dots + \beta_{13}(t)X_{13i} \\ &= (\beta_1^* + \beta_1^{**}t)X_{1i} \\ &\quad + (\beta_2^* + \beta_2^{**}t)X_{2i} + \dots \\ &\quad + (\beta_{13}^* + \beta_{13}^{**}t)X_{13i}. \end{aligned} \tag{7.4}$$

where  $X_{1i} \dots X_{13i}$  are as in 7.3 (similarly for Model 2.c). For the covariates that did satisfy the assumption of proportional hazards, the co-efficients  $\beta_j^{**}$  for the interaction of the covariate with time were set to zero. The estimated hazard ratios relating to the main effects of the covariates in Models 1.c and 2.c (denoted  $e^{\hat{\beta}_j^*}$  in Table 7.4) now refer to the effect of these covariates at time zero, i.e. at or immediately after the later of the start of TB treatment or enrolment for ART. These effects (for covariates for which the coefficient of the time interaction was not set to zero) are now greater (the estimated hazard ratios are further from one) than the effects from Models 1.b and 2.b (denoted  $e^{\hat{\beta}_j}$ ) that were averaged over the entire time period (other than for the effect of a patient having a missing baseline anaemia classification). The estimated hazard ratios for the interaction terms (denoted  $e^{\hat{\beta}_j^{**}}$ ) measure the time-dependent effect modification.

In the models including time interactions, the hazard ratios for the time from the start of TB treatment to enrolment are further above one, while the hazard ratios for the interaction of analysis time with the time from start of TB treatment to enrolment for ART are lower than one. This indicates that the hazard of death is higher initially for patients who have a longer time between starting TB treatment and enrolling for ART, but that this effect decreases over time. For example, at baseline, a patient who enrolled for ART before starting TB treatment has a lower hazard of death than a

patient who has just enrolled for ART. As time since the later of starting TB treatment or enrolment for ART increases, however, this survival benefit of having enrolled earlier (longer before start of TB treatment or sooner after) decreases.

The hazard ratios for the CD4 count categories ‘50-99’, ‘100-199’, ‘ $\geq 200$ ’ and ‘missing’ at baseline are all lower (further below one) than their averages over all analysis time in the model not including time interactions. The hazard ratios for their interactions with time are all above one. This indicates that those with higher or missing baseline CD4 counts have a much lower hazard of death at baseline, but that the beneficial effects on survival of having these baseline covariates is no longer as great at later times.

The same applies to those who are not anaemic at baseline compared to those who are. For those who did not have haemoglobin measurements at baseline, the hazard of death is greater at baseline than for those who had haemoglobin measurements indicating that they were anaemic. The hazard ratio for the interaction of a missing anaemia classification at baseline with time is also greater than one, indicating that the harmful effect on survival at baseline for those with no haemoglobin measurement becomes even greater at later times. Neither of these hazard ratios (for either missing anaemia classification or its interaction with time) are significantly different from one.

### 7.3 Model checking for unweighted models

In this section I show some results from model checking for the unweighted models 1.a to 2.c (see Tables 7.3, 7.4 and 7.5). I also include for comparison Models 1.d and 2.d, which used the same covariate values as Models 1.b and 2.b, but additionally included the time-updated values for CD4 count category, anaemia classification and weight category, as well as the interaction of the time from the start of TB treatment to enrolment for ART with the analysis time (see Table 7.6). These models cannot be used to estimate an unbiased causal effect of ART on the hazard of death (in the presence of time-dependent confounders whose future values are affected by past treatment history), but could be used for predictions of the time to death. Their inclusion demonstrates (unsurprisingly) that for the purpose of obtaining good model fit, including the most recent values of time-varying variables is preferable to including time-varying coefficients for the baseline values of these variables.

Table 7.6: Unweighted Cox regressions that include most recent categories for time-updated covariates

|                                                                                                       | Model 1.d ‘On ART’  |              | Model 2.d ‘Weeks not on ART’ |              |
|-------------------------------------------------------------------------------------------------------|---------------------|--------------|------------------------------|--------------|
|                                                                                                       | $e^{\hat{\beta}_j}$ | 95% CI       | $e^{\hat{\beta}_j}$          | 95% CI       |
| Function of Treatment                                                                                 | 0.65                | 0.42 - 1.00  | 1.01                         | 0.99 - 1.04  |
| Months from start of TB treatment to Enrolment for ART <sup>a</sup>                                   | 1.11                | 1.01 - 1.22  | 1.11                         | 1.01 - 1.22  |
| Age (years)                                                                                           | 1.02                | 1.00 - 1.04  | 1.02                         | 1.00 - 1.04  |
| WHO Clinical Stage                                                                                    |                     |              |                              |              |
| I-III                                                                                                 | 1.00                | -            | 1.00                         | -            |
| IV                                                                                                    | 1.41                | 0.99 - 2.01  | 1.40                         | 0.98 - 2.00  |
| Baseline CD4 count                                                                                    |                     |              |                              |              |
| 0-49                                                                                                  | 1.00                | -            | 1.00                         | -            |
| 50-99                                                                                                 | 1.59                | 0.60 - 4.17  | 1.51                         | 0.57 - 4.01  |
| 100-199                                                                                               | 2.45                | 0.90 - 6.67  | 2.30                         | 0.82 - 6.48  |
| $\geq 200$                                                                                            | 4.60                | 1.37 - 15.48 | 4.19                         | 1.14 - 15.33 |
| missing                                                                                               | 1.82                | 0.40 - 8.26  | 1.75                         | 0.38 - 8.01  |
| Anaemic at baseline                                                                                   |                     | -            |                              | -            |
| yes                                                                                                   | 1.00                | -            | 1.00                         | -            |
| no                                                                                                    | 1.38                | 0.80 - 2.36  | 1.34                         | 0.79 - 2.29  |
| missing                                                                                               | 1.88                | 0.78 - 4.52  | 1.84                         | 0.76 - 4.44  |
| Baseline weight                                                                                       |                     |              |                              |              |
| (0; 50)                                                                                               | 1.00                | -            | 1.00                         | -            |
| [50; 60)                                                                                              | 1.51                | 0.73 - 3.10  | 1.44                         | 0.71 - 2.93  |
| [60; $\infty$ )                                                                                       | 1.28                | 0.39 - 4.23  | 1.24                         | 0.38 - 4.07  |
| missing                                                                                               | 4.11                | 0.68 - 24.82 | 3.43                         | 0.56 - 21.09 |
| Interaction of Analysis Time with Months from start of TB treatment to Enrolment for ART <sup>a</sup> | 0.99                | 0.98 - 1.00  | 0.99                         | 0.98 - 1.00  |
| Most recent CD4 count                                                                                 |                     | -            |                              | -            |
| 0-49                                                                                                  | 1.00                | -            | 1.00                         | -            |
| 50-99                                                                                                 | 0.23                | 0.09 - 0.63  | 0.25                         | 0.09 - 0.68  |
| 100-199                                                                                               | 0.11                | 0.04 - 0.30  | 0.12                         | 0.04 - 0.33  |
| $\geq 200$                                                                                            | 0.07                | 0.02 - 0.23  | 0.09                         | 0.03 - 0.29  |
| missing                                                                                               | 0.31                | 0.06 - 1.63  | 0.34                         | 0.06 - 1.76  |
| Anaemic at most recent measurement                                                                    |                     |              |                              |              |
| yes                                                                                                   | 1.00                | -            | 1.00                         | -            |
| no                                                                                                    | 0.30                | 0.18 - 0.52  | 0.31                         | 0.18 - 0.52  |
| missing                                                                                               | 1.06                | 0.38 - 2.94  | 1.12                         | 0.41 - 3.07  |
| Most recent weight                                                                                    |                     |              |                              |              |
| (0; 50)                                                                                               | 1.00                | -            | 1.00                         | -            |
| [50; 60)                                                                                              | 0.26                | 0.13 - 0.55  | 0.26                         | 0.13 - 0.54  |
| [60; $\infty$ )                                                                                       | 0.19                | 0.06 - 0.56  | 0.18                         | 0.06 - 0.54  |
| missing                                                                                               | 0.25                | 0.04 - 1.47  | 0.28                         | 0.05 - 1.74  |

<sup>a</sup> The date of enrolment for ART refers to the date that the screening process for ART initiation began

### 7.3.1 Assessment of model fit

All models appear to have reasonable fit according to the plots of the Cox-Snell residuals against the cumulative hazard, although the fit is not that good in the right-hand tails (see Figure 7.2). Model 2.d (which uses the treatment function ‘Weeks not on ART’ and includes the most recent values for time-varying covariates) appears to have the best fit.

### 7.3.2 Deviance Residuals

Figure 7.3 shows plots of the deviance residuals against the linear predictors for each of the models 1.a to 2.d. The markers in the plot are labelled according to whether a patient died (1) or was censored (0). If the fit is adequate, the uncensored observations should be randomly scattered, whilst the censored observations would be expected to be clustered around zero (Klein and Moeschberger, 1997).

For model 1.a, which includes only whether a patient was on ART or not, the linear prediction is the same for all patients who were on ART (not on ART) at the last observation before they died or were censored. For models 1.b, 1.c, 2.a, 2.b and 2.c, the deviance residuals for patients who died do not appear to be randomly scattered, but instead show a downward trend. This indicates that the models are under-predicting the risk of death for many of the patients who die. Including the time-updated covariates (as for models 1.d and 2.d) appears to alleviate this problem, since the plots of the deviance residuals against the linear predictors for these two models are closer to random scatter (particularly Model 1.d). If the purpose of the models was to predict the survival time of patients, then these models may be more appropriate. Since the aim, however, is to estimate the causal effect of ART initiation on mortality it is not appropriate to include the time-updated covariate values. This is because these values reflect the effect of the patient having been on ART such that the estimated effect of being on ART is smaller when time-updated covariates (that are affected by past treatment) are included in the model.

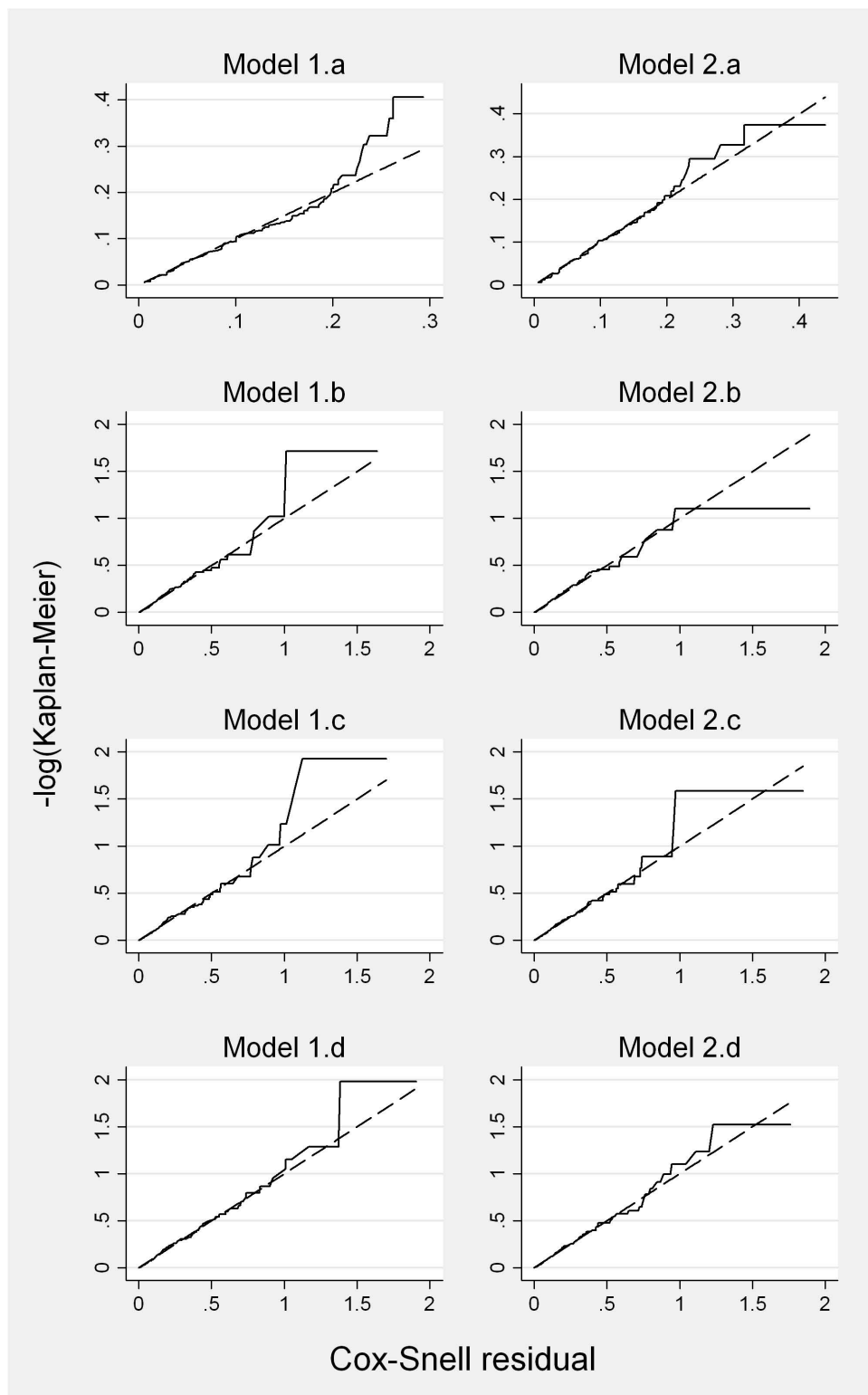


Figure 7.2: Cox-Snell residual plots for Models 1.a to 2.d  
 Model labels correspond to those used in Tables 7.3-7.6

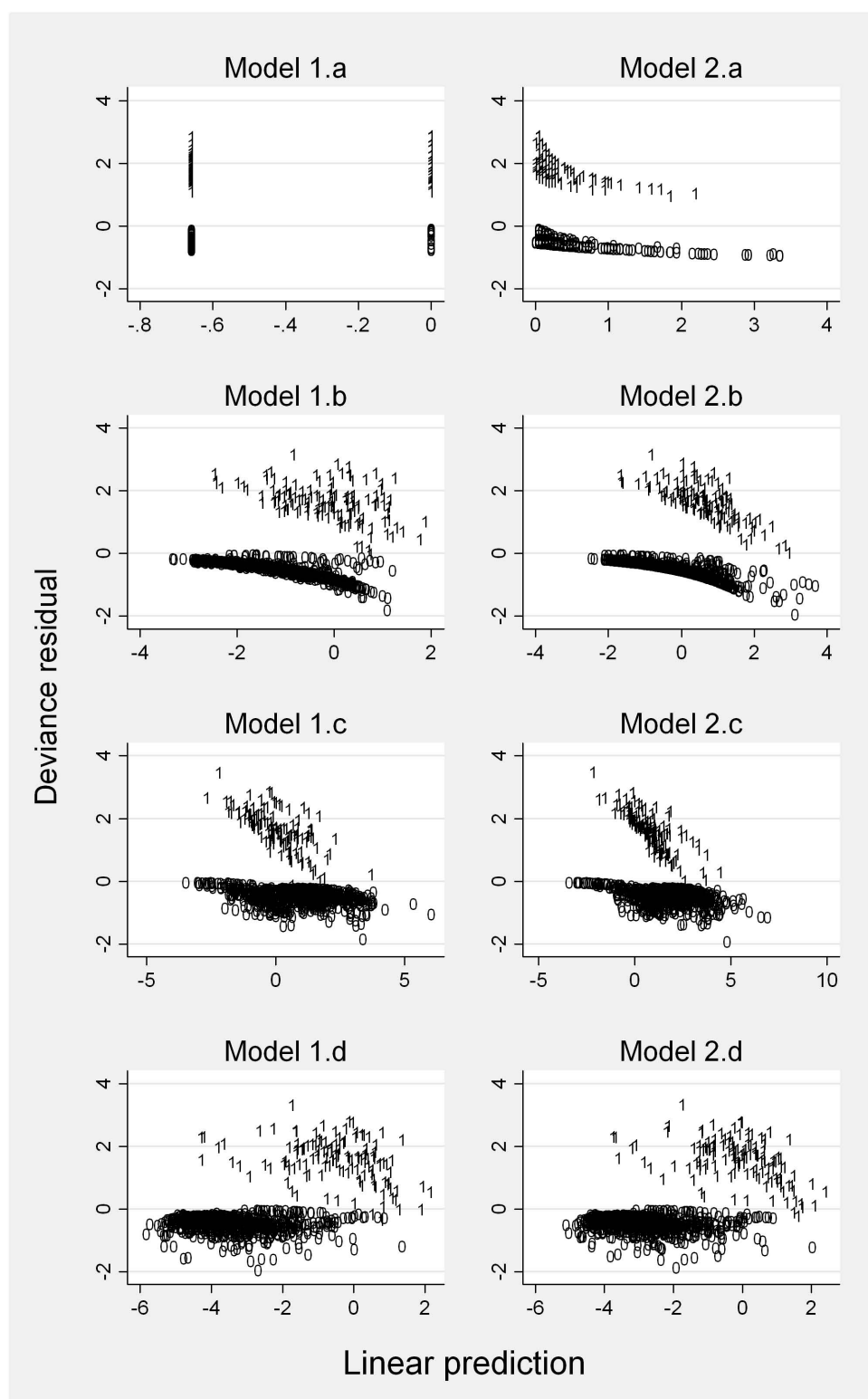


Figure 7.3: Deviance residual plots for Models 1.a to 2.d  
 Model labels correspond to those used in Tables 7.3-7.6

In addition to the lack of random scatter, there are also many deviance residuals that are fairly large. No patients had deviance residuals less than -2, but the number of deviance residuals greater than 2 were 42, 33, 37 and 33 for models 1.a-1.d respectively and 41, 31, 36 and 33 for models 2.a-2.d respectively. All of these patients were observed to die within 18 months of the later of starting TB treatment and enrolling for ART. The highest deviance residual corresponded to the same patient for all eight models (sometimes tied with other patients). This patient died within one week of enrolling for ART, despite this being less than a month after starting TB treatment and being recorded as being in WHO clinical stage III of HIV, having a CD4 count over 100 and not being anaemic. The patient with the lowest deviance residual in models 1.b-1.d and 2.b-2.d was still alive 18 months after enrolling for HIV care despite this being more than three months after starting TB treatment and being recorded as having WHO stage IV of HIV disease and no recorded baseline values for CD4 count, weight or haemoglobin level. Deleting either of these patients did not substantially alter results.

### 7.3.3 Sensitivity of model estimates to influential patients

Figure 7.4 shows the likelihood displacement values versus survival time for each of the models 1.a to 2.d. The patient numbers have been used as markers in the plots. Patients who have a high likelihood displacement value are expected to have a large influence on the estimates of the model.

The patient who had the highest likelihood displacement value for Model 1.a (patient number 402) was 31 years old, was recorded as being in WHO stage III or earlier of HIV disease, had CD4 count of less than 50 cells/ $\mu$ L, was classified as anaemic, had no recorded weight measurement near the baseline time, enrolled for ART six weeks after starting TB treatment and died two weeks thereafter, despite initiating ART four days after enrolment for ART. Excluding this patient from the model results in a slightly lower (further from one and therefore greater effect) hazard ratio of death for patients on ART.

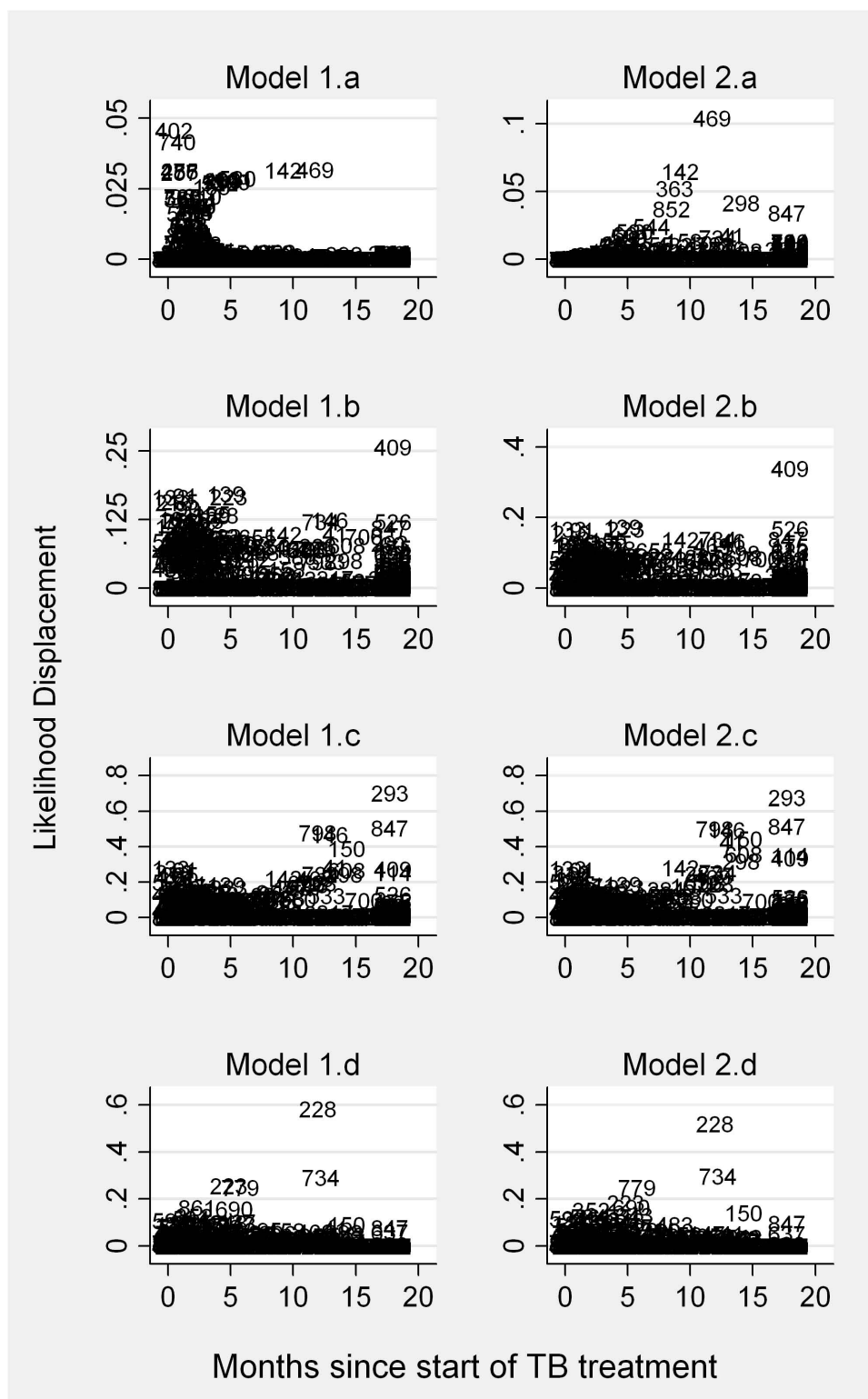


Figure 7.4: Likelihood displacement value plots for Models 1.a to 2.d  
Model labels correspond to those used in Tables 7.3-7.6

The patient who had the highest likelihood displacement value for Model 2.a (patient number 469) was 47 years old, was recorded as being in WHO stage III or earlier of HIV disease, had a CD4 count of  $\geq 200$  cells/ $\mu\text{L}$ , was not classified as anaemic and had no recorded value for weight near the baseline time, enrolled for ART more than five months after starting TB treatment and died almost a year thereafter, without having initiated ART. Excluding this patient from the model results in a slightly lower (closer to one and therefore smaller effect) hazard ratio of death for patients who have a one week additional delay in ART initiation.

The patient who had the highest likelihood displacement value for Models 1.b and 2.b (patient number 409) was 34 years old, was recorded as being in WHO stage IV of HIV disease, had no recorded CD4 count, haemoglobin level or weight near the baseline time, enrolled for ART more than three months after starting TB treatment and initiated ART almost seven months thereafter and was still in care 18 months after enrolling for ART. Excluding this patient does result in slightly lower (further from one and therefore greater effect) estimated hazard ratios of death for patients on ART and higher (also further from one and therefore greater effect) estimated hazard ratios of death for patients whose ART initiation is delayed by an additional week for Models 1.b and 2.b respectively. For both models the estimated hazards ratios of death increase slightly (further from one and therefore greater effect) for patients who had a longer time between starting TB treatment and enrolling for ART, those who were in WHO stage IV of HIV disease, those who had a missing anaemia classification and those who had a missing weight value at baseline. The estimated hazard ratio for death for patients who had a no CD4 count recorded near the baseline time also increased, but was closer toward one indicating a smaller effect.

The patient who had the highest likelihood displacement value for Models 1.c and 2.c (patient number 293) was 47 years old, was recorded as being in WHO stage III or earlier of HIV disease, had a CD4 count of  $< 50$  cells/ $\mu\text{L}$ , was not classified as anaemic and had a baseline weight  $\geq 60$  kg near the baseline time, enrolled for ART three days after starting TB treatment and initiated ART more than four months thereafter and was recorded to die just less than 18 months after enrolling for ART. Excluding this patient resulted in lower estimated hazard ratios of death (further from one and therefore greater effect) for the baseline effects and higher estimated hazard ratios of death (also further from one and therefore greater effect) for the time-varying effects of having a CD4 counts  $\geq 50$  cells/ $\mu\text{L}$  or no CD4 count near baseline for both Models 1.c and 2.c.

The patient who had the highest likelihood displacement value for Models 1.d and 2.d (patient number 228) was 23 years old, was recorded as being in WHO stage IV of HIV disease, had a CD4 count of 100-199 cells/ $\mu$ L, was classified as anaemic and had a baseline weight  $< 50$  kg near the baseline time, enrolled for ART 16 days before starting TB treatment and initiated ART about three months after starting TB treatment. Despite having a CD4 count of  $\geq 200$  cells/ $\mu$ L, no or mild anaemia and a weight of  $\geq 60$  kg at the most recent measurement times prior to death, this patient died about a year after starting TB treatment. Excluding this patient resulted in lower estimated hazard ratios for death (closer to one and therefore less effect) for patients who were in stage IV of HIV and those who had a CD4 count in the category 100-199 cells/ $\mu$ L near baseline, higher estimated hazard ratios of death (further from one and therefore greater effect) for patients having CD4 counts  $\geq 200$  cells/ $\mu$ L, mild or no anaemia or no haemoglobin measurement, weight  $\geq 50$  kg or no weight measurement near baseline and lower estimated hazard ratios of death (further from one and therefore more effect) for patients having most recent weight values  $\geq 50$  kg or no recent weight value.

Although the above show some sensitivity of the magnitude of estimated effects to the exclusion of individual patients, the direction of effect estimates and hence the overall message of the model results (in particular that being on ART or initiating ART sooner is beneficial) remains the same.

## 7.4 Weighted Models

The models presented in Tables 7.4 and 7.5 adjusted the crude effect of ART initiation for possible confounding due to the baseline values of the selected covariates. To also adjust for time-updated values, I use marginal structural models. This requires the estimation of inverse probability of treatment and censoring weights. It is only the effect estimate for the causal effect of interest that should be interpreted from a weighted model. I have indicated this by drawing a horizontal line below the effect estimate of the treatment function of interest for each of the weighted models. I have chosen to use the treatment function ‘On ART’ to demonstrate the individual effects of inverse probability of treatment and inverse probability of censoring weights respectively and will only show tables and discuss results for the weighted analysis including the treatment function ‘Weeks not on ART’ that uses the combined effect of

the inverse probability of treatment and censoring weights. Unless otherwise stated, all models use the stabilized inverse probability weights.

### 7.4.1 Inverse probability of treatment weights

In this section I describe the process of estimating the inverse probability of treatment weights. I then show the results from including these weights in the Cox regression models that estimate the effect that being on ART has on the time to death.

#### Estimating inverse probability of treatment weights

The denominator for the inverse probability of treatment weights (IPTWs) was estimated using a pooled logistic regression (PLR) for the probability of initiating ART. All observations for a patient up to the first interval in which the patient initiated ART were included in the regression. The predicted value for being on ART for each of those observations was obtained using the equation:

$$\begin{aligned}
 \text{logit} [\text{Pr}(\text{on ART})] = & \underbrace{\beta_1 X_{1i}}_{\text{Male}} + \underbrace{\beta_2 X_{2i}}_{\text{Months from start of TB treatment to enrolment for ART}} \\
 & + \underbrace{\beta_3 X_{3i} + \dots + \beta_6 X_{6i}}_{\substack{\text{CD4 categories} \\ \text{vs } < 50 \text{ cells}/\mu\text{L} \\ \text{(baseline)}}} + \underbrace{\beta_7 X_{7i} + \beta_8 X_{8i}}_{\substack{\text{not anaemic} \\ \& \text{ missing} \\ \text{vs anaemic} \\ \text{(baseline)}}} + \underbrace{\beta_9 X_{9i} + \dots + \beta_{11} X_{11i}}_{\substack{\text{weight categories} \\ \text{vs } < 50 \text{ kg} \\ \text{(baseline)}}} \\
 & + \underbrace{\beta_{12} X_{12i} + \dots + \beta_{15} X_{15i}}_{\substack{\text{CD4 categories} \\ \text{vs } < 50 \text{ cells}/\mu\text{L} \\ \text{(updated)}}} + \underbrace{\beta_{16} X_{16i} + \beta_{17} X_{18i}}_{\substack{\text{not anaemic} \\ \& \text{ missing} \\ \text{vs anaemic} \\ \text{(updated)}}} + \underbrace{\beta_{19} X_{19i} + \dots + \beta_{21} X_{21i}}_{\substack{\text{weight categories} \\ \text{vs } < 50 \text{ kg} \\ \text{(updated)}}} \\
 & + \underbrace{\beta_{22} X_{22i} + \dots + \beta_{25} X_{25i}}_{\text{Spline Variables}},
 \end{aligned} \tag{7.5}$$

from which it follows that the predicted probability of being on ART for observation  $k$  of patient  $i$  is

$$\hat{p}_i(k) = \frac{\exp[\mathbf{X}_i \boldsymbol{\beta}_i(k)]}{1 + \exp[\mathbf{X}_i \boldsymbol{\beta}_i(k)]}.$$

Observations after the first interval in which a patient initiated ART were assigned a value of one, since patients who initiate ART are assumed to continue receiving it. The denominator of the inverse probability of treatment weight for patient  $i$  in interval  $k$  is then estimated as (Hernán et al., 2000):

$$\bar{\hat{p}}_i(k) = \begin{cases} \prod_{u=0}^k [1 - \hat{p}_i(u)] & \text{if the patient has not initiated ART by interval } k \\ \hat{p}_i(t) \prod_{u=0}^{t-1} [1 - \hat{p}_i(u)] & \text{if the patient initiated ART at interval } t, \end{cases} \quad (7.6)$$

where  $t \leq k$  and  $\hat{p}_i(u)$  is the predicted probability that the patient was on ART in interval  $u$ , as estimated in (7.5). The estimated numerator values were calculated in a similar way, but the most recent covariate ( $X_{12}$  to  $X_{21}$ ) values were not included in the prediction equation.

Table 7.7 shows pooled logistic regression approximations to the hazard ratios of ART initiation ( $\text{HR}^*$ ), as well as the robust 95% confidence intervals (CI) for each covariate included in the regressions to estimate the denominator and numerator probabilities of treatment used to create IPTWs.

The stabilized IPTWs were estimated as the value of the numerator divided by the value of the denominator at each time interval for each patient (Hernán et al., 2000):

$$\begin{aligned} sw_i(k) &= \prod_{u=0}^k \frac{\Pr[A_i(u) = a_i(u) | \bar{A}_i(u-1) = \bar{a}_i(u-1), L_i(0) = l_i(0), C_i(k) = 0]}{\Pr[A_i(u) = a_i(u) | \bar{A}_i(u-1) = \bar{a}_i(u-1), \bar{L}_i(u) = \bar{l}_i(u), C_i(k) = 0]} \\ &= \frac{\bar{\hat{p}}_i^N(k)}{\bar{\hat{p}}_i^D(k)}, \end{aligned}$$

where  $sw_i(k)$  is similar to Equation 6.10, but focuses on the  $i^{th}$  patient and takes into account that treatment assignment is modelled as occurring at discrete intervals; and  $\bar{\hat{p}}_i^N(k)$  and  $\bar{\hat{p}}_i^D(k)$  are the estimated values of the numerator and denominator (respectively) for the stabilized weight for patient  $i$  up to interval  $k$ . The estimated stabilized treatment weights ranged from 0.13 to 10.95 and had a mean of 1.01 (standard deviation (sd) 0.43).

