Addressing Selection Bias in Observational Event History Data, with Application to HIV Data from Western Kenya

by

Ann Wanjiru Mwangi

B.Sc., Jomo Kenyatta University of Agriculture and Technology, Kenya, 2001
M.Sc. Applied Statistics, Limburgs Universitair Centrum, Belgium, 2003
M.Sc. Biostatistics, Limburgs Universitair Centrum, Belgium, 2004

A Dissertation Submitted in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy
in the Division of Biology and Medicine at Brown University

Providence, Rhode Island
May 2011

© Copyright

by

Ann Wanjiru Mwangi

2011

This dissertation by Ann Wanjiru Mwangi is accepted in its present form by
the Division of Biology and Medicine as satisfying the
dissertation requirement for the degree of
Doctor of Philosophy

Date ....
Joseph W. Hogan, Advisor

Recommended to the Graduate Council

Date ....
Tao Liu, Reader

Date ....
Crystal Linkletter, Reader

Date ....
Rami Kantor, Reader

Date ....
E. Jane Carter, Reader

Approved by the Graduate Council

Date ....
Peter Weber, Dean of the Graduate School

The Vita of Ann Wanjiru Mwangi

Ann Mwangi was born on 31st July 1977 in Nyeri, Kenya. She attended Kiamuiru Primary school, Kangubiri Girls High school. She attained her Bachelor of Science degree in Statistics and Computing from Jomo Kenyatta University in 2001 with First class honours. She holds two masters degree from Limburgs Universitair, Belgium one in Applied Statistics and the other in Biostatistics. Before coming to Brown in 2007 she worked with USAID-AMPATH as a Biostatistician.

Acknowledgments

I wish to express my sincere gratitude to my mentor and advisor, Professor Joseph W. Hogan, for his excellent guidance, support and constant encouragement during my time here at Brown. Without his advice and persistent help and his believing in me, this thesis would not be possible. I also would like to express my gratitude to my committee members, Professors Tao Liu, Crystal Linkletter, E. Jane Carter and Rami Kantor, for their helpful discussions and comments to improve the quality of my research. I am also thankful to Dr. Abraham Siika from Moi University, AMPATH for his guidance in the research content and Prof.Constantin Yiannoutsos from Indiana University.

Many thanks to my friends and fellow graduate students for numerous discussions. I would like to thank Allison Delong and Eileen Wright from Fogarty for her administrative help which made my student life so much easier.

Special thanks to my husband Sum and daughter Chebet for their unconditional love, endurance and patience when I spent time away from them. Thanks to my family for their support and unconditional love.

This work was sponsored by Brown-Fogarty AITRP, and the data is from USAID-AMPATH program. I gratefully acknowledged Fogarty and AMPATH for their support.

Dedication

To my husband, Sum Kipyego and daughter Marie Chebet

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Observational data

Randomized studies are considered to have the highest degree of evidence in the hierarchy of research designs for drawing inferences about exposure effects. However, these trials are often expensive, can raise ethical concerns and have limited follow up and sample size. Several sources of observational data are being compiled on HIV treatment, outcomes, and modes of care, such as AMPATH Medical Records System (AMRS) ((Tierney et al., 2007), International epidemiologic Databases to Evaluate AIDS (IeDEA), HIV Epidemiology Research Study (HERS) (Smith et al., 1997) and Multicenter AIDS Cohort Study (MACS) (Mellors et al., 1997). Data from these observational studies, might play a major role for some clinical decisions and monitoring and evaluation (M&E) of programs. However, appropriate methods are required in the use of these datasets. This project will address two major complications in the use of large observational studies both for treatment decision and M&E:

1. Non-random allocation to treatment and exposure. In observational studies the patients' risk profiles prior to start of treatment might not be comparable between treatment groups. There

may also be unmeasurable clinical and social factors in the pathway between diagnosis and treatment since we are not able to observe all the factors clinicians take into account in their decision to treat.

2. High loss to follow-up (LTFU) rates in studies where mortality is the primary outcome. Frequently for both research and M&E, mortality is a primary outcome.

1.2 Data source: Description of AMPATH

The AMPATH program was initiated in 2001 as a joint partnership between Moi University School of Medicine, Moi Teaching and Referral Hospital (MTRH) and Indiana University School of Medicine to fight HIV. In later years several medical schools from North America have joined the consortium including Brown University. The initial goal of the consortium was to establish a HIV care program to serve both rural and urban patients and assess outcomes of antiretroviral therapy as well as barriers to treatment. The program started with two HIV clinics in Western Kenya: one urban (MTRH) and one rural (Mosoriot Rural Health Centre). More details on the development of the program have been described by Einterz et al. (2007). Since 2001 the program has expanded and now includes 18 HIV/AIDS clinics funded by a $60 million, 5-year grant from USAID. AMPATH is the largest HIV/AIDS comprehensive care program in Kenya and one of the largest in all of sub-Saharan Africa. To date, the program has enrolled more than 75,000 HIV infected adults and children, 35,000 of whom are eligible for and are receiving cART. All HIV and TB related care and treatment are provided free through AMPATH and Kenya National Leprosy, Tuberculosis and Lung Disease Program.

The HIV clinical care protocol (Wools-Kaloustian et al., 2006) used in AMPATH are consistent with the World Health Organization (WHO) guidelines for care of HIV infected patients (WHO,

2006a) and Kenya Ministry of Health's National AIDs and STI Control Program (NASCOP) guidelines. Patients who are in AMPATH are seen by a clinical officer or a physician. Patients who are not on cART are seen every 1-3 months depending on their clinical status. Patients on cART are seen 2 weeks after initiation of cART, and then monthly thereafter. WHO has guidelines for initiation of cART in HIV/TB coinfected adults (WHO, 2006a). AMPATH developed their own guideline in consultation with NASCOP based on observed clinical outcomes of HIV/TB coinfected patients they were seeing in the care program. Both protocols are shown in Tables 2.1.

AMPATH has an outreach program that traces patients who fail to adhere to their clinic visits through telephone and home visit. For patients newly initiated on cART the tracing start on the 2nd day of their defaulted scheduled appointments. However, for those already on cART and stable this might take 2-4 weeks. The main goals of outreach are to determine the patients vital status and encourage return to the treatment program. The program is described in detail by Wools-Kaloustian et al. (2006) and Einterz et al. (2007).

Clinical visit information is recorded in the AMPATH Medical Records System (AMRS), a computerized database designed for clinical data management; (Tierney et al., 2007). Data emerging from this program can be used to guide in assessing barriers and outcomes of cART and other prophylaxis treatments offered to HIV as well as TB patients. This can also be used for program evaluation and to guide future care and provision of treatment for other infectious diseases.

1.3 Summary of projects

The aims of this project are:

1. *Apply methods for marginal structural models to investigate the causal effect of timing of combination antiretroviral treatment (cART)on mortality among adult HIV patients co-infected with tuberculosis (TB)*: HIV treatment is complicated by TB due to high pill burden, drug

to drug interaction and immune reconstitution syndrome (IRIS). The decision on when to initiate cART in this population is crucial both for survival and for slowing AIDS progression. Evaluation of treatment outcomes for different initiation times can be based on observational data adjusting for non-random allocation to treatment and LTFU using marginal structural models applying the inverse probability weighting (IPW) technique. We also adjust for missing expsoure due to drop out and mortality using regression imputation.

2. *Assess the causal effect of cART timing on the rate of occurence of opportunistic infections (OIs)*: The rate of occurence of OIs in HIV infected patients is a key indicator of HIV progression. We aim to assess whether the rate of occurence of OIs differ depending on when they are initiated. One complication with recurrent event is termination due to death and LTFU. To adjust for termination due to death we make use of IPW while for death we consider partially conditional model where we condition on surviving.

3. *Estimate survival distribution from double-sampled data by modeling the informative censoring process* : Censoring is a common feature in both clinical trials and observational studies. Different assumptions can be made and analysis performed to evaluate the effect of treatment on a survival outcome. Most of these assumptions cannot be verified using observed data and hence sensitivity analysis is key. When historical data or expert opinion is available on the relationship between censoring and failure, this can be used to formulate models that allow adjustment for censoring in these analyses. We propose to incorporate information from a double sampling study of drop outs for estimation of mortality and regression models.

Chapter 2

Causal Inference for Survival Times with Informative Censoring & Missing Exposure, with Application to Treatment of TB/HIV Coinfection in Western Kenya

Abstract

Objectives. The optimal timing for initiation of combination anti-retroviral treatment (cART) is a major concern in the care of patients who are co-infected with HIV and TB. Data from randomized trials are still limited. We use observational data from a cohort of HIV/TB co-infected patients to quantify the causal effect of cART initiation time on one-year mortality. The data come from the USAID-AMPATH medical records system (AMRS), which has data on >100,000 individuals in

HIV care in Western Kenya. Between March 2004 and April 2008 a total of 8,810 non-pregnant HIV-infected adults initiated TB treatment and subsequently initiated cART at varying times relative to TB treatment initiation

Methods. Complications associated with observational data include nonrandom allocation to treatments and informative drop out. Our cohort had drop out leading to censoring of exposure, outcome or both. We categorized the patients into 5 groups based on timing of cART initiation following anti-tuberculosis treatment initiation: (0,2], (2,8], (8,16], (16,35] weeks and, $(35,\infty)$. We use a marginal structural proportional hazards model and applied inverse probability weighting to control for nonrandom allocation of cART initiation time and informative dropout. Confounders used to make the adjustments included baseline CD4, weight, age, presence of AIDS defining illness, WHO stage, gender, marital status, post-primary education and rural vs. urban clinic attendance. For those with censored cART initiation time, we imputed timing of treatment based on the conditional probability of initiating cART at the remaining intervals.

Results. The final analysis included a total of 4908 HIV/TB co-infected adults enrolled in the program between March 2004 and April 2008. Our model finds that rate of mortality increases with delay in initiation of cART. The hazard of dying in the first year among those that initiated cART in the intervals (2,8] , (8,16],(16,35],$(35,\infty)$ compared to those initiating in the interval (0,2] were 1.12, 1.42, 1.76, and 2.02, respectively. The marginal model gives dramatically different results compared to standard regression that does not account for confounding. Our model finds that hazard for death increases with delay in initiation of cART. Our method allows 1060 individuals with missing treatment time to be included.

Conclusions. Early initiation of cART leads to better survival among HIV/TB co-infected patients compared to delayed initiation.

Keywords. HIV/TB; Inverse Probability weighting; Informative censoring; Cox proportional hazards model.

2.1 Introduction

The rapid growth of data from electronic medical records (EMR) systems is opening new windows of opportunity to study the effects of medications, interventions, and clinical practices on health outcomes. Methods for drawing causal inference from observational data, including sampling, model formulation, and inferential techniques necessarily play a central role in in analysis of EMR data.

An important outcome for gauging effectiveness of clinical practice, especially in the area of chronic disease, is mortality. Characterizing the causal effect of an exposure on hazard of death – or any event time for that matter – requires methodology that is specifically tailored to event history data.

This paper is concerned with drawing causal inferences about the effect of an exposure on mortality when the exposure itself is an event time. Specifically, we are interested to estimate the causal effect of the timing of HIV therapy initiation for those who are coinfected with HIV and tuberculosis (TB) and are initiating TB treatment. Because of potentially dangerous interactions between HIV and TB medications, the timing of HIV therapy initiation is a major concern, especially in the developing world where HIV/TB coinfection is relatively common.

We use data on over 8,000 individuals with HIV/TB coinfection drawn from the EMR system of the Academic Model Providing Access to Healthcare (AMPATH), a PEPFAR-funded program delivering care to over 100,000 HIV-infected individuals in western Kenya.

The methods we propose here must accommodate several complexities associated with longitudinal observational data: (i) the timing of HIV therapy initiation is not randomized, and indeed is strongly associated with HIV disease severity at the time of TB diagnosis; (ii) mortality outcomes are subject to possibly informative right censoring; and (iii) the exposure time also is right-censored, in some cases by the outcome itself. Our strategy is to use marginal structural models estimated by inverse probability weighting. The innovation in our approach is the handling of censored or missing exposure times, which has received little attention in the extant statistical literature. We show how censored exposure times can be handled by carefully defining casual effects and by conducting appropriate sensitivity analyses. We also show the importance of including the partial information from censored exposure times by comparing our analyses to an analysis that omits these data. Our analysis suggests that early initiation of HIV therapy is highly beneficial.

2.2 HIV/TB coinfection

2.2.1 Overview

Globally, it is estimated that 33.4 million people are infected with HIV, the greatest number of whom reside in lower income countries, with 67% in sub-Saharan Africa alone (UNAIDS/WHO, 2009). HIV is the strongest risk factor for developing tuberculosis (TB) infection. Not only does HIV increase the risk of developing TB, it also increases the risk of rapid TB progression soon after TB infection or re-infection. In resource-poor countries, TB is responsible for 12 to 50% of all AIDS related deaths (WHO, 2009). In a study carried out in Eldoret, Kenya using admissions data in early antiretroviral era, TB accounted for 27% of the AIDS related deaths (Siika et al., 2008).

Combination antiretroviral treatment (cART) has proven to be effective for the treatment HIV, and has been shown to reduce TB recurrence by 50% (Golub et al., 2008). In patients who are HIV/TB coinfected, administration of HIV treatment is complicated by the presence of TB because of high pill burden, drug toxicity and drug interactions between cART and anti-TB treatment (Burman, 2005). Another complication is immune reconstitution inflammatory syndrome (IRIS), a spectrum of clinical symptoms resulting from the restored ability to mount an inflammatory response associated with immune recovery (Robertson et al., 2006).

TB treatment has two main phases: Phase I lasts for two months; patients are treated with isoniazid (INH), rifampicin and ethambutol. Phase II lasts for 6 months; patients are treated with INH and ethambutol. For patients who are HIV/TB coinfected, the World Health Organization (WHO) recommends initiation of TB treatment immediately and initiation of cART after patients have stabilized on TB treatment (WHO, 2006a).