Table 7.7: Pooled logistic regressions (PLR) used to estimate the inverse probability of treatment weights (IPTWs)

| Covariate                                                                 | Denominator |        |       | Numerator |             |
|---------------------------------------------------------------------------|-------------|--------|-------|-----------|-------------|
|                                                                           | HR*         | 95% CI |       | HR*       | 95% CI      |
| Gender                                                                    |             |        |       |           |             |
| Female                                                                    | 1.00        | -      |       | 1.00      | -           |
| Male                                                                      | 0.86        | 0.74 - | 1.01  | 0.88      | 0.75 - 1.03 |
| Months from start of TB treatment to enrolment for ART                    | 1.03        | 1.00 - | 1.06  | 1.03      | 1.00 - 1.05 |
| Baseline CD4 count                                                        |             |        |       |           |             |
| 0-49                                                                      | 1.00        | -      |       | 1.00      | -           |
| 50-99                                                                     | 0.99        | 0.24 - | 4.03  | 0.72      | 0.58 - 0.88 |
| 100-199                                                                   | 0.83        | 0.24 - | 2.90  | 0.70      | 0.57 - 0.85 |
| $\geq 200$                                                                | 1.55        | 0.35 - | 6.79  | 0.35      | 0.26 - 0.47 |
| missing                                                                   | 2.28        | 0.15 - | 34.20 | 0.46      | 0.28 - 0.77 |
| Anaemic at baseline                                                       |             |        |       |           |             |
| yes                                                                       | 1.00        | -      |       | 1.00      | -           |
| no                                                                        | 0.92        | 0.51 - | 1.67  | 0.82      | 0.69 - 0.97 |
| missing                                                                   | 2.05        | 0.22 - | 19.24 | 0.94      | 0.68 - 1.32 |
| Baseline Weight                                                           |             |        |       |           |             |
| (0; 50)                                                                   | 1.00        | -      |       | 1.00      | -           |
| [50; 60)                                                                  | 0.72        | 0.41 - | 1.26  | 1.08      | 0.85 - 1.37 |
| [60; $\infty$ )                                                           | 0.49        | 0.24 - | 1.00  | 1.34      | 1.04 - 1.72 |
| missing                                                                   | 4.68        | 0.84 - | 25.91 | 1.15      | 0.90 - 1.46 |
| Most recent CD4 count                                                     |             |        |       |           |             |
| 0-49                                                                      | 1.00        | -      |       |           | -           |
| 50-99                                                                     | 0.74        | 0.18 - | 3.03  |           | -           |
| 100-199                                                                   | 0.87        | 0.25 - | 3.03  |           | -           |
| $\geq 200$                                                                | 0.22        | 0.05 - | 0.96  |           | -           |
| missing                                                                   | 0.19        | 0.01 - | 2.95  |           | -           |
| Anaemic at most recent haemoglobin measurement                            |             |        |       |           |             |
| yes                                                                       | 1.00        | -      |       |           | -           |
| no                                                                        | 0.86        | 0.47 - | 1.56  |           | -           |
| missing                                                                   | 0.44        | 0.05 - | 4.17  |           | -           |
| Most recent Weight                                                        |             |        |       |           |             |
| (0; 50)                                                                   | 1.00        | -      |       |           | -           |
| [50; 60)                                                                  | 1.45        | 0.82 - | 2.55  |           | -           |
| [60; $\infty$ )                                                           | 2.91        | 1.43 - | 5.92  |           | -           |
| missing                                                                   | 0.25        | 0.04 - | 1.40  |           | -           |
| Spline for time from later of start of TB treatment and enrolment for ART |             |        |       |           |             |
| Variable 1                                                                | 1.01        | 1.00 - | 1.01  | 1.01      | 1.00 - 1.01 |
| Variable 2                                                                | 0.88        | 0.84 - | 0.91  | 0.89      | 0.86 - 0.93 |
| Variable 3                                                                | 1.49        | 1.30 - | 1.71  | 1.42      | 1.25 - 1.62 |
| Variable 4                                                                | 0.62        | 0.50 - | 0.77  | 0.65      | 0.52 - 0.80 |

HR\*: pooled logistic approximation to Cox proportional hazards regression

In a multiple logistic regression using observations of patients up to their first week on ART and including a spline for the time variable, as well as the baseline and most recent values for each of the categorical variables CD4 count, whether patient was anaemic and weight, the only ‘optional’ variables that had pooled logistic regression approximations to the hazard ratios that were significantly different from one at the 10% level were gender and time between the start of TB treatment and enrolment for ART. All other variables (age, whether patient was diagnosed for TB before 2006, whether patient was recorded as having WHO clinical stage IV of HIV disease (versus stages I to III), whether a patient had a new TB diagnosis (versus retreatment) and whether they had smear-positive pulmonary or extra-pulmonary TB (versus smear-negative pulmonary TB)) were therefore omitted from the model.

As an illustration of what the weights actually do, I have included Figure 7.5, which shows box plots of the stabilized IPTWs by baseline CD4 count category at four selected time points.

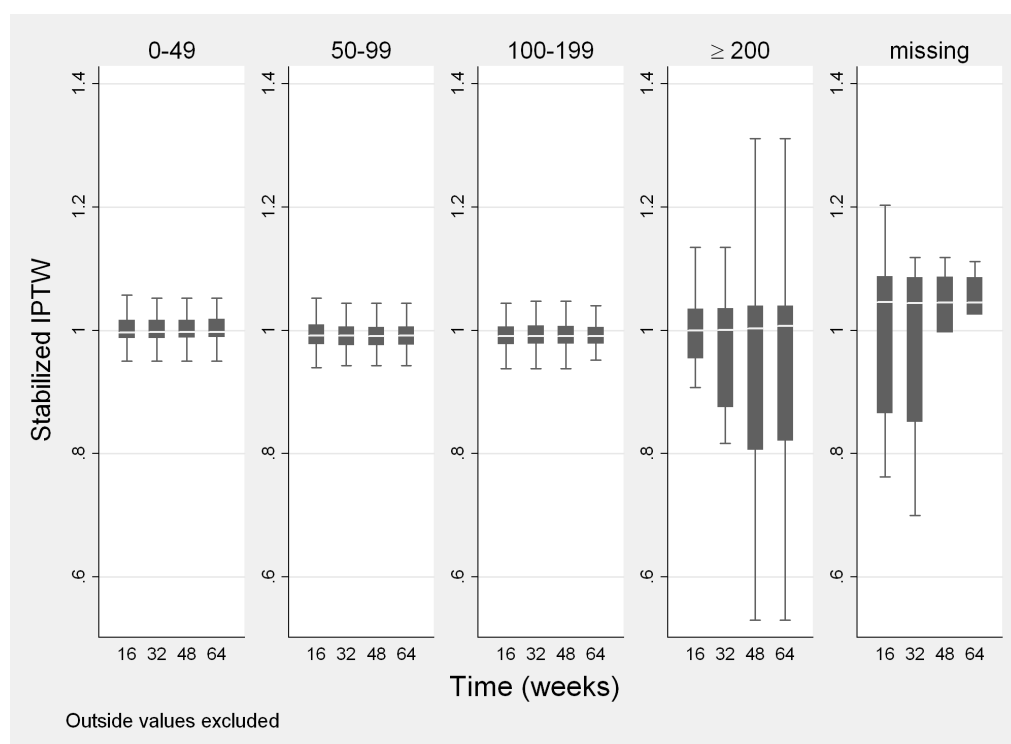


Figure 7.5: Box plots of stabilized IPTWs by baseline CD4 count category

For patients who had baseline CD4 counts in the categories 0-49, 50-99 and 100-199 cells/ $\mu$ L, the median and interquartile ranges of the stabilized

IPTWs were close to one for each of the selected four time points. The lower quartile for the stabilized IPTWs at later times for patients who had a baseline CD4 count categories  $\geq 200$  cells/ $\mu\text{L}$  were a below one, indicating that those patients contributed less than a complete patient to the pseudo-population. This is unsurprising, since it is expected that patients who had higher baseline CD4 count would initiate ART at later times, such that there would be a disproportionately large number of patients in this CD4 count category initiating ART at later times and so they would be down-weighted. The median stabilized IPTWs were slightly above one for patients who had no CD4 count recorded near baseline, probably because the overall proportion of these patients was very low and thus they needed to be up-weighted. However, the lower quartiles at earlier times are lower than one, indicating down-weighting for some patients in this group and up-weighting for others. Similar graphs and interpretation could have been given for box plots of the stabilized IPTWs by each of the other covariates included in the models for estimating the IPTWs.

### **Inverse probability of treatment weighted models**

Table 7.8 shows the estimated hazard ratios and 95% confidence intervals for adjusted models weighted using the stabilized IPTWs that have the same covariates as in models 1.b and 1.c reported in Table 7.4. The estimated hazard ratio of death for patients on ART is slightly greater (indicating a smaller survival benefit from being on ART) in the inverse probability of treatment weighted models compared to the unweighted models.

Table 7.8: Cox regressions using IPTW to assess effect on hazard of death of being on ART

|                                                              | Model 1.b IPTW      |             | Model 1.c IPTW        |             |                          |             |
|--------------------------------------------------------------|---------------------|-------------|-----------------------|-------------|--------------------------|-------------|
|                                                              | $e^{\hat{\beta}_j}$ | 95% CI      | $e^{\hat{\beta}_j^*}$ | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI      |
| On ART                                                       | 0.54                | 0.33 - 0.90 | 0.53                  | 0.33 - 0.86 | 1.00                     | -           |
| Months from start<br>of TB treatment to<br>Enrolment for ART | 1.02                | 0.97 - 1.08 | 1.12                  | 1.03 - 1.21 | 0.98                     | 0.98 - 0.99 |
| Age (years)                                                  | 1.02                | 1.00 - 1.04 | 1.02                  | 1.00 - 1.04 | 1.00                     | -           |
| WHO Clinical Stage                                           |                     |             |                       |             |                          |             |
| I-III                                                        | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| IV                                                           | 1.48                | 1.04 - 2.11 | 1.47                  | 1.03 - 2.10 | 1.00                     | -           |
| Baseline CD4 count                                           |                     |             |                       |             |                          |             |
| 0-49                                                         | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| 50-99                                                        | 0.43                | 0.26 - 0.71 | 0.23                  | 0.11 - 0.49 | 1.17                     | 1.01 - 1.35 |
| 100-199                                                      | 0.35                | 0.21 - 0.57 | 0.14                  | 0.06 - 0.32 | 1.24                     | 1.06 - 1.44 |
| $\geq 200$                                                   | 0.55                | 0.29 - 1.02 | 0.22                  | 0.08 - 0.61 | 1.23                     | 1.03 - 1.46 |
| missing                                                      | 0.87                | 0.37 - 2.05 | 0.53                  | 0.16 - 1.80 | 1.14                     | 0.91 - 1.43 |
| Anaemic at baseline                                          |                     |             |                       |             |                          |             |
| yes                                                          | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| no                                                           | 0.65                | 0.43 - 0.99 | 0.36                  | 0.19 - 0.67 | 1.14                     | 1.03 - 1.27 |
| missing                                                      | 1.90                | 1.03 - 3.52 | 1.44                  | 0.60 - 3.42 | 1.07                     | 0.91 - 1.26 |
| Baseline Weight                                              |                     |             |                       |             |                          |             |
| (0; 50)                                                      | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| [50; 60)                                                     | 0.58                | 0.34 - 0.97 | 0.56                  | 0.33 - 0.94 | 1.00                     | -           |
| [60; $\infty$ )                                              | 0.26                | 0.13 - 0.52 | 0.25                  | 0.13 - 0.50 | 1.00                     | -           |
| missing                                                      | 1.18                | 0.75 - 1.87 | 1.15                  | 0.73 - 1.82 | 1.00                     | -           |

$e^{\hat{\beta}_j}$ ,  $e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 7.4.

## 7.4.2 Inverse probability of censoring weights

In this section, I describe the process for estimating inverse probability of censoring weights and present the results from Cox regression models weighted using stabilized inverse probability of censoring weights.

### Estimating inverse probability of censoring weights

The outcome for patients who are censored is not known. Selection bias will be present if the probability of being censored is affected by prior treatment and/or covariate values. Inverse probability of censoring weights can be used to create a pseudo-population in which the probability of being censored is

not associated with prior treatment and covariate values. For calculating censoring weights, I considered patients who were permanently deferred for ART, transferred out or lost to follow up as censored. Administrative censoring (occurring 18 months after later of starting TB treatment or enrolling for ART) was not included in the calculation of IPCWs. Inverse probability of censoring weights (IPCWs) are created analogously to inverse probability of treatment weights, but the outcome for the pooled logistic regressions used to calculate them is whether the patient is censored, rather than whether they are on ART (which is included as a covariate). In addition, all observations for patients (not only those up to the first observation that the patient was on ART) are included in the pooled logistic regressions used for calculating IPCWs.

I began with a logistic regression model (using all observations of patients), which contained a spline for the time variable, as well as the baseline and most recent values for each of the variables CD4 count category, whether patient was anaemic and weight category and whether they were on ART. I then included other variables thought to potentially be predictive of a patient being permanently deferred for ART, transferring out or being lost to follow up (age, gender, WHO clinical stage IV of HIV disease, months from start of TB treatment to enrolment for ART, whether TB treatment was recorded as being completed or cured, calendar period of starting TB treatment, whether patient was being retreated for TB or had a new TB infection, whether patient had extra-pulmonary, smear-positive pulmonary or smear negative pulmonary TB). I found that the only variable that had an approximated hazard ratio significantly different from one when included in this model was the variable denoting calendar period of starting TB treatment (denominated in years and centred at 2005). It had an approximated hazard ratio of 1.47 (95% CI: 1.24 to 1.76). The selected denominator and numerator models are shown in Table 7.9.

Table 7.9: Pooled logistic regressions (PLR) used to estimate the inverse probability of censoring weights (IPCWs)

| Covariate                                                                 | Denominator |              | Numerator |             |
|---------------------------------------------------------------------------|-------------|--------------|-----------|-------------|
|                                                                           | HR*         | 95% CI       | HR*       | 95% CI      |
| On ART                                                                    | 0.94        | 0.54 - 1.62  | 0.97      | 0.57 - 1.64 |
| Year of starting TB treatment                                             | 1.48        | 1.24 - 1.76  | 1.47      | 1.24 - 1.74 |
| Baseline CD4 count                                                        |             |              |           |             |
| 0-49                                                                      | 1.00        | -            | 1.00      | -           |
| 50-99                                                                     | 0.64        | 0.35 - 1.16  | 0.59      | 0.37 - 0.95 |
| 100-199                                                                   | 0.89        | 0.53 - 1.48  | 0.94      | 0.65 - 1.38 |
| $\geq 200$                                                                | 0.94        | 0.47 - 1.89  | 1.07      | 0.63 - 1.83 |
| missing                                                                   | 2.49        | 0.65 - 9.60  | 1.52      | 0.48 - 4.77 |
| Anaemic at baseline                                                       |             |              |           |             |
| yes                                                                       | 1.00        | -            | 1.00      | -           |
| no                                                                        | 0.87        | 0.59 - 1.30  | 0.97      | 0.69 - 1.36 |
| missing                                                                   | 0.51        | 0.17 - 1.52  | 0.52      | 0.21 - 1.32 |
| Baseline Weight                                                           |             | -            |           | -           |
| (0; 50)                                                                   | 1.00        | -            | 1.00      | -           |
| [50; 60)                                                                  | 0.93        | 0.51 - 1.70  | 0.88      | 0.55 - 1.41 |
| [60; $\infty$ )                                                           | 0.87        | 0.41 - 1.86  | 0.81      | 0.49 - 1.33 |
| missing                                                                   | 1.30        | 0.15 - 11.17 | 0.58      | 0.35 - 0.96 |
| Most recent CD4 count                                                     |             |              |           |             |
| 0-49                                                                      | 1.00        | -            |           | -           |
| 50-99                                                                     | 0.80        | 0.37 - 1.75  |           | -           |
| 100-199                                                                   | 1.02        | 0.53 - 1.96  |           | -           |
| $\geq 200$                                                                | 1.16        | 0.55 - 2.45  |           | -           |
| missing                                                                   | 0.22        | 0.02 - 2.81  |           | -           |
| Anaemic at most recent haemoglobin measurement                            |             |              |           |             |
| yes                                                                       | 1.00        | -            |           | -           |
| no                                                                        | 1.34        | 0.80 - 2.25  |           | -           |
| missing                                                                   | 1.08        | 0.16 - 7.13  |           | -           |
| Most recent Weight                                                        |             |              |           |             |
| (0; 50)                                                                   | 1.00        | -            |           | -           |
| [50; 60)                                                                  | 0.92        | 0.45 - 1.88  |           | -           |
| [60; $\infty$ )                                                           | 0.87        | 0.37 - 2.04  |           | -           |
| missing                                                                   | 0.41        | 0.04 - 3.75  |           | -           |
| Spline for time from later of start of TB treatment and enrolment for ART |             |              |           |             |
| Variable 1                                                                | 0.99        | 0.98 - 0.99  | 0.99      | 0.98 - 1.00 |
| Variable 2                                                                | 1.05        | 0.99 - 1.11  | 1.05      | 0.99 - 1.11 |
| Variable 3                                                                | 0.94        | 0.80 - 1.10  | 0.93      | 0.79 - 1.09 |
| Variable 4                                                                | 0.96        | 0.80 - 1.16  | 0.97      | 0.81 - 1.17 |

HR\*: pooled logistic approximation to Cox proportional hazards regression

Again, I have chose to illustrate what the weights do using box plots of the stabilized IPCWs by CD4 count category at four selected time points as shown in Figure 7.6.

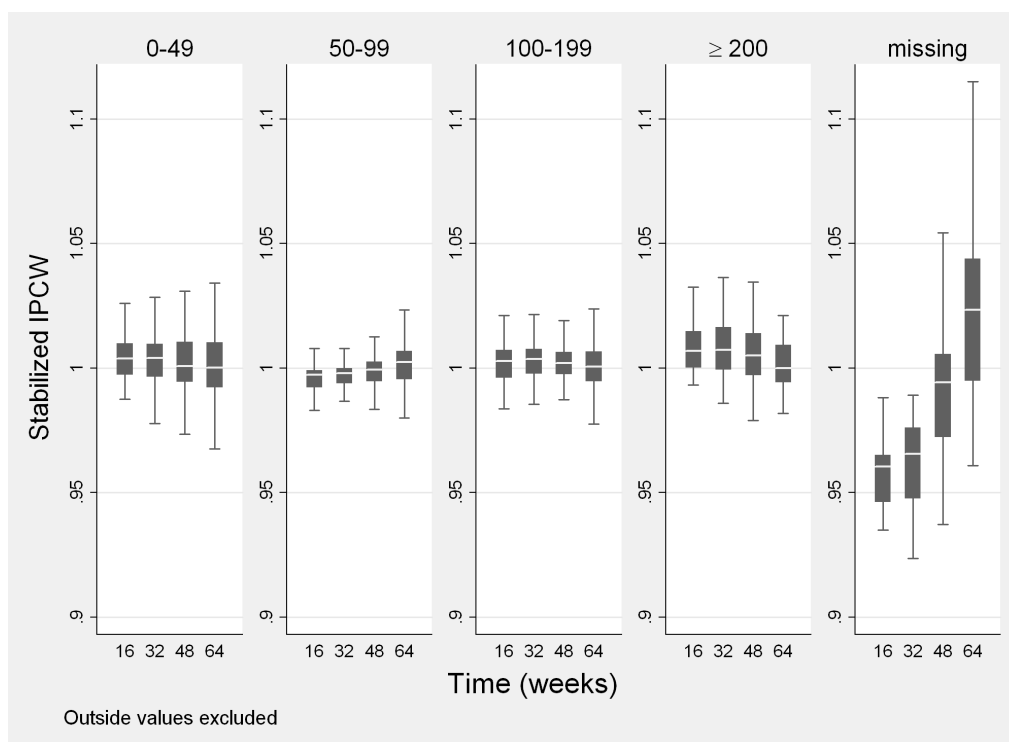


Figure 7.6: Box plots of stabilized IPCWs by baseline CD4 count category

The median and interquartile ranges of the stabilized inverse probability of censoring weights are close to one for most categories of baseline and time-updated CD4 counts. For patients who had no recorded baseline CD4 count, the median and interquartile ranges were below one at earlier times and higher at later times. This reflects down-weighting of these patients in the pseudo-population created by weighting patients by their IPCWs at earlier times and up-weighting at later times.

Figure 7.7 shows the median stabilized IPTWs and IPCWs over time (red line), as well as their interquartile ranges (light grey band) and the area between the 5th and 95th percentiles (dark grey). The medians remain at one over the duration of the analysis time. The distance between the interquartile values expands slightly and then remains fairly constant over time. The ranges from 5th to 95th percentile fan outward over time, indicating that

most of the weights are closer to one at earlier times and then become further from one, some below and others above. This is expected, since fewer patients have been treated or censored at earlier times. The distribution of the stabilized weights is more symmetric for the IPCWs than the IPTWs (for which the weights at the 5th percentile are further below than the distance above one for the weights at the 95th percentile).

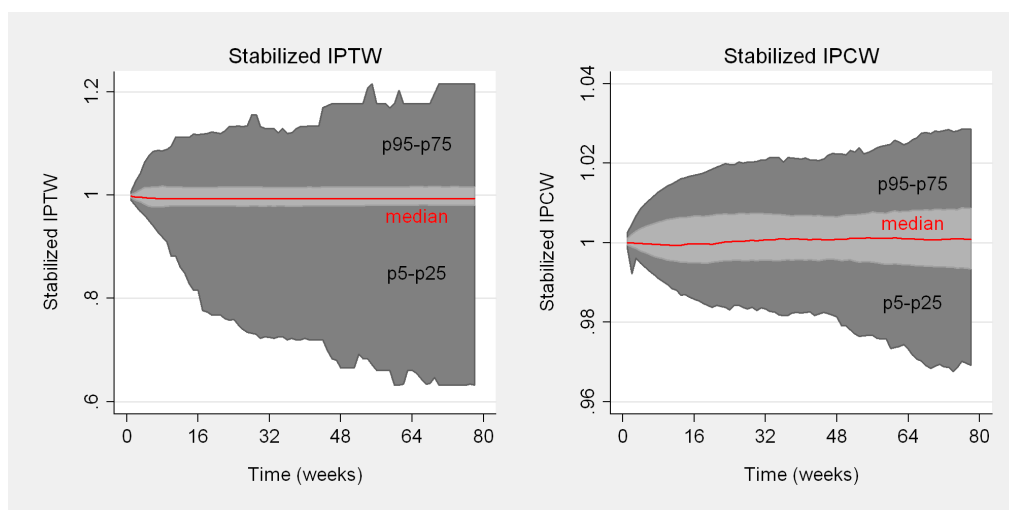


Figure 7.7: Median, interquartile range and 5th and 95th percentiles of stabilized IPTWs and IPCWs

### Inverse probability of censoring weighted models

Table 7.10 shows the results from models weighted using the stabilized inverse probability of censoring weights. The same variables are used as in the unweighted models 1.b and 1.c of Table 7.4. The hazard ratios of death for patients on ART are the same (to two decimal places) as those of the unweighted models, though the confidence intervals are slightly wider (as a result of using robust confidence intervals for the weighted model). This indicates that there is probably not selection bias by measured confounders.

Table 7.10: Cox regressions using IPCW to assess effect on hazard of death of being on ART

|                                                              | Model 1.b IPCW      |             | Model 1.c IPCW        |             |                          |             |
|--------------------------------------------------------------|---------------------|-------------|-----------------------|-------------|--------------------------|-------------|
|                                                              | $e^{\hat{\beta}_j}$ | 95% CI      | $e^{\hat{\beta}_j^*}$ | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI      |
| On ART                                                       | 0.50                | 0.30 - 0.82 | 0.48                  | 0.29 - 0.78 | 1.00                     | -           |
| Months from start<br>of TB treatment to<br>Enrolment for ART | 1.03                | 0.97 - 1.09 | 1.12                  | 1.04 - 1.21 | 0.98                     | 0.98 - 0.99 |
| Age (years)                                                  | 1.02                | 1.00 - 1.04 | 1.02                  | 1.00 - 1.04 | 1.00                     | -           |
| WHO Clinical Stage                                           |                     |             |                       |             |                          |             |
| I-III                                                        | 1.00                | -           | 1.00                  | -           |                          | -           |
| IV                                                           | 1.44                | 1.01 - 2.05 | 1.44                  | 1.01 - 2.05 | 1.00                     | -           |
| Baseline CD4 count                                           |                     |             |                       |             |                          |             |
| 0-49                                                         | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| 50-99                                                        | 0.41                | 0.25 - 0.68 | 0.23                  | 0.11 - 0.49 | 1.16                     | 1.01 - 1.34 |
| 100-199                                                      | 0.34                | 0.21 - 0.55 | 0.14                  | 0.06 - 0.32 | 1.22                     | 1.05 - 1.42 |
| $\geq 200$                                                   | 0.53                | 0.29 - 0.97 | 0.23                  | 0.08 - 0.62 | 1.21                     | 1.02 - 1.43 |
| missing                                                      | 0.83                | 0.36 - 1.91 | 0.48                  | 0.14 - 1.64 | 1.15                     | 0.92 - 1.43 |
| Anaemic at baseline                                          |                     |             |                       |             |                          |             |
| yes                                                          | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| no                                                           | 0.66                | 0.44 - 1.00 | 0.36                  | 0.19 - 0.66 | 1.15                     | 1.04 - 1.28 |
| missing                                                      | 1.95                | 1.06 - 3.58 | 1.44                  | 0.61 - 3.39 | 1.08                     | 0.92 - 1.27 |
| Baseline Weight                                              |                     |             |                       |             |                          |             |
| (0; 50)                                                      | 1.00                | -           | 1.00                  | -           | 1.00                     | -           |
| [50; 60)                                                     | 0.54                | 0.32 - 0.91 | 0.53                  | 0.32 - 0.89 | 1.00                     | -           |
| [60; $\infty$ )                                              | 0.27                | 0.14 - 0.53 | 0.27                  | 0.13 - 0.53 | 1.00                     | -           |
| missing                                                      | 1.22                | 0.77 - 1.91 | 1.20                  | 0.76 - 1.88 | 1.00                     | -           |

$e^{\hat{\beta}_j}$ ,  $e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 7.4.

### 7.4.3 Inverse Probability of Treatment and Censoring Weighted Models

Tables 7.11 and 7.12 show the result of weighting by the combined stabilized inverse probability of treatment and censoring weights (obtained by multiplying the stabilized IPTWs and IPCWs). Again the same variables as in Models 1.b and 1.c of Table 7.4 are used. The results are very similar to those of the unweighted models (Tables 7.4 and 7.5) and almost identical to those of the models weighted only using the IPTWs (Table 7.8 for treatment function ‘On ART’; results not shown for treatment function ‘Weeks not on ART’).

Table 7.11: Cox regressions using IPTCW to assess effect on hazard of death of being on ART

|                                                              | Model 1.b IPTCW     |        |        | Model 1.c IPTCW       |             |                          |             |
|--------------------------------------------------------------|---------------------|--------|--------|-----------------------|-------------|--------------------------|-------------|
|                                                              | $e^{\hat{\beta}_j}$ | 95% CI |        | $e^{\hat{\beta}_j^*}$ | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI      |
| On ART                                                       | 0.54                | 0.33   | - 0.89 | 0.53                  | 0.32 - 0.86 | 1.00                     | -           |
| Months from start<br>of TB treatment to<br>Enrolment for ART | 1.02                | 0.96   | - 1.08 | 1.12                  | 1.03 - 1.21 | 0.98                     | 0.98 - 0.99 |
| Age (years)                                                  | 1.02                | 1.00   | - 1.04 | 1.02                  | 1.00 - 1.04 | 1.00                     | -           |
| WHO Clinical Stage                                           |                     |        |        |                       |             |                          |             |
| I-III                                                        | 1.00                | -      |        | 1.00                  | -           | 1.00                     | -           |
| IV                                                           | 1.49                | 1.04   | - 2.12 | 1.47                  | 1.03 - 2.10 | 1.00                     | -           |
| Baseline CD4 count                                           |                     |        |        |                       |             |                          |             |
| 0-49                                                         | 1.00                | -      |        | 1.00                  | -           | 1.00                     | -           |
| 50-99                                                        | 0.43                | 0.26   | - 0.71 | 0.23                  | 0.11 - 0.49 | 1.17                     | 1.01 - 1.35 |
| 100-199                                                      | 0.35                | 0.21   | - 0.57 | 0.14                  | 0.06 - 0.32 | 1.23                     | 1.06 - 1.44 |
| $\geq 200$                                                   | 0.54                | 0.29   | - 1.02 | 0.22                  | 0.08 - 0.61 | 1.23                     | 1.03 - 1.46 |
| missing                                                      | 0.86                | 0.37   | - 2.01 | 0.51                  | 0.15 - 1.75 | 1.14                     | 0.91 - 1.44 |
| Anaemic at baseline                                          |                     |        |        |                       |             |                          |             |
| yes                                                          | 1.00                | -      |        | 1.00                  | -           | 1.00                     | -           |
| no                                                           | 0.66                | 0.43   | - 0.99 | 0.36                  | 0.19 - 0.67 | 1.14                     | 1.03 - 1.27 |
| missing                                                      | 1.92                | 1.05   | - 3.54 | 1.45                  | 0.61 - 3.41 | 1.08                     | 0.91 - 1.26 |
| Baseline Weight                                              |                     |        |        |                       |             |                          |             |
| (0; 50)                                                      | 1.00                | -      |        | 1.00                  | -           | 1.00                     | -           |
| [50; 60)                                                     | 0.58                | 0.34   | - 0.97 | 0.56                  | 0.33 - 0.93 | 1.00                     | -           |
| [60; $\infty$ )                                              | 0.26                | 0.13   | - 0.52 | 0.26                  | 0.13 - 0.51 | 1.00                     | -           |
| missing                                                      | 1.19                | 0.75   | - 1.89 | 1.16                  | 0.73 - 1.83 | 1.00                     | -           |

$e^{\hat{\beta}_j}$ ,  $e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 7.4.

Omitting the variables that were not significant at the 10% level from the IPT and IPC weighting models did not affect the mean and standard deviation of the overall IPTC weight, nor the estimated hazard ratios for the treatment functions, nor their standard errors (to two decimal places). Thus these models were not sensitive to the exclusion of variables for which effect estimates were not significant in the pooled logistic regression models.

Table 7.12: Cox regressions using IPTCW to assess effect on hazard of death of the number of weeks not on ART

|                                                              | Model 2.b IPTCW     |        |        | Model 2.c IPTCW       |             |                          |             |
|--------------------------------------------------------------|---------------------|--------|--------|-----------------------|-------------|--------------------------|-------------|
|                                                              | $e^{\hat{\beta}_j}$ | 95% CI |        | $e^{\hat{\beta}_j^*}$ | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI      |
| Weeks not on ART                                             | 1.04                | 1.02   | - 1.07 | 1.04                  | 1.01 - 1.07 | 1.00                     | -           |
| Months from start<br>of TB treatment to<br>Enrolment for ART | 1.03                | 0.96   | - 1.09 | 1.12                  | 1.03 - 1.21 | 0.98                     | 0.98 - 0.99 |
| Age (years)                                                  | 1.02                | 1.00   | - 1.04 | 1.02                  | 1.00 - 1.04 | 1.00                     | -           |
| WHO Clinical Stage                                           |                     |        |        |                       |             |                          |             |
| I-III                                                        | 1.00                | -      |        | 1.00                  | -           | 1.00                     | -           |
| IV                                                           | 1.55                | 1.08   | - 2.21 | 1.51                  | 1.06 - 2.16 | 1.00                     | -           |
| Baseline CD4 count                                           |                     |        |        |                       |             |                          |             |
| 0-49                                                         | 1.00                | -      |        | 1.00                  | -           | 1.00                     | -           |
| 50-99                                                        | 0.42                | 0.26   | - 0.70 | 0.25                  | 0.12 - 0.51 | 1.15                     | 0.99 - 1.34 |
| 100-199                                                      | 0.35                | 0.22   | - 0.58 | 0.14                  | 0.06 - 0.33 | 1.23                     | 1.06 - 1.43 |
| $\geq 200$                                                   | 0.52                | 0.27   | - 0.99 | 0.27                  | 0.10 - 0.75 | 1.16                     | 0.96 - 1.41 |
| missing                                                      | 0.85                | 0.37   | - 1.97 | 0.57                  | 0.17 - 1.92 | 1.12                     | 0.88 - 1.42 |
| Anaemic at baseline                                          |                     |        |        |                       |             |                          |             |
| yes                                                          | 1.00                | -      |        | 1.00                  | -           | 1.00                     | -           |
| no                                                           | 0.65                | 0.43   | - 0.99 | 0.36                  | 0.19 - 0.67 | 1.14                     | 1.03 - 1.26 |
| missing                                                      | 1.95                | 1.06   | - 3.61 | 1.40                  | 0.59 - 3.35 | 1.08                     | 0.92 - 1.29 |
| Baseline Weight                                              |                     |        |        |                       |             |                          |             |
| (0; 50)                                                      | 1.00                | -      |        | 1.00                  | -           | 1.00                     | -           |
| [50; 60)                                                     | 0.58                | 0.35   | - 0.99 | 0.54                  | 0.32 - 0.91 | 1.00                     | -           |
| [60; $\infty$ )                                              | 0.27                | 0.14   | - 0.54 | 0.25                  | 0.13 - 0.49 | 1.00                     | -           |
| missing                                                      | 1.23                | 0.78   | - 1.95 | 1.15                  | 0.73 - 1.80 | 1.00                     | -           |

$e^{\hat{\beta}_j}$ ,  $e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 7.4.

#### 7.4.4 Investigation of the distribution of the weights to check the weight model specification

Cole and Hernán (2008) suggest investigating the distribution of the weights to check the weight model specification, as well as assessing the sensitivity of the effect estimates to truncation of extreme weights. Truncating the weight estimates at the first and 99<sup>th</sup> percentile (replacing weights smaller and larger than the values of those for the first and 99<sup>th</sup> percentile with the values estimated for the first and the 99<sup>th</sup> percentile) did substantially reduce the standard deviation of the weights, as well as their range and brought the mean closer to 1. The hazard ratio of death for those on ART decreased slightly (further from one and closer toward the unweighted estimate). Truncating weights at increasing lower and decreasing upper percentiles continued

to result in smaller standard deviations and ranges of the estimated weights, however the effect estimates of being on ART or an additional weeks delay of ART and their standard errors were the same as that for the unweighted model (to two decimal places) when the weights were truncated at the 10<sup>th</sup> and 90<sup>th</sup> percentiles (truncation at the 50<sup>th</sup> percentiles is equivalent to not weighting). Unsurprisingly (since the weighted and unweighted effect estimates themselves are so similar) truncation of weights did not have a large effect.

Table 7.13: Analysis of precision and variance trade-off by truncating weights

| Truncation<br>Percentiles | Estimated Weights |        |              | On ART   |        | Weeks On ART |        |
|---------------------------|-------------------|--------|--------------|----------|--------|--------------|--------|
|                           | Mean              | (SD)   | Min / Max    | Estimate | (SE)   | Estimate     | (SE)   |
| 0, 100                    | 1.02              | (0.43) | 0.12 / 11.16 | 0.53     | (0.13) | 1.04         | (0.01) |
| 1, 99                     | 1.00              | (0.16) | 0.40 / 1.78  | 0.51     | (0.13) | 1.04         | (0.01) |
| 5, 95                     | 0.99              | (0.08) | 0.73 / 1.13  | 0.49     | (0.12) | 1.05         | (0.01) |
| 10, 90                    | 1.00              | (0.03) | 0.94 / 1.06  | 0.48     | (0.12) | 1.05         | (0.01) |
| 25, 75                    | 1.00              | (0.02) | 0.98 / 1.02  | 0.48     | (0.12) | 1.05         | (0.01) |
| 50, 50                    | 1.00              | (0.00) | 1.00 / 1.00  | 0.48     | (0.12) | 1.05         | (0.01) |

SD: Standard Deviation, SE: Standard Error. Truncation involves replacing very small and large weights with those of the lower and upper truncation percentiles shown. Truncation at the percentiles 0 and 100 implies no truncation of weights (i.e. estimates are those of the weighted model), while truncation at the 50<sup>th</sup> percentiles implies no weighting (i.e. estimates are those of the unweighted models).

## 7.5 Multiple Imputation

The different models in the previous sections focused on reducing possible biases due to time-dependent confounding and patient drop-out. In this section I focus on the bias that can arise as a result of missing data. I consider multiple imputation as a method to estimate effects that (under assumptions explained below) will not be biased by missing data and compare the sensitivity of effect estimates to different assumptions and methods for the handling of missing data.

I have chosen to show detailed model output only for unweighted models using the treatment function ‘On ART’ and including time-varying coefficients. Similar comparisons between complete case versus using missing categories and multiply imputed versus not multiply imputed were obtained when using the treatment function ‘Weeks not on ART’ and when using weighted models.

The first line of Table 7.14 shows that the estimated hazard ratio of death ( $e^{\hat{\beta}_j^*}$ ) for patients on ART compared to those not on ART is very different when patients not having baseline values are included by using missing categories versus when they are excluded and a complete case analysis is performed. Using multiple imputation allowed me to estimate the effect of the treatment function of interest on mortality using all available information on patients who were missing baseline values for one or more covariates, without having to use missing categories. These results can be compared with models that use missing categories and with estimates obtained using only complete cases.

Table 7.14: Comparison of unweighted Cox Regressions using treatment function ‘On ART’ using all patients (861) versus only complete cases (542)

|                                                        | All patients included |             |                          | Complete Cases |                          |             |
|--------------------------------------------------------|-----------------------|-------------|--------------------------|----------------|--------------------------|-------------|
|                                                        | $e^{\hat{\beta}_j^*}$ | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI         | $e^{\hat{\beta}_j^{**}}$ | 95% CI      |
| On ART                                                 | 0.48                  | 0.31 - 0.75 | 1.00                     | 0.45 - 2.24    | 1.00                     | -           |
| Months from start of TB treatment to Enrolment for ART | 1.12                  | 1.02 - 1.23 | 0.98                     | 0.97 - 0.99    | 1.16                     | 1.00 - 1.35 |
| Age (years)                                            | 1.02                  | 1.00 - 1.04 | 1.00                     | -              | 1.02                     | 0.99 - 1.05 |
| WHO Clinical Stage                                     |                       |             |                          |                |                          |             |
| I-III                                                  | 1.00                  | -           | 1.00                     | -              | 1.00                     | -           |
| IV                                                     | 1.43                  | 1.01 - 2.04 | 1.00                     | -              | 1.30                     | 0.74 - 2.28 |
| Baseline CD4 count                                     |                       |             |                          |                |                          |             |
| 0-49                                                   | 1.00                  | -           | 1.00                     | -              | 1.00                     | -           |
| 50-99                                                  | 0.23                  | 0.11 - 0.49 | 1.16                     | 1.01 - 1.33    | 0.22                     | 0.07 - 0.74 |
| 100-199                                                | 0.14                  | 0.06 - 0.31 | 1.22                     | 1.07 - 1.40    | 0.25                     | 0.08 - 0.78 |
| $\geq 200$                                             | 0.23                  | 0.09 - 0.60 | 1.21                     | 1.04 - 1.41    | 0.07                     | 0.00 - 1.03 |
| missing                                                | 0.49                  | 0.15 - 1.60 | 1.14                     | 0.93 - 1.41    | -                        | -           |
| Anaemic at baseline                                    |                       |             |                          |                |                          |             |
| yes                                                    | 1.00                  | -           | 1.00                     | -              | 1.00                     | -           |
| no                                                     | 0.36                  | 0.19 - 0.67 | 1.15                     | 1.03 - 1.28    | 0.22                     | 0.08 - 0.61 |
| missing                                                | 1.43                  | 0.62 - 3.31 | 1.08                     | 0.93 - 1.26    | -                        | -           |
| Baseline Weight                                        |                       |             |                          |                |                          |             |
| (0; 50)                                                | 1.00                  | -           | 1.00                     | -              | 1.00                     | -           |
| [50; 60)                                               | 0.54                  | 0.32 - 0.90 | 1.00                     | -              | 0.55                     | 0.30 - 1.01 |
| [60; $\infty$ )                                        | 0.27                  | 0.14 - 0.52 | 1.00                     | -              | 0.31                     | 0.14 - 0.69 |
| missing                                                | 1.19                  | 0.76 - 1.85 | 1.00                     | -              | -                        | -           |

$e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 7.4.

### 7.5.1 Predictors of missingness

My first step was to investigate whether observed values for any measured covariate were associated with a patient having a missing value for another variable. Table 7.15 shows crude estimates from logistic regressions for whether a patient has a baseline CD4 count, haemoglobin or weight value. The estimated odds ratios and their corresponding 95% confidence intervals have been coloured blue when the confidence intervals exclude one. Variables that had a marked effect on the presence or absence of a baseline CD4 count included the presence or absence of a baseline haemoglobin level and the indicators of death (pre-ART or after ART initiation) within 18 months of the baseline time. No patients who had missing baseline CD4 counts were recorded to transfer out within 18 months of the baseline time. Variables that had a marked effect on the presence or absence of a baseline haemoglobin value were the presence or absence of a baseline CD4 count, the calendar year of starting TB treatment and the indicators of death (pre-ART or after ART initiation) within 18 months of the baseline time. Variables that had a marked effect on the presence or absence of a baseline weight value were the value of the baseline haemoglobin level, the time from the start of TB treatment to enrolment for ART, the time from the later of starting TB treatment or enrolling for ART to ART initiation, the calendar year of starting TB treatment, the indicator of a new TB infection and the indicators of death (pre-ART or after ART initiation) within 18 months of the baseline time.

Table 7.15: Logistic regressions for indicators of missing baseline values

|                                                        | Measurement for which there are missing baseline values |                |                   |                |        |             |
|--------------------------------------------------------|---------------------------------------------------------|----------------|-------------------|----------------|--------|-------------|
|                                                        | CD4 count                                               |                | Haemoglobin level |                | Weight |             |
|                                                        | cOR                                                     | 95% CI         | cOR               | 95% CI         | cOR    | 95% CI      |
| No baseline value                                      |                                                         |                |                   |                |        |             |
| CD4 count                                              |                                                         | -              | 101.90            | 34.58 - 300.22 | 0.64   | 0.28 - 1.50 |
| Haemoglobin                                            | 101.90                                                  | 34.58 - 300.22 |                   | -              | 0.69   | 0.40 - 1.17 |
| Weight                                                 | 0.64                                                    | 0.28 - 1.50    | 0.69              | 0.40 - 1.17    |        | -           |
| Baseline values                                        |                                                         |                |                   |                |        |             |
| CD4 count (50 cells/ $\mu$ L) <sup>1</sup>             |                                                         | -              | 0.99              | 0.84 - 1.18    | 1.01   | 0.92 - 1.10 |
| Haemoglobin (g/dL) <sup>2</sup>                        | 1.30                                                    | 0.80 - 2.10    |                   | -              | 0.88   | 0.81 - 0.95 |
| Weight (kg) <sup>3</sup>                               | 1.00                                                    | 0.97 - 1.04    | 1.00              | 0.98 - 1.02    |        | -           |
| Months from start of TB treatment to Enrolment for ART | 1.00                                                    | 0.89 - 1.13    | 1.07              | 0.98 - 1.16    | 0.95   | 0.90 - 1.00 |
| Time from baseline to ART initiation                   |                                                         |                |                   |                |        |             |
| 1st quartile (baseline)                                | 1.00                                                    | -              | 1.00              | -              | 1.00   | -           |
| 2nd quartile                                           | 1.05                                                    | 0.36 - 3.06    | 1.27              | 0.64 - 2.50    | 0.67   | 0.44 - 1.03 |
| 3rd quartile                                           | 0.74                                                    | 0.23 - 2.38    | 1.19              | 0.60 - 2.36    | 0.58   | 0.38 - 0.90 |
| 4th quartile                                           | 1.25                                                    | 0.44 - 3.50    | 0.88              | 0.42 - 1.84    | 0.52   | 0.33 - 0.82 |
| Did not initiate ART                                   | 2.06                                                    | 0.67 - 6.33    | 2.09              | 0.98 - 4.45    | 1.40   | 0.84 - 2.34 |
| Gender                                                 |                                                         |                |                   |                |        |             |
| Female (baseline)                                      | 1.00                                                    | -              | 1.00              | -              |        | -           |
| Male                                                   | 0.80                                                    | 0.39 - 1.66    | 1.19              | 0.76 - 1.88    | 1.08   | 0.80 - 1.45 |
| Age (years)                                            | 0.96                                                    | 0.92 - 1.01    | 1.00              | 0.97 - 1.03    | 0.99   | 0.97 - 1.01 |
| WHO Stage of HIV                                       |                                                         |                |                   |                |        |             |
| Stage III (baseline)                                   | 1.00                                                    | -              | 1.00              | -              | 1.00   | -           |
| Stage IV                                               | 1.26                                                    | 0.61 - 2.59    | 1.24              | 0.77 - 1.98    | 1.26   | 0.92 - 1.72 |
| Year of TB treatment                                   | 0.78                                                    | 0.58 - 1.04    | 0.65              | 0.54 - 0.79    | 2.13   | 1.79 - 2.53 |
| Type of TB treatment                                   |                                                         |                |                   |                |        |             |
| New Infection                                          | 0.73                                                    | 0.35 - 1.50    | 0.88              | 0.54 - 1.42    | 1.54   | 1.10 - 2.16 |
| Retreatment (baseline)                                 | 1.00                                                    | -              | 1.00              | -              | 1.00   | -           |
| Type of TB                                             |                                                         |                |                   |                |        |             |
| SNP (baseline)                                         | 1.00                                                    | -              | 1.00              | -              | 1.00   | -           |
| SPP                                                    | 1.81                                                    | 0.68 - 4.79    | 1.39              | 0.76 - 2.55    | 1.04   | 0.70 - 1.54 |
| Extra-pulmonary                                        | 1.27                                                    | 0.47 - 3.38    | 1.11              | 0.61 - 2.01    | 1.15   | 0.79 - 1.67 |
| Outcome <sup>4</sup>                                   |                                                         |                |                   |                |        |             |
| In Care (baseline)                                     | 1.00                                                    | -              | 1.00              | -              | 1.00   | -           |
| Died pre-ART                                           | 4.20                                                    | 1.46 - 12.11   | 3.84              | 1.91 - 7.72    | 3.37   | 1.87 - 6.11 |
| Died on ART                                            | 3.89                                                    | 1.60 - 9.49    | 2.13              | 1.12 - 4.05    | 1.78   | 1.11 - 2.86 |
| Permanently Deferred for ART                           | 1.09                                                    | 0.14 - 8.48    | 0.31              | 0.04 - 2.33    | 0.95   | 0.44 - 2.08 |
| Transferred Out                                        | omitted                                                 |                | 0.74              | 0.22 - 2.48    | 0.97   | 0.49 - 1.92 |
| Lost to follow up                                      | 2.03                                                    | 0.66 - 6.27    | 0.58              | 0.20 - 1.66    | 0.84   | 0.48 - 1.47 |

N = 861 for all models not including baseline values of measurements (1: N=828, 2: N=777, 3: N=610), except the model for missing baseline CD4 count regressed on outcome (4) (as no patients with a missing baseline CD4 count were transferred out) Blue colouring indicates odds ratios for which the 95% CI excludes one (when not rounded to two decimal places), cOR: crude Odds Ratio, CI: Confidence Interval.

### 7.5.2 Missing data mechanisms

The associations seen between baseline CD4 count, haemoglobin and weight values being missing and the observed values of other variables suggest that the missing values cannot be assumed to be MCAR. It may be reasonable to assume that the values are MAR given the observed values of other variables, but this assumption cannot be tested (Schafer, 1997). This is because I cannot show that the probability of values being missing is independent of unobserved values (i.e. not MNAR) without having these unobserved values. I will therefore include as many relevant observed variables as possible (particularly those that are associated with the missingness of one or more variables) in the model used for imputation, in order to capture the predictive ability of observed variables. If the assumption of missing at random is appropriate, then conditional on the observed variables, those with a missing value for a particular variable will not be systematically different from those for whom a value was observed for that variable.

### 7.5.3 Implementation of Imputation

#### Transformation of variables

Figure 7.8 shows the distributions of the baseline CD4 cell counts, haemoglobin levels and weights (left column) and the logit transformation used for imputation for each of these values (right column) using histograms. The baseline CD4 counts are clearly not normally distributed, while the logit of the baseline CD4 counts show a slightly closer approximation to the normal distribution. I have chosen to use the logit transformation for CD4 counts, as this can be used to constrain imputed values to be between 0 and 600 cells/ $\mu$ L by dividing the observed baseline values by 600 before performing the logit transformation and then reversing these steps after imputation. This also avoids the problem of squaring negative imputed square-root of CD4 values, which may bias imputation results. For haemoglobin values, I subtracted 5 from the baseline value and then divided the results by 13, before using the logit transformation on these adjusted values, in order to constrain the baseline haemoglobin values that are imputed (after reversing the above steps on the imputed values) to be between 5 and 18 g/dL. Similarly, for weight values, I subtracted 25 from the baseline value and then divided it by 95, before using the logit transformation on these adjusted values, in order to constrain the imputed baseline weight values to be between 25 and 120 kg.

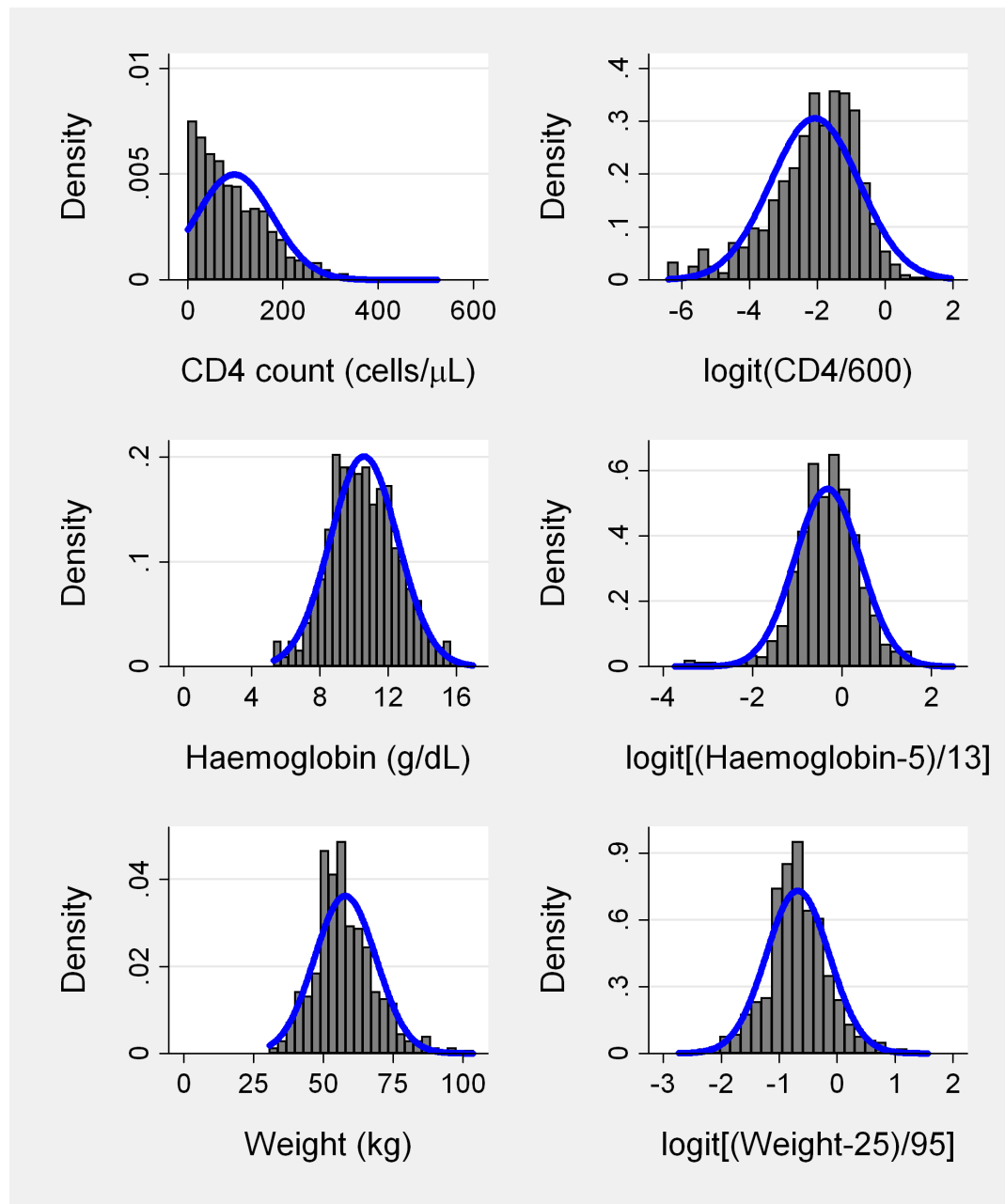


Figure 7.8: Distribution of variables

### Imputation methods and model

Two procedures for multiple imputation were performed using Stata version 11.2 (both using the same imputation equation as outlined below):

- I. Multivariate normal (MVN) data augmentation (DA) using the `mi` command suite in Stata
- II. Multiple Imputation by Chained Equations using the `ice` program developed for Stata by Royston (2005)

The underlying methodologies of these procedures were described in Section 6.8.3. The imputation equation used contained the logit transformation of the adjusted baseline CD4 count, haemoglobin level and weight as variables to be predicted. Variables used as predictors were as follows:

- Time from the start of TB treatment to enrolment for ART (months)
- Time from the baseline date to ART initiation (categorized into quartiles, with a fifth category for patients who do not initiate ART within 18 months)
- The number of days between the baseline date and the date on which each baseline measurement for the time-updated health indicators was taken (set to zero for patients for whom values for these indicators are being predicted)
- The age of the patient at baseline (years)
- The gender of the patient (male or female)
- The calendar year of starting TB treatment
- Whether the patient was being retreated for TB or had a new TB infection
- Whether the patient was in WHO clinical stage IV of HIV, versus stage III
- Whether the patient was diagnosed with smear positive pulmonary (SPP) TB or smear negative pulmonary (SNP) TB or extra-pulmonary (EP) TB
- The outcome category of the patient (death pre-ART, permanent deferral, death on ART, transferral to another clinic, loss to follow up or still in care 18 months after the baseline date)

- The Nelson-Aalen estimator for the cumulative hazard (using only one observation per patient) as an approximation of the outcome variable for the patient (as suggested by White and Royston (2009) - see Section 6.8.4)

The categorical variables were coded as dummy variables. Since 562 (63%) of 861 patients had baseline values recorded for all three measurements, the proportion of incomplete cases was just under 40% and I therefore chose to do 40 imputations, in line with the suggestion of White et al. (2011). A random seed of 1 was set in order to allow repetition of results. For the MVN data augmentation, the length of the burn-in period and the number of iterations between each value taken as an imputed value for  $\mathbf{Y}_{\text{mis}}^{(t)}$  was chosen to be 500 (see the explanation and discussion in Section 6.8.3 of the process to determine the required number of iterations between draws to achieve independence when using data augmentation).

#### 7.5.4 Multiple Imputation Results

The mean values for the baseline variables including imputed values (using either multivariate normal data augmentation or chained equations) were very close to the means for the patients having recorded baseline values. For weight, the mean baseline value was slightly lower when imputed values were included compared to using only recorded weight values. The 95% confidence intervals for the baseline values including imputed values of all three variables did include the mean baseline values calculated using only recorded values (see Table 7.16).

Table 7.16: Comparison of means before and after imputation

|                                                 | N   | Mean  | 95% CI         |
|-------------------------------------------------|-----|-------|----------------|
| <b>Observed Data</b>                            |     |       |                |
| CD4 count (cells/ $\mu$ L)                      | 828 | 98.33 | 92.87 - 103.79 |
| Haemoglobin (g/dL)                              | 777 | 10.58 | 10.45 - 10.72  |
| Weight (kg)                                     | 610 | 57.81 | 56.93 - 58.68  |
| <b>Data Imputed using MVN Data Augmentation</b> |     |       |                |
| CD4 count (cells/ $\mu$ L)                      | 861 | 98.33 | 92.81 - 103.85 |
| Haemoglobin (g/dL)                              | 861 | 10.59 | 10.45 - 10.74  |
| Weight (kg)                                     | 861 | 57.42 | 56.45 - 58.39  |
| <b>Data Imputed using Chained Equations</b>     |     |       |                |
| CD4 count (cells/ $\mu$ L)                      | 861 | 98.41 | 92.83 - 103.99 |
| Haemoglobin (g/dL)                              | 861 | 10.59 | 10.45 - 10.74  |
| Weight (kg)                                     | 861 | 57.38 | 56.48 - 58.28  |

Since categorical versions of these variables are used in the time to death models, I present a comparison of the proportions of patients in each category (of the variables for which multiple imputation was performed) among those having recorded values versus among the imputed values for those not having recorded values in Table 7.17. For CD4 count, there were greater proportions of patients in the highest and lowest categories among the imputed values compared to the recorded values. There was no change in the proportions of patients in different anaemia categories. For weight, there was a greater proportion of patients in the lowest category and there were lower proportions of patients in the other two categories, reflecting the slight decrease of the average mean weight for the imputed data sets compared to the mean for the recorded weight values.

Table 7.17: Comparison of proportions of patients in each category in recorded versus imputed data\*

|                     | Recorded Values |        | Imputed Values* |        |
|---------------------|-----------------|--------|-----------------|--------|
|                     | N               | (%)    | N               | (%)    |
| Baseline CD4 count  |                 |        |                 |        |
| 0-49                | 274             | (33%)  | 552             | (42%)  |
| 50-99               | 205             | (25%)  | 297             | (22%)  |
| 100-199             | 260             | (31%)  | 277             | (21%)  |
| $\geq 200$          | 89              | (11%)  | 194             | (15%)  |
| All                 | 828             | (100%) | 1320            | (100%) |
| Anaemic at baseline |                 |        |                 |        |
| Yes                 | 365             | (47%)  | 1589            | (47%)  |
| No                  | 412             | (53%)  | 1771            | (53%)  |
| All                 | 777             | (100%) | 3360            | (100%) |
| Baseline Weight     |                 |        |                 |        |
| (0; 50)             | 130             | (21%)  | 3106            | (31%)  |
| [50; 60)            | 253             | (41%)  | 3345            | (33%)  |
| [60; $\infty$ )     | 227             | (37%)  | 3589            | (36%)  |
| All                 | 610             | (100%) | 10040           | (100%) |

\*Only includes patients for whom measures were not recorded and combines all 40 data sets imputed using multivariate normal data augmentation

Table 7.18 compares results from unweighted Cox models for the observed data set versus the combined imputed data sets that were created using multivariate normal data augmentation.

Table 7.18: Comparison of unweighted Cox Regressions using treatment function ‘On ART’ for non-MI and MI data

|                                                        | Unweighted; observed data |             |                          | Unweighted; imputed using MVN DA |                          |             |
|--------------------------------------------------------|---------------------------|-------------|--------------------------|----------------------------------|--------------------------|-------------|
|                                                        | $e^{\hat{\beta}_j^*}$     | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI                           | $e^{\hat{\beta}_j^{**}}$ | 95% CI      |
| On ART                                                 | 0.48                      | 0.31 - 0.75 | 1.00                     | -                                | 0.47                     | 0.30 - 0.74 |
| Months from start of TB treatment to Enrolment for ART | 1.12                      | 1.02 - 1.23 | 0.98                     | 0.97 - 0.99                      | 1.12                     | 1.02 - 1.23 |
| Age (years)                                            | 1.02                      | 1.00 - 1.04 | 1.00                     | -                                | 1.01                     | 0.99 - 1.03 |
| WHO Clinical Stage                                     |                           |             |                          |                                  |                          |             |
| I-III                                                  | 1.00                      | -           | 1.00                     | -                                | 1.00                     | -           |
| IV                                                     | 1.43                      | 1.01 - 2.04 | 1.00                     | -                                | 1.42                     | 1.00 - 2.03 |
| Baseline CD4 count                                     |                           |             |                          |                                  |                          |             |
| 0-49                                                   | 1.00                      | -           | 1.00                     | -                                | 1.00                     | -           |
| 50-99                                                  | 0.23                      | 0.11 - 0.49 | 1.16                     | 1.01 - 1.33                      | 0.25                     | 0.12 - 0.52 |
| 100-199                                                | 0.14                      | 0.06 - 0.31 | 1.22                     | 1.07 - 1.40                      | 0.15                     | 0.07 - 0.33 |
| $\geq 200$                                             | 0.23                      | 0.09 - 0.60 | 1.21                     | 1.04 - 1.41                      | 0.22                     | 0.08 - 0.59 |
| missing                                                | 0.49                      | 0.15 - 1.60 | 1.14                     | 0.93 - 1.41                      | -                        | -           |
| Anaemic at baseline                                    |                           |             |                          |                                  |                          |             |
| yes                                                    | 1.00                      | -           | 1.00                     | -                                | 1.00                     | -           |
| no                                                     | 0.36                      | 0.19 - 0.67 | 1.08                     | 0.93 - 1.26                      | 0.41                     | 0.22 - 0.76 |
| missing                                                | 1.43                      | 0.62 - 3.31 | 1.15                     | 1.03 - 1.28                      | -                        | -           |
| Baseline weight                                        |                           |             |                          |                                  |                          |             |
| (0; 50)                                                | 1.00                      | -           | 1.00                     | -                                | 1.00                     | -           |
| [50; 60)                                               | 0.54                      | 0.32 - 0.90 | 1.00                     | -                                | 0.51                     | 0.32 - 0.81 |
| [60; $\infty$ )                                        | 0.27                      | 0.14 - 0.52 | 1.00                     | -                                | 0.29                     | 0.15 - 0.57 |
| missing                                                | 1.19                      | 0.76 - 1.85 | 1.00                     | -                                | -                        | -           |

$e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 7.4.

The largest differences were observed for categories of baseline anaemia and weight, which is not surprising, since they are the variables for which there were more missing baseline values. The hazard ratio of death at baseline for patients who were not anaemic increased (closer towards one) indicating a smaller reduction in the hazard of death for these patients compared to those who are anaemic when imputed values were used rather than a missing category. The 95% confidence interval was also slightly wider and higher. The hazard ratio for the time-varying effect of anaemia, however, decreased (closer toward one) and the 95% confidence interval limit became lower (closer toward one), thus indicating a smaller dampening effect over time on the survival benefit for patients who were not anaemic at baseline compared to those who were anaemic. The hazard ratio for patients having baseline weights in the category  $[50; 60)$  decreased (further from one) indicating a slightly increased reduction in the hazard of death for patients in this higher category compared to the category  $(0; 50)$ . The upper limit of the 95% confidence interval also decreased. For patients with weights in the category  $[60; \infty)$ , the hazard ratio of death increased slightly (closer toward one) indicating a slightly smaller reduction in the hazard of death for patients in this category compared to patients in the category  $(0; 50)$  when using the multiply imputed data rather than missing categories. The upper and lower limits of the 95% confidence interval both increased slightly, but the upper limit by more such that the interval became slightly wider.

Table 7.19 shows the results of models using continuous versions of the variables CD4 count, haemoglobin and weight for a complete case analysis and for the multiply imputed data. The estimated hazard ratio of death for patients on ART is very close to one in the complete case analysis, but is much lower than one and has a 95% confidence interval also far below one for the model estimated using the multiply imputed data. The hazard ratio of death for patients having an increased time from the start of TB treatment to enrolment for ART decreased slightly (closer toward one). The hazard ratio of death for patients who are in WHO clinical stage IV of HIV disease increased (further from one), but was still not significantly different from one. The hazard ratio of death per one g/dL increase in haemoglobin level increased (became closer to one), while the hazard ratios of death for an increase in CD4 count and weight hardly changed.

The results for the multiply imputed data shown used multivariate normal data augmentation, but results using chained equations were very similar. The results were also robust against small changes to the imputation equation (such as leaving out one of the variables).