In developing countries, HIV treatment decisions are often based on both clinical (WHO staging) and immunological markers (CD4 count). The WHO system classifies HIV disease on the basis of clinical manifestations that can be recognized and treated by clinicians in diverse settings, including resource-constrained settings, and by clinicians with varying levels of HIV expertise and training. Patients in the 4-point WHO scale stage 3 or 4, or with a CD4 count below 250, are eligible for cART at initiation of TB treatment; furthermore, because TB is one of the defining diseases for WHO Stage 3, cART is indicated.

At least two randomized trials comparing cART initiation time are either ongoing or completed in developing countries. The SAPIT trial (Starting Antiretroviral Therapy at 3 Points in TB) was carried out in Durban, South Africa between June, 2005 and July, 2008 and included 642 patients who were HIV positive, cART naive and smear positive for TB. Interim analysis showed that mortality rates were significantly higher among patients who initiated antiretroviral therapy after completion of TB treatment (11.6 per 100 person-year) compared to patients who started during TB treatment (5.1 per 100 person-year). Results from this trial suggest that cART should be initiated during TB treatment (Baleta, 2008; Karim et al., 2009).

An ongoing trial is the Adult AIDS Clinical Trials Group (AACTG) network study A5221. Patients with TB/HIV having CD4 count <200 are randomized into two arms: cART initiated within 2 weeks of TB treatment and cART initiated within 8 to 12 weeks of TB treatment. This multi-site trial enrolling 800 individuals is currently taking place in both developed and developing countries (Blanc et al., 2007). Results from an observational study of 1217 HIV/TB co-infected individuals carried out in Spain showed improved survival rate among patients who initiated cART and TB treatment simultaneously compared to those that did not initiate both treatments simultaneously (Velasco et al., 2009).

2.2.2 Data source: AMPATH Medical Records System

The AMPATH program was initiated in 2001 as a joint partnership between Moi University School of Medicine, Moi Teaching and Referral Hospital (MTRH) and Indiana University School of Medicine (Einterz et al., 2007). AMPATH now includes 28 HIV/AIDS clinics, is the largest HIV/AIDS comprehensive care program in Kenya, and is one of the largest in all of sub-Saharan Africa. All HIV and TB related care and treatment are provided free through AMPATH and the Kenya National Leprosy, Tuberculosis and Lung Disease Program. Clinical visit information is recorded in the AMPATH Medical Records System (AMRS), an electronic medical records database (Tierney et al., 2007). Data from the AMRS has been used to monitor and evaluate a number of intervention and treatment programs in the fight against HIV in western Kenya (Nyandiko et al., 2006; Braitstein et al., 2009, 2010).

The PEPFAR-funded AMPATH Program provides care to well over 100,000 HIV-infected in-

dividuals in western Kenya. It has developed its own guideline in consultation with the Kenyan National AIDS and STI Control Program (NASCOP); the guidelines are based on observed clinical outcomes of TB/HIV coinfected patients in the AMPATH care program. The AMPATH protocol is based primarily on the CD4 count at start of TB treatment (Table 2.1).

To investigate causal effects of cART timing on mortality, we use data from the AMRS on a cohort of TB/HIV infected patients receiving care in AMPATH. Between March 1, 2004 and April 18, 2008, 8,810 non-pregnant, HIV+ patients aged 18 years and older were receiving treatment for TB in AMPATH, and had initiated cART at varying times relative to initiation of TB treatment. The dataset provides individual level information on demographic characteristics such as age, marital status, education level, gender and the clinic site (rural/ urban). In addition, clinical covariates such as weight, CD4 count, WHO staging, and presence of AIDS defining illness (ADE) were recorded at initiation of TB treatment and at subsequent follow-up visits.

Clinicians in AMPATH did not strictly adhere to TB/HIV protocol, leading to heterogeneity in the initiation of cART (Figure 4.4). This feature is helpful in identifying causal effects. However, there are high rates of loss to follow up (40%), censoring both the outcome and the time of cART initiation.

2.3 Framing the causal question

Our goal is to estimate the causal effect of an exposure A on an outcome T where both A and T are event times. The exposure A is time from TB treatment initiation to cART initiation, and the endpoint T is time from TB treatment initiation to death. In a randomized controlled trial, one might randomize individuals to mutually exclusive time intervals for cART initiation, and compare mortality rates between the groups. Our approach is to define the full data in these terms.

Suppose the time axis is partitioned into $Z + 1$ mutually exclusive intervals, each defining a period of time where cART could be initiated. Then each individual is assumed to have potential death times $T_0, \ldots, T_Z$, where T_z is the death time corresponding to cART initiation in interval z. This formulation suggests modeling causal effects with a structural model for event times. We use a structural proportional hazards model, although the methods described below are applicable to a

wide variety of model choices. Our model takes the general form

$$\lambda^{T_z}(t) \quad = \quad \lambda^{T_0}(t)\exp(\beta_z), , z = 1, ..., Z \tag{2.1}$$

where $\lambda^{T_0}(t)$ is the hazard corresponding to being treated in interval $z = 0$ and β_z is the causal log hazard ratio that measures hazard rate for interval z relative to interval $z = 0$. The model can be elaborated to include baseline covariates and stratification variables, described in more detail in Sections 2.4.2 and 2.6.

Even when A and T are fully observed, inference about causal effects from observational data must take into account potential confounding that arises from nonrandom allocation of treatment. In the AMPATH data, timing of cART initiation depends on a patient's clinical condition at the time of TB diagnosis, which is related to mortality rate. Methods that can be used for model fitting include propensity score risk adjustment or matching (Rosenbaum & Rubin, 1984), inverse probability weighting (Robins et al., 2000; Hernan et al., 2000), and methods based on instrumental variables (Hogan & Lancaster, 2004; Bijwaard & Ridder, 2005).

The structural model (2.1) shares similarities with others that have been used to estimate causal effects from event time data, but has some key differences. For example, Hernan et al. (2000) use a structural proportional hazards model formulated in terms of treatment histories $\overline{a}_t = \{a_s : 0 < s < t\}$, where treatment status is as a 0-1 counting process $\{a_t : t > 0\}$ that takes value 0 until treatment is initiated, at which time it takes value 1. The general form of the model in Hernan et al. (2000) is

$$\lambda_{\overline{a}_t}(t) \quad = \quad \lambda_0(t)\exp\{\beta g(\overline{a}_t)\},$$

where $\lambda_{\overline{a}_t}(t)$ is the hazard of death under treatment history $\overline{a}_t$ and g is a known function that maps treatment history to a scalar or finite-dimensional vector. The version used by Hernan et al. (2000) sets $g(\overline{a}_t) = a_t$, so that β represents the causal log hazard ratio measuring the effect of receiving treatment at t relative to not receiving it. The coefficient β captures only the effect of receiving treatment, not its timing. Potential confounding by time-varying factors that influence both mortality and receipt of treatment can be controlled by fitting the model using inverse probability weights for

treatment status at a given time t (see also Hernan et al., 2001).

Our model is also distinct from those concerned with evaluating dynamic treatment regimes. A dynamic regime depends on evolving information such as measured biomarkers or observable symptoms; the decision is made when one or more thresholds is crossed. Structural models can be used to compare dynamic regimes by applying inverse probability weighting in an appropriate manner. For example, Hernan et al. (2006) uses MSM to compare the following two dynamic regimes for asymptomatic treatment-naive individuals enrolled into HIV care: (i) start cART the first time the CD4 count drops below 500, versus (ii) start cART the first time the CD4 count drops below 200. Individuals who are not observed to follow either of these regimes are artificially censored; IPW is used to adjust selection bias introduced by the artificial censoring. Robins et al. (2008) proposed generalizations of the method to determine the optimal CD4 value at which to initiate treatment. Cain et al. (2010) derived methods for using observational data to compare regimes having the form 'initiate treatment within a certain time period of a time varying covariate crossing a threshold'. The AMPATH data are not particularly well suited to evaluating dynamic regimes due to the infrequency of CD4 monitoring during the short duration of TB treatment; however this is a possible extension of our approach. Specifically we could extend our method to determine optimal *timing*, rather than optimal CD4 count. The methods of Robins et al. (2008) would have to be extended to accommodate censoring of the initiation times.

Our paper frames the comparison of cART initiation times in terms of a static regime, where the timing of cART would be determined when TB therapy is started. Aside from the nonrandom allocation to cART initiation time, we must address various types of right censoring that lead to incompletely observed data on some individuals. Figure 4.1 shows four configurations of censoring. In the ideal case we would observe cART initiation at time A, followed by death at time T (Case I). However, it is possible that cART time is observed but death time is censored (Case II); death occurs before initiation of cART (thereby censoring A, Case III); or, censoring prevents observation of both A and T (Case IV).

Three distinct issues must be addressed: (i) nonrandom allocation to treatment initiation time interval, which is handled using inverse probability of treatment weighting (IPTW); (ii) informative censoring death times due to dropout, which is handled using inverse probability of censoring weights (IPCW); and (iii) informative censoring of treatment initiation time, which leads to only

partial information about which time interval cART would have been initiated. To address this third issue, we use the model for treatment initiation to compute the probability associated with initiating cART in the remaining potential intervals. We also outline a strategy for conducting sensitivity analysis to assumptions required to handle censoring of the treatment times.

The primary innovation of our method is handling missing exposure, which has received scant attention in published literature on this topic. Shortreed & Forbes (2010) compared different methods for handling missing data in the exposure of interest when using marginal structural models to estimate the causal effect of cumulative exposure history on an outcome. They considered different missing data scenarios: missing completely at random, missing at random and missing not at random. They compared different methods of inference: multiple imputation, last value carried forward analysis, available case analysis as well as complete case analysis. In a specific simulation study, they observed that complete case analysis provided the least bias. In the Section 6, we compare a complete-case analysis to our method for handling missing exposure and find substantial differences in the conclusions about treatment effect.

2.4 Structural models: Specification and inference

2.4.1 Notation

For an event time variable T, we use the following conventions: $N^T(t)$ is the associated 0-1 counting process, $S^T(t) = P(T > t)$ is the survival function, $\lambda^T(t)$ is the hazard function, and $\Lambda^T(t)$ is the cumulative hazard. Survival and hazard functions can be conditioned on covariates V, denoted (for example) by $\lambda^T(t|V)$. We are interested in drawing inference about treatment effects over a fixed period $(0, \tau]$, where τ is set in advance.

2.4.2 Model specification

Following the structure of current randomized trials of HIV/TB coinfection, we treat the timing of cART as a non-dynamic regime in the sense that it is decided at TB treatment initiation. The following time intervals, measured in weeks from start of TB treatment, are used to characterize Z non-overlapping treatment initiation categories wherein cART is started in that interval and contin-

ued thereafter. The intervals are denoted by

$$(l_0, u_0], (l_1, u_1], \ldots, (l_Z, u_Z],$$

where $l_0 = 0$ and $u_Z = \infty$. Let T_z denote death time when treatment is started in interval $(l_z, u_z]$, which we refer to hereafter as interval z. The causal model therefore defines a set $\{T_0, \ldots, T_Z\}$ of potential death times for each individual, one corresponding to each cART initiation interval. To capture the causal effect of cART timing on mortality, we use the marginal structural proportional hazards model (2.1) described above,

$$\lambda^{T_z}(t|V) \quad = \quad \lambda^{T_0}(t) \ \exp(\beta_z),$$

where $\lambda^{T_z}(t)$ is the hazard function associated with T_z, $\lambda^{T_0}(t)$ is the hazard function for $z = 0$, β_z is the causal log hazard ratio for interval z relative to interval 0. We also allow for stratified baseline hazard function, illustrated in our analysis in Section 2.6. Model can be elaborated to include baseline covariates. For purposes of clarity, from this point forward, we describe inference for the model without covariates V; so long as these baseline covariates, the needed modifications are straightforward.

2.4.3 Parameter estimation when treatment initiation time is always observed

2.4.3.0.1 Case 0: All potential outcomes observed, no censoring. It is instructive to consider this hypothetical case first in order to build the estimating equations associated with the more complex setting where only one of the potential outcomes is observed and event times can be censored. Suppose $T_0, \ldots, T_Z$ were actually observed for each individual. Let x_z denote the $Z \times 1$ vector that indicates treatment group, such that $x_0 = (0, 0, \ldots, 0)^{\mathrm{T}}$, $x_1 = (1, 0, \ldots, 0)^{\mathrm{T}}$, $x_2 = (0, 1, 0, \ldots, 0)^{\mathrm{T}}$, and so forth. Furthermore, let $R_z(t) = I(T_z \geq t)$ denote the at-risk indicator under regimen z. Then a consistent estimator $\widehat{\beta}$ of β would be the root of the estimating

function

$$U_0(\beta) \;=\; \sum_{i=1}^{n}\sum_{z=0}^{Z}\int_0^{\tau}\{x_z - \overline{x}(t;\beta)\}\, dN_i^{T_z}(t), \tag{2.2}$$

where

$$\overline{x}(t;\beta) \;=\; \frac{\sum_i \sum_z x_z R_{zi}(t)\exp(x_z^{\mathrm{T}}\beta)}{\sum_i \sum_z R_{zi}(t)\exp(x_z^{\mathrm{T}}\beta)}. \tag{2.3}$$

2.4.3.0.2 Case 1: Randomized treatment, no censoring of exposure time or outcome. In reality, only one of the potential death times can be observed. Consider the scenario where treatment initiation time is always observed and there is no censoring of T. Denote observed initiation time by A_i. Define a vector $\Delta_i = (\Delta_{0i}, \Delta_{1i}, \ldots, \Delta_{Zi})^{\mathrm{T}}$ having elements $\Delta_{zi} = I\{A_i \in (l_z, u_z]\}$. Observed death time is therefore $T_i = \sum_z \Delta_{zi} T_{zi}$, and the corresponding treatment covariate vector is $X_i = \sum_z \Delta_{zi} x_{zi}$. If treatment initiation time were randomly assigned (i.e., no confounding), then the root $\widehat{\beta}$ of the estimating function

$$\begin{aligned} U_1(\beta) \;&=\; \sum_{i=1}^{n}\sum_{z=0}^{Z}\int_0^{\tau}\Delta_{zi}\,\{x_z - \overline{x}(t;\beta)\}\, dN_i^{T_z}(t) \\ &=\; \sum_{i=1}^{n}\int_0^{\tau}\left\{X_i - \overline{X}(t;\beta)\right\} dN_i^{T}(t) \end{aligned}$$

is consistent for the causal effect β. Here,

$$\begin{aligned} \overline{X}(t;\beta) \;&=\; \frac{\sum_i X_i R_i(t)\exp(X_i^{\mathrm{T}}\beta)}{\sum_i R_i(t)\exp(X_i^{\mathrm{T}}\beta)} \\ &=\; \sum_i X_i W_{1i}(t;\beta) \end{aligned}$$

is a weighted average of the (observed) X_i with weights

$$W_{1i}(t;\beta) \;=\; \frac{R_i(t)\exp(X_i^{\mathrm{T}}\beta)}{\sum_i R_i(t)\exp(X_i^{\mathrm{T}}\beta)}.$$

In other words, if treatment is randomized and there is no censoring, the casual effect β is the solution to the standard Cox proportional hazards partial likelihood equations (Cox, 1975).

2.4.3.0.3 Case 2: Nonrandomized treatment, censored event times, uncensored exposure times. We now consider nonrandom allocation to treatment initiation time and potential censoring of T. Censoring of A introduces additional complexities and is addressed in Section 2.5; for now we continue to assume information on observed treatment time A is fully available. We also assume there is an vector-valued covariate process $\{V(t) : t > 0\}$, observed for each individual at a finite set of time points during follow up; $V(t)$ may contain both time-varying and time-stationary covariates.

Let $C > 0$ denote a right-censoring time, and let $T^* = \min(T, C)$ denote observed follow up time for mortality. Our inference procedures for Case 2 rely on four assumptions:

(i) *Among those at risk at any time t, censoring is non-informative in the sense that the hazard of censoring at t is independent of failure time conditionally on a subset $V_1(t)$ of $V(t)$;*

(ii) *the model $\lambda^C\{t \mid V_1(t); \gamma\}$ for hazard of censoring is known up to a finite-dimensional parameter γ;*

(iii) *conditionally on a subset $V_2(t)$ of baseline covariates $V(t)$, the rate (hazard) of treatment initiation at t is independent of the potential outcomes $T_0, \ldots, T_Z$;* and

(iv) *the model $\lambda^A\{t \mid V_2(t), \eta\}$ is known up to a finite-dimensional parameter η.*

To streamline notation, denote individual-specific hazard functions for censoring and treatment initiation as $\lambda_i^C(t; \gamma) = \lambda\{t \mid V_{1i}(t); \gamma\}$ and $\lambda_i^A(t; \eta) = \lambda\{t \mid V_{2i}(t); \eta\}$; corresponding survival and cumulative hazard functions are denoted similarly.

The probability of cART initiation in interval z is

$$S_i^A(l_z; \eta) - S_i^A(u_z; \eta) \quad = \quad \exp\{-\Lambda_i^A(l_z; \eta)\} - \exp\{-\Lambda_i^A(u_z; \eta)\}. \tag{2.4}$$

Under assumptions (i)–(iv), and with (γ, η) fixed at their true values (γ^*, η^*), the root $\widehat{\beta}$ of $U_2(\beta)$

consistently estimates β from the observed data, where

$$U_2(\beta) \;=\; \sum_{i=1}^{n}\sum_{z=0}^{Z}\int_0^{\tau}\left[\frac{\Delta_{zi}}{S^A(l_z;\eta^*)-S^A(u_z;\eta^*)}\cdot\frac{\{X_i-\sum_i X_i W_{2i}(t;\beta,\gamma^*)\}}{S_i^C(t;\gamma^*)}\right]dN_i^T(t) \tag{2.5}$$

and where the weight function $W_{2i}(t;\beta,\gamma)$ is

$$W_{2i}(t;\beta,\gamma) \;=\; \frac{R_i(t)\exp(X_i^{\mathrm{T}}\beta)\;/\;S_i^C(t;\gamma)}{\sum_i R_i(t)\exp(X_i^{\mathrm{T}}\beta)\;/\;S_i^C(t;\gamma)}.$$

The estimator $\widehat{\beta}$ remains consistent when (γ^*,η^*) are replaced by consistent estimators $(\widehat{\gamma},\widehat{\eta})$, both of which can be obtained by suitably specified models for $\lambda^C(t;\gamma)$ and $\lambda^A(t;\eta)$ (Robins, 1999b).

2.4.4 Estimating treatment and censoring weights

In Assumption (iv) we assume time to treatment initiation follows the proportional hazards model

$$\lambda^A\{t \mid V_2(t)\} \;=\; \lambda_0^A(t)\exp\{V_2^{\mathrm{T}}(t)\eta\},$$

where $\lambda_0^A(t)$ is the (possibly stratified) baseline hazard. Let $\widehat{\lambda}_0^A(t)$ denote the Nelson estimator and $\widehat{\eta}$ the maximum partial likelihood estimator of the baseline hazard and regression parameter, respectively, obtained by fitting the Cox proportional hazards model to the treatment initiation times.

Conditionally on covariates $V_{2i}(\tau)$, the probability of initiating treatment in the interval $(l_z,u_z]$ is given by (2.4). From Assumptions (iii) and (iv), the hazard functions $\Lambda_i^A(t;\eta)$ can be consistently estimated by

$$\widehat{\Lambda}_i^A(t;\widehat{\eta}) \;=\; \int_0^t \widehat{\lambda}_0^A(s)\exp\{V_{2i}^{\mathrm{T}}(s)\widehat{\eta}\}\;d\overline{N}^A(s), \tag{2.6}$$

where $\overline{N}^A(s)=\sum_i N_i^A(s)$ In our application we stratified the baseline hazard for treatment initiation time on CD4 count category at TB treatment initiation (0–50, 50–200, 200–350).

Stabilized weights generally yield confidence intervals that not only are narrower but also have actual coverages rates that are closer to their nominal values (Hernan et al., 2000). To stabilize the

treatment weights we use numerator weights estimated from a stratified Kaplan Meier estimator $\widehat{P}_k^A(t)$ for time to treatment initiation, where $k = 1, 2, 3$ indexes baseline CD4 stratum (see above). Weights corresponding to probability of initiating cART in interval $(l_z, u_z]$ for an individual in CD4 stratum k are calculated as $\widehat{P}_k^A(l_z) - \widehat{P}_k^A(u_z)$.

A similar approach is used to obtain estimated censoring probabilities $\widehat{S}_i^C(t) = S^C\{t \mid V_{1i}(t); \widehat{\gamma}\}$. We assume the hazard of dropout follows $\lambda^C(t \mid V_1(t); \gamma) = \lambda_0^C(t) \exp\{V_1^{\mathrm{T}}(t)\gamma\}$; hence $\widehat{S}_i^C(t) = \exp\{-\widehat{\Lambda}_i^C(t)\}$, and is computed as in (2.6). As with the treatment weights, we stabilize the censoring weights by including a numerator derived from a stratified Kaplan Meier estimator $\widehat{P}_k^C(t)$ applied to the censoring times (again stratifying on baseline CD4 count). Note that evaluation of (2.6) and hence (2.5) requires estimating $S_i^C(t)$ for every individual at every event time.

2.5 Handling censored exposure times

In the AMPATH data, some individuals either die or are lost to follow up before initiating cART. We assume that, prior to the beginning of follow up, each of these individuals could have been assigned to any of the five treatment groups; hence, we use information about the length of follow up time without receiving cART to allocate the indicator variable Δ_{zi} across the remaining potential categories of treatment initiation time.

2.5.0.0.4 Case 3: Censored exposure times. For those who die or are lost to follow up prior to initiating cART, the initiation time is censored and A is missing. Let $A^* = \min(A, C, T)$, which is the actual cART initiation time when $A^* = A$ and is a minimum observed time without receiving cART when $A^* < A$ (i.e., when it is censored by death or loss to follow up).

Our approach to incorporating partial information on cART initiation time for those with censored values of A is to replace $\Delta_{zi} = I\{A_i \in (l_z, u_z]\}$ in the estimating function (2.5) with (an estimate of) its expected value,

$$\begin{aligned} \Delta_{zi}^* &= E\{\Delta_z \mid A > A_i^*, V_i(\tau)\} \\ &= \Pr\{A \in (l_z, u_z] \mid A > A_i^*, V_i(t)\}, \end{aligned}$$

for $z = 0, 1, \ldots, Z$. This distributes the indicator variables such that $\sum_z \Delta^*_{zi} = 1$; because we are conditioning on $A > A^*_i$, intervals where $u_z < A^*_i$ have $\Delta^*_{zi} = 0$.

It is important to realize that Δ^*_{zi} is not identifiable from the observed data without making additional assumptions. If we assume that censoring of A at time t is non-informative conditional on $V(t)$, then Δ^*_{ai} can be computed using the fitted hazard model (2.6) for cART initiation that is used to construct the inverse probability weights. Specifically, under the assumption

(v) *Among those at risk at t and who have $A > t$, the hazard of either death or censoring at t is independent of A conditionally on covariates $V_2(t)$,*

we have

$$\begin{aligned}
\widehat{\Delta}^*_{zi} &= \frac{\widehat{\Pr}\{A \in (l_z, u_z],\ A > A^*_i \mid V_i(\tau)\}}{\widehat{\Pr}\{A > A^*_i \mid V_i(\tau)\}} \\
&= \begin{cases} 0 & \text{if } A^*_i > u_z \\ \dfrac{\widehat{S}^A\{A^*_i \mid V_i(A^*_i)\} - \widehat{S}^A\{u_z \mid V_i(A^*_i)\}}{\widehat{S}^A\{A^*_i \mid V_i(A^*_i)\}} & \text{if } l_z \le A^*_i \le u_z \\ \dfrac{\widehat{S}^A\{l_z \mid V_i(A^*_i)\} - \widehat{S}^A\{u_z \mid V_i(A^*_i)\}}{\widehat{S}^A\{A^*_i \mid V_i(A^*_i)\}} & \text{if } A^*_i < l_z. \end{cases}
\end{aligned}$$

Other assumptions are of course possible. At one extreme, we can set $\widehat{\Delta}^*_{zi} = I\{A^*_i \in (l_z, u_z]\}$, which assumes treatment would have been initiated with probability one during the interval where censoring or death took place; the opposite extreme is to assign all probability mass to the interval $(l_Z, u_Z]$, which means cART would not have been initiated during the period of TB treatment. It is also possible to allocate probabilities uniformly, for example by assigning $\widehat{\Delta}^*_{zi} = 1/m_i$, where $m_i = \sum_z I(A^*_i > l_z)$ is the number of intervals within which cART could still be initiated; mass can also be assigned in relative proportion to interval length. More formally, a sensitivity parameter can be introduced to control the relative allocation of $\widehat{\Delta}^*_{zi}$ over the remaining intervals.

Let $\widetilde{\Delta}_{zi} = \Delta_{zi} I(A_i = A^*_i) + \widehat{\Delta}^*_{zi} I(A_i > A^*_i)$, which uses the observed interval indicator when A_i is observed and the estimated one when A_i is censored. With this modification, and replacing

functionals with their estimates, the estimating function (2.5) becomes

$$U_3(\beta) = \sum_{i=1}^{n}\sum_{z=0}^{Z}\int_0^\tau \frac{\widehat{P}_{k_i}^A(l_z) - \widehat{P}_{k_i}^A(u_z)}{\widehat{S}_i^A(l_a) - \widehat{S}_i^A(u_a)} \cdot \frac{\widehat{P}_{k_i}^C(t)}{\widehat{S}_i^C(t)}\ \widetilde{\Delta}_{zi}\ \{x_z - \textstyle\sum_i x_z W_{3i}(t;\beta,\widehat{\gamma})\}\ dN_i^T(t), \quad (2.7)$$

where

$$W_{3i}(t;\beta,\gamma) \;=\; \frac{\sum_z \widetilde{\Delta}_{zi} R_i(t)\exp(x_z^{\mathrm{T}}\beta)\{\widehat{P}_{k_i}^C \;/\; S_i^C(t;\gamma)\}}{\sum_i\sum_z \widetilde{\Delta}_{zi} R_i(t)\exp(x_z^{\mathrm{T}}\beta)\{\widehat{P}_{k_i}^C \;/\; S_i^C(t;\gamma)\}}.$$

2.6 Analysis of AMPATH Data

2.6.1 Data summaries

Of the 8810 HIV positive adult patients who initiated TB treatment between March 1, 2004 and April 18, 2008, we excluded 3039 who had missing data on one or more key baseline covariates and 863 who had a CD4 count greater than 350 (the latter based on the criterion used for cART initiation in resource constrained setting). Our analysis dataset therefore contains 4908 patients on TB treatment initiating cART at different times in the course of their TB treatment. Of these, 3506 had observed cART initiation times during TB treatment period (8 months), and 342 were observed not to start cART during TB treatment. The remaining 1060 patients were either administratively censored, lost to follow up, or had died prior to initiation of cART; all therefore have right censored treatment initiation times. Of the 1060 patients with missing exposure, 184 (17%) were confirmed to have died before initiating cART, median time to death 48 days, indicating that censoring due to death tends to happen early in the TB treatment period.