Table 7.19: Comparison of unweighted Cox Regressions using treatment function ‘On ART’ and continuous values of baseline variables for non-MI (complete case,  $n = 542$ ) and MI ( $n = 861$ ) data

|                                                        | Unweighted; observed data |             |                          |             | Unweighted; imputed using MVN DA |             |                          |             |
|--------------------------------------------------------|---------------------------|-------------|--------------------------|-------------|----------------------------------|-------------|--------------------------|-------------|
|                                                        | $e^{\hat{\beta}_j^*}$     | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI      | $e^{\hat{\beta}_j^*}$            | 95% CI      | $e^{\hat{\beta}_j^{**}}$ | 95% CI      |
| On ART                                                 | 0.96                      | 0.43 - 2.13 | 1.00                     | -           | 0.44                             | 0.28 - 0.70 | 1.00                     | -           |
| Months from start of TB treatment to Enrolment for ART | 1.15                      | 1.00 - 1.33 | 0.99                     | 0.97 - 1.01 | 1.12                             | 1.03 - 1.22 | 0.99                     | 0.98 - 1.00 |
| Age (years)                                            | 1.03                      | 1.00 - 1.06 | 1.00                     | -           | 1.02                             | 1.00 - 1.04 | 1.00                     | -           |
| WHO Clinical Stage                                     |                           | -           | -                        | -           |                                  | -           | -                        | -           |
| I-III                                                  | 1.00                      | -           | 1.00                     | -           | 1.00                             | -           | 1.00                     | -           |
| IV                                                     | 1.27                      | 0.72 - 2.23 | 1.00                     | -           | 1.35                             | 0.95 - 1.94 | 1.00                     | -           |
| Baseline CD4 count (50 cells/ $\mu$ L)                 | 0.60                      | 0.42 - 0.85 | 1.06                     | 1.02 - 1.10 | 0.60                             | 0.49 - 0.75 | 1.06                     | 1.03 - 1.09 |
| Baseline Haemoglobin level (g/dL)                      | 0.70                      | 0.56 - 0.89 | 1.03                     | 0.99 - 1.07 | 0.79                             | 0.68 - 0.93 | 1.02                     | 0.99 - 1.05 |
| Baseline Weight (kg)                                   | 0.95                      | 0.92 - 0.98 | 1.00                     | -           | 0.95                             | 0.92 - 0.97 | 1.00                     | -           |

$e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 7.4.

### 7.5.5 Model checking using imputed data-sets

I plotted Cox-Snell residuals versus the cumulative hazard and deviance residuals versus the linear predictor for a random sample of the imputed data sets for Models 1.c and 2.c. These plots (not shown) were similar to those obtained using the unimputed data set.

### 7.5.6 Sensitivity of results to assumptions about missing values

I also did a brief sensitivity analysis to explore the effect on results of assuming that missing values were all very low (CD4 count  $< 50$  cells/ $\mu$ L, anaemic, weight  $< 50$  kg, i.e. patients have poor health) or all very high (CD4 count  $\geq 200$  cells/ $\mu$ L, not anaemic, weight  $\geq 60$  kg, i.e. patients health is relatively good), as well as replacing the CD4 count category  $\geq 200$  cells/ $\mu$ L with the category 100-199 cells/ $\mu$ L for the high assumption (since it has been noted elsewhere that patients having CD4 counts  $\geq 200$  cells/ $\mu$ L who are enrolled for ART are often quite sick in some way that is not represented by their CD4 count. Thus in this dataset, patients who have CD4 counts in the category 100-199 cells/ $\mu$ L may actually be healthier at baseline than those who are recorded as having CD4 counts  $\geq 200$  cells/ $\mu$ L), which will be referred to as high\*. The hazard ratios and corresponding standard errors are shown in Table 7.20.

Compared to the effect estimates from the model using missing categories, the effect estimates from the model for which patients who had missing values were grouped with patients who had values in the lowest category for that measurement (sickest patients) were similar, but the effect estimates from the models for which patients who had missing values were grouped with patients who had higher values for those measurements (indicating relatively healthy patients) differed quite a bit. The effect of ART on the relative hazard of death was insensitive to the assumption made about missing values, except for a slightly lower hazard ratio when missing CD4 counts were assumed to be in the 100-199 range (and missing haemoglobin measurements were assumed to reflect no anaemia and missing weights assumed to be  $\geq 60$  kg).

Table 7.20: Comparison of unweighted Cox Regressions using treatment function ‘On ART’ for different assumptions about missing values

|                                                              | Observed<br>$e^{\hat{\beta}_j^*}$ ( SE ) | Low<br>$e^{\hat{\beta}_j^*}$ ( SE ) | High<br>$e^{\hat{\beta}_j^*}$ ( SE ) | High*<br>$e^{\hat{\beta}_j^*}$ ( SE ) |
|--------------------------------------------------------------|------------------------------------------|-------------------------------------|--------------------------------------|---------------------------------------|
| On ART                                                       | 0.48 (0.11)                              | 0.48 (0.11)                         | 0.48 (0.11)                          | 0.44 (0.10)                           |
| Months from start<br>of TB treatment to<br>Enrolment for ART | 1.12 (0.05)                              | 1.13 (0.05)                         | 1.10 (0.05)                          | 1.10 (0.05)                           |
| Age (years)                                                  | 1.02 (0.01)                              | 1.02 (0.01)                         | 1.01 (0.01)                          | 1.01 (0.01)                           |
| WHO Clinical Stage                                           |                                          |                                     |                                      |                                       |
| I-III                                                        | 1.00 ( - )                               | 1.00 ( - )                          | 1.00 ( - )                           | 1.00 ( - )                            |
| IV                                                           | 1.43 (0.26)                              | 1.45 (0.26)                         | 1.59 (0.28)                          | 1.60 (0.28)                           |
| Baseline CD4 count                                           |                                          |                                     |                                      |                                       |
| 0-49                                                         | 1.00 ( - )                               | 1.00 ( - )                          | 1.00 ( - )                           | 1.00 ( - )                            |
| 50-99                                                        | 0.23 (0.09)                              | 0.26 (0.10)                         | 0.24 (0.09)                          | 0.23 (0.09)                           |
| 100-199                                                      | 0.14 (0.06)                              | 0.16 (0.06)                         | 0.15 (0.06)                          | 0.28 (0.09)                           |
| $\geq 200$                                                   | 0.23 (0.11)                              | 0.25 (0.12)                         | 0.47 (0.18)                          | 0.18 (0.08)                           |
| missing                                                      | 0.49 (0.30)                              | -                                   | -                                    | -                                     |
| Anaemic at baseline                                          |                                          |                                     |                                      |                                       |
| yes                                                          | 1.00 ( - )                               | 1.00 ( - )                          | 1.00 ( - )                           | 1.00 ( - )                            |
| no                                                           | 0.36 (0.11)                              | 0.35 (0.11)                         | 0.46 (0.13)                          | 0.48 (0.13)                           |
| missing                                                      | 1.43 (0.61)                              | -                                   | -                                    | -                                     |
| Baseline weight                                              |                                          |                                     |                                      |                                       |
| (0; 50)                                                      | 1.00 ( - )                               | 1.00 ( - )                          | 1.00 ( - )                           | 1.00 ( - )                            |
| [50; 60)                                                     | 0.54 (0.14)                              | 0.48 (0.10)                         | 0.51 (0.13)                          | 0.53 (0.14)                           |
| [60; $\infty$ )                                              | 0.27 (0.09)                              | 0.26 (0.08)                         | 0.74 (0.16)                          | 0.76 (0.17)                           |
| missing                                                      | 1.19 (0.27)                              | -                                   | -                                    | -                                     |
|                                                              | $e^{\hat{\beta}_j^{**}}$ ( SE )          | $e^{\hat{\beta}_j^{**}}$ ( SE )     | $e^{\hat{\beta}_j^{**}}$ ( SE )      | $e^{\hat{\beta}_j^{**}}$ ( SE )       |
| Months from start<br>of TB treatment to<br>Enrolment for ART | 0.98 (0.01)                              | 0.99 (0.01)                         | 0.99 (0.01)                          | 0.99 (0.01)                           |
| Baseline CD4 count                                           |                                          |                                     |                                      |                                       |
| 0-49                                                         | 1.00 ( - )                               | 1.00 ( - )                          | 1.00 ( - )                           | 1.00 ( - )                            |
| 50-99                                                        | 1.16 (0.08)                              | 1.13 (0.07)                         | 1.17 (0.08)                          | 1.17 (0.08)                           |
| 100-199                                                      | 1.22 (0.08)                              | 1.18 (0.07)                         | 1.22 (0.08)                          | 1.15 (0.07)                           |
| $\geq 200$                                                   | 1.21 (0.09)                              | 1.17 (0.09)                         | 1.17 (0.08)                          | 1.27 (0.08)                           |
| missing                                                      | 1.14 (0.12)                              | -                                   | -                                    | -                                     |
| Anaemic at baseline                                          |                                          |                                     |                                      |                                       |
| yes                                                          | 1.00 ( - )                               | 1.00 ( - )                          | 1.00 ( - )                           | 1.00 ( - )                            |
| no                                                           | 1.15 (0.06)                              | 1.11 (0.05)                         | 1.12 (0.06)                          | 1.11 (0.06)                           |
| missing                                                      | 1.08 (0.08)                              | -                                   | -                                    | -                                     |

$e^{\hat{\beta}_j^*}$  and  $e^{\hat{\beta}_j^{**}}$  are as in Table 7.4, SE: standard error.

Missing values replaced as follows:

Low: CD4 count  $< 50$  cells/ $\mu$ L, anaemic and weight  $< 50$  kg

High: CD4 count  $\geq 200$  cells/ $\mu$ L, not anaemic and weight  $\geq 60$  kg

High\*: CD4 count 100-199 cells/ $\mu$ L, not anaemic and weight  $\geq 60$  kg

The largest change under the ‘low’ assumptions (missing values correspond to sickest patients) were a decreased baseline hazard ratio of death for patients who had baseline weight values in the category 50-59 kg (suggesting a greater survival benefit of being in this category when the reference is the combined group of those who had no baseline weight value or weight  $< 50$  kg compared to the reference category being only those who had weight  $< 50$  kg). The values for the time-varying components of the hazard ratios for CD4 counts  $\geq 50$  cells/ $\mu$ L and those who were not anaemic were also lower, suggesting a smaller and thus slower dampening effect over time of being in these higher categories compared to the reference category when those who had no recorded values were included in the reference category.

Under the ‘high’ assumptions (missing values correspond to relatively healthy patients), the hazard ratio of death for patients in WHO stage IV (compared to those in WHO stages I-III) is greater than in the model using missing categories. The hazard ratios of death are also greater (and the time-varying components smaller where these are modelled) for the highest categories of CD4 count, anaemia classification and weight when these are grouped with the missing values for each of these variables, compared to when missing categories are used, indicating that the effect of these categories have now been diluted due to the addition of the patients with the missing values.

The changes in hazard ratios under the ‘high\*’ assumptions (missing values correspond to healthier patients, with CD4 counts in the range 100-199 cells/ $\mu$ L) were similar to (though usually greater than) those under the high assumptions, but differed for the hazard ratio of death for those on ART and the hazard ratios corresponding to CD4 counts. The hazard ratio of death for patients in the CD4 count category 100-199 cells/ $\mu$ L increased when patients who had no value for CD4 count were grouped into this category and the corresponding time varying effect became smaller, while the hazard ratio of death for patients in the CD4 count category  $\geq 200$  cells/ $\mu$ L decreased (and the corresponding time-varying effect increased). Again, the addition of missing values to a category dilutes the effect of that category.

This sensitivity analysis shows that it is unlikely that the missing values all belong to one of the other categories and also that they are probably not homogenous. It is for this reason that grouping them all together as a ‘missing’ category could create bias. Thus multiple imputation is probably the best approach. However, results (particularly for the effect of ART) are not very sensitive to the assumptions made for the missing values.

### 7.5.7 Summary of estimated hazard ratios of ART under varying modelling assumptions

To summarize and compare all the models fitted (using the same covariates as for Model 1.c of Table 7.4), Table 7.21 shows the unweighted and weighted estimated hazard ratios of death and their corresponding standard errors for the treatment functions ‘On ART’ and ‘Weeks not on ART’. I have used three decimal places in this table such that small differences in effect estimates can be seen.

Table 7.21: Comparison of unweighted and weighted Cox Regressions

|                                          | On ART                  |                       | Weeks not on ART        |                       |
|------------------------------------------|-------------------------|-----------------------|-------------------------|-----------------------|
|                                          | Unweighted<br>HR ( SE ) | Weighted<br>HR ( SE ) | Unweighted<br>HR ( SE ) | Weighted<br>HR ( SE ) |
| Missing categories used                  | 0.483 (0.109)           | 0.526 (0.130)         | 1.045 (0.012)           | 1.037 (0.015)         |
| MVN, categorical                         | 0.474 (0.108)           | 0.527 (0.129)         | 1.046 (0.012)           | 1.035 (0.014)         |
| MICE, categorical                        | 0.473 (0.107)           | 0.527 (0.129)         | 1.047 (0.012)           | 1.035 (0.014)         |
| Missing low <sup>†</sup> , categorical   | 0.481 (0.107)           | 0.524 (0.125)         | 1.048 (0.012)           | 1.038 (0.014)         |
| Missing high <sup>†</sup> , categorical  | 0.478 (0.109)           | 0.529 (0.135)         | 1.043 (0.012)           | 1.035 (0.013)         |
| Missing high* <sup>†</sup> , categorical | 0.444 (0.101)           | 0.496 (0.124)         | 1.042 (0.012)           | 1.034 (0.012)         |
| Categorical complete case                | 1.003 (0.410)           | 1.169 (0.566)         | 1.020 (0.022)           | 1.006 (0.019)         |
| Continuous complete case                 | 0.957 (0.390)           | 1.092 (0.524)         | 1.014 (0.022)           | 1.003 (0.022)         |
| MVN, continuous                          | 0.444 (0.102)           | 0.491 (0.120)         | 1.045 (0.012)           | 1.034 (0.012)         |
| MICE, continuous                         | 0.442 (0.101)           | 0.489 (0.119)         | 1.046 (0.012)           | 1.034 (0.012)         |

Covariates as for Model 1.c of Table 7.4, SE: standard error, MVN: Multivariate Normal imputation, MICE: Multiple Imputation by Chained Equations.

<sup>†</sup>Missing values replaced as follows:

Low: CD4 count < 50 cells/ $\mu$ L, anaemic and weight < 50 kg

High: CD4 count  $\geq$  200 cells/ $\mu$ L, not anaemic and weight  $\geq$  60 kg

High\*: CD4 count 100-199 cells/ $\mu$ L, not anaemic and weight  $\geq$  60 kg

In general, the effects of ART are estimated to be slightly greater in the unweighted models than in the weighted models. The effect estimates are similar under different model assumptions for both of the treatment functions, except for the complete case analyses, for which ART was estimated to have a very small (and sometimes harmful, but not significant) effect and the standard errors were very large.

The largest effect estimate for patients being on ART (smallest hazard ratio of death) was for the unweighted model using continuous values after multiple imputation by chained equations (the effect estimates were very close using multivariate normal imputation and under the high\* assumptions for missing

values). The largest effect estimate for an additional weeks delay in ART was for the unweighted model using categorical variables and the low assumptions for missing variables (the effect estimates were close for unweighted models using multiple imputation by chained equations or multivariate normal imputation with categorical or continuous versions of variables and for the model using missing categories).

The smallest estimated effect for being on ART (excluding complete cases models) was obtained for the weighted model using the high assumptions for missing values (effect estimates were very close for the models using categorical variables after multivariate normal imputation or multiple imputations by chained equations and for the model using missing categories). The smallest effect estimate for one weeks additional delay of ART was obtained for the model using the high\* assumptions for missing values (and was very similar for the models using continuous versions of variables after multivariate normal imputation or multiple imputation by chained equations).

In conclusion, all analyses indicate a beneficial effect of early ART and the estimated magnitude of this effect is similar (except for complete case analysis).

## 7.6 Plausibility of assumptions required for valid causal inference

In this section I describe how likely it is that the assumptions required for valid causal inference are satisfied. Although this discussion relates to the analysis I have performed - using weighted regressions (or marginal structural models) - these assumptions are required for valid causal inference from any study (whether or not there is a time-dependent treatment and whether weighting or alternative methods are used). For randomized controlled trials, the identifiability conditions (conditional exchangeability, positivity and consistency) can usually be assumed to hold (unless there is informative censoring), but for observational studies more care needs to be taken to ensure that these conditions hold (at least approximately) (Hernán and Robins, 2012). For both randomized controlled trials and observational studies, care must be taken to correctly specify models as far as is possible and to report sensitivity of results to model assumptions. Accurate information also cannot be guaranteed in either randomized controlled or observational studies.

Hernán and Robins (2012) state that it cannot ever be expected that these assumptions hold perfectly (for observational studies) and that a discussion of the assumptions is more important than a discussion about the methods used to obtain the causal effect estimates.

### 7.6.1 Conditional exchangeability

The assumption of conditional exchangeability requires that there be both no unmeasured confounding and no selection bias. No unmeasured confounding implies that sufficient variables that are predictors of both ART initiation and death are included in models such that any association between ART initiation and death is a result of the causal effect of ART on death and not a result of any common causes of ART initiation and death (Cole and Hernán, 2008). This assumption cannot be empirically verified, but researchers can use expert knowledge to include as many appropriate variables as possible to try to ensure that it holds at least approximately (thereby reducing any confounding bias in estimates as much as possible).

Lawn et al. (2011d) make the claim that patients who have more advanced immune suppression are more likely to start ART sooner during TB treatment and that observational data cannot fully adjust for how this decision is made. Although updated measures such as CD4 count, haemoglobin and weight may be reasonable proxies of the underlying health status of a patient, they are not able to fully capture it. It is quite possible that other unrecorded factors are also taken into account by the clinician who makes the decision about when a patient initiates ART. In the data set I am using it is unlikely that this assumption does hold (particularly given the lack of repeated measurements for values such as CD4 count prior to ART initiation). There may therefore be some bias in estimates as a result of residual confounding, but the adjustment for measured covariates should reduce the magnitude of overall confounding bias.

Cole and Hernán (2008) explain that it is not always optimal to include as many potential confounder variables as possible, as this may introduce selection bias (if non-confounding collider variables are stratified on - see Section 6.3.2) or finite-sample bias (if too many variables relative to the number of observations are included) and may decrease the statistical efficiency of the effect estimate (if variables that are actually not confounders are included in the model for the weights). The possibility of introducing finite-sample bias is related to the bias arising from non-positivity, as dis-

cussed below.

Other than conditioning on (adjusting for or stratifying on) the common effect of ART and death (or variables they are associated with), selection bias may arise through informative censoring. For example, patients who are lost to follow-up might not be well-represented by patients who remain in care. It is unlikely that there is no informative censoring for this study. I did attempt to adjust for this using inverse-probability of censoring weights, but these barely changed effect estimates. It is therefore likely that there is still residual bias from informative censoring that cannot be adjusted for using measured covariates.

### 7.6.2 No structural non-positivity

No structural non-positivity means that there should not be any combination of covariate levels for which it is guaranteed that either all patients will or will not receive treatment. Even if there is no structural non-positivity, non-positivity bias can still occur if there are many random zeroes (Cole and Hernán (2008) state that estimates from weighted regressions are more sensitive to random zeroes than usual regression estimates). There is a trade-off between reducing unmeasured confounding bias and the creation of bias through non-positivity. If many variables are used in an attempt to adjust for any potential confounding bias, this makes it less likely that there are patients both who do and do not receive treatment for every covariate level, particularly if many variables are continuous.

In Tables 7.22 and 7.23, I explore the possible extent of non-positivity for the data and models I am using. This is by no means an exhaustive analysis, but simply attempts to highlight some potential sources of bias. Table 7.22 shows the numbers and percentages of patients who initiated ART during the study period and the numbers and percentages of weeks that patients were on ART during the study. This table indicates that there are likely to be positivity problems associated with the missing categories, particularly for the missing CD4 count category, as there were very few patients who had no CD4 count recorded near the baseline time.

Table 7.22: Exploration of positivity by category

| Baseline<br>CD4 count<br>(cells/ $\mu$ L) | Baseline<br>Anaemia<br>Classification | Baseline<br>Weight<br>(kg) | Patients |    | Initiated ART<br>(%) | Patient-weeks |             |        |
|-------------------------------------------|---------------------------------------|----------------------------|----------|----|----------------------|---------------|-------------|--------|
|                                           |                                       |                            | All      | N  |                      | In study      | On ART<br>N | (%)    |
| 0-49                                      | anaemic                               | (0;50)                     | 31       | 30 | (97%)                | 1492          | 1282        | (86%)  |
|                                           |                                       | [50;60)                    | 37       | 33 | (89%)                | 1943          | 1747        | (90%)  |
|                                           |                                       | $\geq 60$                  | 19       | 19 | (100%)               | 1172          | 1111        | (95%)  |
|                                           |                                       | missing                    | 49       | 38 | (78%)                | 2091          | 1878        | (90%)  |
|                                           | no anaemia                            | (0;50)                     | 16       | 15 | (94%)                | 936           | 860         | (92%)  |
|                                           |                                       | [50;60)                    | 33       | 31 | (94%)                | 2246          | 2026        | (90%)  |
|                                           |                                       | $\geq 60$                  | 41       | 38 | (93%)                | 2650          | 2430        | (92%)  |
|                                           |                                       | missing                    | 31       | 27 | (87%)                | 1550          | 1426        | (92%)  |
|                                           | missing                               | (0;50)                     | 1        | 0  | (0%)                 | 10            | 0           | (0%)   |
|                                           |                                       | [50;60)                    | 4        | 3  | (75%)                | 213           | 197         | (92%)  |
|                                           |                                       | $\geq 60$                  | 6        | 5  | (83%)                | 315           | 299         | (95%)  |
|                                           |                                       | missing                    | 6        | 5  | (83%)                | 217           | 200         | (92%)  |
| 50-99                                     | anaemic                               | (0;50)                     | 15       | 15 | (100%)               | 915           | 820         | (90%)  |
|                                           |                                       | [50;60)                    | 33       | 31 | (94%)                | 2002          | 1819        | (91%)  |
|                                           |                                       | $\geq 60$                  | 14       | 14 | (100%)               | 1051          | 912         | (87%)  |
|                                           |                                       | missing                    | 28       | 25 | (89%)                | 1726          | 1532        | (89%)  |
|                                           | no anaemia                            | (0;50)                     | 12       | 12 | (100%)               | 765           | 677         | (88%)  |
|                                           |                                       | [50;60)                    | 26       | 24 | (92%)                | 1701          | 1450        | (85%)  |
|                                           |                                       | $\geq 60$                  | 39       | 38 | (97%)                | 2840          | 2536        | (89%)  |
|                                           |                                       | missing                    | 22       | 18 | (82%)                | 1450          | 1248        | (86%)  |
|                                           | missing                               | (0;50)                     | 1        | 1  | (100%)               | 78            | 73          | (94%)  |
|                                           |                                       | [50;60)                    | 4        | 4  | (100%)               | 312           | 256         | (82%)  |
|                                           |                                       | $\geq 60$                  | 7        | 7  | (100%)               | 528           | 504         | (95%)  |
|                                           |                                       | missing                    | 4        | 4  | (100%)               | 243           | 227         | (93%)  |
| 100-199                                   | anaemic                               | (0;50)                     | 19       | 19 | (100%)               | 1121          | 962         | (86%)  |
|                                           |                                       | [50;60)                    | 33       | 32 | (97%)                | 2265          | 2060        | (91%)  |
|                                           |                                       | $\geq 60$                  | 11       | 10 | (91%)                | 636           | 579         | (91%)  |
|                                           |                                       | missing                    | 39       | 38 | (97%)                | 2582          | 2295        | (89%)  |
|                                           | no anaemia                            | (0;50)                     | 14       | 13 | (93%)                | 859           | 751         | (87%)  |
|                                           |                                       | [50;60)                    | 42       | 35 | (83%)                | 2383          | 2027        | (85%)  |
|                                           |                                       | $\geq 60$                  | 53       | 50 | (94%)                | 3633          | 3194        | (88%)  |
|                                           |                                       | missing                    | 33       | 30 | (91%)                | 2163          | 1919        | (89%)  |
|                                           | missing                               | (0;50)                     | 5        | 3  | (60%)                | 167           | 110         | (66%)  |
|                                           |                                       | [50;60)                    | 5        | 4  | (80%)                | 296           | 269         | (91%)  |
|                                           |                                       | $\geq 60$                  | 4        | 4  | (100%)               | 294           | 276         | (94%)  |
|                                           |                                       | missing                    | 2        | 2  | (100%)               | 156           | 133         | (85%)  |
| $\geq 200$                                | anaemic                               | (0;50)                     | 3        | 3  | (100%)               | 234           | 166         | (71%)  |
|                                           |                                       | [50;60)                    | 12       | 11 | (92%)                | 825           | 674         | (82%)  |
|                                           |                                       | $\geq 60$                  | 5        | 4  | (80%)                | 338           | 239         | (71%)  |
|                                           |                                       | missing                    | 16       | 9  | (56%)                | 560           | 279         | (50%)  |
|                                           | no anaemia                            | (0;50)                     | 4        | 3  | (75%)                | 237           | 211         | (89%)  |
|                                           |                                       | [50;60)                    | 15       | 15 | (100%)               | 1063          | 777         | (73%)  |
|                                           |                                       | $\geq 60$                  | 15       | 10 | (67%)                | 790           | 686         | (87%)  |
|                                           |                                       | missing                    | 13       | 9  | (69%)                | 751           | 547         | (73%)  |
|                                           | missing                               | (0;50)                     | 2        | 1  | (50%)                | 87            | 70          | (80%)  |
|                                           |                                       | [50;60)                    | 1        | 1  | (100%)               | 78            | 75          | (96%)  |
|                                           |                                       | $\geq 60$                  | 2        | 2  | (100%)               | 156           | 114         | (73%)  |
|                                           |                                       | missing                    | 1        | 1  | (100%)               | 78            | 71          | (91%)  |
| missing                                   | anaemic                               | (0;50)                     | 0        | 0  | -                    | 0             | 0           | -      |
|                                           |                                       | [50;60)                    | 1        | 1  | (100%)               | 20            | 13          | (65%)  |
|                                           |                                       | $\geq 60$                  | 0        | 0  | -                    | 0             | 0           | -      |
|                                           |                                       | missing                    | 0        | 0  | -                    | 0             | 0           | -      |
|                                           | no anaemia                            | (0;50)                     | 0        | 0  | -                    | 0             | 0           | -      |
|                                           |                                       | [50;60)                    | 0        | 0  | -                    | 0             | 0           | -      |
|                                           |                                       | $\geq 60$                  | 2        | 2  | (100%)               | 98            | 90          | (92%)  |
|                                           |                                       | missing                    | 1        | 1  | (100%)               | 78            | 78          | (100%) |
|                                           | missing                               | (0;50)                     | 7        | 4  | (57%)                | 230           | 150         | (65%)  |
|                                           |                                       | [50;60)                    | 7        | 6  | (86%)                | 276           | 226         | (82%)  |
|                                           |                                       | $\geq 60$                  | 9        | 7  | (78%)                | 549           | 440         | (80%)  |
|                                           |                                       | missing                    | 6        | 6  | (100%)               | 402           | 314         | (78%)  |

Table 7.23: Exploration of positivity by CD4 category and weeks in study

| Baseline<br>CD4 count<br>(cells/ $\mu$ L) | Week | Patients |             |       |
|-------------------------------------------|------|----------|-------------|-------|
|                                           |      | In study | On ART<br>N | (%)   |
| 0-49                                      | 1    | 274      | 11          | (4%)  |
|                                           | 2    | 271      | 43          | (16%) |
|                                           | 3    | 261      | 79          | (30%) |
|                                           | 4    | 257      | 122         | (47%) |
|                                           | 5    | 250      | 158         | (63%) |
|                                           | 6    | 240      | 171         | (71%) |
|                                           | 7    | 236      | 183         | (78%) |
|                                           | 8    | 231      | 184         | (80%) |
|                                           | 9    | 228      | 188         | (82%) |
|                                           | 10   | 226      | 192         | (85%) |
|                                           | 11   | 221      | 193         | (87%) |
|                                           | 12   | 219      | 196         | (89%) |
| 50-99                                     | 1    | 205      | 6           | (3%)  |
|                                           | 2    | 204      | 19          | (9%)  |
|                                           | 3    | 202      | 32          | (16%) |
|                                           | 4    | 201      | 68          | (34%) |
|                                           | 5    | 200      | 98          | (49%) |
|                                           | 6    | 200      | 121         | (61%) |
|                                           | 7    | 199      | 133         | (67%) |
|                                           | 8    | 199      | 143         | (72%) |
|                                           | 9    | 199      | 154         | (77%) |
|                                           | 10   | 197      | 160         | (81%) |
|                                           | 11   | 197      | 162         | (82%) |
|                                           | 12   | 197      | 165         | (84%) |
| 100-199                                   | 1    | 260      | 4           | (2%)  |
|                                           | 2    | 257      | 13          | (5%)  |
|                                           | 3    | 250      | 22          | (9%)  |
|                                           | 4    | 248      | 71          | (29%) |
|                                           | 5    | 247      | 112         | (45%) |
|                                           | 6    | 245      | 128         | (52%) |
|                                           | 7    | 244      | 149         | (61%) |
|                                           | 8    | 243      | 164         | (67%) |
|                                           | 9    | 243      | 174         | (72%) |
|                                           | 10   | 241      | 183         | (76%) |
|                                           | 11   | 237      | 187         | (79%) |
|                                           | 12   | 235      | 193         | (82%) |
| $\geq 200$                                | 1    | 89       | 1           | (1%)  |
|                                           | 2    | 89       | 2           | (2%)  |
|                                           | 3    | 83       | 2           | (2%)  |
|                                           | 4    | 82       | 13          | (16%) |
|                                           | 5    | 79       | 22          | (28%) |
|                                           | 6    | 78       | 28          | (36%) |
|                                           | 7    | 78       | 31          | (40%) |
|                                           | 8    | 76       | 34          | (45%) |
|                                           | 9    | 75       | 41          | (55%) |
|                                           | 10   | 73       | 43          | (59%) |
|                                           | 11   | 73       | 45          | (62%) |
|                                           | 12   | 73       | 48          | (66%) |
| missing                                   | 1    | 33       | 2           | (6%)  |
|                                           | 2    | 32       | 3           | (9%)  |
|                                           | 3    | 30       | 4           | (13%) |
|                                           | 4    | 29       | 10          | (34%) |
|                                           | 5    | 29       | 14          | (48%) |
|                                           | 6    | 29       | 17          | (59%) |
|                                           | 7    | 29       | 18          | (62%) |
|                                           | 8    | 28       | 18          | (64%) |
|                                           | 9    | 28       | 18          | (64%) |
|                                           | 10   | 28       | 19          | (68%) |
|                                           | 11   | 27       | 19          | (70%) |
|                                           | 12   | 27       | 19          | (70%) |

Table 7.23 shows the numbers and percentages of patients on ART at each interval for the first 12 weeks of the study. For CD4 count  $\geq 200$  cells/ $\mu$ L and CD4 count missing near baseline, there are few patients who initiate ART in the first few weeks of the study. At later weeks, there are often few patients initiating ART in any category. This may cause positivity problems since there are a number of other baseline variables included in the model, so for some covariate levels there are likely to be no patients who are on ART in earlier weeks and no patients initiating ART in later weeks.

Table 7.24 shows a comparison of the means, standard deviations, minimum and maximum values for the weights and the weighted estimates and their standard errors for the two treatment functions using different weight model specifications.

If only baseline and time-updated CD4 count category are included in weight calculations (in addition to spline variables for time and for censoring weights whether the patient is on ART, Specification 3), the weighted regression estimates for the hazard ratios of ART on death and of the delay of ART for an additional week on death are very similar to those of the corresponding unweighted models.

Using only the baseline and time-updated weight category (Specification 4) rather than the CD4 count category to estimate weights yields weighted effect estimates very similar to those of the models weighted using the weights shown in Tables 7.7 and 7.9.

Using no covariates (other than the spline function of time and for the censoring model whether a patient is on ART) for the numerators of the weights gives weighted effect estimates that are similar to each other for models using either the original denominator (from Tables 7.7 and 7.9, Specification 5) or including in the denominator only baseline and time-updated CD4 counts in addition to the variables in the numerator (Specification 6). The effect estimates indicate a greater effect of being on ART or an additional week's delay of ART on the hazard of death. The mean weights for these models are further from one and have higher standard deviations and larger ranges than the weights that included baseline variables in the numerator. Using no extra variables in the numerator and only baseline and time-updated weight values in the denominator (Specification 7) results in weighted effect estimates that suggest a smaller effect of ART. The mean for these weight is closer to one and the standard deviation and range of the weights smaller than those of the model that included the baseline weight in the numerator.

Table 7.24: Analysis of bias and variance trade-off in different weighting model specifications

| No. | Specification<br>Description of variables included in IPWs                                   | Estimated Weights |        |             | On ART   |        | Weeks On ART |        |
|-----|----------------------------------------------------------------------------------------------|-------------------|--------|-------------|----------|--------|--------------|--------|
|     |                                                                                              | Mean              | (SD)   | Min / Max   | Estimate | (SE)   | Estimate     | (SE)   |
| 1   | Unweighted                                                                                   | 1.00              | (0.00) | 1.00 / 1.00 | 0.48     | (0.12) | 1.05         | (0.01) |
| 2   | Numerator and denominator for IPTWs and IPCWs<br>as shown in Tables 7.7 and 7.9 respectively | 1.02              | (0.43) | 0.12/11.16  | 0.53     | (0.13) | 1.04         | (0.01) |
| 3   | Numerators include baseline CD4<br>Denominators include baseline and updated CD4             | 1.00              | (0.18) | 0.08/ 6.81  | 0.48     | (0.12) | 1.04         | (0.01) |
| 4   | Numerators include baseline weight<br>Denominators include baseline and updated weight       | 1.02              | (0.46) | 0.13/11.85  | 0.54     | (0.13) | 1.04         | (0.01) |
| 5   | Numerators include no extra variables<br>Denominators as shown in Tables 7.7 and 7.9         | 1.03              | (0.83) | 0.06/20.71  | 0.44     | (0.13) | 1.04         | (0.01) |
| 6   | Numerators include no extra variables<br>Denominators include baseline and updated CD4       | 1.01              | (0.49) | 0.04/14.08  | 0.45     | (0.13) | 1.05         | (0.01) |
| 7   | Numerators include no extra variables<br>Denominators include baseline and updated weight    | 1.02              | (0.42) | 0.19/11.01  | 0.52     | (0.13) | 1.03         | (0.01) |

All models include a spline for the time variable. Models for IPCWs all include the variable indicating whether a patient is on ART or not. Descriptions only describe variables in addition to these. IPWs: Inverse Probability Weights, SD: Standard Deviation, SE: Standard Error. For specifications 4-7, where baseline variables were not included in the numerator, those baseline variables used in the denominator were excluded from the analysis model in order to avoid double adjustment.

The weighting model specification that gives weights with a mean closest to one and having the smallest standard deviation and range is Specification 3. This model, however, probably does not sufficiently adjust for time-dependent confounding, as a result of the lack of repeated CD4 count measures pre-ART. It could possibly be contended that the best weighting model would be Specification 7, which has a low standard deviation and range of weight values and is a simple model and possibly does adjust for time-dependent confounding, since it adjusts for time-updated weight values, which were the time-updated measurements for which there were the greatest number of values per patient pre-ART. I have, however, chosen to use Specification 2, which is a more comprehensive model that includes variables that were expected to be time-dependent confounders by clinicians, as well as other baseline variables, but results in similar mean and standard deviation estimates for the weights and similar weighted regression estimates and corresponding standard errors to those of Specification 7.

### 7.6.3 Consistency

The assumption of consistency implies that the observed outcome for a patient who initiates ART at a given time is the same as the outcome that would be obtained had that patient been assigned to initiate ART at the time they did initiate it, and relies on the initiation of ART being clearly defined. This seems a reasonable assumption for this study. Cole and Hernán (2008) affirm that assuming that consistency holds “is a reasonable assumption when estimating the effect of medical treatments.”

### 7.6.4 No measurement error

The assumption of no measurement error or no information bias relies on accurate recording of whether and when a patient initiates ART and dies, as well as accurate measurement of potential confounders (baseline and time-dependent). It is unlikely that there is no measurement error and the extent of the bias caused by any measurement error cannot be determined from the data. The model checking section (7.3) nonetheless shows that the results are not that sensitive to the exclusion of the most influential observations.

An assumption related both to measurement error and consistency is that once a patient initiates ART, they continue to take ART until the end of the

study. This would usually be true (for patients who are not lost to follow-up), but there may be some cases where there are treatment interruptions or poor adherence. If patients who are not receiving ART are assumed to be receiving it, this is likely to cause the estimated effect of ART initiation to be lower than its actual effect (biased towards the null).

### 7.6.5 No model misspecification

Valid causal inference requires that the weighting models (for both treatment (ART initiation) and censoring) and the analysis model (model that will be weighted) are correctly specified. I have attempted to choose the most appropriate models and also show (or report) results of varying forms of these models. Although effect estimates do change slightly, the same overall message was obtained in all models (other than analysis models based on patients with complete data). The analysis models using the continuous versions of CD4 and haemoglobin (and thus excluding patients who had missing values for those variables), but either excluding weight or using weight as a categorical variable with a missing category gave similar results to those models that used all three variables as categorical variables having missing categories.

Included in this assumption (and also related to the assumption of no unmeasured confounding) is the assumption that given the baseline and most recent values of time-updated variables, no other timing of these variables would be useful to predict ART initiation and death. Including the 26-week lagged CD4 count, anaemia classification or weight category in the denominators for the IPTWs and IPCWs hardly changed results (similarly for including lags for both CD4 and anaemia or both anaemia and weight - estimates from models including lags for both CD4 and weight could not be obtained).

## 7.7 Extent to which potential disadvantages of MSMs may affect my analysis

Some of the potential disadvantages of MSMs are the possibilities of very large weights for some patients (Robins and Hernán, 2009) or very variable and skewed distributions of weights, as well as the requirement to correctly specify the probability of a patient being treated or censored at each time interval conditional on the values of their past treatment and covariate histories (Hernán et al., 2000). For the analysis I performed, the range of stabilized

inverse probability of treatment and censoring weights was fairly small and there were no extremely large or small weights (see line 1 of Table 7.13). Although the standard deviation of the weights was not that large, the distribution of the inverse probability of treatment weights was slightly skewed (see Figure 7.7). This may mean that the true parameter will not always be included in the 95% confidence interval 95% of the time. It is possible that there is some bias as a result of model misspecification, although effect estimates did not vary greatly under different weighting model specifications (see Table 7.24), nor for weighted versus unweighted models (see Table 7.21). It is thus likely that the greatest residual source of bias is as a result of unmeasured variables.

## Chapter 8

# Conclusion for Outcome Analysis

In this section of the thesis, I demonstrated the use of marginal structural models using inverse probability of treatment and censoring weights to adjust for potential time-dependent confounding and selection bias. I also performed multiple imputation for baseline covariates and showed comparisons of different methods of adjusting for baseline covariates for which there are missing values. In all models other than the complete case models (for which many patients were excluded), the use of ART was estimated to have a beneficial effect on survival.

For the inverse probability of treatment and censoring weighted model that included time-varying covariates and used missing categories for unrecorded baseline measurements, the hazard ratio of death was 0.53 (95% CI: 0.32 to 0.86), which was very close to the crude unweighted effect estimate of 0.52 (95% CI: 0.33 to 0.81). Multiply imputing values for unrecorded measurements did not alter this effect estimate when categorical versions of baseline variables were used, but did allow estimation of a model using continuous versions of baseline variables, for which the hazard ratio of death was slightly smaller (0.49, 95% CI: 0.30 to 0.79), thus indicating a slightly greater survival benefit for patients on ART.

The hazard ratio of death for an additional week not on ART was 1.04 (95% CI: 1.01 to 1.07) in the IPTC weighted model including time-varying covariates and using missing categories for unrecorded baseline measurements, indicating an increased hazard of death for patients who had a longer delay

to ART initiation. Results were almost identical in the imputed model (1.04, 95% CI: 1.01 to 1.06), the unweighted models using missing categories or imputed values (1.05, 95% CI: 1.02 to 1.07) and the crude unweighted model (1.04, 95% CI: 1.02 to 1.07).

For the unweighted model allowing interactions of covariates with time, the following was observed: Patients who had a longer time from the start of TB treatment to ART initiation had a higher hazard of death. This effect increased over time (as indicated by the time-varying coefficient lower than one). Patients who had higher baseline CD4 counts and weight values and those who were not classified as anaemic at baseline had lower hazards of death at baseline compared to those who had CD4 counts  $< 50$  cells/ $\mu$ L, weight  $< 50$  kg or who were anaemic. Patients who had missing CD4 counts had lower hazards of death than patients who had CD4 counts  $< 50$  cell/ $\mu$ L, while those who had missing haemoglobin levels or weight measurements had higher hazards of death than patients who were classified as anaemic or had weight  $< 50$  kg. Over time, the baseline covariates CD4 count category and anaemia classification no longer had as much of an effect on the hazard of death (indicated in the models by time-varying coefficients greater than one). Patients who were in WHO clinical stage IV of HIV disease at baseline and those who were older had higher hazards of death.

Weighting models using inverse probability of treatment and censoring weights did not substantially change the estimates for the effect of ART. This suggests that there may not be time-dependent confounding nor selection bias by measured variables. Using multiply imputed values for missing observations also did not substantially alter results compared to including missing categories for variables for which measurements were not all recorded. It was nonetheless important to weight models by inverse probability of treatment and censoring weights and to perform multiple imputation in order to show that these methods were not required to reduce bias. Including missing categories may be a more informative way of presenting results to clinicians, since there may be interest in the outcomes of these patients. Reporting the relatively poor outcomes observed for such patients may help to encourage greater efforts to ensure that measurements are recorded for all patients (and that patients remain in care).

## 8.1 Problems arising from using an observational cohort for analysis

One of the main draw-backs of using an observational cohort is that the values of time-varying covariates (such as weight, CD4 count and WHO clinical stage of HIV) are not always available at time points of interest (such as the start of TB treatment or ART initiation). Nor are they measured as frequently or as regularly as would be desired. For patients who did not initiate ART, there were very few time-updated measurements available (often only one per variable). For patients who did initiate ART, there was also often only one measurement pre-ART per variable. Thus although CD4 count, weight and haemoglobin may in theory be reasonable proxies for the health state of the patient, in this data set they may not have been able to adequately represent the health status of patients as a result of the infrequent measurements. The large window interval during which a recorded measurement would be considered near to baseline (up to six months before or after the baseline date and not later than a week after initiating ART) may also introduce some bias. Comparisons of including patients who have no measurements recorded near baseline using missing categories versus using multiple imputation were presented above. Results were not very sensitive to which of these two methods was used, nor to the use of continuous versus categorical variables.

Defining the baseline time of interest and the time at which a patient was considered to be ‘at risk of ART initiation’ was not trivial. This is because ideally clinicians would like to know the optimal time after starting TB treatment that a patient should initiate ART, but in our study many patients had not yet enrolled for ART at the time of starting TB treatment and thus were not at risk of initiating ART in the time between starting TB treatment and enrolling for ART. In an ideal randomized controlled trial, all patients included in the study would be ‘at risk of initiating ART’ whether or not they met the current eligibility criteria (assuming equipoise surrounding this issue). Two options for dealing with this problem for the data I used were to ignore the information on when a patient enrolled for ART (thereby assuming that all patients could initiate ART at any time after starting TB treatment) and use the time of starting TB treatment as the baseline time, or to use the later of the date of starting TB treatment and enrolling for ART as the baseline time. I chose to use the latter.

The results from this study only apply to patients who both started TB

treatment and began the screening process for ART initiation (enrolled for ART). There is no information for patients who died or were censored after experiencing only one of these two events. Thus although the variable ‘Time from start of TB treatment to enrolment for ART’ is included in analyses, attempts to predict the time to death from the time of starting TB treatment or from the time of enrolment for ART using model estimates would probably be biased.

## 8.2 Ideal clinical trial

Although I have attempted to use statistical methods to overcome some of the biases inherent in observational studies (such as differential treatment assignment and lack of recorded values at desired time intervals), these methods cannot replace the need for adequate data. The gold standard method to measure causal effects of treatments on outcomes is to conduct a randomized controlled trial.

To determine the causal effect of the timing of ART initiation on survival for patients co-infected with HIV and TB that are being treated for TB, the following ideal randomized controlled trial design would be appropriate: Patients co-infected with TB and HIV that were eligible for ART at the time of TB diagnosis would be randomized to initiate ART two, four, six and eight weeks after starting TB treatment. Baseline patient characteristics (such as age, height and WHO clinical stage of HIV) would be recorded at the time of TB diagnosis. All patients would be seen at regular intervals irrespective of whether they had initiated ART or not. Relevant time-updated measurements such as CD4 count, haemoglobin level and weight would be taken at each visit (or at regular intervals). Clinicians would be requested to examine patients and record all observations, diagnoses, WHO clinical stage and Karnofsky score (a measure of functional ability that is of prognostic value). Dates of clinic visits, measurements (if different to dates of clinic visits) and other important events (such as ART initiation, development of TB IRIS and death) would be recorded. Information on the type of TB and on subsequent smears and cultures would be recorded, as well as the final TB outcome. Any other important information (such as details on adverse events and their management, or reasons for switching treatment) would be recorded. In an ideal trial, there would also be no patients who transfer out or are lost to follow-up. This cannot be guaranteed in the real world, but effort can be made to track patients lost to follow up and to continue to

obtain information about outcomes of patients who transfer out.

# Chapter 9

## Conclusion

I address the following:

1. The suitability of selected methods to cope with the design and data
2. Other methods that could have been used
3. The clinical objective and aims of the thesis

### 9.1 Suitability of selected methods to cope with the design and data

Part I of this thesis was focused on an exploratory and descriptive analysis (as opposed to a causal analysis in Part II), so the lack of randomization and values recorded for measurements at regular intervals was not problematic. I chose to categorize CD4 counts and created a missing category for patients for whom there were no recorded values near the time of starting TB treatment. This facilitated looking specifically at the group of patients for whom baseline values were unknown and comparing their times to ART initiation (or death) to that of patients for whom CD4 counts were known and were within specific categorizations commonly used by clinicians as part of the criteria for deciding when to initiate patients on ART. I used Cox regressions, parametric survival models and competing risks analysis to investigate the effect of different patient characteristics on the time to ART initiation.

The assumption of proportional hazards was not satisfied, so the standard Cox proportional hazards model could not be relied upon. Cox regressions

using time varying covariates and parametric regressions such as the Weibull (in proportional hazards and accelerated failure time frameworks) and the generalized Gamma accelerated failure time regressions were implemented instead. Cox regressions using time-varying covariates remove the need for the assumption of proportional hazards and are able to show both the effect at baseline and how this effect changes over time for each level of a covariate. This means that there can be differing slopes for different values of a covariate. The parametric models could accommodate the non-proportionality of the hazards by allowing shape (or ancillary) parameters to vary by covariate value, thus allowing differing shapes of the hazard function for different covariate values. Similar results were obtained for all models. The Cox regressions using time-varying covariates were perhaps easiest to interpret out of all models (not assuming proportional hazards) when results were presented in tabular form. The results from parametric models and particularly the generalized Gamma accelerated failure time model (which has two shape parameters) were easier to understand when displayed graphically than in tables. The model with the best fit was the generalized Gamma accelerated failure time model.

The probability of a patient initiating ART, however, could be affected by the probabilities of competing events (such as death pre-ART, permanent deferral for ART or the end of TB treatment, which result in patients no longer being ‘at risk’ of ART initiation before the end of TB treatment). I therefore decided to consider these competing risks explicitly. The competing risks analysis consisted of graphical comparisons of cause-specific hazards and cumulative incidence functions for each of the competing risks, as well as a comparison of regressions on the cause-specific hazards and sub-distribution hazards and a comparison of cumulative incidence functions under different proportionality assumptions. Although competing risks analysis did provide better understanding of the data, it was perhaps not essential in this analysis, since the number of competing events occurring was very small compared to the number of patients initiating ART. Had the time to death pre-ART been the focus of this analysis, exploration of the impact of competing risks would have been more critical, as would the choice of whether to estimate effects of covariates on the relative risk of death at a given point in time (using regression on the cause-specific hazard) or on the cumulative incidence of death before ART initiation, the end of TB treatment or permanent deferral (using regression on the sub-distribution hazard).

For Part II of this thesis, the aim was to obtain a valid causal estimate for the effect of the timing of ART initiation on the survival of patients. I

demonstrated the use of DAGs in aiding the understanding of causal structures. This helped me to illustrate the causal structure that I assumed for the data I am using and the potential sources of bias.

I thought that patients whose health deteriorated faster before ART initiation would be initiated on ART sooner than those whose health did not deteriorate so quickly. I hoped to be able to adjust for this potential time-dependent confounding bias by using inverse probability of treatment weights. Few patients, however, had more than one value for a specific measurement prior to ART initiation. I also expected that censoring might not be random and that the use of inverse probability of censoring weights would be able to adjust for this bias.

The combination of the lack of recorded values for specific measurements taken near the time(s) of interest and the lack of randomization of treatment assignment made the task of obtaining a valid causal estimate for the effect of the timing of ART initiation on survival complex. My first step in attempting to deal with the lack of random treatment assignment was to include baseline covariates that are indicators of patients' health status (and are used by clinicians in their decisions about treatment assignment) in the model used to estimate the effect of the timing of ART initiation on patient survival. This was only possible to the extent that values of such measures were recorded.

In order to deal with the lack of known values of measurements for patients at specific times of interest, I used fairly wide intervals around these times to reduce the number of patients who would be regarded as having no value available for a specific measurement near the time of interest. Changing the width of this interval did cause small changes in the estimated effects for other covariates, but did not substantially change the estimated effect of the timing of ART initiation on survival of patients.

To facilitate the inclusion of patients who did not have baseline values (for one or more of the baseline measurements) and compare their survival times to those of patients who did have baseline values, I used the relevant continuous measures as categorical variables and created missing categories. I also used multiple imputation as an alternative to using missing categories for patients who did not have values for measurements recorded near the baseline time. The results using multiple imputation were similar to those that used a missing category and did not substantially change the estimate for the effect of the timing of ART initiation on survival. It was nonetheless important to

perform multiple imputation and present results of analyses using multiply imputed data in order to demonstrate that the use of missing categories for unrecorded observations did not substantially bias effect estimates.

The assumption of proportional hazards was once again not satisfied in the model including baseline covariates and I chose to deal with this by including time-varying covariates for the variables for which this assumption was not satisfied (months from the start of TB treatment to enrolment for ART, CD4 count category and anaemia classification).

The inclusion of baseline covariates did result in a slight decrease (further from one) in the estimated hazard ratio of death for patients on ART versus not on ART (indicating a greater treatment effect than in the unadjusted analyses). The hazard ratio of death for patients on ART estimated using a model including baseline covariates and weighted using the inverse probability of treatment weights was slightly higher (closer to one) than the unadjusted treatment effect. The estimated hazard ratio of death for patients on ART using a Cox regression model weighted by the inverse probability of censoring weights hardly differed from that of the corresponding unweighted model, indicating either that there was no selection bias, or that the recorded values for patient measurements were not sufficient to account for it (the latter is more likely). The estimated hazard ratio of death for patients on ART estimated using a Cox regression model including continuous versions of baseline covariates (using the values imputed by multivariate Normal data augmentation for patients whose values were missing) and weighted using the stabilized inverse probability of treatment and censoring weights was slightly lower than both the crude unweighted hazard ratio and the hazard ratio for the similarly weighted model that used missing categories.

## 9.2 Other methods that could have been used

Possibilities for further research include the development and implementation of methods for competing risks analysis in the parametric framework. Joint modelling and dynamic prediction by landmarking are two further methodologies that were not within the scope of this thesis, but could potentially be useful for the analysis of similar data.

Joint modelling involves the fitting of two models, one for longitudinal data and the other for the distribution of the time to an event, which both depend

on a common set of latent random effects (Tsiatis and Davidian, 2004). This can be done using a variety of methods, such as commonly used two-stage methods (that are generally fairly simple to understand and implement, but may lead to biases), or the more standard, but more computationally intense likelihood methods, or Bayesian methods (Wu et al., 2011).

Landmark analysis is useful for comparing the estimated survival of two groups of patients from a pre-specified time post-baseline (landmark time), when patients may belong to different groups at different times (Dafni, 2011). For example, in a comparison of patients on ART and not on ART, (most) patients will initially not be on ART and some may initiate ART sooner than others. Comparing all patients who ever initiated ART would give biased results, as some patients may die before having the opportunity to initiate ART. Van Houwelingen (2007) proposes the use of dynamic survival prediction using landmarking. Essentially, Cox regressions are performed for a range of landmark times and effect estimates smoothed by maximising a pseudo-log-likelihood obtained by adjusting the partial-log-likelihood for landmarking.

Doubly robust methods described by Bang and Robins (2005) and mentioned briefly in Section 6.4.2 could also be considered in conjunction with marginal structural models to reduce the likelihood of bias resulting from possible model-misspecification. An alternative to the use of inverse probability weighting to obtain parameter estimates of marginal structural models is the use of G-estimation to obtain parameter estimates of nested structural AFT models (Hernán et al., 2005).

Finally, returning to the schematic diagram presented in Figure 1.5, it would be of interest to perform a formal path analysis, in which the predicted delays to ART initiation from the models in Part I are used as the main covariate of interest in the models from Part II that estimate the effect of ART (and its timing) on the time to death.

### 9.3 Clinical objective and aims

In this section I consider the two main clinical questions of interest and discuss the results that I obtained in attempting to answer them using the methods described above.

### 9.3.1 Effect of predictors of earlier ART initiation on the time to ART initiation

The time from the start of TB treatment to ART initiation varied between clinic sites, but has decreased over the years and was shorter for patients who were diagnosed with TB in the ART clinic as a part of the screening process for ART initiation (rather than in a TB clinic before enrolling for HIV care). Time to ART initiation was also shorter for patients who had a CD4 count measurement near the time of starting TB treatment and for whom the measured value was lower. The actual effect of each covariate on the time to ART initiation depends on the values for the other covariates. Nonetheless the same pattern of results was observed for all models.

### 9.3.2 Impact of early ART initiation on survival

Using inverse probability of treatment and censoring weighted models that allowed for time-varying effects (and used missing categories), I was able to show a 47% reduced hazard of death for patients on ART in a given interval and a 4% increased hazard of death per additional week that a patient did not initiate ART. Although I attempted to eliminate time-dependent confounding and selection bias and to choose the most appropriate models, there may still be some residual bias in these estimates resulting from time dependent confounding or selection bias by unmeasured variables, or from measurement error or model misspecification.

In summary, notwithstanding the few repeated measured values for time-updated values pre-ART and the fact that statistical methods cannot completely adjust for all the resulting biases, the analyses in this thesis confirm those reported in Section 1.2, which indicates that early ART initiation has a protective impact on patient mortality. Due to the lack of adequate data, I was unable to ascertain whether this comes at the cost of higher incidences of IRIS or other severe adverse events.

# Appendix A

## Abbreviations

|          |                                                                    |
|----------|--------------------------------------------------------------------|
| ADIs     | AIDS Defining Illnesses                                            |
| AHR      | Adjusted Hazard Ratio                                              |
| AIDS     | Acquired Immune Deficiency Syndrome                                |
| AFT      | Accelerated Failure Time                                           |
| ART      | Anti-Retroviral Therapy                                            |
| ASHR     | Adjusted Sub-distribution Hazard Ratio                             |
| CI       | Confidence Interval                                                |
| CIF      | Cumulative Incidence Function                                      |
| col      | column                                                             |
| CSH      | Cause-Specific Hazard                                              |
| CR       | Competing Risk                                                     |
| cOR      | crude Odds Ratio                                                   |
| DA       | Data Augmentation                                                  |
| DAG      | Directed Acyclic Graph                                             |
| EM       | Expectation Maximization                                           |
| HAART    | Highly Active Anti-Retroviral Therapy                              |
| HF       | Hazard Function                                                    |
| HIV      | Human Immunodeficiency Virus                                       |
| HR       | Hazard Ratio                                                       |
| IeDEA    | International epidemiological Databases for the Evaluation of AIDS |
| IeDEA-SA | IeDEA Southern Africa                                              |
| IPCW     | Inverse Probability of Censoring Weight                            |
| IPTCW    | Inverse Probability of Treatment and Censoring Weight              |
| IPTW     | Inverse Probability of Treatment Weight                            |
| IPW      | Inverse Probability Weight                                         |
| IRIS     | Immune Reconstitution Inflammatory Syndrome                        |
| IRR      | Incidence-Rate Ratio                                               |
| IQR      | InterQuartile Range                                                |
| ISPM     | Institute of Social and Preventive Medicine                        |
| LR       | Logistic Regression                                                |
| MAR      | Missing At Random                                                  |
| MCAR     | Missing Completely At Random                                       |
| MCMC     | Monte-Carlo Markov Chain                                           |
| MDR      | MultiDrug-Resistant                                                |
| MI       | Multiple Imputation                                                |
| MICE     | Multiple Imputation by Chained Equations                           |
| MNAR     | Missing Not At Random                                              |
| MSM      | Marginal Structural Model                                          |
| MVN      | MultiVariate Normal                                                |
| NIAID    | National Institutes of Allergy and Infectious Diseases             |
| OR       | Odds Ratio                                                         |
| PH       | Proportional Hazards                                               |
| SAPIT    | Starting Antiretroviral Therapy at Three Points in Tuberculosis    |
| SHR      | Sub-distribution Hazard Ratio                                      |
| SIPTCW   | Stabilized Inverse Probability of Treatment and Censoring Weight   |
| SNP      | Smear-negative pulmonary                                           |
| SPP      | Smear-positive pulmonary                                           |
| TB       | Tuberculosis                                                       |
| trt      | treatment                                                          |
| WHO      | World Health Organization                                          |
| XDR      | eXtensively Drug-Resistant                                         |

# Appendix B

## Papers Published

During the preliminary analysis stages for my thesis, I co-authored two papers:

S. D. Lawn, L. Campbell, R. Kaplan, A. Boulle, M. Cornell, B. Kerschberger, C. Morrow, F. Little, M. Egger and R. Wood. Time to initiation of antiretroviral therapy among patients with HIV-associated tuberculosis in Cape Town, South Africa *Journal of Acquired Immune Deficiency Syndrome*, 57:136-140, 2011

S. D. Lawn, L. Campbell, R. Kaplan, F. Little, C. Morrow and Robin Wood. Delays in starting antiretroviral therapy in patients with HIV-associated tuberculosis accessing non-integrated clinical services in a South African township. *BMC Infectious Diseases*, 11:258, 2011

My contributions to these papers were the statistical analyses. I have not included these papers to avoid infringing on copyright.

# Appendix C

## Additional Analyses, Graphs and Tables

### C.1 Sensitivity Analyses

#### C.1.1 Consideration of inclusion of stage and type of TB in time to ART models

The WHO clinical stage of HIV (hereafter stage) is an indicator of HIV progression (with higher stages indicating patients who are more sick). In theory, patients who are classified as stage IV of HIV disease should perhaps be fast-tracked compared to patients in stage III (if all other characteristics are the same). In practice, patients who are classified as stage IV have more severe immune-suppression and this may lead to them being more sick and having other complications that lead to ART initiation needing to be delayed (or even to death pre-ART).

Clinicians often also take more notice of a patient's CD4 count (which is included in the model) and their appearance (which is not recorded) in terms of deciding how quickly to try to start a patient on ART (Lawn, personal communication).

A further problem with using the stage of a patient is that it is ascertained when a patient enrolls for HIV care, rather than at the time of TB diagnosis, which is the baseline time for this analysis. This means that depending on whether a patient is diagnosed with TB before or after enrolling for HIV care

and how long the time is between these events, the stage classification may either not yet be ascertained, or may be outdated at the time of TB diagnosis.

This same draw-back also applies to CD4 count, which is usually taken close to the time of enrolling for HIV care. Inclusion of a variable that measures the time from the start of TB treatment to the CD4 count (where this is positive) and its interaction with the CD4 count, as was done by Franke et al. (2011) could be considered for further analyses.

In theory, type of TB should not influence the time to ART initiation, as the South African guidelines for when to start ART that were available during the time of the study (National Department of Health, 2004) do not differentiate the recommended treatment of patients with TB for different types of TB.

In practice, clinicians may consider patients with extra-pulmonary TB to be more severe cases than those diagnosed with pulmonary TB and thus consider that initiating ART is more urgent for them. Similarly, clinicians may suspect that patients diagnosed with smear-positive pulmonary TB are not as immune-suppressed as those who did not have a positive smear for TB. They may thus have a tendency to try to fast-track patients with smear-negative pulmonary TB compared to those with smear-positive pulmonary TB. The same comments as for stage, however, apply with regards to patients who are more immune-suppressed possibly experiencing delays for starting ART either as a result of complications that arise before ART initiation or the concern about the possible development of IRIS if ART is initiated soon after the start of TB treatment.

### **Results from including Stage and type of TB in selected time to ART models**

Neither stage nor type of TB had hazard ratios significantly different from one in the Cox proportional hazards model and their inclusion did not significantly alter the hazard ratios ( $e^{\hat{\beta}_j}$ ) for the other variables included in the model (see Table C.1). The assumption of proportional hazards was satisfied for the stage variable and for the category ‘extra-pulmonary TB’, but not for the category ‘smear-negative pulmonary TB’, where smear-positive pulmonary TB was the baseline category (PH test  $p$ -value = 0.001). In a Generalized Gamma model including stage and the type of TB and allowing the ancillary parameters to vary by the type of TB (see Model 9s in

Table C.2), the decelerators ( $\exp(\hat{\beta}_{\mu_j})$ ) were not significantly different from one for stage or category of TB and the estimated values of  $\hat{\beta}_{\sigma_j}$  and  $\hat{\beta}_{\kappa_j}$  for the TB categories were not significantly different from zero (at 5% level of significance). The inclusion of these variables also did not significantly change the estimated decelerators or values of  $\hat{\beta}_{\sigma_j}$  and  $\hat{\beta}_{\kappa_j}$  for other variables. These variables were therefore not included in the comparison of Cox and parametric regression models.

Table C.1: Cox proportional hazards regression including stage and type of TB as covariates

|                                | Model 1s            |             | PH test    |
|--------------------------------|---------------------|-------------|------------|
|                                | $e^{\hat{\beta}_j}$ | 95% CI      | $p$ -value |
| Location of Clinic             |                     |             |            |
| Site 1                         | 0.67                | 0.57 - 0.79 | 0.68       |
| Site 2                         | 0.35                | 0.29 - 0.42 | 0.05       |
| Site 3 (baseline)              |                     | -           |            |
| Year in which TB diagnosed     |                     | -           |            |
| 2002-2005                      | 0.63                | 0.56 - 0.70 | 0.01       |
| 2006-2007                      |                     | -           |            |
| Clinic in which TB diagnosed   |                     | -           |            |
| TB clinic                      | 0.57                | 0.51 - 0.64 | 0.00       |
| ART Clinic                     |                     | -           |            |
| CD4 count <sup>a</sup>         |                     | -           |            |
| 0-49 (baseline)                |                     | -           |            |
| 50-99                          | 0.72                | 0.62 - 0.83 | 0.00       |
| 100-149                        | 0.61                | 0.52 - 0.72 | 0.00       |
| 150-199                        | 0.51                | 0.42 - 0.62 | 0.00       |
| $\geq 200$                     | 0.40                | 0.31 - 0.51 | 0.00       |
| missing                        | 0.19                | 0.16 - 0.22 | 0.00       |
| Type of TB                     |                     | -           |            |
| Smear-positive pulm (baseline) |                     | -           |            |
| Smear-negative pulm            | 0.99                | 0.86 - 1.14 | 0.00       |
| Extra-pulm                     | 0.95                | 0.83 - 1.08 | 0.67       |
| WHO Clinical Stage of HIV      |                     | -           |            |
| Stages I to III                |                     | -           |            |
| Stage IV                       | 0.98                | 0.87 - 1.11 | 0.27       |
| Global PH test                 |                     |             | 0.00       |

CI, Confidence Interval

<sup>a</sup> CD4 count taken closest start of TB treatment that satisfied criteria of being taken not earlier than 183 days before or 91 days after start of TB treatment and not more than 7 days after start of ART

Table C.2: Generalized Gamma AFT, Models 8s &amp; 9s

|                                | Model 8s <sup>(1)</sup>   |             | Model 9s <sup>(2)</sup>   |             |                          |               |
|--------------------------------|---------------------------|-------------|---------------------------|-------------|--------------------------|---------------|
|                                | $e^{\hat{\beta}_{\mu j}}$ | 95% CI      | $e^{\hat{\beta}_{\mu j}}$ | 95% CI      | $\hat{\beta}_{\sigma j}$ | 95% CI        |
| Location of Clinic             |                           |             |                           |             |                          |               |
| Site 1                         | 1.20                      | 1.10 - 1.31 | 1.05                      | 0.94 - 1.16 | 0.09                     | -0.06 - 0.24  |
| Site 2                         | 1.67                      | 1.53 - 1.83 | 1.51                      | 1.35 - 1.69 | 0.00                     | -0.16 - 0.15  |
| Site 3 (baseline)              | 1.00                      | -           | 1.00                      | -           | 0.00                     | -             |
| Year in which TB diagnosed     |                           |             |                           |             |                          |               |
| 2002-2005                      | 1.27                      | 1.20 - 1.34 | 1.20                      | 1.12 - 1.28 | -0.05                    | -0.14 - 0.04  |
| 2006-2007                      | 1.00                      | -           | 1.00                      | -           | 0.00                     | -             |
| Clinic in which TB diagnosed   |                           |             |                           |             |                          |               |
| TB clinic                      | 1.42                      | 1.33 - 1.51 | 1.41                      | 1.27 - 1.55 | -0.41                    | -0.51 - -0.32 |
| ART Clinic                     | 1.00                      | -           | 1.00                      | -           | 0.00                     | -             |
| CD4 count (cells/ $\mu$ L)     |                           |             |                           |             |                          |               |
| 0-49 (baseline)                | 1.00                      | -           | 1.00                      | -           | 0.00                     | -             |
| 50-99                          | 1.25                      | 1.16 - 1.36 | 1.38                      | 1.22 - 1.56 | -0.34                    | -0.47 - -0.22 |
| 100-149                        | 1.35                      | 1.24 - 1.47 | 1.54                      | 1.36 - 1.75 | -0.37                    | -0.50 - -0.24 |
| 150-199                        | 1.49                      | 1.34 - 1.64 | 1.51                      | 1.30 - 1.75 | -0.40                    | -0.54 - -0.26 |
| $\geq 200$                     | 1.68                      | 1.48 - 1.91 | 1.76                      | 1.52 - 2.05 | -0.53                    | -0.72 - -0.34 |
| missing                        | 2.34                      | 2.15 - 2.55 | 2.83                      | 2.50 - 3.19 | -0.55                    | -0.69 - -0.42 |
| Type of TB                     |                           |             |                           |             |                          |               |
| Smear-positive pulm (baseline) | 1.00                      | -           | 1.00                      | -           | 0.00                     | -             |
| Smear-negative pulm            | 0.99                      | 0.93 - 1.06 | 1.04                      | 0.95 - 1.14 | 0.11                     | 0.00 - 0.21   |
| Extra-pulm                     | 1.02                      | 0.95 - 1.09 | 1.03                      | 0.95 - 1.12 | 0.03                     | -0.07 - 0.13  |
| WHO Clinical Stage of HIV      |                           |             |                           |             |                          |               |
| Stages I to III                | 1.00                      | -           | 1.00                      | -           | -                        | -             |
| Stage IV                       | 1.00                      | 0.94 - 1.06 | 0.99                      | 0.94 - 1.04 | -                        | -             |

<sup>(1)</sup> For Model 8s, the ancillary parameters  $\sigma$  and  $\kappa$  are assumed to be the same for all covariate patterns, i.e.  $\hat{\beta}_{\sigma j} = \hat{\beta}_{\kappa j} = 0$  for all  $j \neq 0$ , such that

$\hat{\sigma} = \exp(\hat{\beta}_{\sigma 0}) = \exp(-0.67) = 0.51$  for all patients (95% CI for  $\hat{\sigma}$ : 0.49 to 0.53) and  $\hat{\kappa} = \hat{\beta}_{\kappa 0} = 0.67$  for all patients (95% CI: 0.57 to 0.78).