Because CD4 count is a key marker used in deciding when to initiate cART, we categorized the CD4 count at initiation of TB treatment into the following categories: 0–50, 50–200, and 200–350. The weight model for time to initiation of cART used these categories to stratify the baseline hazard function because the crude hazards in the 3 groups were not found to be proportional. Figure 4.4 shows the Kaplan Meier plot of time to cART initiation in weeks by CD4 count categories. Clearly, those with low CD4 count upon starting TB treatment initiated cART earlier compared to those with high CD4 count.

Table 4.1 shows the distribution across categories of cART initiation time within CD4 count category. Despite the AMPATH protocol, there was considerable variability in start times, which is likely attributable to physicians exercising clinical judgment or using unrecorded patient information.

2.6.2 Summary of approaches being compared

We generate inferences about the effect of cART initiation time by fitting versions of model (2.1) using four different approaches. The first three exclude those with censored cART initiation times. We define the cART initiation time intervals as $(0, 2]$, $(2, 8]$, $(8, 16]$, $(16, 35]$, $(35, \infty)$, and label them as $z = 0, 1, 2, 3, 4$. We fit the first three models as follows: (i) fit an unweighted model to the death times using $U_1(\beta)$; (ii) fit an unweighted proportional hazards regression, but include baseline covariates V; and (iii) fit a weighted model using $U_2(\beta)$, where parameters η and γ in the weight functions are replaced with their maximum partial likelihood estimators. Our fourth set of inferences — model (iv) — includes information on the 1060 individuals with censored exposure times; the model is fit using $U_3(\beta)$, given by (2.7).

Stabilized weights for models (iii) and (iv) are calculated as described in Section 2.4.4, where $V_1 = V_2 = V$ and V includes the following baseline covariates measured at initiation of TB treatment: CD4 count (category used to stratify baseline hazard; actual value included in linear predictor), clinic site (urban/rural), gender (male/female), presence of AIDS defining illness other than TB (yes/no), weight in kg, age in years, WHO staging (three categories as follows: stage 1& 2, stage 3 & 4, and missing), attainment of post primary education (yes/no), and marital status (yes/no). In fitting the model for hazard of loss to follow up (C), we treated mortality as a 'censoring' events in the sense that it can right-censor C. For the exposure time model, we treated both mortality and loss to follow up as censoring events for A.

We also carried out estimation for models (iii) and (iv) by incorporating time-varying covariates into the weight models; in particular, we used most recent measures of CD4 count, weight, and presence of opportunistic infection (yes/no) to fit the hazard model for cART initiation time (2.6). The inclusion of time-varying covariates did not affect the treatment effect estimates to any measurable degree, but added considerably to the time required for model fitting. Hence we present the results

for the model using baseline covariates only.

In all cases, bootstrap re-sampling with 500 replicates was used for confidence interval estimation, to properly account for uncertainty in estimation of both the weights and, for model (iv), the $\widetilde{\Delta}_{zi}$. We report 95% confidence intervals based on the lower and upper 2.5 percentiles from the ordered 500 bootstrap estimates. The stabilized weights ranged from .00219 to 41.0, with 25th, 50th and 75th percentiles .244, .544, and .836 respectively.

2.6.3 Summary of results

Results from the fitted models are summarized in Table 4.3. The first three columns show the results for the inferences that exclude the 1060 individuals with censored cART initiation times. The unadjusted model reflects a strong effect of confounding by indication, in that patients with the highest eventual mortality rates are more likely to initiate cART earlier in the course of TB therapy. Adjusting for confounders either through regression (column 2) or via an MSM fit to the subset with observed cART initiation times attenuates the confounding effect to some degree.

The last column in Table 4.3 shows the results for the MSM fitted to the full sample, *including* those with censored cART initiation time. Compared to the first three models, the treatment effect dramatically reverses direction, indicating that early cART initiation reduces mortality. The confidence intervals suggest that mortality rates associated with initiating cART after 16 weeks are significantly higher compared to starting in the first two weeks.

The substantial discrepancy in results can be at least partly attributed to the high mortality rate among those with censored treatment initiation times. Those whose treatment initiation time is censored by death are assigned an initiation time beyond the observed death time, which has the effect of attributing those deaths to later initiation times. This is consistent with using the IPW strategy in a way that treats initiation time as something that is assigned at baseline, before any deaths take place. Put another way, in a trial where cART initiation times are randomly allocated at baseline, deaths can occur before the treatment is initiated; in the treatment comparison, those deaths would be attributed to the assigned treatment initiation time.

To further illustrate the importance of incorporating information on censored cART initiation times, consider the comparison presented in Figure 2.3. The top panel shows unweighted Kaplan-

Meier estimates applied to the dataset that excludes those with censored initiation times, and the bottom panel shows survival curves computed using weighted Nelson-Aalen estimators.

When those with censored cART initiation times are excluded, all observed deaths occur after observed initiation times, leading to estimated survival probability of 1.0 prior to first observed initiation time within each group. Hence any analysis of a dataset that excludes censored treatment times — even an analysis that uses IPW or some other adjustment — will yield ordered mortality rates that favor later initiation times.

2.7 Discussion

Several sources of observational data are being compiled on HIV treatment, outcomes, and modes of care, in the developing world; two prominent examples are the AMPATH Medical Records System (AMRS) (Tierney et al., 2007) and the International epidemiologic Databases to Evaluate AIDS (IeDEA) (Forster et al., 2008). Through use of appropriate methods, data from these observational databases can play a major role in the evaluation of clinical practices and programs, and can form the basis for evidence-based recommendations for treatment and care.

Electronic medical records (EMR) data are becoming very important because of large sample sizes and mirroring of clinical practice, but can be very complex due to observational nature and multiple types of attrition. In observational studies, unlike with clinical trials, treatment allocation is not random and might be a associated with factors that are associated with the outcome of interest. In this paper, we use marginal structural models to adjust for potential selection bias due to censoring and non random allocation to treatment under some assumptions. We develop a principled method to include individuals with partially observed exposure, in this case a censored time to cART initiation. Inclusion of these individuals makes a dramatic difference in the conclusions of the analysis, bringing the treatment effect ordering more into line with results from recent clinical trials. Our findings are consistent with the available results from the two clinical trials referenced in Section 2.2.1; both are based on smaller sample sizes, but both suggest that cART should be initiated earlier in the course of TB treatment.

The results from the MSM relies on assumption of no unmeasured confounding. This assumption might not hold if there are some other factors that clinicians take into account in making the

decision to initiate cART in this population. We propose in a follow up paper to carry out a sensitivity analysis to unmeasured confounders. For the missing exposure we described how to compare different scenarios as a form of sensitivity analysis. As a follow up, we plan to develop a more formal sensitivity analysis that distributes mass between the two extremes using a sensitivity parameter that indexes an appropriately constructed multinomial distribution. Another assumption made for this analysis is that a patient was considered to have TB if they were on TB treatment. We did not have data on sputum results or chest x-rays for everyone because many are too ill to provide a specimen. We also did not have data on the type of TB — whether extrapulmonary or miliary — a key indicator of TB severity and potential response to cART.

Our work can be extended in several possible directions. Currently we are modifying these methods for settings where the primary outcome is a recurrent event such as occurrence of HIV-related opportunistic infection. Another important direction concerns optimizing the timing of cART, which was discussed in Section 2.3. Finally, missing CD4 count at baseline limited the sample we used for analysis; incorporating an appropriate adjustment would allow an expanded sample size for further investigation.

Table 2.1: AMPATH guidelines for initiation of cART in TB/HIV coinfected adults.

Criteria at TB treatment initiation time	**Timing of cART initiation**
CD4 count $<$50 and/or presence of opportunistic infection (OI)	After 2 weeks of TB treatment
CD4 count 50–200	After 1 month of TB treatment
CD4 count $>$ 200 or clinically stable	After 2 months of TB treatment

Table 2.2: Cross tabulation of *observed* cART initiation time by baseline CD4 category. Percentages are row percents.

Baseline CD4 Count	Time to cART in weeks						
	[0-2]	(2-8]	(8-16]	(16-35]	(35,∞]	Censored	Total
$\leq$50	475	448	242	66	26	285	1542
	(31)	(29)	(16)	(4)	(2)	(19)	
50–200	463	605	551	225	98	460	2402
	(19)	(25)	(23)	(9)	(4)	(19)	
200–350	96	91	124	120	218	315	964
	(10)	(9)	(13)	(12)	(23)	(33)	
Total	1034	1144	917	411	342	1060	4908
	(22)	(23)	(19)	(8)	(7)	(22)	

Table 2.3: Covariate distribution at baseline, stratified by time to cART initiation. For binary variables in the top part of the table, percentages are listed in parentheses.

	Observed time to cART in weeks					
	(0-2]	(2-8]	(8-16]	(16-35]	(35,∞)	Censored
	$N = 1034$	$N = 1144$	$N = 917$	$N = 411$	$N = 342$	$N = 1060$
Male	551 (53)	583 (51)	452 (49)	182 (44)	168 (49)	499 (47)
Married	610 (59)	637 (56)	506 (55)	228 (56)	181 (53)	543 (51)
Post primary education	530 (51)	455 (40)	360 (39)	162 (39)	129 (38)	327 (31)
Urban clinic	611 (59)	598 (52)	473 (52)	224 (55)	215 (63)	555 (52)
AIDS defining Event	126 (12)	102 (9)	67 (7)	14 (3)	12 (3)	97 (9)
WHO Stage 1 or 2	28 (3)	19 (2)	45 (5)	18 (4)	17 (5)	28 (3)
WHO Stage 3 or 4	697 (67)	669 (58)	518 (56)	230 (56)	187 (55)	554 (52)
Missing WHO Stage	309 (30)	456 (40)	354 (39)	163 (40)	138 (40)	478 (45)
Confirmed Deaths	112 (11)	110 (10)	80 (9)	23 (6)	34 (10)	184 (17)
LTFU	279 (27)	270 (24)	180 (20)	69 (17)	54 (16)	581 (55)
CD4 count at baseline						
Median	58	71	100	139	228	116.5
Range	(0, 350)	(0, 350)	(1, 349)	(4, 341)	(1, 350)	(0, 350)
Weight in kg at baseline						
Median	52	51	51	52	53	50
Range	(29, 99)	(29, 101)	(26, 104)	(29, 92)	(32, 98)	(24, 87)

Table 2.4: Hazard ratios and 95% bootstrap confidence intervals for the effect of cART initiation time on mortality rate. First three models exclude those with right-censored cART initiation times. Fourth model (MSM-2) includes them.

	Excluding censored cART times			
cART initiation time	Unadjusted	Regression-adjusted	MSM-1	MSM-2
[0-2]	1	1	1	1
(2-8]	$_{0.53}0.88_{1.44}$	$_{0.52}0.89_{1.48}$	$_{0.50}0.89_{1.56}$	$_{0.67}1.12_{1.85}$
(8-16]	$_{0.36}0.65_{1.10}$	$_{0.39}0.71_{1.16}$	$_{0.38}0.74_{1.29}$	$_{0.86}1.42_{2.26}$
(16,35]	$_{0.18}0.43_{0.88}$	$_{0.21}0.53_{1.15}$	$_{0.21}0.56_{1.39}$	$_{1.07}1.76_{3.01}$
(35,∞)	$_{0.08}0.28_{0.72}$	$_{0.11}0.39_{1.03}$	$_{0.12}0.69_{3.18}$	$_{1.07}2.02_{4.49}$

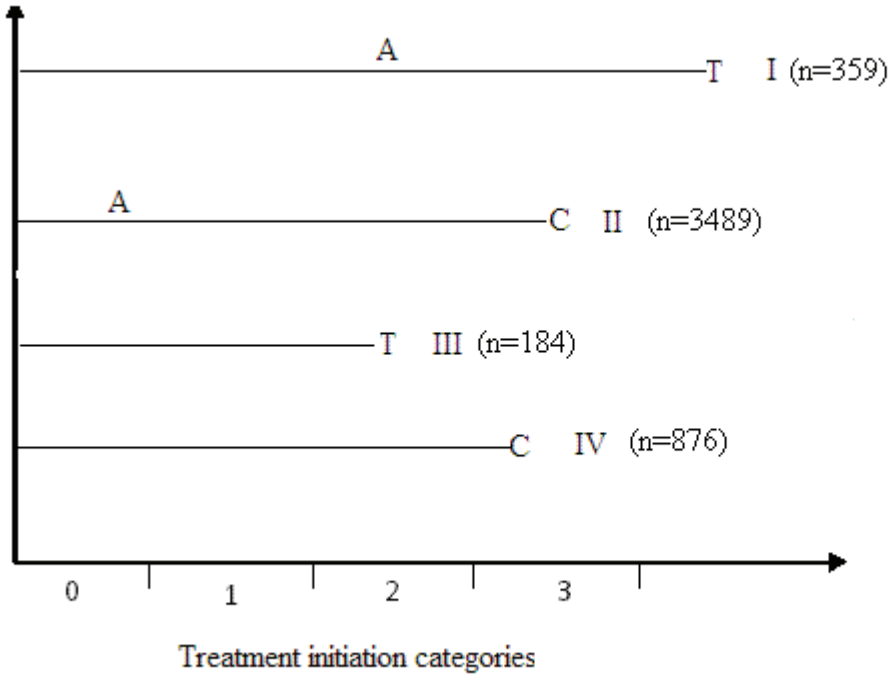


Figure 2.1: Patterns of observed information on cART initiation time A and death time T. Either or both may be right-censored at time C.

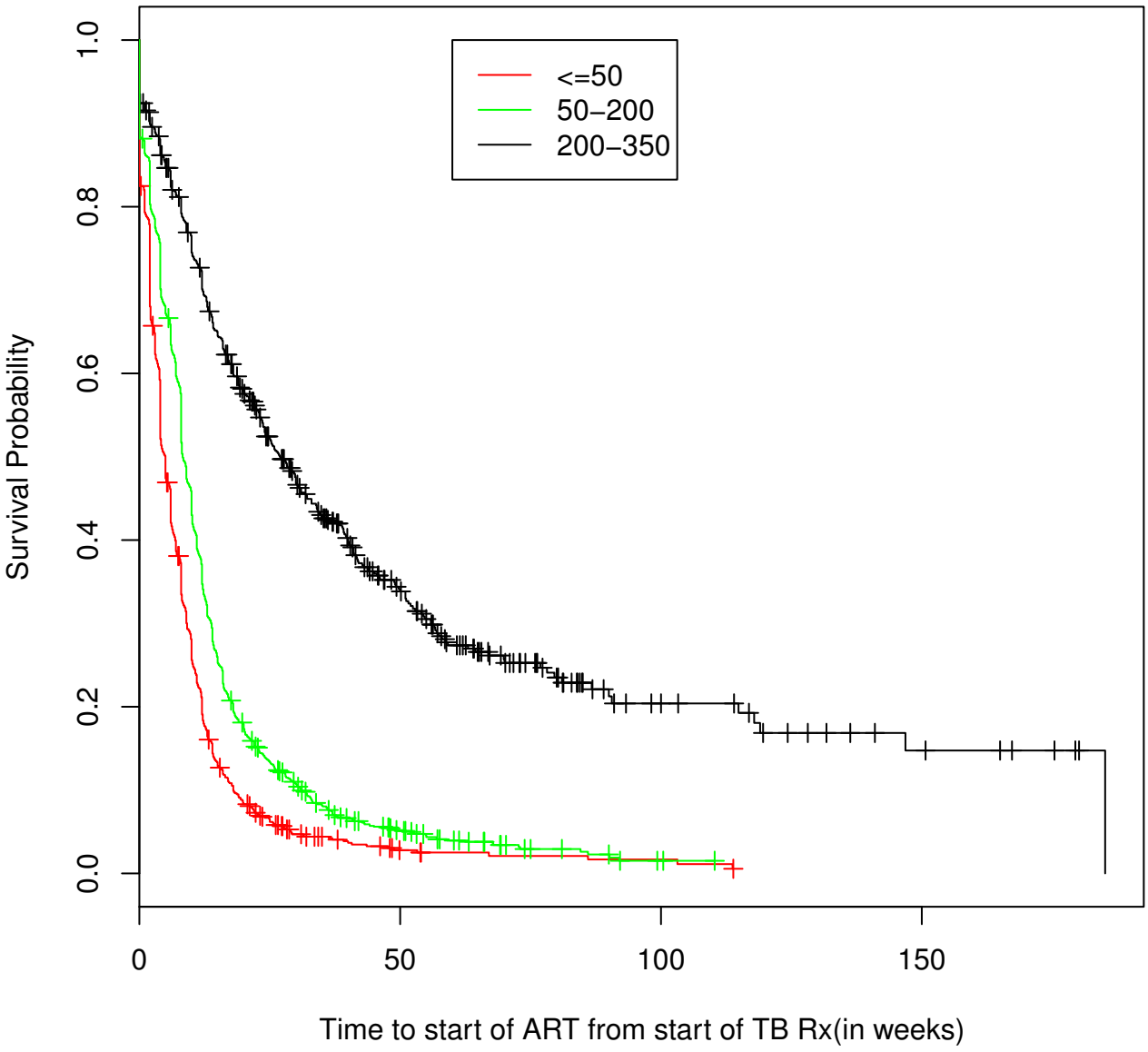


Figure 2.2: Kaplan Meier estimates of time to cART initiation, stratified by baseline CD4 count category.

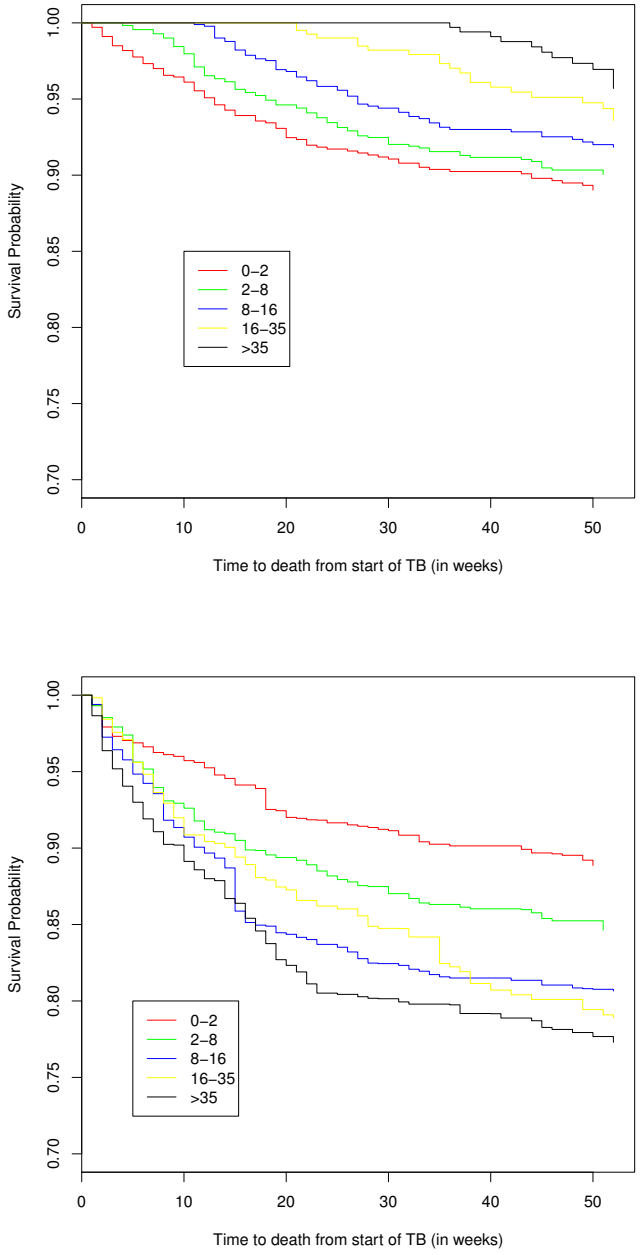


Figure 2.3: Estimated survival curves, excluding (top panel) and including (bottom panel) those with censored exposure, stratified by cART initiation time. When those with censored cART initiation times are excluded, all observed deaths occur after observed initiation times, leading to estimated survival rates of 1.0 prior to first observed initiation time within each group.

Chapter 3

Causal inference for recurrent events in the presence of terminating events, loss to follow up, and censored exposure

3.1 Introduction

3.1.1 HIV/TB Coinfection

Human immunodeficiency virus/acquired immune deficiency syndrome (HIV/AIDS) and tuberculosis (TB) are overlapping epidemics that cause an immense burden of disease in sub-Saharan Africa. To overcome the burden, the World Health Organization (WHO) recommends prevention of TB in people living with HIV and treatment of HIV in patients with TB (WHO, 2007). HIV infected patients are also vulnerable to infections and malignancies that are called opportunistic infections (OI) other than TB due to weakened immune system. The most common HIV-related OIs and diseases include bacterial pneumonia, toxoplasmosis, microsporidiosis, cryptosporidiosis, PCP, candidia-

sis, cryptococcosis, cytomegalovirus, herpes simplex and herpes zoster virus, Kaposis sarcoma and lymphoma, among others (CDC, 2009). Prevention and treatment of opportunistic infections not only helps HIV-positive people to live longer, healthier lives, but can also help prevent TB and other transmissible opportunistic infections from spreading to others. The WHO recommends the use of co-trimoxazole as a prophylaxis to prevent OIs in HIV-positive adults and adolescents who have mild, advanced or severe symptoms of HIV disease, or a CD4 count below 350 cells per ml, in resource-limited settings (WHO, 2006b). For patients who are HIV/TB co-infected, WHO recommends treatment for TB and opportunistic infections (OIs) to start immediately (Harries et al., 2009).

In the era of combination antiretroviral treatment (cART) there has been a significant improvement in the outcome of HIV/AIDS, in terms of prevention of OIs as well as mortality. Results from clinical trials (Baleta, 2008; Karim et al., 2009) and observational studies (Velasco et al., 2009; Mwangi et al., 2010) have shown that early initiation of cART leads to better survival among HIV-TB coinfected adults. The relative hazard of developing presenting and secondary OIs has also declined dramatically (Palella et al., 1998). However, this benefit might not be realized in patients with TB due to complications in the use of cART mainly due to drug interactions between cART and TB treatment.

In a previous paper (Mwangi et al., 2010), we observed that early initiation of cART reduces the risk of death among HIV/TB co-infected adults using observational data from Academic Model Providing Access to Health Care (AMPATH). The aim of this study is to assess whether early cART also supresses other OIs among HIV/TB infected since the main goal is to have patients living longer and healthier lives.

3.1.2 Recurrent events

In many health related studies involving longitudinal follow up of subjects, the event of interest may recur on the same subject. Multiple opportunistic infections in HIV infected patients, tumor recurrence in cancer patients, seizure recurrence in study of epileptic seizures, and repeated attacks of asthma are examples. There are various approaches for analysis recurrent events. Analyses

can be based on time to first episode of an opportunistic infection; the main shortcoming of this type of analysis is the fact that we fail to make use of the information on opportunistic infection occurring after the first one. Another approach could be based on number of opportunistic infections observed during a fixed period; however, this type of analysis does not address the issue of censoring and death. Analysis could also be based on the number of events per person years, these are also problematic because they do not distinguish between death, loss to follow up and end of study. Thus methods are needed that address possible association between recurrent event process and survival process. Recurrent event data can be regarded as a specific type of correlated data with the feature that event times of a subject are ordered and correlated and could be terminated by end of study, death or loss to follow up.

In this paper we consider models for recurrent event in the presence of termination due to death and drop out using observational data from electronic medical records of patients being treated by AMPATH program. The main aim is to draw causal comparison of rates of occurrence of OI's associated with cART initiation times in patients with HIV/TB.

We are interested in comparing the occurrence of opportunistic infections among patients initiated on cART at different times in the course of TB treatment with termination due to death and drop out. As stated by Cook et al. (2009), analysis based on expected counts are appealing when comparing treatment effects. The complications in the data analysis include presence of death that leads to termination of the event of interest, nonrandom allocation to treatment, and informative censoring and drop out leading to missing exposure.

We estimate event rates directly, and make use of structural models for drawing causal comparisons between treatment groups defined by their cART initiation time. The structural models are estimated assuming that the rates follow a Poisson process that may or may not be time homogeneous. To handle censoring of the event process and termination due to death, we follow methods established by Cook et al. (2009). We then extend the methods to handle non-random allocation to treatment group and missingness in the treatment assignment. Estimation is handled using inverse probability treatment weighting for nonrandom allocation to treatment, and an imputation-based method for missing treatment.

3.1.3 Description of the data

We use data from the AMPATH program which is the largest HIV/AIDS comprehensive care program in Kenya and one of the largest in all of sub-Saharan Africa (Einterz et al., 2007). To date, the program has enrolled more than 100,000 HIV infected adults and children, 40,000 of whom are receiving cART. All HIV and TB related care and treatment are provided free through AMPATH and the Kenya National Leprosy, Tuberculosis and Lung Disease Program.

In AMPATH HIV infected patients are started on cotrimaxole prophylaxis to prevent OI if they have a CD4 count less than 200. HIV/TB coinfected patients are started on TB treatment immediately they are diagnosed to have TB. TB diagnosis is mainly based on chest X ray, TB history, TB exposure and sputum test where possible. TB treatment has two main phases Phase I lasts for two months; patients are treated with Isoniazid (INH), Rifampicin and Ethambutol. Phase II lasts for 6 months; patients are treated with INH and Ethambutol. Patients on TB treatment are seen every 2 weeks during phase I and monthly during phase II.

At AMPATH patients initiating cART are seen after 2 weeks and monthly thereafter. At every clinic visit, presence of OI is assessed based on whether the patients has any of the following illnesses: PCP, Lymphoma, Kaposis Sarcoma, HIV Encephalopathy, Atypical Mycobateriosis, Toxoplasmosis, PML, Cryptococcosis, Cryptosporidiosis, CMV, Mycosis or Salmonella. Adherance to medication and weight are also monitored at every visit. CD4 count which is an immunological marker of disease progression is monitored every 6 months. Clinical visit information is recorded in the AMPATH Medical Records System (AMRS), an electronic medical records database designed for clinical data management (Tierney et al., 2007). Data from the AMRS has been used to monitor and evaluate a number of intervention and treatment programs in the fight against HIV in western Kenya. The dataset provides individual level information on demographic characteristics such as age, marital status, education level, gender and the clinic site (rural/ urban). In addition clinical covariates such as weight, CD4 count, WHO staging, and presence of AIDS defining illness are recorded at initiation of TB treatment and at subsequent follow-up visits.

Between March 1, 2004 and April 18, 2008, 8,810 non-pregnant, HIV+ patients aged 18 years and older were receiving treatment for TB in AMPATH, and initiated cART at varying times relative to initiation of TB treatment. Table(4.1) shows the distribution of OIs by cART timing over the

course of one year from initiation of TB treatment.

3.2 Statistical methods for recurrent events

Here we give a brief accounting of models and methods for recurrent data, with focus on estimating the rate of events as a function of covariates. We then discuss modifications of standard approaches that accommodate censoring by a terminal event such as death. Finally, we give a precise description of the causal quantity we are interested to estimate.