<sup>(2)</sup> For Model 9s,  $\hat{\beta}_{\sigma 0} = -0.25$  (95% CI: -0.43 to -0.08) and  $\hat{\beta}_{\kappa 0} = 1.05$  (95% CI: 0.54 to 1.56).

### C.1.2 Stratified Kaplan-Meier survival plots excluding patients who have no CD4 count near the start of TB treatment

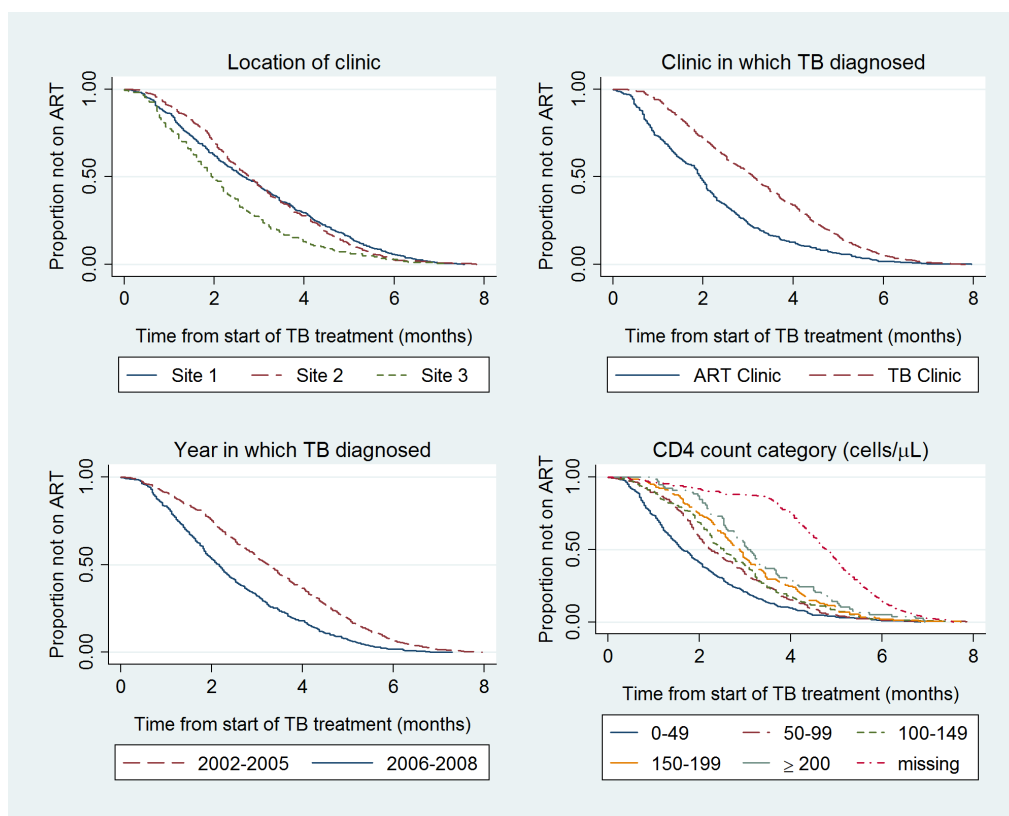


Figure C.1: Kaplan-Meier survival plots for time from start of TB treatment to ART initiation for those who had a CD4 count near the time of starting TB treatment, N=1162

### C.1.3 Sensitivity of generalized Gamma models to assumptions about missing CD4 counts

Table C.3: Comparison of  $\hat{\sigma}$  for models making different assumptions about the appropriate treatment of patients with no CD4 count near the start of TB treatment

| Location of Clinic       | Model GG1                    |            | Model GG2                    |            | Model GG3                    |           | Model GG4                    |           |
|--------------------------|------------------------------|------------|------------------------------|------------|------------------------------|-----------|------------------------------|-----------|
|                          | missing category             | 95% CI     | missing excluded             | 95% CI     | missing $\geq 200$           | 95% CI    | missing $< 50$               | 95% CI    |
|                          | $e^{\hat{\beta}_{\sigma j}}$ |            | $e^{\hat{\beta}_{\sigma j}}$ |            | $e^{\hat{\beta}_{\sigma j}}$ |           | $e^{\hat{\beta}_{\sigma j}}$ |           |
| Site 1                   | 1.11                         | 0.95- 1.29 | 1.18                         | 1.00- 1.39 | 1.06                         | 0.91-1.24 | 0.92                         | 0.80-1.06 |
| Site 2                   | 1.04                         | 0.89- 1.21 | 0.99                         | 0.84- 1.16 | 1.02                         | 0.88-1.19 | 0.90                         | 0.78-1.03 |
| Site 3 (baseline)        | 1.00                         | -          | 1.00                         | -          | 1.00                         | -         | 1.00                         | -         |
| TB diagnosed before 2006 | 0.96                         | 0.88- 1.04 | 0.93                         | 0.85- 1.03 | 0.94                         | 0.86-1.03 | 0.86                         | 0.79-0.95 |
| Referred from TB Clinic  | 0.67                         | 0.61- 0.73 | 0.69                         | 0.62- 0.76 | 0.67                         | 0.61-0.74 | 0.69                         | 0.63-0.77 |
| CD4 count                |                              |            |                              |            |                              |           |                              |           |
| 0-49 (baseline)          | 1.00                         | -          | 1.00                         | -          | 1.00                         | -         | 1.00                         | -         |
| 50-99                    | 0.72                         | 0.64- 0.81 | 0.71                         | 0.63- 0.80 | 0.72                         | 0.64-0.81 | 0.69                         | 0.61-0.78 |
| 100-149                  | 0.69                         | 0.60- 0.79 | 0.69                         | 0.60- 0.79 | 0.69                         | 0.61-0.79 | 0.67                         | 0.59-0.77 |
| 150-199                  | 0.67                         | 0.58- 0.77 | 0.65                         | 0.57- 0.75 | 0.67                         | 0.58-0.77 | 0.62                         | 0.53-0.71 |
| $\geq 200$               | 0.60                         | 0.50- 0.72 | 0.58                         | 0.48- 0.70 | 0.64                         | 0.57-0.72 | 0.61                         | 0.51-0.73 |
| missing                  | 0.58                         | 0.51- 0.66 |                              |            |                              |           |                              |           |
| Constant                 | 0.78                         | 0.66- 0.93 | 0.78                         | 0.65- 0.93 | 0.80                         | 0.68-0.95 | 0.98                         | 0.83-1.15 |

Table C.4: Comparison of  $\hat{\kappa}$  for models making different assumptions about the appropriate treatment of patients with no CD4 count near the start of TB treatment

|                          | Model GG1        |                              |        | Model GG2        |                              |        | Model GG3          |                              |        | Model GG4      |                              |        |
|--------------------------|------------------|------------------------------|--------|------------------|------------------------------|--------|--------------------|------------------------------|--------|----------------|------------------------------|--------|
|                          | missing category | $e^{\hat{\beta}_{\kappa j}}$ | 95% CI | missing excluded | $e^{\hat{\beta}_{\kappa j}}$ | 95% CI | missing $\geq 200$ | $e^{\hat{\beta}_{\kappa j}}$ | 95% CI | missing $< 50$ | $e^{\hat{\beta}_{\kappa j}}$ | 95% CI |
| Location of Clinic       |                  |                              |        |                  |                              |        |                    |                              |        |                |                              |        |
| Site 1                   | -0.56            | -0.98--0.13                  |        | -0.41            | -0.87-0.05                   |        | -0.44              | -0.86--0.01                  |        | 0.01           | -0.39- 0.42                  |        |
| Site 2                   | -0.45            | -0.88--0.02                  |        | -0.41            | -0.85-0.04                   |        | -0.51              | -0.94--0.09                  |        | -0.17          | -0.55- 0.22                  |        |
| Site 3 (baseline)        | 1.00             | -                            |        | 1.00             | -                            |        | 1.00               | -                            |        | 1.00           | -                            |        |
| TB diagnosed before 2006 | -0.16            | -0.39- 0.07                  |        | 0.00             | -0.26-0.27                   |        | -0.15              | -0.38- 0.09                  |        | 0.05           | -0.21- 0.31                  |        |
| Referred from TB Clinic  | -0.26            | -0.52- 0.01                  |        | -0.12            | -0.40-0.15                   |        | -0.18              | -0.45- 0.08                  |        | -0.17          | -0.44- 0.11                  |        |
| CD4 count                |                  |                              |        |                  |                              |        |                    |                              |        |                |                              |        |
| 0-49 (baseline)          | 1.00             | -                            |        | 1.00             | -                            |        | 1.00               | -                            |        | 1.00           | -                            |        |
| 50-99                    | 0.18             | -0.18- 0.54                  |        | 0.14             | -0.21-0.48                   |        | 0.16               | -0.19- 0.51                  |        | -0.13          | -0.47- 0.22                  |        |
| 100-149                  | 0.36             | 0.00- 0.72                   |        | 0.24             | -0.11-0.60                   |        | 0.33               | -0.02- 0.68                  |        | -0.05          | -0.38- 0.28                  |        |
| 150-199                  | -0.31            | -0.78- 0.16                  |        | -0.39            | -0.84-0.06                   |        | -0.25              | -0.71- 0.20                  |        | -0.51          | -0.93--0.08                  |        |
| $\geq 200$               | -0.09            | -0.64- 0.46                  |        | -0.14            | -0.72-0.44                   |        | 0.28               | -0.06- 0.61                  |        | -0.53          | -1.17- 0.11                  |        |
| missing                  | 0.43             | 0.08- 0.79                   |        |                  |                              |        |                    |                              |        |                |                              |        |
| Constant                 | 1.17             | 0.71- 1.63                   |        | 1.00             | 0.52-1.47                    |        | 1.12               | 0.67- 1.57                   |        | 0.99           | 0.55- 1.42                   |        |

## C.2 Assessment of proportional hazards

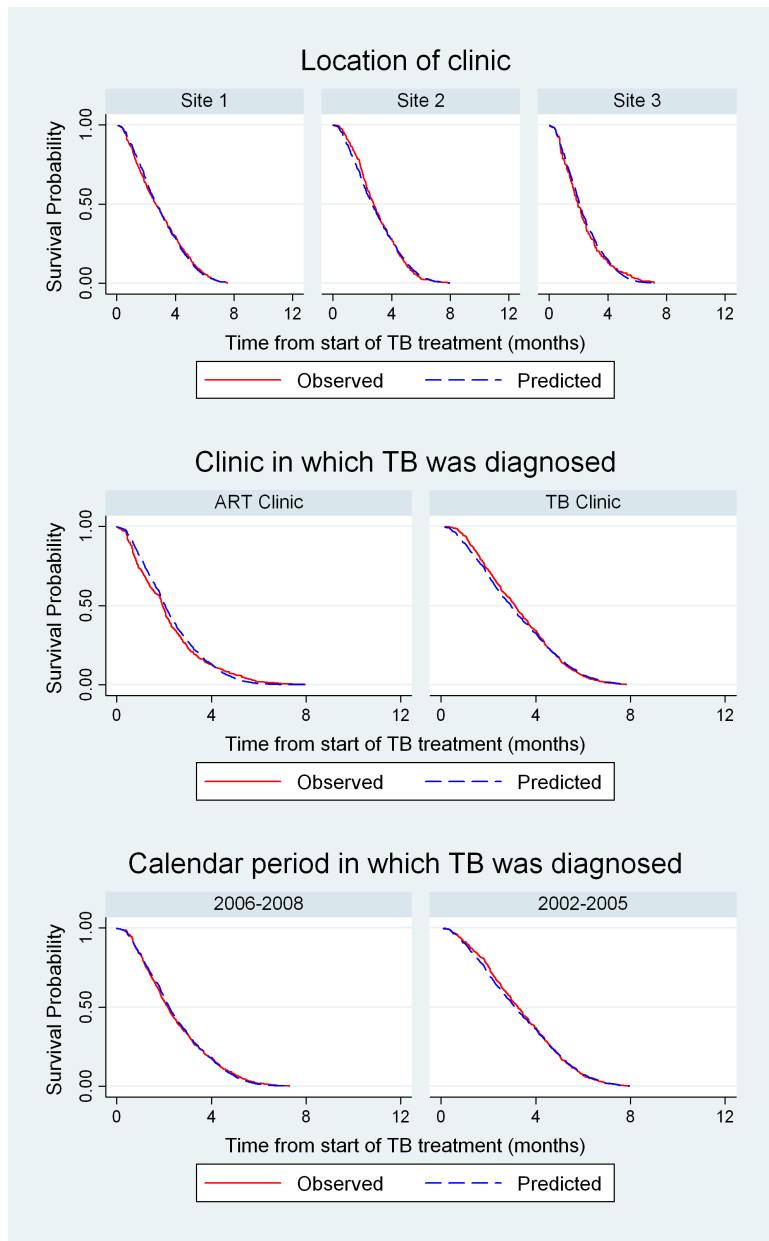


Figure C.2: Comparison of KM vs Cox survival curves by covariate, N=1444

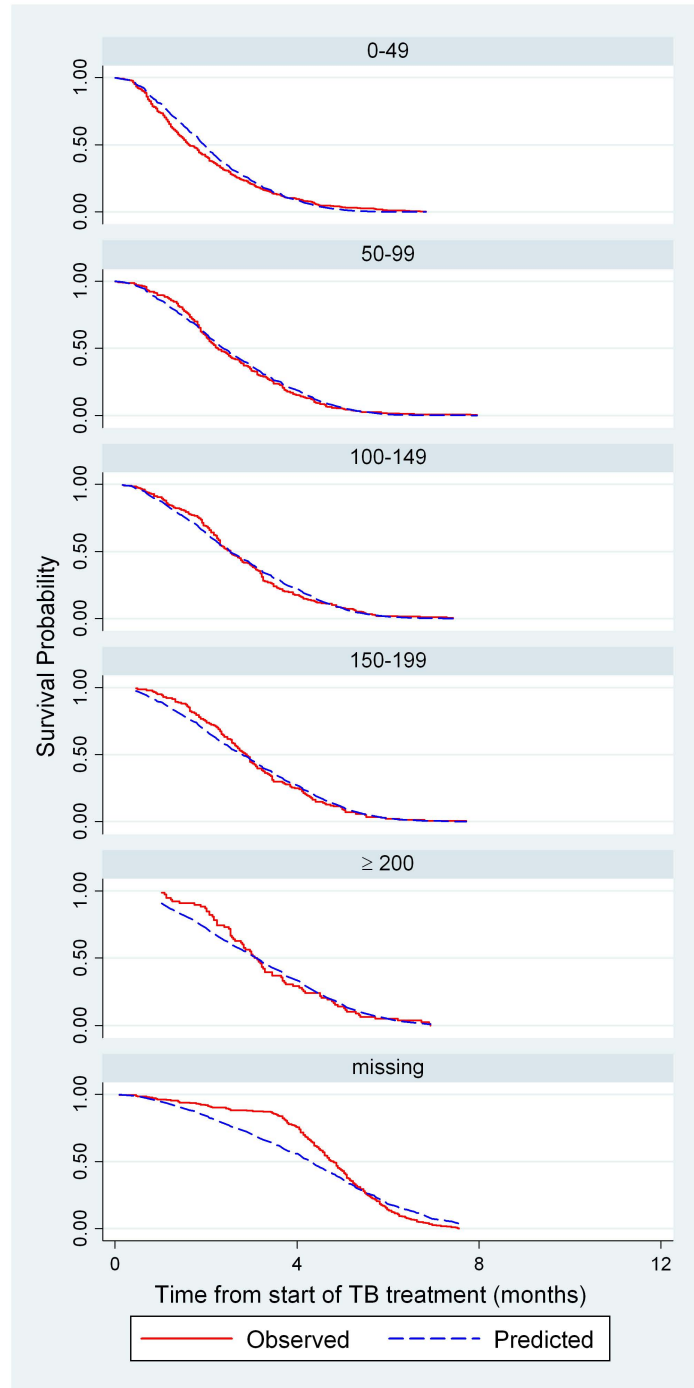


Figure C.3: Comparison of KM vs Cox survival curves by CD4 count, N=1444

### C.3 Log-logistic and log-Normal Models

For the Log-logistic model, the probability density, survival and hazard functions are:

$$f(t) = \frac{\lambda^{\frac{1}{\gamma}} t^{\frac{1}{\gamma}-1}}{\gamma \left\{ 1 + (\lambda t)^{\frac{1}{\gamma}} \right\}^2} \quad (\text{C.1})$$

$$S(t) = \left\{ 1 + (\lambda t)^{\frac{1}{\gamma}} \right\}^{-1} \quad (\text{C.2})$$

and

$$h(t) = \frac{\frac{1}{\gamma} \lambda^{\frac{1}{\gamma}} t^{\frac{1}{\gamma}-1}}{\left\{ 1 + (\lambda t)^{\frac{1}{\gamma}} \right\}}. \quad (\text{C.3})$$

For the Log-Normal model, the probability density, survival and hazard functions are:

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp \left[ \frac{-1}{2\sigma^2} \{\log(t) - \mu\}^2 \right], \quad (\text{C.4})$$

$$S(t) = 1 - \Phi \left\{ \frac{\log(t) - \mu}{\sigma} \right\} \quad (\text{C.5})$$

and

$$h(t) = \frac{\frac{1}{t\sigma\sqrt{2\pi}} \exp \left[ \frac{-1}{2\sigma^2} \{\log(t) - \mu\}^2 \right]}{1 - \Phi \left\{ \frac{\log(t) - \mu}{\sigma} \right\}}, \quad (\text{C.6})$$

where  $\Phi(z)$  is the standard normal cumulative distribution function.

Figure C.4 shows the flexibility of the Log-logistic and log-Normal hazard functions. The log-logistic model is fit using (C.2), where

$$\begin{aligned} \lambda &= \exp(-[\alpha_0 + \alpha_1 X_{1i} + \beta_2 X_{2i} + \dots + \alpha_9 X_{9i}]) \\ \log(\gamma) &= \beta_{\gamma 0} + \beta_{\gamma 1} X_{1i} + \beta_{\gamma 2} X_{2i} + \dots + \beta_{\gamma 9} X_{9i} \end{aligned}$$

and  $X_{1i} \dots X_{9i}$  are as in (4.1). The log-Normal model is fit using (C.5), where

$$\begin{aligned} \mu &= \alpha_{\mu 0} + \alpha_{\mu 1} X_{1i} + \alpha_{\mu 2} X_{2i} + \dots + \alpha_{\mu 9} X_{9i} \\ \log(\sigma) &= \beta_{\sigma 0} + \beta_{\sigma 1} X_{1i} + \beta_{\sigma 2} X_{2i} + \dots + \beta_{\sigma 9} X_{9i} \end{aligned}$$

and  $X_{1i} \dots X_{9i}$  are as in (4.1).

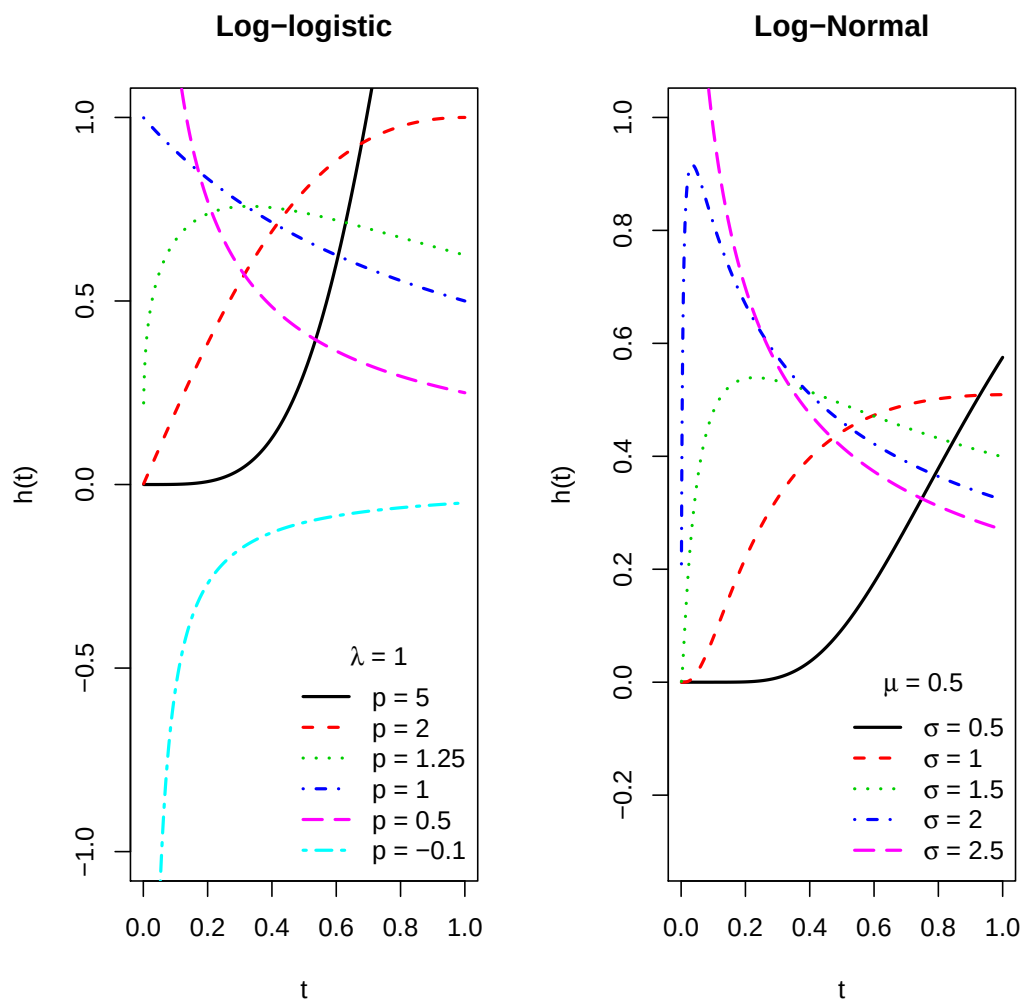


Figure C.4: Some possible shapes of hazard curves for the Log-logistic and Log-Normal hazard functions

Table C.5: Log-logistic AFT regressions, Models 10 &amp; 11

|                              | Model 10             |             | Model 11             |             |
|------------------------------|----------------------|-------------|----------------------|-------------|
|                              | $e^{\hat{\alpha}_j}$ | 95% CI      | $e^{\hat{\alpha}_j}$ | 95% CI      |
| Location of Clinic           |                      |             |                      |             |
| Site 1                       | 1.17                 | 1.07 - 1.28 | 1.18                 | 1.09 - 1.27 |
| Site 2                       | 1.67                 | 1.53 - 1.83 | 1.63                 | 1.50 - 1.77 |
| Site 3 (baseline)            | 1.00                 | -           | 1.00                 | -           |
| Year in which TB diagnosed   |                      |             |                      |             |
| 2002-2005                    | 1.28                 | 1.21 - 1.35 | 1.25                 | 1.19 - 1.31 |
| 2006-2008 (baseline)         | 1.00                 | -           | 1.00                 | -           |
| Clinic in which TB diagnosed |                      |             |                      |             |
| TB Clinic                    | 1.63                 | 1.53 - 1.74 | 1.58                 | 1.47 - 1.69 |
| ART Clinic (baseline)        | 1.00                 | -           | 1.00                 | -           |
| CD4 count <sup>a</sup>       |                      |             |                      |             |
| 0-49 (baseline)              | 1.00                 | -           | 1.00                 | -           |
| 50-99                        | 1.41                 | 1.31 - 1.53 | 1.39                 | 1.28 - 1.51 |
| 100-149                      | 1.55                 | 1.42 - 1.69 | 1.51                 | 1.38 - 1.64 |
| 150-199                      | 1.73                 | 1.57 - 1.91 | 1.69                 | 1.55 - 1.85 |
| $\geq 200$                   | 1.98                 | 1.75 - 2.24 | 1.87                 | 1.69 - 2.08 |
| missing                      | 2.75                 | 2.53 - 2.99 | 2.72                 | 2.52 - 2.95 |

<sup>a</sup> CD4 count taken closest start of TB treatment that satisfied criteria of being taken not earlier than 183 days before or 91 days after start of TB treatment and not more than 7 days after start of ART

$$\log(\hat{\gamma}) = \hat{\beta}_{\gamma 0} + \hat{\beta}_{\gamma 1}X_{1i} + \hat{\beta}_{\gamma 2}X_{2i} + \dots + \hat{\beta}_{\gamma 9}X_{9i}.$$

For Model 10, the ancillary parameter  $\gamma$  is assumed to be the same for all covariate patterns, i.e.  $\hat{\beta}_{\gamma j} = 0$  for all  $j \neq 0$ , such that

$$\hat{\gamma} = \exp(\hat{\beta}_{\gamma 0}) = \exp(-1.19) = 0.30 \text{ for all patients (95\% CI for } \hat{\gamma}: 0.29 \text{ to } 0.32)$$

For Model 11,  $\hat{\beta}_{\gamma 0} = -0.51$  (95% CI: -0.67 to -0.35).

Table C.6: Log-Normal AFT regressions, Models 12 &amp; 13

|                              | Model 12             |             | Model 13             |             |
|------------------------------|----------------------|-------------|----------------------|-------------|
|                              | $e^{\hat{\alpha}_j}$ | 95% CI      | $e^{\hat{\alpha}_j}$ | 95% CI      |
| Location of Clinic           |                      |             |                      |             |
| Site 1                       | 1.23                 | 1.11 - 1.35 | 1.21                 | 1.11 - 1.33 |
| Site 2                       | 1.76                 | 1.59 - 1.95 | 1.66                 | 1.51 - 1.83 |
| Site 3 (baseline)            | 1.00                 | -           | 1.00                 | -           |
| Year in which TB diagnosed   |                      |             |                      |             |
| 2002-2005                    | 1.30                 | 1.22 - 1.39 | 1.26                 | 1.20 - 1.33 |
| 2006-2008 (baseline)         | 1.00                 | -           | 1.00                 | -           |
| Clinic in which TB diagnosed |                      |             |                      |             |
| TB Clinic                    | 1.71                 | 1.60 - 1.83 | 1.63                 | 1.51 - 1.75 |
| ART Clinic (baseline)        | 1.00                 | -           | 1.00                 | -           |
| CD4 count <sup>a</sup>       |                      |             |                      |             |
| 0-49 (baseline)              | 1.00                 | -           | 1.00                 | -           |
| 50-99                        | 1.39                 | 1.27 - 1.52 | 1.37                 | 1.26 - 1.49 |
| 100-149                      | 1.54                 | 1.40 - 1.70 | 1.47                 | 1.34 - 1.61 |
| 150-199                      | 1.76                 | 1.57 - 1.96 | 1.70                 | 1.56 - 1.87 |
| $\geq 200$                   | 2.01                 | 1.74 - 2.32 | 1.86                 | 1.68 - 2.07 |
| missing                      | 2.64                 | 2.40 - 2.90 | 2.67                 | 2.46 - 2.89 |

<sup>a</sup> CD4 count taken closest start of TB treatment that satisfied criteria of being taken not earlier than 183 days before or 91 days after start of TB treatment and not more than 7 days after start of ART

$$\log(\hat{\sigma}) = \hat{\beta}_{\sigma 0} + \hat{\beta}_{\sigma 1}X_{1i} + \hat{\beta}_{\sigma 2}X_{2i} + \dots + \hat{\beta}_{\sigma 9}X_{9i}.$$

$$\hat{\kappa} = \hat{\beta}_{\kappa 0} + \hat{\beta}_{\kappa 1}X_{1i} + \hat{\beta}_{\kappa 2}X_{2i} + \dots + \hat{\beta}_{\kappa 9}X_{9i}.$$

For Model 12, the ancillary parameter  $\sigma$  is assumed to be the same for all covariate patterns, i.e.  $\hat{\beta}_{\sigma j} = 0$  for all  $j \neq 0$ , such that

$$\hat{\sigma} = \exp(\hat{\beta}_{\sigma 0}) = \exp(-0.52) = 0.59 \text{ for all patients (95\% CI for } \hat{\sigma}: 0.57 \text{ to } 0.62).$$

For Model 13,  $\hat{\beta}_{\sigma 0} = 0.17$  (95% CI: 0.04 to 0.31).

# Appendix D

## Code

### D.1 Time to ART Analysis for patients from all 3 clinic sites who do initiate ART

#### D.1.1 Tables 4.3-4.6, B.5 and B.6

```
*Model 1: Cox regression
*that does not allow interactions of covariates with time
xi: stcox i.clinic1 yeartbB2006 TBpreEnrol i.cd4tbc2m

*Model 2: Cox regression
*including interactions of covariates with analysis time
xi: stcox i.clinic1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
    tvc(i.clinic1 yeartbB2006 TBpreEnrol i.cd4tbc2m)

*Model 3 Cox regression
*including interactions of covariates with
*the natural logarithm of analysis time
xi: stcox i.clinic1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
    tvc(i.clinic1 yeartbB2006 TBpreEnrol i.cd4tbc2m) texp(ln(_t))

*Model 4: Weibull model in PH framework
*that does not allow shape parameter to vary by covariate value
xi: streg i.clinic1 yeartbB2006 TBpreEnrol i.cd4tbc2m, d(w)
```

```
*Model 5: Weibull model in PH framework
*that does allow shape parameter to vary by covariate value
xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
d(w) anc(i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m)

*Model 6: Weibull model in AFT framework
*that does not allow shape parameter to vary by covariate value
xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
d(w) ti tr

*Model 7: Weibull model in AFT framework
*that does allow shape parameter to vary by covariate value
xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
d(w) anc(i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m) ti

*Model 8: Generalized Gamma AFT model that
*does not allow ancillary parameters to vary by covariate value
xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
d(gamma) ti tr

*Model 9: Generalized Gamma AFT model that
*does allow ancillary parameters to vary by covariate value
xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
d(gamma) anc(i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m) ///
anc2(i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m)

*Model 10: Log-logistic AFT model that
*does allow shape parameter to vary by covariate value
xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, d(11)

*Model 11: Log-logistic AFT model that
*does not allow shape parameter to vary by covariate value
xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
d(11) anc(i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m)

*Model 12: Log-Normal AFT model that
*does allow shape parameter to vary by covariate value
xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
d(ln)
```

```

*Model 13: Log-Normal AFT model that
*does not allow shape parameter to vary by covariate value
xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
d(ln) anc(i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m)

```

### D.1.2 Example of code to create plot of Cox-Snell residuals against the estimated cumulative hazard

The following code was adapted for my data from that of StataCorp (2009b) pages 176-178. Line 1 sets the data for survival analysis, using the time from the start of TB treatment to ART initiation as the analysis time and ART initiation as the failure event. Line 2 shows the regression for the generalized Gamma AFT model that allows ancillary parameters to vary by covariate value. The Cox-Snell residuals are predicted in line 3. The data is then reset for survival analysis, using the Cox-Snell residual estimates as the analysis time. The Kaplan-Meier survival function is estimated in line 5 and used to estimate the cumulative hazard in line 6. Line 7 shows the command to plot the Cox-Snell residuals against the estimated cumulative hazard and create a straight line through the origin with a gradient of one.

```

1 stset tbtoartm1, fail(arvoutcome1==1)
2 xi: streg i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
  d(gamma) ///
  anc(i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m) ///
  anc2(i.clinicc1 yeartbB2006 TBpreEnrol i.cd4tbc2m)
3 predict double cs, csnell
4 stset cs, fail(arvoutcome1==1)
5 sts gen km=s
6 gen double H=-ln(km)
7 line H cs cs if cs<=12, sort legend(off) xtitle("") ///
  title("Generalized Gamma") xlab(0 (2) 12) ylab(0 (2) 12)

```

## D.2 Competing risks analysis

### D.2.1 Figure 3.4

#### a) Cox proportional hazards curves

Line 1 sets the data for survival analysis using the time from the start of TB treatment to the earliest of ART initiation, death pre-ART, permanent deferral or the end of TB treatment as the analysis time and using ART initiation as the failure event. A crude Cox regression for CD4 count is performed in line 2. Hazard curves stratified by covariate value are plotted in line 3.

```
1 stset tbtoaddasstbend, fail(outcomeb4asstbend==1) scale(30.4375)
2 stcox i.cd4tbc2m, nolog
3 stcurve , haz at1(cd4tbc2m=0) at2(cd4tbc2m=50) ///
  at3(cd4tbc2m=100) at4(cd4tbc2m=150) ///
  at5(cd4tbc2m=200) at6(cd4tbc2m=999) range(0 8) ///
  legend(label(1 "0-49") label(2 "50-99") label(3 "100-149") ///
  label(4 "150-199") label(5 "{&ge} 200") ///
  label(6 "missing ") col(3)) ///
  xtitle(" " "Time from start of TB treatment (months)") ///
  ytitle("Hazard" " ") scheme(s2color) title("") ///
  lpat(solid dash shortdash longdash longdash_dot dash_dot) ///
  text(-0.425 -0.35 "CD4 count") ///
  text(-0.525 -0.35 "category") ///
  text(-0.625 -0.35 "(cells/{&mu}L)")
```

#### b) Cumulative Incidence curves using (3.32)

In lines 4 and 5, the cumulative incidence curves for ART initiation (by CD4 count category) are estimated treating death pre-ART, permanent deferral and the end of TB treatment as competing risks. These are plotted in line 6.

```
4 stcompet stcompetci3cd4tbc2m = ci , ///
  compet1(0) compet2(2) compet3(4) by(cd4tbc2m)
5 gen cumincART3cd4tbc2m=stcompetci3cd4tbc2m ///
  if outcomeb4asstbend==1
6 line cumincART3cd4tbc2m _t if _t<=8 & cd4tbc2m==0,///
```

```

lpat(solid) || ///
line cumincART3cd4tbc2m _t if _t<=8 & cd4tbc2m==50,///
lpat(dash) || ///
line cumincART3cd4tbc2m _t if _t<=8 & cd4tbc2m==100, ///
lpat(shortdash) || ///
line cumincART3cd4tbc2m _t if _t<=8 & cd4tbc2m==150,///
lpat(longdash_dot) || ///
line cumincART3cd4tbc2m _t if _t<=8 & cd4tbc2m==200, ///
lpat(longdash) || ///
line cumincART3cd4tbc2m _t if _t<=8 & cd4tbc2m==999, ///
lpat(dash_dot) ///
legend(label(1 "0-49") label(2 "50-99") label(3 "100-149") ///
label(4 "150-199") label(5 "{&ge} 200") label(6 "missing") ///
col(3)) xtitle("Time from start of TB treatment (months)") ///
ytitle("Cumulative Incidence " " ") scheme(s2color) ///
text(-0.27 -0.3 "CD4 count") ///
text(-0.32 -0.3 "category") text(-0.37 -0.3 "(cells/{&mu}L)")

```

### c) Cumulative incidence curves using (3.36)

In line 7, a competing risks regression on the sub-distribution hazard is performed, where ART initiation is the event of interest and death pre-ART, permanent deferral and the end of TB treatment are competing risks. Line 8 plots the corresponding cumulative incidence curves by CD4 count.

```

7 stcrreg i.cd4tbc2m, compet(outcomeb4asstbend=0 2 4)
8 stcurve , cif at1(cd4tbc2m=0) at2(cd4tbc2m=50) ///
at3(cd4tbc2m=100) at4(cd4tbc2m=150) ///
at5(cd4tbc2m=200) at6(cd4tbc2m=999) range(0 8) ///
legend(label(1 "0-49") label(2 "50-99") ///
label(3 "100-149") label(4 "150-199") label(5 "{&ge} 200") ///
label(6 "missing") col(3)) ///
xtitle(" " "Time from start of TB treatment (months)") ///
ytitle("Cumulative Incidence " " ") scheme(s2color) ///
lpat(solid dash shortdash longdash longdash_dot dash_dot) ///
title("") text(-0.3 -0.3 "CD4 count") ///
text(-0.36 -0.3 "category") text(-0.42 -0.3 "(cells/{&mu}L)")

```

## D.2.2 Figures 3.5-3.7 and 3.12

### Cause-specific hazard functions

Line 1 sets the data for survival analysis using the time (in months) from the start of TB treatment to the earliest of ART initiation, death pre-ART, permanent deferral or the end of TB treatment as the analysis time and using ART initiation as the failure event. In line 2, the data is sorted by analysis time and a variable denoting the corresponding observation time number is generated. The cause-specific hazard function for ART initiation is generated in line 3. Line 4 smooths over the analysis time using a kernel density estimator to create the variables required to plot smoothed hazard curves for the failure event 'ART initiation'. This process (lines 1,3 and 4) are repeated for each competing event, resetting the data for survival analysis using the same analysis time, but changing the failure event (code not shown). Line 5 plots smoothed hazard curves for each competing event.

```
1 stset tbtoaddasstbendm, fail(outcomeb4asstbend==1)
2 bysort _t: gen tj3=_n
3 sts gen h_art3=h
4 kdensity(_t) [iweight=h_art3] if tj3==1, ///
  gen(x_art3 dens_art3) at(_t) nograph
5 line dens_death3 _t if _t<=8, sort xlab(0 (2) 8) ///
  lpat(solid) || line dens_defer3 _t if _t<=8, sort ///
  lpat(dash)|| line dens_t bend2 _t if _t<=8, sort ///
  lpat(shortdash)|| line dens_art3 _t if _t<=8, sort ///
  xtitle(" " "Time from start of TB treatment (months)") ///
  ytitle("Hazard" " ") ///
  subtitle("1.a) Cause-Specific Hazard Functions")
  scheme(s2color) lpat(longdash_dot) ///
  legend(ring(0) pos(10) col(1) label(1 "Death") ///
  label(2 "Deferral") label(3 "End of TB trt") ///
  label(4 "ART initiation") symxsize(10))
```

### Stacked cause-specific hazard functions

Lines 6-8 set the failure event to be permanent deferral or death pre-ART and estimate smoothed hazard curve for this combined event. Smoothed hazard curves for the combined events of death pre-ART, permanent deferral and the end of TB treatment, as well as for all competing events (including ART

initiation) were also estimated (code not shown). Line 9 creates a variable with value zero (to be used in colouring the regions between lines on plots). In line 10, the ‘stacked’ cause-specific hazard curves are plotted, beginning with the hazard curve for death pre-ART only, followed by the estimated hazard for death pre-ART or permanent deferral, then death pre-ART, permanent deferral or end of TB treatment and lastly hazard of any event.

```

6 stset tbtoaddasstbendm, fail(outcomeb4asstbend==0 2)
7 sts gen h_deathanddefer3=h
8 kdensity(_t) [iweight=h_deathanddefer3] if tj3==1, ///
   gen(x_deathanddefer3 dens_deathanddefer3) at(_t) nograph
9 gen zero=0
10 twoway rarea zero dens_death3 _t if _t<=8 & _t>=0, ///
   sort || rarea dens_death3 dens_deathanddefer3 _t ///
   if _t<=8 & _t>=0, sort || ///
   rarea dens_deathanddefer3 dens_dde3 _t ///
   if _t<=8 & _t>=0, sort || ///
   rarea dens_dde3 dens_adde3 _t if _t<=8 & _t>=0, sort ///
   xlab(0 (2) 8) ///
   xtitle(" " "Time from start of TB treatment (months)", ) ///
   ytitle("Hazard" " ", ) ///
   subtitle("1.b) Stacked Cause-Specific Hazard Functions") ///
   scheme(s2color) legend(order(1 2 3 4) label(1 "Death") ///
   label(2 "Deferral") label(3 "End of TB trt") ///
   label(4 "ART initiation " " ) symxsize(7))

```

### Stacked relative cause-specific hazard functions

The relative densities to facilitate plots of relative hazard curves are calculated in lines 11-14 using the density for the combined event of any competing risk as the denominator. In line 15, the relative hazard curves are plotted (and the region between these curves is coloured). The first curve is the relative hazard of death to all events, the second is the relative hazard of death or deferral to all events, the third is the relative hazard of death, deferral or end of TB treatment to all events and the final curve is horizontal at one.

```

11 gen dens_death_rel3=dens_death3/dens_adde3
12 gen dens_deathanddefer_rel3=dens_deathanddefer3/dens_adde3
13 gen dens_dde_rel3=dens_dde3/dens_adde3
14 gen dens_adde_rel3=dens_adde3/dens_adde3

```

```

15 twoway rarea zero dens_death_rel3 _t if _t<=8 & _t>=0, ///
    sort || rarea dens_death_rel3 dens_deathanddefer_rel3 _t ///
    if _t<=8 & _t>=0, sort || ///
    rarea dens_deathanddefer_rel3 dens_dde_rel3 _t ///
    if _t<=8 & _t>=0, sort || ///
    rarea dens_dde_rel3 dens_adde_rel3 _t if _t<=8 & _t>=0, ///
    sort xlab(0 (2) 8) ///
    xtitle("Time from start of TB treatment (months)") ///
    ytitle("Relative Contributions" " ", ) ///
    subtitle("1.c) Stacked Relative Cause-Specific ///
    Hazard Functions") scheme(s2color) ||///
    line dens_adde_rel3 dens_adde_rel3 dens_adde_rel3 ///
    dens_adde_rel3 dens_adde_rel3 _t if _t<=8 & _t>=0, ///
    lc(edkblue maroon forest_green dkorange dkorange ) ///
    lpat(solid dash shortdash longdash_dot solid) ///
    legend(order(1 2 3 4) label(1 "Death") ///
    label(2 "Deferral") label(3 "End of TB trt") ///
    label(4 "ART initiation      ") symxsize(7))

```

### Cause-specific cumulative incidence functions

In line 16, the failure event is set to be death pre-ART. Lines 17-21 estimate the cause-specific cumulative incidence functions for each of the competing risks ART initiation, deferral, death pre-ART and the end of TB treatment (treating the other events as competing risks). The cause-specific cumulative incidence functions are plotted in line 22.

```

16 stset tbtoaddasstbendm, fail(outcomeb4asstbend==1)
17 stcompet stcompetci3 = ci , compet1(0) compet2(2) compet3(4)
18 gen cumincART3=stcompetci3 if outcomeb4asstbend==1
19 gen cumincDef3=stcompetci3 if outcomeb4asstbend==0
20 gen cumincDeath3=stcompetci3 if outcomeb4asstbend==2
21 gen cumincTBend3=stcompetci3 if outcomeb4asstbend==4
22 line cumincDeath3 cumincDef3 cumincTBend3 cumincART3 ///
    _t if _t<=8, ytitle("Cumulative Incidence" " ") ///
    xtitle(" " "Time from start of TB treatment (months)") ///
    subtitle("2.a) Cause-specific cumulative incidence ///
    Functions") scheme(s2color) ///
    lpat(solid dash shortdash longdash_dot) ///
    legend(ring(0) pos(10) col(1) label(1 "Death") ///

```

```
label(2 "Deferral") label(3 "End of TB trt") ///
label(4 "ART initiation") symxsize(10))
```

### Stacked cause-specific cumulative incidence functions

Lines 23-24 generate a categorical variable that treats death pre-ART and deferral as a combined event. Lines 25-26 generate a categorical variable that treats death pre-ART, deferral and the end of TB treatment as a combined event. Line 27 uses the variable for which death pre-ART and permanent deferral have been combined as one event and sets ART initiation as the failure event. Lines 28-29 estimate the cumulative incidence for the combined event of death pre-ART and permanent deferral (in the presence of the competing risks of ART initiation and the end of TB treatment). Line 30 uses the variable for which death pre-ART, permanent deferral and the end of TB treatment have been combined as one event and sets ART initiation as the failure event. Lines 31-32 estimate the cumulative incidence for the combined event of death pre-ART, permanent deferral and the end of TB treatment (in the presence of the competing risk of ART initiation). Line 33 sets the failure event to be any of the competing risks. Lines 34-35 estimate the cumulative incidence of any event as one minus the Kaplan-Meier estimate for survival. The stacked cumulative incidence curves are then plotted (colouring area between curves). The curve for the cumulative incidence of death is plotted first, followed by the curve for the cumulative incidence of death pre-ART and permanent deferral, then the cumulative incidence of death pre-ART, permanent deferral or the end of TB treatment and lastly the cumulative incidence for any of the competing events (line 36). Note that the plot has been restricted to 8 months and so the total cumulative incidence does not sum to one, but would reach one if time was left unrestricted.

```
23 gen outcomeb4asstbend02 = outcomeb4asstbend
24 replace outcomeb4asstbend02 = 2 if outcomeb4asstbend==0
25 gen outcomeb4asstbend024 = outcomeb4asstbend02
26 replace outcomeb4asstbend024 = 2 if outcomeb4asstbend02==4
27 stset tbtoaddasstbendm, fail(outcomeb4asstbend02==1)
28 stcompet st3competci02 = ci , compet1(2) compet3(4)
29 gen cumincDD3=st3competci02 if outcomeb4asstbend==2
30 stset tbtoaddasstbendm, fail(outcomeb4asstbend024==1)
31 stcompet st3competci024 = ci , compet1(2)
32 gen cumincDDE3=st3competci024 if outcomeb4asstbend024==2
33 stset tbtoaddasstbendm, fail(outcomeb4asstbend02==0 1 2 4)
```

```

34 sts gen kmADDE3 = s
35 gen cumincADDE3=1-kmADDE3
36 twoway rarea zero cumincDeath3 _t if _t<=7.5, sort || ///
    rarea cumincDeath3 cumincDD3 _t if _t<=7.5, sort || ///
    rarea cumincDD3 cumincDDE3 _t if _t<=7.5, sort || ///
    rarea cumincDDE3 cumincADDE3 _t if _t<=7.5, sort ///
    ytitle("Cumulative Incidence" " ") ///
    xtitle(" " "Time from start of TB treatment (months)") ///
    scheme(s2color) ///
    subtitle("2.b) Stacked cause-specific cumulative incidence ///
Functions") legend(label(1 "Death") label(2 "Deferral") ///
label(3 "End of TB trt") ///
label(4 "ART initiation      ") symxsize(7))

```

### Stacked relative cause-specific cumulative incidence functions

Lines 37-40 generate the relative cumulative incidences using the cumulative incidence for all events as the denominator. The relative cumulative incidence curves are plotted in line 41.

```

37 gen relcumincDeath3=cumincDeath3/cumincADDE3
38 gen relcumincDD3=cumincDD3/cumincADDE3
39 gen relcumincDDE3=cumincDDE3/cumincADDE3
40 gen relcumincADDE3=cumincADDE3/cumincADDE3
41 twoway rarea zero relcumincDeath3 _t if _t<=7.5, sort || ///
    rarea relcumincDeath3 relcumincDD3 _t if _t<=7.5, sort || ///
    rarea relcumincDD3 relcumincDDE3 _t if _t<=7.5, sort || ///
    rarea relcumincDDE3 relcumincADDE3 _t if _t<=7.5, sort ///
    ytitle("Relative Contributions" " ") ///
    xtitle(" " "Time from start of TB treatment (months)") ///
    legend(label(1 "Death") label(2 "Deferral") ///
label(3 "End of TB trt") label(4 "ART initiation      ") ///
symxsize(7)) scheme(s2color) subtitle("2.c) Stacked ///
Relative Cause-specific cumulative incidence functions")

```

The plots for Figure 3.12: ‘Cause-specific hazard and cumulative incidence functions’ were obtained in a very similar manner to above.

### D.2.3 Models for competing risks analysis, Table 4.10

The data is set for survival analysis using the time from the start of TB treatment to the first competing event as the analysis time and ART initiation as the failure event (code not shown).

```
*Model 14: Regression on sub-distribution hazards,
*no interactions of covariates with time are included
xi: stcrreg yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
compete(outcomeb4asstbend==0 2 4) nolog
```

```
*Model 15: Regression on sub-distribution hazards,
*interactions of covariates with analysis time are included
xi: stcrreg yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
compete(outcomeb4asstbend==0 2 4) nolog ///
tvc(yeartbB2006 TBpreEnrol i.cd4tbc2m)
```

```
*Model 16: Regression on sub-distribution hazards,
*interactions of covariates with the
*natural logarithm of analysis time are included
xi: stcrreg yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
compete(outcomeb4asstbend==0 2 4) nolog ///
tvc(yeartbB2006 TBpreEnrol i.cd4tbc2m) texp(ln(_t))
```

```
*Model 17: Regression on sub-distribution hazards,
*no interactions of covariates with time are included
xi: stcox yeartbB2006 TBpreEnrol i.cd4tbc2m, nolog
```

```
*Model 18: Regression on sub-distribution hazards,
*interactions of covariates with analysis time are included
xi: stcox yeartbB2006 TBpreEnrol i.cd4tbc2m, ///
nolog tvc(yeartbB2006 TBpreEnrol i.cd4tbc2m)
```

```
*Model 19: Regression on sub-distribution hazards,
*interactions of covariates with the
*natural logarithm of analysis time are included
xi: stcox yeartbB2006 TBpreEnrol i.cd4tbc2m , //
nolog tvc(yeartbB2006 TBpreEnrol i.cd4tbc2m) texp(ln(_t))
```

### D.2.4 Figures 4.13 and 4.15

The data was set for survival analysis, using the time from the start of TB treatment to the first competing risk as the analysis time and ART initiation as the failure event. A Cox regression including covariates for the CD4 count category, the calendar period of TB diagnosis (2002-2005 versus 2006-2008) and whether the patient was diagnosed with TB before enrolment for ART (versus as a part of the screening process for ART initiation).

```
stset tbtoaddasstbendm, fail(outcomeb4asstbend==1)
xi: stcox i.cd4tbcatt2m yeartbB2006 TBpreEnrol
```

First plot of Figure 4.13: Comparison of log-log survival curves by covariate. This plot shows a comparison of the smoothed log-log survival curves for patients diagnosed with TB in 2002-2005 versus 2006-2008, where the failure event is ART initiation.

```
stphplot, by(yeartbB2006) adjust(cd4tbcatt2m TBpreEnrol ) z ///
legend(label(1 "2006-2008") label(2 "2002-2005") ///
order(2 1) col(1) symxsize(7) ring(0) pos(2)) ///
subtitle("Calendar period of TB diagnosis") ///
scheme(s2color) plot1opts(m(i)) plot2opts(m(i) lpat(dash)) ///
yttitle("-log[-log(S(t))]" " ") xttitle(" " "log(t)")
```

First plot of Figure 4.15: Smoothed cause-specific hazard curves by calendar period (1) and clinic (2) of TB diagnosis for each competing risk. This plot shows a comparison of the smoothed cause-specific hazard curves for patients diagnosed with TB in 2002-2005 versus 2006-2008, where the failure event is ART initiation.

```
stcurve, haz at1(yeartbB2006=0, TBpreEnrol=0, cd4tbcatt2m=0) ///
at2(yeartbB2006=1, TBpreEnrol=0, cd4tbcatt2m=0) range(0 8) ///
scheme(s2color) lpat(solid dash) legend(off) ///
subtitle("1.a) ART initiation") xttitle("") title("") ///
yttitle("Hazard")
```

### D.2.5 Figure 4.16 plot 1.a)

The code for lines 1-14 was adapted for my data from that of StataCorp (2009b) pages 207-211, while the code for lines 15-19 was adapted from that of StataCorp (2009b) pages 226-227. In line 1, I set the data for survival analysis, using the time from the start of TB treatment to the first competing

risk as the analysis time and ART initiation as the failure event (line 1). Line 2 performs a univariate Cox regression for the calendar period of TB diagnosis (2002-2005 versus 2006-2008). Line 3 predicts the hazard contribution for patients whose TB was diagnosed in 2006 or later. StataCorp (2009b) explain that the hazard contributions ‘represent the increments of the cumulative hazard at each event time’. Line 4 sorts the data in ascending order for analysis time (earlier times first) and descending order for the events (patients who experienced the event of interest at a given time are placed above those who experienced a competing risk at that time). Line 5 replaces the hazard contribution variable with a missing value for the second and subsequent patients experiencing an event at a specific time. Line 6 generates the hazard contribution at each time for patients whose TB was diagnosed before 2006. Generate the hazard contribution for patients for whom TB was diagnosed before 2006. Lines 1-6 are then repeated for each competing event (changing the failure event and labelling of hazard contributions appropriately, code not shown).

```

1  stset tbtoaddasstbendm, fail(outcomeb4asstbend==1)
2  stcox yeartbB2006, nolog
3  predict h_ART_0, basehc
4  gsort _t-_d
5  by _t: replace h_ART_0= . if _n>1
6  gen h_ART_1 = h_ART_0*exp(_b[yeartbB2006])

```

Observations for which there was no non-missing hazard contribution for any of the competing events were deleted (line 7). These corresponded to times at which there was more than one patient who experienced a particular event and for which the hazard contribution for all patients but the first was set to missing. In line 8, the value of the variable for the hazard contribution corresponding to ART initiation for patients who were diagnosed with TB in 2006-2008 was replaced with zero (if it had previously been set to zero). Such observations corresponded to patients who experienced a competing event other than ART initiation. Similarly for line 9. Lines 8 and 9 were repeated for each competing event (code not shown). The data were then sorted by analysis time (line 10). In lines 11-14, the survival and cumulative incidence curves (based on cause-specific hazards) were calculated by covariate value using the hazard contributions already estimated.

```

7  drop if missing(h_ART_0) & missing(h_DEA_0) ///
   & missing(h_DEF_0) & missing(h_END_0)
8  replace h_ART_0=0 if missing(h_ART_0)
9  replace h_ART_1=0 if missing(h_ART_1)

```

```

10 sort _t
11 gen S_0 = exp(sum(log(1-h_ART_0-h_DEA_0-h_DEF_0-h_END_0)))
12 gen S_1 = exp(sum(log(1-h_ART_1-h_DEA_1-h_DEF_1-h_END_1)))
13 gen cif_ART_0 = sum(S_0[_n-1]*h_ART_0)
14 gen cif_ART_1 = sum(S_1[_n-1]*h_ART_1)

```

In line 15 the data was reset as per line 1 (as this would have changed when hazard contributions for other competing risks were estimated). Regression on the sub-distribution hazards (not including any covariates) was then performed for patients who were diagnosed with TB in 2006-2008 (line 16) and the corresponding cumulative incidence function obtained (line 17). This was repeated for patients who were diagnosed with TB in 2002-2005 (lines 18-19). The cumulative incidence functions obtained in this way correspond to non-parametric cumulative incidence functions, which make no proportionality assumption and are simply stratified by covariate.

```

15 stset tbtoaddasstbendm, fail(outcomeb4asstbend==1)
16 stcrreg if !yeartbB2006, compete(outcomeb4asstbend==0 2 4)
17 predict basecif_ART_0, basecif
18 stcrreg if yeartbB2006, compete(outcomeb4asstbend==0 2 4)
19 predict basecif_ART_1, basecif

```

In line 20, a regression on the sub-distribution hazards for all patients that included the calendar period of TB treatment was performed. Line 21 plots the cumulative incidence functions by calendar period of TB diagnosis that correspond to the assumption that covariates have proportional sub-distribution hazards for the event ART initiation (blue), as well as the previously calculated cumulative incidence functions that assume that the cause-specific hazards of all events are proportional (cif\_ART\_0 and cif\_ART\_1 in red) and the cumulative incidence functions that make no proportionality assumption (basecif\_ART\_0 and basecif\_ART\_1 in black).

```

20 stcrreg yeartbB2006, compete(outcomeb4asstbend==0 2 4)
21 stcurve, cif at1(yeartbB2006=0) at2(yeartbB2006=1) ///
    title("1.a) ART Initiation", col(black)) ///
    range(0 8) ylab(0 (0.2) 1) ///
    addplot( line cif_ART* basecif_ART* _t if _t<8, ///
    connect(J J) sort title("") ///
    ytitle("Cumulative Incidence" " ", col(black)) ///
    xtitle(" " "Time from start of TB treatment (months)") ///
    lpat( solid shortdash solid shortdash) ///
    lcol(red red black black) ///
    legend(label(5 "2006-2008") label(6 "2002-2005" " ") ///

```

```
order(5 6))) lpat( solid shortdash ) lcol(blue blue)  ///
scheme(s2color)
```

## D.3 Inverse Probability Weighting

The following code is adapted for my data from that of Fewell et al. (2001). Line 1 generates a cubic spline with 5 knot points. Line 2 sets the data for survival analysis. The analysis time for each patient starts at the later of the date of starting TB treatment or enrolling for ART and ends when the patient dies or is censored (through permanent deferral for ART, transferral to another clinic, loss to follow up or administrative censoring after 78 weeks (approximately 18 months) of follow-up). This analysis time is split up into 1 week intervals. The failure event is death.

```
1 mkspline laterofsplineM=lateroftoorder, cubic di
2 stset lateroftoorderm, id(newpid) fail(dead)
```

### D.3.1 Estimating Stabilized Inverse Probability of Treatment Weights (SIPTWs)

#### Estimating the denominator of the SIPTWs

Line 3 is the pooled logistic regression used to calculate the denominator of the Stabilized Inverse Probability of Treatment Weights (IPTWs). It includes observations for patients up to the first interval for which they are on ART. Time-updated variables are included in this logistic regression. Line 4 generates the predicted probability that a patient is on ART. Line 5 assigns a value of one to patients who are already on ART (since I assume that patients remain on ART after having initiated ART). In line 6, the probability variable `ponart` is replaced by the probability of a patient receiving the treatment that they actually received at that interval (i.e. the probability of not being on ART for patients not on ART and the probability of being on ART for patients on ART). Line 7 sorts the data by the patient identifier and the time variable. In line 8, the probability variable is replaced by the product of the probability of the treatment received in that interval and the probabilities (of receiving the treatment that was received) for all previous intervals. This probability variable now represents the probability at each time of the patient having the treatment history that they were observed

to have. The probability variable is renamed in line 9 to reflect that it corresponds to the denominator for the SIPTW.

```

3 logistic onart male tbtoenrolm i.cd4laterof1cat5 ///
  i.weightlaterof1cat2 i.hblaterof1cat3 i.cd4bfc5cat5 ///
  i.hbbfc5cat3 i.weightbfc5cat2 laterofsplineM* ///
  if lateroftoorder<=minlaterofonart
4 predict ponart if e(sample)
5 replace ponart=1 if lateroftoorder>minlaterofonart
6 replace ponart=ponart*onart+(1-ponart)*(1-onart)
7 sort newpid lateroftoorder
8 by newpid: replace ponart=ponart*ponart[_n-1] if _n>1
9 rename ponart denomonart

```

### Estimating the numerator of the SIPTWs

Lines 10-16 are very similar to lines 3-9. The differences are that the pooled logistic regression to calculate the numerator of the SIPTWs does not include time-updated variables (line 10) and that the probability variable is renamed to reflect that it corresponds to the numerator for the SIPTW (line 16).

```

10 logistic onart male tbtoenrolm i.cd4laterof1cat5 ///
  i.weightlaterof1cat2 i.hblaterof1cat3 laterofsplineM* ///
  if lateroftoorder<=minlaterofonart
11 predict ponart if e(sample)
12 replace ponart=1 if lateroftoorder>minlaterofonart
13 replace ponart=ponart*onart+(1-ponart)*(1-onart)
14 sort newpid lateroftoorder
15 by newpid: replace ponart=ponart*ponart[_n-1] if _n>1
16 rename ponart numonart

```

### Generating the SIPTW

Line 17 calculates the SIPTW as the quotient of the numerator and denominator for the SIPTWs and line 18 displays a summary of this variable.

```

17 gen stabweightonart=numonart/denomonart
18 summ stabweightonart, d

```

### D.3.2 Estimating the Stabilized Inverse Probability of Censoring Weights (SIPCWs)

Line 1 sorts the data by the patient identifier variable and the time variable. Line 2 generates a variable containing the latest time at which a patient is still under observation. Lines 3-5 generate variables that indicate whether a patient was censored (by being lost to follow up, transferred to another clinic or permanently deferred for ART) before the time of administrative censoring (546 days (approximately 18 months) after the later of the start of TB treatment and enrolment for ART). Line 6 generates a variable that indicates whether a patient was censored by any of the three events mentioned above before the time of administrative censoring. Line 7 centres the calendar year of starting TB treatment at the year 2005.

```

1 sort newpid lateroftoorder
2 by newpid: egen maxlateroftoorder=max(lateroftoorder)
3 gen ltfuC=ltfudate!=. & maxlateroftoorder==lateroftoorder ///
  & maxlateroftoorder<546
4 gen transferoutC=transferoutdate!=. ///
  & maxlateroftoorder==lateroftoorder & maxlateroftoorder<546
5 gen rxpermdeferC=rxpermdeferdate!=. ///
  & maxlateroftoorder==lateroftoorder & maxlateroftoorder<546
6 gen ltrC=ltfuC==1 | transferoutC==1 | rxpermdeferC==1
7 gen yeartbC2005=yeartb-2005

```

#### Estimating the denominator of the SIPCWs

Line 8 shows the command to perform the pooled logistic regression used to estimate the denominator for the SIPCW. It includes all observations for patients (up to 546 days after the later of the start of TB treatment or enrolment for ART). Time-updated variables are included in this logistic regression. In line 9, the probability of being censored is predicted using the results from the logistic regression. In line 10, the probability variable **pcens** is replaced with the probability of a patient being censored if they were censored by that interval or remaining uncensored if they were not yet censored at that interval. Line 11 sorts the data by the patient identifier and time variables. Line 12 calculates the product of the estimated probabilities of being censored or uncensored at each interval up to the given interval, such that the probability variable now reflects the probability of the censoring history of the patient up to that interval. In line 13, the censoring variable

is renamed to reflect that it estimates the denominator for the SIPCW.

```

8 logistic ltrC onart yeartbC2005 i.cd4laterof1cat5 ///
  i.hblaterof1cat3 i.weightlaterof1cat2 i.cd4bfc5cat5 ///
  i.hbbfc5cat3 i.weightbfc5cat2 laterofsplineM*
9 predict pcens
10 replace pcens=pcens*ltrC+(1-pcens)*(1-ltrC)
11 sort newpid lateroftoorder
12 by newpid: replace pcens=pcens*pcens[_n-1] if _n>1
13 rename pcens denomcens

```

### Estimating the numerator of the SIPCWs

Lines 14-19 are similar to lines 8-13. The differences are that the logistic regression to estimate the numerator for the SIPCWs does not include time-updated variables (line 14) and the probability variable is renamed to reflect that it is the estimated value for the numerator of the SIPCW in line 19.

```

14 logistic ltrC onart yeartbC2005 i.cd4laterof1cat5 ///
  i.hblaterof1cat3 i.weightlaterof1cat2 laterofsplineM*
15 predict pcens
16 replace pcens=pcens*ltrC+(1-pcens)*(1-ltrC)
17 sort newpid lateroftoorder
18 by newpid: replace pcens=pcens*pcens[_n-1] if _n>1
19 rename pcens numcens

```

### Generating the SIPCW and SIPTCW

Line 20 calculates the SIPTW as the quotient of its numerator and denominator and line 21 displays a summary of this variable. Line 22 calculates the Stabilized Inverse Probability of Treatment and Censoring Weight (SIPTCW) as the product of the SIPTW and SIPCW and line 23 displays a summary of this variable.

```

20 gen stabweightcens=numcens/denomcens
21 summ stabweightcens, d
22 gen stabweightonartcens=stabweightonart*stabweightcens
23 summ stabweightonartcens, d

```

### D.3.3 Figure 7.7: Median, interquartile range and 5th and 95th percentiles of stabilized IPTWs and IPCWs

Lines 1-6 calculate (respectively) the mean, median, 25th percentile, 75th percentile, 5th percentile and 95th percentile of the SIPTWs for each time interval. Line 7 plots the median in red and colours the interquartile range light grey and the area between the 5th and 95th percentiles dark grey. Similar code (replacing `stabweightonart` with `stabweightcens` and changing labels appropriately) was used to calculate and plot the corresponding values for the SIPCWs (not shown).

```

1 bysort lateroftoorderwk: egen ///
  meanstabweightonart=mean(stabweightonart)
2 bysort lateroftoorderwk: egen ///
  medstabweightonart=median(stabweightonart)
3 bysort lateroftoorderwk: egen ///
  p25stabweightonart=pctile(stabweightonart), p(25)
4 bysort lateroftoorderwk: egen ///
  p75stabweightonart=pctile(stabweightonart), p(75)
5 bysort lateroftoorderwk: egen ///
  p5stabweightonart=pctile(stabweightonart), p(5)
6 bysort lateroftoorderwk: egen ///
  p95stabweightonart=pctile(stabweightonart), p(95)
7 twoway rarea p5stabweightonart p95stabweightonart ///
  lateroftoorderwk || rarea p25stabweightonart ///
  p75stabweightonart lateroftoorderwk || ///
  line medstabweightonart lateroftoorderwk, lcol(red) ///
  lpat(solid) lw(medium) legend(off) ///
  xtitle(" " "Time (weeks)") ///
  ytitle("Stabilized IPTW" " ") ///
  xlab(0 (16) 80) text(0.85 65 "p5-p25") ///
  text(0.96 65 "median" ,col(red)) ///
  text(1.1 65 "p95-p75") subtitle("Stabilized IPTW")

```

### D.3.4 Figure 7.5: Box plots of stabilized IPTWs by baseline CD4 count category

Line 1 creates the time variable including only the selected times (in weeks) for which the box plots were plotted. Line 2 shows the command for generating a box plot for the baseline CD4 count category 0-49 cells/ $\mu$ L. Similar plots were generated for each of the CD4 count categories and then combined into one figure (code not shown). Similar code was also used to obtain Figure 7.6 (replacing `stabweightonart` with `stabweightcens`).

```
1 gen bovertime=lateroftoorderwk if lateroftoorderwk==16 | ///
  lateroftoorderwk==32 | lateroftoorderwk==48 | ///
  lateroftoorderwk==64
2 graph box stabweightonart if cd4laterof1cat5==0, ///
  over(bovertime) ytitle("") title("0-49") noout note("")
```

### D.3.5 Code for performing weighted Cox regressions in Stata when there are multiple observations per patient

The code shown below to perform weighted Cox regressions in Stata when there are multiple observations per patient was made available by Francesca Little. In line 1, the analysis time and failure event are set and multiple observations are grouped by a patient identifier (`idvariable`). Lines 2-4 capture the start times, end times and event indicators for each interval of observation for every patient. The values of these variables are then used to set the data for survival analysis (line 5) treating as independent observations each interval for which a patient was observed and thus allowing weights that vary by observation interval for each patient.

```
1 stset timevariable , fail(failure) id(idvariable)
2 gen starttime=_t0
3 gen stime=_t
4 gen event=_d
5 stset stime [pw=sIPTCW], fail(event) time0(starttime)
```

## D.4 Marginal Structural Models

The data was set for weighted survival analysis (adapting the code shown in Section D.3.5 for my data). The analysis time was from the later of the start of TB treatment and enrolment for ART until death or censoring. It was split into one week intervals. The failure event was death. The patient observations were weighted using the estimated SIPTCWs.