3.2.1 Estimation of rate functions

For an event of interest, let $\{N(t) : t > 0\}$ denote a counting process such that $N(t)$ denotes the number of times the event has occurred since time $t = 0$. Following Cook & Lawless (2007), we assume $N(t)$ is a Poisson process, and that the number of events in non-overlapping time intervals are statistically independent. Let $H(t)$ denote the history of $N(t)$. The *rate function* is

$$\rho(t|H(t)) = \lim_{\Delta t \to 0} \frac{P\{\Delta N(t) = 1 \mid H(t)\}}{\Delta t},$$

where $\Delta N(t) = N(t + \Delta t^-) - N(t^-)$ is the number of events occurring in the interval $[t, t + \Delta t)$, and where it is assumed that only one event can occur at a fixed time t. The rate is defined conditionally on event history; hence, standard likelihood- and moment-based estimation can be implemented even in the presence of censoring, provided that the censoring mechanism only depends on the event process through its history $H(t)$ (Cook et al., 2009).

The *marginal mean*, denoted by $\mu(t) = E\{N(t)\}$, is related to the rate process via

$$\mu(t) = E\{N(t)\} = \int_0^t \rho(s)ds.$$

This implies $d\mu(t) = \rho(t)dt$. The process is said to be time homogeneous when $\rho(t) = \rho$.

In the absence of censoring, the mean function can be consistently using nonparametric moment estimators. However, when censoring is present, consistent estimation requires the censoring process to be completely independent of the recurrent event process, a condition that is unlikely to hold in practice. Here we consider right censoring. Let C denote a right censoring time, and let $Y^c(t) = I(C > t)$ be an at-risk indicator. In the presence of censoring (but absence of a terminal event such as death), the mean functions can be estimated nonparametrically as the solution to the estimation equation

$$\sum_{i=1}^{n} Y_i^c(t)\{dN_i(t) - d\mu(t)\} = 0, \tag{3.1}$$

which yields

$$d\widehat{\mu}(t) = \frac{\sum_i Y_i^c(t) dN_i(t)}{\sum_i Y_i^c(t)},$$

and therefore $\widehat{\mu}(t) = \int_0^t d\widehat{\mu}(t)$. Importantly, unbiasedness of the estimating equation (3.1) relies on a independence between $Y^c(t)$ and $dN(t)$. This assumption can be relaxed using an inverse probability of censoring weighted (IPCW) estimator that solves the estimating equation

$$\sum_{i=1}^{n} \frac{Y_i^c(t)}{S_i^c(t)}\{dN_i(t) - d\mu(t)\} = 0, \tag{3.2}$$

where $S_i^c(t) = P\{C > t \mid H_i(t), V_i(t)\}$ is the probability of remaining uncensored at t, $H_i(t)$ is the history of the event process as before, and $V_i(t)$ is a possibly vector-valued set of auxiliary covariates. In practice, $S_i^c(t)$ must be estimated from data, but the weighted estimator remains consistent so long as the weights themselves are consistently estimated (Robins & Finkelstein, 2000; Cook et al., 2009); in practice, this can be accomplished by fitting an appropriate model for $S_i^c(t)$, and using estimated censoring probabilities $\widehat{S}_i^c(t)$ in (3.2). To be more precise, consistency of the IPCW estimator requires the following two assumptions:

(A) *Censoring at t is conditionally independent of $N(t)$, given $H(t)$ and $V(t)$;* and

(B) $P\{C > t \mid H_i(t), V_i(t)\}$ *is bounded away from 0 for all i.*

3.2.2 Regression models

Effects of covariates on recurrent event processes can be captured via regression modeling. To estimate treatment effects, Cook et al. (2009) suggest the use of the mean function because intensity based models that condition on $H(t)$ make the interpretation of treatment effects difficult; in particular treatment effect can be confounded with prior event history.

In the presence of a binary covariate Z, Cook & Lawless (2007) considered intensities of the form

$$\rho(t|Z) = \rho_0(t)\exp(Z\beta)$$

where ρ_0 is the baseline rate or intensity corresponding to an individual for whom $Z = 0$. This multiplicative model is also referred to as log-linear model and is often used to model count data. If the baseline rate is an arbitrary positive-valued function then the model is sometimes called Andersen-Gill (AG) model (Andersen & Gill, 1982). The mean function corresponding to the AG model is

$$\begin{aligned} \mu(t|Z) &= E\{N(t)|Z\} \\ &= \mu_0(t)\exp(Z\beta) \end{aligned} \tag{3.3}$$

where $\mu_0(t)$ is the baseline mean function. This model relies on the assumption that the mean and rate functions for any two individuals are proportional.

The AG model is a generalization of the Cox proportional hazards model, and relates the intensity function of the recurring event to the covariates multiplicatively. The AG model assumes that the recurring event follows a non-homogenous Poisson process, i.e the risk of a recurring event follows the usual proportional hazards assumption and is unaffected by earlier occurrences. This assumption can be relaxed by introducing time dependent covariates in the model such as the number of prior recurrences. Parameters for the AG model can be estimated using partial likelihood. Cook & Lawless (2007) summarize several other models that have been proposed for dealing with recurrent events; these include parameterizations in terms of time between events (Prentice et al., 1981) and methods for individual-level recurrence (Wei et al., 1989).

The models described above are concerned with estimating covariate effects on the rate, av-

eraged over some time period. It is also possible to use Poisson regression and other generalized linear models to characterize the effect of covariates on rates over a fixed interval. For example, if we are concerned with event rates on the interval $(0, \tau]$, we can use the Poisson regression

$$\begin{aligned} \log E\{N(\tau|Z)\} &= \log \mu_0(\tau) + \beta Z \\ &= \alpha + \beta Z \end{aligned} \tag{3.4}$$

Note that $\log \mu_0(\tau) = \alpha$ is a constant here. Compared to (3.3), this model ignores the functional form of change in the mean over the interval $(0, \tau]$, but may be appropriate when the focus is on inference about number of events at a fixed time point.

3.2.3 Inference in the presence of a terminating event

Several analysis techniques have been proposed when dealing with longitudinal data subject to a terminating event. Our focus in this paper is on settings where the terminating event is death. Kurland et al. (2009) provide a comprehensive overview, and divide modeling approaches into the following categories: *unconditional*, *fully conditional*, *partly conditional*, and *joint distribution of the outcome and death time*. The choice of model must be properly aligned with the inferential objective.

In terms of drawing inference about recurrent events, we can match objectives with models as follows. Let U denote death time. The *unconditional model* captures $E\{N(t)\}$, or the mean number of events had deaths not occurred. The *fully conditional model* describes individual response trajectories conditional on fixing death time; specifically, it captures $E\{N(t) \mid U = s\}$ for some $s < t$. The *partly conditional model* describes the mean number of events at time t conditional on being alive at time t; i.e., $E\{N(t) \mid U > t\}$. The *joint model* is the most general, and is designed to capture the evolving health status of the entire cohort by modeling both the recurrent events and the survival process as a single joint distribution. In general it is necessary to make some assumptions about the joint distribution of $\{N(t), U\}$ to fit any of the models described above. The assumptions used for our application are listed in Section 3.3.

Many existing methods in the literature implicitly fit an unconditional model. For example, Li & Lagakos (1997) proposed a modification of Wei et al. (1989) by treating death as a censoring variable for the recurrent event. Ghosh & Lin (2002) extended this method by focusing on the marginal mean of the cumulative number of recurrent events over time. They then incorporated the fact that once a subject dies they cannot experience further events. To do this they made use of the inverse probability of censoring weights. This approach leads to estimation of an unconditional mean under the assumption that data missing on death is missing at random.

When the terminating event is death, a consistent estimate for the partly conditional mean $\mu^*(t) = E\{N(t) \mid U > t\}$ can be obtained by solving the estimating equation

$$\sum_{i=1}^{n} \frac{Y_i^c(t)}{\widehat{S}_i^c(t)} Y_i^u(t)\{dN_i(t) - d\mu(t)\} = 0, \tag{3.5}$$

where $Y_i^u(t) = I(U_i > t)$ (Cook et al., 2009). Similarly, we can define an intensity model

$$E\{N(t) \mid Z, U > t\} \quad = \quad \mu_0^*(t) \exp(\beta^* Z) \tag{3.6}$$

and a loglinear model

$$\log E\{N(t') \mid Z, U > t'\} \quad = \quad \alpha^* + \theta^* Z. \tag{3.7}$$

Estimators for regression parameters are constructed similarly; in particular, a consistent estimator for β^* is the root of the estimating function

$$U(\beta^*) \quad = \quad \sum_{i=1}^{n} \int_0^\infty \frac{Y_i^c(t) Y_i^u(t)}{\widehat{S}_i^c(t)} W_i(t; \beta^*) dN_i(t), \tag{3.8}$$

where

$$W_i(t; \beta^*) \quad = \quad Z_i - \frac{\sum_j Z_j Y_j^c(t) Y_j^u(t) \exp(\beta^* Z_j) / \widehat{S}_j^c(t)}{\sum_j Y_j^c(t) Y_j^u(t) \exp(\beta^* Z_j) / \widehat{S}_j^c(t)}.$$

A consistent estimator of θ^* can be obtained as the root of the weighted estimating function

$$U(\alpha^*, \theta^*) = \sum_{i=1}^{n} \begin{pmatrix} 1 \\ Z_i \end{pmatrix} \frac{Y_i^c(t')Y_i^u(t')}{\widehat{S}_i^c(t')} \{N_i(t') - \exp(\alpha^* + \theta^* Z_i)\}.$$

3.2.4 Objectives for this paper

The foregoing section has illustrated principles of drawing inference in the presence of informative censoring and a terminating event. In the next section, we introduce additional methods for drawing causal contrasts between treatment groups who are defined in terms of timing of treatment initiation. Specifically, our goals include drawing inferences about the following quantities: (i) the event rate among survivors, $d\mu^*(t)$; (ii) rate ratio between treatment groups and among survivors, in terms of time-averaged rate ratio and rate ratio at a fixed time (as in models (3.6) and (3.7), respectively).

3.3 Causal inferences about treatment effects with uncensored exposure times

3.3.1 Framing the causal question

Our goal is to estimate the causal effect of time to exposure, denoted by A, on rate of occurrence of a counting process $\{N(t) : t > 0\}$. In our application, the time to exposure is timing of cART following initiation of TB treatment, and $N(t)$ represents the number of new opportunistic infections at time t. The target population is those who are alive at time t.

In a randomized controlled trial it would be difficult to randomized exact times of treatment initiation. Instead, we define mutually exclusive time intervals $\mathscr{I}_0, \mathscr{I}_1, \ldots, \mathscr{I}_Z$, where $\mathscr{I}_z = [a_z, b_z)$ for $z = 0, \ldots, Z$ and $\mathscr{I}_0 = (0, b_0)$. We further stipulate that a_Z is the last possible time for receiving treatment and $b_Z = \infty$, which implies those in interval $\mathscr{I}_Z$ do not receive the treatment during

the observed time period. Finally, we assume that each individual has $Z + 1$ *potential* outcomes processes

$$\{N_z(t) : t > 0\}, \quad z = 0, \ldots, Z$$

where $N_z(t)$ is the number of events that would occur if treatment was in fact received during interval $\mathscr{I}_z$. Because each individual has only one treatment time, we can only observe the process $N(t) = \{N_z(t) : A \in \mathscr{I}_z\}$; the unobserved processes are sometimes referred to as *counterfactual*. In this section, we provide methods for estimating causal treatment effects when observed treatment initiation times are uncensored (i.e., observed on each individual). The case of censored initiation times is handled in Section 3.4.

Because the subjects are not randomized into the treatment initiation times, to compare the occurrences of OIs in the various groups, we need to ensure that any imbalances in the treatment specific covariate distributions have been factored out. In observational studies where the goal is to estimate treatment effect, confounding is a key issue. Decisions about when to initiate treatment are made on the basis of a patient's risk profile, which tends also to be related to the actual outcome. Several methods exist for dealing with non random allocation into treatment. These include: multivariable risk adjustment via regression model, propensity score risk adjustment, propensity score matching, inverse probability weighting (IPW) and methods based on instrumental variables; see Rosenbaum & Rubin (1984); Robins et al. (2000); D'Agostino (1998); and Hogan & Lancaster (2004), among others.

As discussed in the previous section, occurrence of OIs could be terminated by death or censored by drop out. The death and dropout processes may also censor time to treatment initiation, leading to missing information on the exposure of interest. We will address each of these issues below. Figure 3.1 shows the possible observed-data patterns.

The objectives of our analyses are as follows: (i) estimation of the rate function $d\mu_z^*(t)$ and mean function $\mu_z^*(t)$, for $z = 0, \ldots, Z$; (ii) inference about the *marginal mean ratios* $\mu_z^*(t)/\mu_0^*(t)$ for $z = 1, \ldots, Z$. To meet objective (i), we will modify (3.2) to accommodate nonrandom allocation to treatment initiation intervals. To meet objective (ii), we will modify the estimating functions associated with the Anderson-Gill model (3.6) and the Poisson regression model (3.7).

3.3.2 Estimation of the rate functions

Recall that $N_z(t)$ denotes the counting process that records the potential cumulative number of OIs in the interval $(0, t]$ under regimen z. We assume that $\{N_z(t)\}$ follows a Poisson process with mean function $\mu_z(t)$; our target for inference is the mean among survivors, $E\{N_z(t) \mid U > t\}$.

Let $\Delta_z = I(A \in \mathscr{I}_z)$. We first consider a scenario where treatment initiation time is randomly assigned, and therefore independent of each potential outcomes process, censoring time, and death time. In this case, the estimated marginal rate function $d\widehat{\mu}_z^*(t)$ to the estimating equation

$$\sum_{i=1}^{n} \Delta_{zi} \frac{Y_i^c(t)}{\widehat{S}_i^c(t)} Y_i^u(t)\{dN_i(t) - d\mu(t)\} = 0,$$

leading to

$$d\widehat{\mu}_z^*(t) = \frac{\sum_i \Delta_{zi} Y_i^c(t) Y_i^u(t) dN_i(t)/\widehat{S}_i^c(t)}{\sum_i \Delta_{zi} Y_i^c(t) Y_i^u(t)/\widehat{S}_i^c(t)}. \tag{3.9}$$

The mean function is the weighted Nelson-Aalen estimator $\widehat{\mu}_z^*(t) = \int_0^t d\widehat{\mu}_z^*(s)$.