### D.4.1 Table 7.11: Cox regressions using IPTCW to assess effect on hazard of death of being on ART

The code shown below is for performing Cox regressions using the treatment function ‘On ART’. Similar code (replacing `onart` with `cumnotonart`, a variable encoding the number of weeks for which a patient had not been on ART at each interval) was used to estimate the corresponding models for the treatment function ‘Weeks not on ART’. Model 1.b does not allow interactions of covariates with analysis time, while Model 1.c does include interactions of some covariates with analysis time.

```
stset stime [pw=stabweightonartcens], ///
fail(event) time0(starttime)

*Model 1.b IPTCW
xi: stcox onart tbtoenrolm agelyrs stage4 i.cd4laterof1cat5 ///
i.weightlaterof1cat2 i.hblaterof1cat3 , nolog

*Model 1.c IPTCW
xi: stcox onart tbtoenrolm agelyrs stage4 i.cd4laterof1cat5 ///
i.weightlaterof1cat2 i.hblaterof1cat3 , nolog ///
tvc(tbtoenrolm i.cd4laterof1cat5 i.hblaterof1cat3)
```

## D.5 Multiple Imputation

In the first line below, I set the data for survival analysis using only one observation per patient (analysis time is from the later of the start of TB treatment and enrolment for ART until death or censoring and the failure event is death). In the second line, I create a variable that stores the Nelson-Aalen estimator for the cumulative hazard and name it `NA`.

```
stset maxlateroftoorder if n==1, fail(death)
sts gen NA=na
```

### D.5.1 Multivariate Normal Data Augmentation (MVN DA)

In lines 1-4 below, I set the data for multiple imputation, clear previous settings for survival analysis and register variables to be imputed (if they have missing values) or as passive (if they will be used to predict those that have missing values). Line 5 is the command to perform multiple imputation (using MVN DA). I only imputed the baseline values for the first observation for each patient. I used a value of  $k = 500$  as the burn-in period and the number of iterations between each value taken as an imputed value and performed 40 imputations.

```
1 mi set mlong
2 mi stset, clear
3 mi register imputed cd4laterof1logit weightlaterof1logit ///
  hblaterof1logit
4 mi register passive tbtoenrolm lateroftoartcat1 ///
  lateroftoartcat2 lateroftoartcat3 lateroftoartcat9 ///
  lateroftocd4date lateroftohbdate lateroftoweightdate ///
  agelyrs yeartbC2005 newtb stage4 sppulm extrapulm ///
  diedprerx rxpermdeferred diedonrx transferredout ltftu ///
  male NA
5 mi impute mvn cd4laterof1logit hblaterof1logit ///
  weightlaterof1logit = tbtoenrolm lateroftoartcat1 ///
  lateroftoartcat2 lateroftoartcat3 lateroftoartcat9 ///
  lateroftocd4date lateroftohbdate lateroftoweightdate ///
  agelyrs yeartbC2005 newtb stage4 sppulm extrapulm ///
  diedprerx rxpermdeferred diedonrx transferredout ltftu ///
  male NA if n==1, rseed(1) add(40) burnin(500) ///
  burnbetween(500) initm(em, iter(100) nolog)
```

#### Preparation of variables to be used in models that include imputed values

In lines 6-8, I perform the back-transformations in order to obtain the imputed values for variables on their original scale. Line 9 sorts the data by

the patient identifier and the number of the imputation. Lines 10-12 create variables that contain the imputed baseline values (for the variables that had missing values) for every observation of a patient, by replicating the values obtained for the first observation of a patient for a particular imputation set across all observations for that patient. Lines 13-19 create categorical variables from the continuous versions of the imputed variables. Line 20 creates a binary variable for whether the patient had an imputed CD4 value in the range 0-49 cells/ $\mu$ L. In line 21, the binary variable is multiplied by the analysis time. Similar binary variables and their interactions with time were created for every category of the imputed variables. The continuous versions of the variables were also multiplied by the analysis time to allow for the interactions with time to be included in the models. The continuous version of the baseline CD4 values was divided by 50 and the corresponding interaction with time created.

```

6 mi passive: gen cd4laterof1i = ///
  round(600*exp(cd4laterof1logit)/(1+exp(cd4laterof1logit)))
7 mi passive: gen hblaterof1i = ///
  round(13*(exp(hblaterof1logit)/(1+exp(hblaterof1logit)))+5,.1)
8 mi passive: gen weightlaterof1i = ///
  round(95*(exp(weightlaterof1logit)/ ///
    (1+exp(weightlaterof1logit)))+25,.1)
9 sort newpid _mi_m
10 mi passive: by newpid _mi_m: ///
  egen cd4laterof1icf=min(cd4laterof1i)
11 mi passive: by newpid _mi_m: ///
  egen hblaterof1icf=min(hblaterof1i)
12 mi passive: by newpid _mi_m: ///
  egen weightlaterof1icf=min(weightlaterof1i)
13 mi passive: egen cd4laterof1icat5=cut(cd4laterof1icf), ///
  at(0 50 100 200 1500)
14 mi passive: gen hblaterof1icat3=hblaterof1icf>=10 ///
  & male==0 & hblaterof1icf!=.
15 mi passive: replace hblaterof1icat3=1 if hblaterof1icf>=11 ///
  & male==1 & hblaterof1icf!=.
16 mi passive: replace hblaterof1icat3=. if hblaterof1icf==.
17 mi passive: gen weightlaterof1icat2=weightlaterof1icf>=50 ///
  & weightlaterof1icf!=.
18 mi passive: replace weightlaterof1icat2=2 ///
  if weightlaterof1icf>=60 & weightlaterof1icf!=.
19 mi passive: replace weightlaterof1icat2=. if weightlaterof1icf==.

```

```

20 mi passive: gen cd4laterof1i0=cd4laterof1icat5==0 ///
    if cd4laterof1icat5!=.
21 mi passive: gen cd4laterof1i0Xlm = ///
    cd4laterof1i0*lateroftoorderm

```

### Results after performing multiple imputation by MVN DA

Calling line 22 results in the number of patients, the mean, standard deviation, minimum, lower quartile, median, upper quartile and maximum values of the imputed baseline CD4 to be displayed for every imputed data set, as well as for the unimputed data. Line 23 estimates the mean value for the baseline CD4 by combining all imputed data sets according to Rubin's Rules. Lines 22 and 23 can be adjusted to obtain the corresponding values for the baseline haemoglobin and weight of patients. Line 24 sets the multiply imputed data for (unweighted) survival analysis. Lines 25 and 26 perform Cox regressions (which include interactions of some covariates with time) using the multiply imputed data, corresponding to the results shown on the right hand sides of Tables 7.18 and 7.19. Results for models weighted using the SIPTCWs can be obtained by running lines 29-32 and then repeating lines 25 and 26.

```

22 mi xeq: tabstat cd4laterof1i if n==1 & cd4laterof1val==., ///
    stats(n mean sd min q max)
23 mi estimate: mean cd4laterof1i if n==1
24 mi stset lateroftoorderm, id(newpid) fail(dead)
25 mi estimate, eform: stcox onart tbtoenrolm agelyrs stage4 ///
    cd4laterof1i50 cd4laterof1i100 cd4laterof1i200 ///
    hblaterof1i1 i.weightlaterof1icat2 ///
    tbtoenrolmXlateroftoorderm cd4laterof1i50Xlm ///
    cd4laterof1i100Xlm cd4laterof1i200Xlm hblaterof1i1Xlm
26 mi estimate, eform: stcox onart tbtoenrolm agelyrs stage4 ///
    cd4laterof1icfD50 hblaterof1icf weightlaterof1icf ///
    tbtoenrolmXlateroftoorderm cd4laterof1icfD50Xlm ///
    hblaterof1icfXlm
29 gen starttime=_t0
30 gen stime=_t
31 gen event=_d
32 mi stset stime [pw=stabweightonartcensC], ///
    fail(event) time0(starttime)

```

### D.5.2 Multiple Imputation by Chained Equations (MICE)

The variables included in the imputation equation to obtain multiple imputations by chained equations were the same as those included in the imputation equation to obtain multiple imputations by MVN DA. The command to perform MICE is shown below.

```
ice cd4laterof1logit hblaterof1logit weightlaterof1logit ///
tbtoenrolm lateroftoartcat1 lateroftoartcat2 ///
lateroftoartcat3 lateroftoartcat9 lateroftocd4date ///
lateroftohbdate lateroftoweightdate agelyrs ///
yeartbC2005 newtb stage4 sppulm extrapulm ///
diedprerx rxpermdeferred diedonrx transferredout ltfu ///
male NA using "filename" if n==1, ///
m(40) genmiss(M) replace seed(1)
```

# Bibliography

- L. Aaron, D. Saadoun, I. Calatroni, O. Launay, N. Mémain, V. Vincent, G. Marchal, B. Dupont, O. Bouchaud, D. Valeyre, and O. Lortholary. Tuberculosis in HIV-infected patients: a comprehensive review. *European Society of Clinical Microbiology and Infectious Diseases*, 10:388–398, 2004.
- S. Akksilp, O. Karnkawinpong, W. Wattanaamornkiat, D. Viriyakitja, P. Monkongdee, W. Sitti, D. Rienthong, T. Siraprapasiri, C. D. Wells, J. W. Tappero, and J. K. Varma. Antiretroviral therapy during tuberculosis treatment and marked reduction in death rate of HIV-infected patients, Thailand. *Emerging Infectious Diseases*, 13(7):1001–1007, July 2007.
- M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf. Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1):40–49, 2011.
- H. Bang and J. M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(December):962–972, 2005.
- F.-X. Blanc, T. Sok, D. Laureillard, L. Borand, C. Rekacewicz, E. Nerrienet, Y. Madec, O. Marcy, S. Chan, N. Prak, C. Kim, K. K. Lak, C. Hak, B. Dim, C. I. Sin, S. Sun, B. Guillard, B. Sar, S. Vong, M. Fernandez, L. Fox, J.-F. Delfraissy, and A. E. G. for the CAMELIA (ANRS 1295-CIPRA KH001) Study Team. Earlier versus later start of antiretroviral therapy in HIV-infected adults with tuberculosis. *The New England Journal of Medicine*, 365(16):1471–1481, October 2011.
- A. Boule, P. Clayden, K. Cohen, T. Cohen, F. Conradie, K. Dong, N. Geffen, A. Grimwood, R. Hurtado, C. Kenyon, S. Lawn, G. Maartens, G. Meintjes, M. Mendelson, M. Murray, M. Rangaka, I. Sanne, D. Spencer, J. Taljaard, E. Variava, W. D. F. Venter, and D. Wilson. Prolonged deferral of antiretroviral therapy in the SAPIT trial: Did we need a clinical trial to tell us that this would increase mortality? *South African Medical Journal*, 100(9):566–571, 2010a.

- A. Boulle, G. V. Cutsem, K. Hilderbrand, C. Cragg, M. Abrahams, S. Mathee, N. Ford, L. Knight, M. Osler, J. Myers, E. Goemaere, D. Coetzee, and G. Maartens. Seven-year experience of a primary care antiretroviral treatment programme in Khayelitsha, South Africa. *AIDS*, 24:563–572, 2010b.
- W. J. Burman and B. E. Jones. Treatment of HIV-related tuberculosis in the era of effective antiretroviral therapy. *American Journal of Respiratory and Critical Care Medicine*, 164:7–12, 2001.
- M. Cleves, W. Gould, and R. Gutierrez. *An Introduction to Survival Analysis using Stata*. Stata Press, 2nd edition, 2008.
- J. Cohen. Bioethicists assail a celebrated TB/HIV treatment trial. *Science*, 328:799–801, May 2010.
- S. R. Cole and M. A. Hernán. Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168(6): 656–664, 2008.
- S. R. Cole, M. A. Hernán, J. M. Robins, K. Anastos, J. Chmiel, R. Detels, C. Ervin, J. Feldman, R. Greenblatt, L. Kingsley, S. Lai, M. Young, M. Cohen, and A. M. noz. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *American Journal of Epidemiology*, 159(7):687–694, 2003.
- S. R. Cole, M. A. Hernán, J. B. Margolick, M. H. Cohen, , and J. M. Robins. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on cd4 cell count. *American Journal of Epidemiology*, 162(5):471–478, 2005.
- S. R. Cole, M. A. Hernán, K. Anastos, B. D. Jamieson, and J. M. Robins. Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type 1 RNA viral load using a marginal structural left-censored mean model. *American Journal of Epidemiology*, 166(2):219–227, 2007.
- S. R. Cole, L. P. Jacobson, P. C. Tien, L. Kingsley, J. S. Chmiel, and K. Anastos. Using marginal structural measurement-error models to estimate the long-term effect of antiretroviral therapy on incident AIDS or death. *American Journal of Epidemiology*, 171(1):113–122, 2010.
- D. Collett. *Modelling Survival Data in Medical Research*. Texts in Statistical Science. Chapman & Hall/CRC, 2nd edition, 2003.

- L. M. Collins, J. L. Schafer, and C.-M. Kam. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6(4):330–351, 2001.
- D. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.
- U. Dafni. Landmark analysis at the 25-year landmark point. *Circulation Cardiovascular Quality and Outcomes*, 4:363–371, May 2011.
- R. B. D’Agostino, M.-L. Lee, and A. J. Belanger. Relation of pooled logistic regression to time dependent cox regression analysis: the Framingham heart study. *Statistics in Medicine*, 9:1501–1515, 1990.
- G. L. Dean, S. G. Edwards, N. J. Ives, G. Matthews, E. F. Fox, L. Navaratne, M. Fisher, G. P. Taylor, R. Miller, C. B. Taylor, A. de Ruiter, and A. L. Pozniak. Treatment of tuberculosis in HIV-infected persons in the era of highly active antiretroviral therapy. *AIDS*, 16:75–83, 2002.
- K. Dheda, F. C. Lampe, M. A. Johnson, and M. C. Lipman. Outcome of HIV-associated tuberculosis in the era of highly active antiretroviral therapy. *Journal of Infectious Diseases*, 190:1670–1676, 2004.
- L. R. Fairall, M. O. Bachmann, G. M. C. Louwagie, C. van Vuuren, P. Chikobvu, D. Steyn, G. H. Staniland, V. Timmerman, M. Msimanga, C. J. Seebregts, A. Boule, R. Nhiwatiwa, E. D. Bateman, M. F. Zwarenstein, and R. D. Chapman. Effectiveness of antiretroviral treatment in a South African program. *Archives of Internal Medicine*, 168(1):86–93, January 2008.
- Z. Fewell, M. A. Hernán, F. Wolfe, K. Tilling, H. Choi, and J. A. C. Sterne. Controlling for time-dependent confounding using marginal structural models. *The Stata Journal*, 1(1):1–17, 2001.
- J. P. Fine and R. J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.
- M. F. Franke, J. M. Robins, J. Mugabo, F. Kaigamba, L. E. Cain, J. G. Fleming, and M. B. Murray. Effectiveness of early antiretroviral therapy initiation to improve survival among HIV-infected adults with tuberculosis: A retrospective cohort study. *PLoS Medicine*, 8(5):1–8, 2011.

- D. B. Garone, K. Hilderbrand, A. M. Boulle, D. Coetzee, E. Goemaere, G. V. Cutsem, and D. Besada. Khayelitsha 2001-2011: 10 years of primary care HIV and TB programmes. *The Southern African Journal of HIV Medicine*, pages 33–38, December 2011.
- A. Gelman and T. P. Speed. Characterizing a joint probability distribution by conditionals. *Journal of the Royal Statistical Society*, 55(1):185–188, 1993.
- A. Gichangi and W. Vach. The analysis of competing risks data: A guided tour. 2005. URL <http://www.stat.sdu.dk/publications/preprints>.
- E. Girardi, F. Palmieri, A. Cingolani, A. Ammassari, N. Petrosillo, L. Gillini, D. Zinzi, A. D. Luca, A. Antinori, and G. Ippolito. Changing clinical presentation and survival in HIV-associated tuberculosis after highly active antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndrome*, 26:326 – 331, 2001.
- T. A. Gooley, W. Leisenring, J. Crowley, and B. E. Storer. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, 18:695–706, 1999.
- J. W. Graham. Missing data analysis: making it work in the real world. *Annual Review of Psychology*, 60:549 – 576, 2009.
- P. M. Grambsch and T. M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, August 1994.
- T. Gsponer, M. Petersen, M. Egger, S. Phiri, M. H. Maathuis, A. Boulle, P. Musondad, H. Tweya, K. Peter, B. H. Chi, and O. Keiser. The causal effect of switching to second-line ART in programmes without access to routine viral load monitoring. *AIDS*, 26:57–65, 2012.
- O. Harel and X.-H. Zhou. Multiple imputation: Review of theory, implementation and software. *Statistics in Medicine*, 26:3057–3077, 2007.
- F. E. Harrell Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag, 2001.
- D. V. Havlir, M. A. Kendall, P. Ive, J. Kumwenda, S. Swindells, S. S. Qasba, A. F. Luetkemeyer, E. Hogg, J. F. Rooney, X. Wu, M. C. Hosseinipour, U. Lalloo, V. G. Veloso, F. F. Some, N. Kumarasamy, N. Padayatchi, B. R. Santos, S. Reid, J. Hakim, L. Mohapi, P. Mugenyi, J. Sanchez,

- J. R. Lama, J. W. Pape, A. Sanchez, A. Asmelash, E. Moko, F. Sawe, J. Andersen, and I. S. for the AIDS Clinical Trials Group Study A5221. Timing of antiretroviral therapy for HIV-1 infection and tuberculosis. *The New England Journal of Medicine*, 365(16):1482–1491, October 2011.
- M. A. Hernán and J. M. Robins. *Causal Inference*. Chapman & Hall/CRC, February 2012. URL <http://www.hsph.harvard.edu/faculty/miguel-hernan/causal-inference-book/>. (Unpublished).
- M. A. Hernán, B. Brumback, and J. M. Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5):561–570, September 2000.
- M. A. Hernán, B. A. Brumback, and J. M. Robins. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96(454):440–448, June 2001.
- M. A. Hernán, S. Hernández-Díaz, M. M. Werler, and A. A. Mitchell. Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *American Journal of Epidemiology*, 155(2):175–184, 2002.
- M. A. Hernán, S. R. Cole, J. Margolick, M. Cohen, and J. M. Robins. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and drug safety*, 14:477–491, 2005.
- N. J. Horton and K. P. Kleinman. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician*, 61(1):79–90, February 2007.
- Institute of Social and Preventive Medicine, University of Bern. IeDEA Southern Africa, 2010. URL <http://www.iedea-sa.org/>.
- S. S. A. Karim, K. Naidoo, A. Grobler, N. Padayatchi, C. Baxter, A. Gray, T. Gengiah, G. Nair, S. Bamber, A. Singh, C. J. P. Munira Khan, M.B., W. El-Sadr, G. Friedland, and Q. A. Karim. Timing of initiation of antiretroviral drugs during tuberculosis therapy. *The New England Journal of Medicine*, 362(8):697–706, February 2010.
- S. S. A. Karim, K. Naidoo, A. Grobler, N. Padayatchi, C. Baxter, A. L. Gray, T. Gengiah, S. Gengiah, A. Naidoo, N. Jithoo, G. Nair, W. M. El-Sadr,

- G. Friedland, and Q. A. Karim. Integration of antiretroviral therapy with tuberculosis treatment. *The New England Journal of Medicine*, 365(16): 1492–1501, October 2011.
- B. Kerschberger, K. Hilderbrand, A. M. Boule, D. Coetzee, E. Goemaere, V. D. Azevedo, and G. V. Cutsem. The effect of complete integration of HIV and TB services on time to initiation of antiretroviral therapy: A before-after study. *PLoS One*, 7(10), October 2012.
- J. P. Klein and M. L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, 1997.
- S. D. Lawn and R. Wood. Optimum time to initiate antiretroviral therapy in patients with HIV-associated tuberculosis: There may be more than one right answer. *Journal of Acquired Immune Deficiency Syndrome*, 46(1): 121–123, September 2007.
- S. D. Lawn, L. Myer, C. Orrell, L.-G. Bekker, and R. Wood. Early mortality among adults accessing a community-based antiretroviral service in South Africa: implications for programme design. *AIDS*, 19:2141–2148, 2005.
- S. D. Lawn, L. Myer, L.-G. Bekker, and R. Wood. Burden of tuberculosis in an antiretroviral treatment programme in sub-Saharan Africa: impact on treatment outcomes and implications for tuberculosis control. *AIDS*, 20: 1605–1612, 2006.
- S. D. Lawn, L. Myer, L.-G. Bekker, and R. Wood. Tuberculosis-associated immune reconstitution disease: incidence, risk factors and impact in an antiretroviral treatment service in South Africa. *AIDS*, 21:335–341, 2007.
- S. D. Lawn, L. Campbell, R. Kaplan, A. Boule, M. Cornell, B. Kerschberger, C. Morrow, F. Little, M. Egger, and R. Wood. Time to initiation of antiretroviral therapy among patients with HIV-associated tuberculosis in Cape Town, South Africa. *Journal of Acquired Immune Deficiency Syndrome*, 57:136–140, 2011a.
- S. D. Lawn, L. Campbell, R. Kaplan, F. Little, C. Morrow, and R. Wood. Delays in starting antiretroviral therapy in patients with HIV-associated tuberculosis accessing non-integrated clinical services in a South African township. *BMC Infectious Diseases*, 11:258, 2011b.
- S. D. Lawn, A. Fraenzel, K. Kranzer, J. Caldwell, L.-G. Bekker, and R. Wood. Provider-initiated HIV testing increases access of patients with HIV-associated tuberculosis to antiretroviral treatment. *South African Medical Journal*, 101(4):258–262, April 2011c.

- S. D. Lawn, M. E. Török, and R. Wood. Optimum time to start antiretroviral therapy during HIV-associated opportunistic infections. *Current Opinion in Infectious Diseases*, 24:34–42, 2011d.
- R. Little and D. Rubin. *Statistical analysis with missing data*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley, 2002. ISBN 9780471183860.
- W. Manosuthi, S. Chottanapand, S. Thongyen, A. Chaovavanich, and S. Sungkanuparph. Survival rate and risk factors of mortality among HIV/tuberculosis-coinfected patients with and without antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndrome*, 43:42–46, 2006.
- M. Maruza, M. F. P. M. Albuquerque, M. C. Braga, M. T. S. Barbosa, R. Byington, I. Coimbra, L. V. Moura, J. D. L. Batista, G. T. N. Diniz, D. B. Miranda-Filho, H. R. Lacerda, L. C. Rodrigues, and R. A. A. Ximenes. Survival of HIV-infected patients after starting tuberculosis treatment: a prospective cohort study. *The International Journal of Tuberculosis and Lung Disease*, 16(5):618–624, March 2012.
- National Department of Health. *National Antiretroviral Treatment Guidelines*. Jacana, 2004.
- National Department of Health. *Clinical guidelines for the management of HIV & AIDS in adults and adolescents*. National Department of Health, South Africa, 2010.
- S. Philpott and U. Schüklenk. A study that should not have been done. 5 May 2010. URL <http://www.thehastingscenter.org/Bioethicsforum/Post.aspx?id=4626>.
- J. L. Prentice. A log gamma model and its maximum likelihood estimation. *Biometrika*, 61(3):539–544, 1974.
- H. Putter, M. Fiocco, and R. Geskus. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine*, 26:2389–2430, 2007.
- T. E. Raghunathan, J. M. Lepkowski, J. van Hoewyk, and P. Solenberger. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27(1):85–95, June 2001.
- M. Raviglione, A. Harries, R. Msika, D. Wilkinson, and P. Nunn. Tuberculosis and HIV: current status in Africa. *AIDS*, 11 (suppl B):S115–S123, 1997.

- J. M. Robins and M. A. Hernán. *Longitudinal Data Analysis, chapter 23: Estimation of the causal effects of time-varying exposures*. Handbooks of Modern Statistical Methods. Chapman & Hall/CRC, 2009.
- J. M. Robins, M. A. Hernán, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, September 2000.
- P. Royston. Multiple imputation of missing values. *The Stata Journal*, 4(3):227–241, 2004.
- P. Royston. Multiple imputation of missing values: update. *The Stata Journal*, 5(2):188–201, 2005.
- P. Royston. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. *The Stata Journal*, 9(3):466–477, 2009.
- P. Royston and I. R. White. Multiple Imputation by Chained Equations (MICE): Implementation in Stata. *Journal of Statistical Software*, 45(4), December 2011.
- N. Sanguanwongse, K. P. Cain, P. Suriya, S. Nateniyom, N. Yamada, W. Wattanaamornkiat, S. Sumnapan, W. Sattayawuthipong, S. Kaewsard, S. Ingkaseth, and J. K. Varma. Antiretroviral therapy for HIV-infected tuberculosis patients saves lives but needs to be used more frequently in Thailand. *Journal of Acquired Immune Deficiency Syndrome*, 48:181–189, 2008.
- J. L. Schafer. *Analysis of Incomplete Multivariate Data*. Chapman & Hall, 1997.
- StataCorp. *Stata Statistical Software: Release 2011*. College Station, TX: StataCorp LP, 2009a.
- StataCorp. *Stata 11 Base Reference Manual*. College Station, TX: Stata Press, 2009b.
- J. A. C. Sterne, M. A. Hernán, B. Ledergerber, K. Tilling, R. Weber, P. Sendi, M. Rickenbach, J. M. Robins, M. Egger, and the Swiss HIV Cohort Study. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *Lancet*, 366:378–384, 2005.

- S. Sungkanuparph, W. Manosuthi, S. Kiertiburanakul, and A. Vibhagool. Initiation of antiretroviral therapy in advanced AIDS with active tuberculosis: clinical experiences from Thailand. *Journal of Infection*, 52:188–194, 2006.
- S. Tansuphasawadikul, W. Saito, J. Kim, B. Phonrat, J. Dhitavat, S. Chamnanchanan, and P. Pitisuttithum. Outcomes in HIV-infected patients on antiretroviral therapy with tuberculosis. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 38(6):1053–1060, November 2007.
- A. Tsiatis and M. Davidian. Joint modeling of longitudinal data and time-to-event data: An overview. *Statistica Sinica*, 14:809–834, 2004.
- S. Van Buuren, H. Boshuizen, and D. L. Knook. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, 18:681–694, 1999.
- S. Van Buuren, J. P. L. Brand, C. G. M. Groothuis-Oudshoorn, and D. B. Rubin. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, 76(12):1049–1064, December 2006.
- H. C. Van Houwelingen. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, 34:70–85, 2007.
- M. Velasco, V. Castilla, J. Sanz, G. Gaspar, E. Condes, C. Barros, M. Cervero, R. Torres, C. Guijarro, and for the COMESEM Cohort. Effect of simultaneous use of highly active antiretroviral therapy on survival of HIV patients with tuberculosis. *Journal of Acquired Immune Deficiency Syndrome*, 50:148–152, 2009.
- C. Whalen, C. Horsburgh, D. Hom, C. Lahart, and M. S. andr J Ellne. Accelerated course of human immunodeficiency virus infection after tuberculosis. *American Journal of Respiratory and Critical Care Medicine*, 151:129–135, 1995.
- I. R. White and P. Royston. Imputing missing covariate values for the Cox model. *Statistics in Medicine*, 28:1982–1998, 2009.
- I. R. White, P. Royston, and A. M. Wood. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30:377–399, 2011.

- D. Wilson and G. Meintjes. Timing of antiretroviral drugs during tuberculosis therapy. *The New England Journal of Medicine*, 362(22):2137, June 2010.
- R. Wood, R. Kaplan, M. Egger, A. Boulle, and J. Sterne. Exploring when to start antiretrovirals in patients already on treatment for tuberculosis, December 2007. IeDEA TB Concept Sheet (Unpublished).
- L. Wu, W. Liu, G. y. Yi, and Y. Huang. Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*, 2012:1–17, 2011.