Unless A is independent of $\{N_z(t) : z = 0, 1, \ldots, Z\}$, U, and C (i.e., unless there is no confounding of treatment effect), (3.9) will be subject to bias. We propose to use inverse probability of treatment weighting (IPTW) to adjust for non random allocation to treatment under the assumption of that all confounders are measured by a subset of $V_i(t)$. For those who are observed to receive treatment in $\mathscr{I}_z$, the IPTW approach further weights the contribution of each individual inversely by $\pi_{zi} = P\{A \in \mathscr{I}_z \mid H_i(a_z), V_i(a_z)\}$. Specifically, we require the assumptions

(C) *Conditionally on* $\{H(a_z), V(a_z)\}$*, initiating treatment in interval* $\mathscr{I}_z$ *is independent of* $\{N_z(t) : z = 0, \ldots, Z\}$*,* C *and* U*.*

(D) *The probability* π_{zi} *is bounded away from zero for all* i *and* z*.*

With assumptions (C) and (D), and assuming π_{zi} can be consistently estimated by $\widehat{\pi}_{zi}$, a consistent

estimator of the rate function is given by the solution to

$$\sum_{i=1}^{n} \frac{\Delta_{zi} Y_i^c(t)}{\widehat{\pi}_{zi} \widehat{S}_i^c(t)} Y_i^u(t) \{dN_i(t) - d\mu(t)\} = 0, \tag{3.10}$$

which is

$$d\widehat{\mu}_z^*(t) = \frac{\sum_i \Delta_{zi} Y_i^c(t) Y_i^u(t) dN_i(t) / \{\widehat{\pi}_{zi} \widehat{S}_i^c(t)\}}{\sum_i \Delta_{zi} Y_i^c(t) Y_i^u(t) / \{\widehat{\pi}_{zi} \widehat{S}_i^c(t)\}}. \tag{3.11}$$

The mean function is again the weighted Nelson-Aalen estimator $\widehat{\mu}_z^*(t) = \int_0^t d\widehat{\mu}_z^*(s)$.

The associated smoothed estimate of a common rate function for each treatment group can be obtained to provide an insight into the shape of treatment-specific rate functions $\rho^*(t)$. If $\widehat{\mu}_z^*(t)$ is the Nelson-Aalen estimate of $\mu_z^*(t)$, then a smooth nonparametric estimate of $\rho_z^*(t)$ is given by

$$\widehat{\rho}_z^*(t) = m^{-1} \int_{t-m}^{t+m} K\left(\frac{t-s}{m}\right) d\widehat{\mu}_z^*(s), \tag{3.12}$$

where $m > 0$ is a bandwidth and $K(\cdot)$ is a kernel function such as a Gaussian kernel (Cook & Lawless, 2007).

3.3.3 Causal inference about rate ratios using structural models

To capture the causal effect of cART timing on the expected number of OIs, we use a marginal structural model for the mean of $N_z(t)$ among survivors at t,

$$E\{N_z(t) \mid U > t\} \quad = \quad \mu_0^*(t) \exp(\beta_z^*), \tag{3.13}$$

where $\mu_0(t)$ is the expected number of OIs when treatment is started in interval $\mathscr{I}_0$, and the β_z^* are causal log rate ratios for cART initiation in $\mathscr{I}_z$ relative to $\mathscr{I}_0$. This model relies on the assumption that the cART timing has a multiplicative effect on the rate of occurrence of OIs.

Estimation of the β_z^* follows by modifying (3.8) to accommodate multiple treatment groups and

the IPTW. To parameterize in terms of multiple treatment groups, let $\beta^* = (\beta_1^*, \ldots, \beta_Z^*)^{\mathrm{T}}$ denote the $Z \times 1$ vector of treatment effects, and let $x_z = (0, \ldots, 0, 1, 0, \ldots, 0)$ denote the $1 \times Z$ vector with element z equal to 1 and all other entries equal to zero. Now write the estimating function for β^* as

$$U(\beta^*) \quad = \quad \sum_{i=1}^{n} \sum_{z=0}^{Z} \int_0^{\infty} \frac{\Delta_{zi}}{\widehat{\pi}_{zi}} \cdot \frac{Y_i^c(t)}{\widehat{S}_i^c(t)} Y_i^u(t) W_z(t; \beta^*) dN_i(t), \tag{3.14}$$

where

$$W_z(t; \beta^*) \quad = \quad x_z^{\mathrm{T}} - \frac{\sum_j x_z^{\mathrm{T}} \Delta_{zj} Y_j^c(t) Y_j^u(t) \exp(x_z^{\mathrm{T}} \beta^*) / \{\widehat{\pi}_{zj} \widehat{S}_j^c(t)\}}{\sum_j \Delta_{zj} Y_j^c(t) Y_j^u(t) \exp(x_z^{\mathrm{T}} \beta^*) / \{\widehat{\pi}_{zj} \widehat{S}_j^c(t)\}}.$$

Notice here that $W_z(t; \beta^*)$ does not depend on i because there are only Z unique configurations of the design matrix in the structural model. The relative contribution of each individual is dictated by the (estimated) treatment probability $\widehat{\pi}_{zi}$. Under Assumptions (C) and (D), the root $\widehat{\beta}_{\text{IPTW}}^*$ of $U(\beta^*)$ is consistent for the causal parameter β^* (Robins, 1999a).

The special case of (3.13) that assumes constant baseline hazard and focuses on the fixed time interval $(0, \tau]$ reduces to the loglinear structural model

$$\log E\{N_z(\tau) \mid U > \tau\} \quad = \quad \alpha^* + \theta_z^*.$$

In fashion similar to inference for β^* in (3.13), and maintaining Assumptions (C) and (D) on the interval $(0, \tau]$, we can find a consistent estimator of $\theta^* = (\theta_1^*, \ldots, \theta_Z^*)$ as the root of

$$U(\alpha^*, \theta^*) \quad = \quad \sum_{i=1}^{n} \sum_{z=0}^{Z} \frac{\Delta_{zi}}{\widehat{\pi}_{zi}} \begin{pmatrix} 1 \\ x_z^{\mathrm{T}} \end{pmatrix} \frac{Y_i^c(\tau) Y_i^u(\tau)}{\widehat{S}_i^c(\tau)} \{N_i(\tau) - \exp(\alpha^* + x_z^{\mathrm{T}} \theta^*)\}. \tag{3.15}$$

3.3.4 Estimation of treatment and censoring probabilities

Throughout this section we have assumed there exists consistent estimators of $S_i^c(t)$ and π_{zi}. In practice these have to be estimated from data, typically using a model.

We first consider estimation of $S_i^c(t)$. Recall that censoring can depend on both time-varying

covariates $V(t)$ and the event history $H(t)$. In our application below, we assume the hazard of censoring depends linearly on $V(t)$ and on the observed number of events $N(t)$. Hence we assume

$$\lambda^c\{t \mid V(t), H(t)\} \quad = \quad \lambda_0^c(t) \exp\{V(t)\gamma_1 + N(t)\gamma_2\},$$

where $\lambda_0^c(t)$ is the baseline hazard, possibly stratified on baseline covariate categories; γ_1 is a $p \times 1$ vector characterizing the effect of $V(t)$ on the hazard, and γ_2 is a scalar. Consistent estimates $(\widehat{\gamma}_1, \widehat{\gamma}_2)$ can be obtained as maximum partial likelihood estimates (Cox, 1975), and an estimate $\widehat{\lambda}_0^c(t)$ of the baseline hazard follows using Nelson's estimator. In this procedure, we treat censoring times as the 'event' and deaths as censoring times. The individual-specific cumulative hazard of censoring is therefore

$$\widehat{\Lambda}_i^c(t) \quad = \quad \int_0^t \widehat{\lambda}_0^c(s) \exp\{V_i(s)\widehat{\gamma}_1 + N_i(s)\widehat{\gamma}_2\} \, d\overline{N}^c(s), \tag{3.16}$$

where $\overline{N}^c(t) = \sum_i N_i^c(t)$ is the number of censoring events that occur at time t. The estimate of $S_i^c(t)$ used in the estimating functions above is therefore given by $\widehat{S}_i^c(t) = \exp\{-\widehat{\Lambda}_i^c(t)\}$.

In a similar fashion, we assume time to treatment initiation A follows a proportional hazards model

$$\lambda^A\{t \mid V(t), H(t)\} \quad = \quad \lambda_0^A(t) \exp\{V(t)\eta_1 + N(t)\eta_2\}. \tag{3.17}$$

Let $S_i^A(t) = P\{A > t \mid V_i(t), H_i(t)\}$, with functional form implied by (3.17). For individual i, the probability of initiating treatment in $\mathscr{I}_z = (a_z, b_z]$ is therefore given by

$$\begin{aligned} \pi_{zi} \quad &= \quad P\{a_z < A \leq b_z \mid V_i(a_z), N_i(a_z)\} \\ &= \quad S_i^A(a_z) - S_i^A(b_z). \end{aligned}$$

It follows that π_{zi} can be estimated by replacing $S_i^A(\cdot)$ with $\exp\{\widehat{\Lambda}_i^A(\cdot)\}$, where $\widehat{\Lambda}_i^A(t)$ is estimated using an integral similar to (3.16), plugging in relevant analogues from the fitted treatment initiation

time model. In fitting (3.17), we treat treatment initiation time as the 'event', and consider drop-out and death times as censoring events.

3.4 Accommodating censored treatment initiation times

In the dataset some individuals were censored prior to initiation of cART. We assume that each of these individuals could have been assigned to any of the five treatment groups prior to the beginning of follow up. In the AMPATH data, some individuals either die or are lost to follow up before initiating cART. We assume that, prior to the beginning of follow up, each of these individuals could have been assigned to any of the five treatment groups; hence, we use information about the length of follow up time without receiving cART to allocate the indicator variable Δ_{zi} across the remaining potential categories of treatment initiation time. Our approach largely follows Mwangi et al. (2010), and is adapted to the recurrent events setting here.

For those who die or are lost to follow up prior to initiating cART, the initiation time is censored and A is missing. Let $A^* = \min(A, C, U)$, which is the actual cART initiation time when $A^* = A$ and is a minimum observed time without receiving cART when $A^* < A$ (i.e., when it is censored by death or loss to follow up).

Our approach to incorporating partial information on cART initiation time for those with censored values of A is to replace $\Delta_{zi} = I\{A_i \in \mathscr{I}_z\}$ in the estimating functions (3.10), (3.14), and (3.15) with (an estimate of) its expected value,

$$\begin{aligned} \Delta^*_{zi} &= E\{\Delta_z \mid A > A^*_i, U > t, H_i(\infty), V_i(\infty)\} \\ &= \Pr\{A \in \mathscr{I}_z \mid A > A^*_i, U > t, H_i(\infty), V_i(\infty)\}, \end{aligned}$$

for $z = 0, 1, \ldots, Z$. This distributes the indicator variables such that $\sum_z \Delta^*_{zi} = 1$; because we are conditioning on $A > A^*_i$, intervals where $b_z < A^*_i$ have $\Delta^*_{zi} = 0$.

It is important to realize that Δ^*_{zi} is not identifiable from the observed data without making additional assumptions. If we assume that censoring of A at time t is non-informative conditional on $V(t)$ and $H(t)$ (which here we mean to include their histories), then Δ^*_{zi} can be computed using

the fitted hazard model (3.17) for cART initiation that is used to construct the inverse probability weights. Specifically, under the assumption

(E) *Among those in follow up at t and who have $A > t$, the hazard of either death or censoring at t is independent of A conditionally on $\{V(s), N(s) : 0 < s < t\}$,*

we have

$$\begin{aligned}
\widehat{\Delta}^*_{zi} &= \frac{\widehat{\Pr}\{A \in \mathscr{I}_z,\ A > A^*_i \mid U > t, V_i(\infty), H_i(\infty)\}}{\widehat{\Pr}\{A > A^*_i \mid V_i(\infty)\}} \\
&= \begin{cases} 0 & \text{if } A^*_i \geq b_z \\ \dfrac{\widehat{S}^A\{A^*_i \mid V_i(A^*_i), H_i(A^*_i)\} - \widehat{S}^A\{b_z \mid V_i(A^*_i), H_i(A^*_i)\}}{\widehat{S}^A\{A^*_i \mid V_i(A^*_i), H_i(A^*_i)\}} & \text{if } a_z \leq A^*_i < b_z \\ \dfrac{\widehat{S}^A\{a_z \mid V_i(A^*_i), H_i(A^*_i)\} - \widehat{S}^A\{b_z \mid V_i(A^*_i), H_i(A^*_i)\}}{\widehat{S}^A\{A^*_i \mid V_i(A^*_i), H_i(A^*_i)\}} & \text{if } A^*_i < a_z. \end{cases}
\end{aligned}$$

Other assumptions are of course possible. At one extreme, we can set $\widehat{\Delta}^*_{zi} = I\{A^*_i \in \mathscr{I}_z\}$, which assumes treatment would have been initiated with probability one during the interval where censoring or death took place; the opposite extreme is to assign all probability mass to the interval $\mathscr{I}_Z$, which means cART would not have been initiated during the period of TB treatment. It is also possible to allocate probabilities uniformly, for example by assigning $\widehat{\Delta}^*_{zi} = 1/m_i$, where $m_i = \sum_z I(A^*_i > b_z)$, is the number of intervals remaining within which cART could still be initiated; mass can also be assigned in relative proportion to interval length. More formally, a sensitivity parameter can be introduced to control the relative allocation of $\widehat{\Delta}^*_{zi}$ over the remaining intervals.

Let $\widetilde{\Delta}_{zi} = \Delta_{zi} I(A_i = A^*_i) + \widehat{\Delta}^*_{zi} I(A_i > A^*_i)$, which is equal to the observed interval indicator when A_i is observed and the estimated one when A_i is censored. Replacing Δ_{zi} with $\widetilde{\Delta}_{zi}$ yields the appropriate modifications in the estimating functions (3.10), (3.14), and (3.15) that are used to estimate the rate function, the rate ratio under inhomogeneous Poisson process, and homogeneous Poisson process, respectively.

3.5 Analysis of AMPATH data

3.5.1 Summary statistics

We used data for 4908 patients on TB treatment initiating cART at different times in the course of their TB treatment who had data in all the covariates of interest. A total of 3506 had observed cART initiation times during TB treatment period (in the first 8 months of TB treatment), and 342 were observed not to start cART during TB treatment. The remaining 1060 patients were either administratively censored, lost to follow up or died prior to initiation of cART, and therefore have right censored treatment initiation times. Of the 1060 patients with missing exposure, 184 (17%) were confirmed to die before initiating cART with the median time to death of 48 days.

From Table 3.1 we observe that some of the patients were missing data on OIs at 1 year; this was due to death or drop out before one year. Table 3.2 shows the mean number of OIs based on the observed data at 1 year in each of the cART initiation time categories. We observe that the patients initiating cART in the 2-8 weeks interval had the highest mean number of OIs with the never starters having the lowest. In all likelihood, this can be attributable to the 'never-starters' being the healthiest subpopulation among those who initiate TB therapy.

3.5.2 Treatment comparisons

Figures 3.3 and 3.4 show estimates of the rate function $d\mu^*(t)$, both weighted and unweighted, stratified by cART timing. What seems clear from Figure 3.4 is that the rate of OI among survivors decreases in the time periods following cART initiation. For the other intervals the rate fluctuates but seems to hold constant over time. Among those never treated, the OI rate drops sharply following time 0. We found in Mwangi et al. (2010) that this group also has the highest mortality rate, so this could be a healthy survivor effect.

Figure 3.2 shows weighted and unweighted estimators of $E\{N(t) \mid U > t\}$ in the five treatment categories. These plots are somewhat difficult to interpret in the absence of information on survival

distributions for the three groups. Integrating these curves against survival distributions would give an estimate of the total number of OI's expected in each category.

The structural models provide more formal comparisons of OI rates between the treatment groups. Tables (3.3 and 3.4) show the estimated rates of OIs per week over different times from start of TB treatment; these are estimated using the loglinear model structural under the assumption of a homogeneous Poisson process. Comparing estimates from these tables shows the effect of using a weighted versus unweighted estimate. Specifically while the 52-week rate in the unweighted analysis suggests that OI rates are much lower for those who initiate cART earlier, while the weighted model indicates that OI rates are largely the same (with the exception of the last group). Table 3.6 formalizes this comparison using rate ratios and 95% bootstrap confidence intervals, and indicates that over one year, the OI rate among survivors is largely the same, regardless of when treatment is initiated.

Table 3.7 shows results for the Andersen-Gill type models that allow for an inhomogeneous Poisson process. Interestingly, this model suggests that if anything, OI rates among survivors may actually be higher for those who initiate treatment later; however it should be noted that none of the confidence intervals for later treatment times indicates a statistically significant difference from those who initiate in $(0, 2]$.

We used different sets of covariates in the weight models. For the censoring model, we included the following baseline covariates: age, gender, presence of a non-TB AIDS-defining event, CD4 count, and weight. We also included $N(t)$ as a time-varying covariate. For the treatment initiation time model, we included only baseline covariates: age, gender, marital status, indicator of post-primary education, clinic site (urban or rural), CD4 count, weight, presence of a non-TB AIDS-defining event, and WHO disease stage (=1 if stage 3 or 4, =0 otherwise). Both models used stratified baseline hazards, stratifying on baseline CD4 category (0-50, 50-200, 200-350). We also stabilized the weights using Kaplan-Meier estimators of the relevant event time distributions in the numerator; see Mwangi et al. (2010) for details.

To summarize, we have used various methods of estimating and providing causal comparisons of OI rates between those who initiate cART at various times following initiation of TB treatment. In a previous paper (Mwangi et al., 2010), we showed that *mortality* rates depended strongly on cART initiation time, in that those who start early have the lowest mortality rate. In this paper,

our findings suggest that rates of OI among survivors are at best equal between treatment initiation times, and possibly higher among those who start later. The notable exception in our OI analysis is the group who do not receive cART during their period of TB treatment; in all of our analyses, they have demonstrably lower rates of OI. However, they also have the highest mortality rate, so there remains the possibility of a health survivor effect for this group.

3.6 Discussion

In this paper we use observational data from a large electronic medical records system to assess the effect of cART timing on the rate of occurrence of opportunistic infection. We use marginal structural models to adjust for non random allocation to treatment and informative censoring, using inverse probability weighting under some very specific assumptions. We also provide a method for including individuals with censored exposure time using a 'redistribution to the right' type of approach that allocates weight over possible remaining initiation times. This paper builds directly on work by Cook et al. (2009), who developed methods for handling informative censoring and a terminating event. Our methods extend their approach to draw causal inferences about the rates when treatment is not randomly allocated and when the treatment variable itself is censored. We describe methods for estimating smooth rate functions and for inferring causal contrasts for homogeneous and inhomogeneous Poisson processes.

Our estimation of causal contrasts relies on the key assumption of unmeasured confounding conditional on observable information; this assumption might not hold; for example there could be some unknown drug interactions between TB medications and cART that lead to OIs. It might also not be possible to disentangle the effect of cART from the effect of TB medications; hence what we are observing might be due to the immune moderation by TB treatment.

Another limitation of our work relates to the difficulty of diagnosing OIs in resource constrained settings. In our judgment, this probably leads to under-reporting and quite possibly under-estimation of OI rates. Finally, related to subject selection, we did not have data on sputum tests or chest x-rays for everyone (these are standard methods for TB diagnosis in developed countries). Our patient selection is based on the assumption that a patient has TB if they initiated TB medications.

Table 3.1: Number (percent) of observed OIs at 52 weeks following initiation of TB treatment, stratified by observed interval of cART initiation time.

No. of OIs	[0-2) N=734	[2,8) N=852	[8,16) N=733	[16,35) N=345	[35,∞) N=304
0	597 (81.3)	705 (82.7)	588 (80.2)	291 (84.3)	275 (90.5)
1	35 (4.8)	52 (6.1)	63 (8.6)	25 (7.2)	11 (3.6)
2	22 (3.0)	15 (1.8)	14 (1.9)	6 (1.7)	7 (2.3)
3	13 (1.8)	9 (1.1)	7 (1.0)	6 (1.7)	4 (1.3)
4	13 (1.8)	6 (0.7	8 (1.1)	3 (0.9)	3 (1.0)
5	9 (1.2)	4 (0.5)	12 (1.6)	1 (0.3)	1 (0.3)
6	9 (1.2)	9 (1.1)	9 (1.2)	1 (0.3)	0 (0)
7	7 (1.0)	6 (0.7)	4 (0.5)	2 (0.6)	2 (0.6)
8	9 (1.2)	9 (1.1)	10 (1.4)	0 (0)	0 (0)
9	4 (0.5)	15 (1.8)	6 (0.8)	0 (0)	1 (0.3)
$\geq$10	16 (2.2)	22 (2.6)	12 (1.6)	10 (2.9)	0 (0)

Table 3.2: Mean (SD) number of observed OIs, stratified by cART initiation time.

Initiation time (weeks)	Mean (SD)
(0, 2]	0.84 (2.33)
(2, 8]	0.87 (2.63)
(8, 16]	0.77 (2.22)
(16, 35]	0.60 (2.12)
(35, ∞)	0.25 (1.01)

Table 3.3: Estimated rate of OI per week at various fixed time points, assuming homogeneous Poisson process and no adjustments for confounding or informative censoring.

Initiation time (weeks)	8 weeks	16 weeks	35 weeks	52 weeks
[0-2]	0.029	0.021	0.017	0.016
(2-8]	0.034	0.025	0.020	0.017
(8-16]	0.016	0.018	0.017	0.014
(16,35]	0.011	0.011	0.014	0.012
(35, ∞)	0.009	0.005	0.004	0.005

Table 3.4: Estimated rate of OI per week at various fixed time points, assuming homogeneous Poisson process and using IPW estimator to adjust for confounding, informative censoring, and censored cART initiation times.

Initiation time (weeks)	8 weeks	16 weeks	35 weeks	52 weeks
[0-2]	0.026	0.018	0.014	0.014
(2-8]	0.029	0.020	0.016	0.015
(8-16]	0.019	0.019	0.016	0.015
(16,35]	0.024	0.019	0.018	0.015
(35, ∞)	0.025	0.014	0.006	0.006

Table 3.5: Rate ratios and 95% bootstrap confidence intervals for comparing rate of OI at fixed intervals following baseline, under assumption of homogeneous Poisson process. No adjustments for censoring, confounding or censored exposure time.

Initiation time (weeks)	8 weeks	16 weeks	35 weeks	52 weeks
[0-2]	1	1	1	1
(2-8]	$_{0.89}1.19_{1.61}$	$_{0.89}1.18_{1.57}$	$_{0.89}1.16_{1.56}$	$_{0.80}1.04_{1.37}$
(8-16]	$_{0.39}0.56_{0.78}$	$_{0.62}0.84_{1.14}$	$_{0.69}0.95_{1.26}$	$_{0.69}0.92_{1.22}$
(16,35]	$_{0.20}0.37_{0.66}$	$_{0.31}0.53_{0.85}$	$_{0.48}0.80_{1.24}$	$_{0.42}0.70_{1.11}$
(35,∞]	$_{0.15}0.30_{0.55}$	$_{0.12}0.23_{0.40}$	$_{0.14}0.24_{0.38}$	$_{0.16}0.29_{0.48}$

Table 3.6: Rate ratios and 95% bootstrap confidence intervals from marginal structural model, assuming homogeneous Poisson process. The MSM adjusts for censoring and confounding, and includes information on those with censored exposure times.

Initiation time (weeks)	8 weeks	16 weeks	35 weeks	52 weeks
[0-2]	1	1	1	1
(2-8]	$_{0.85}1.14_{1.57}$	$_{0.86}1.13_{1.53}$	$_{0.85}1.17_{1.54}$	$_{0.78}1.07_{1.40}$
(8-16]	$_{0.52}0.73_{1.01}$	$_{0.77}1.04_{1.42}$	$_{0.83}1.15_{1.52}$	$_{0.82}1.11_{1.48}$
(16,35]	$_{0.55}0.91_{1.50}$	$_{0.57}1.01_{1.85}$	$_{0.67}1.22_{2.18}$	$_{0.55}1.07_{1.96}$
(35,∞)	$_{0.50}0.93_{1.80}$	$_{0.40}0.73_{1.42}$	$_{0.20}0.40_{0.76}$	$_{0.22}0.42_{0.75}$

Table 3.7: Rate ratios and 95% bootstrap confidence intervals from unadjusted AG model and marginal structural model, assuming inhomogeneous Poisson process. The MSM adjusts for censoring and confounding, and includes information on those with censored exposure times.

Initiation time (weeks)	Unadjusted	MSM
(0-2]	1	1
(2,8]	$_{0.86}1.10_{1.40}$	$_{0.92}1.23_{1.63}$
(8,16]	$_{0.70}0.91_{1.18}$	$_{0.95}1.25_{1.66}$
(16,35]	$_{0.54}0.77_{1.10}$	$_{0.95}1.39_{2.05}$
(35, ∞)	$_{0.16}0.25_{0.40}$	$_{0.30}0.52_{0.88}$

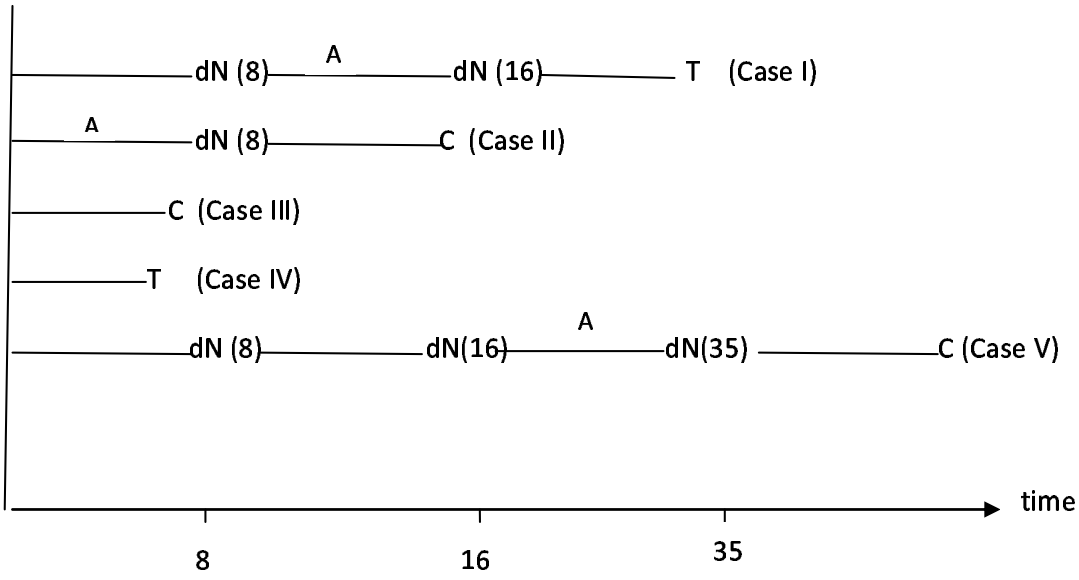
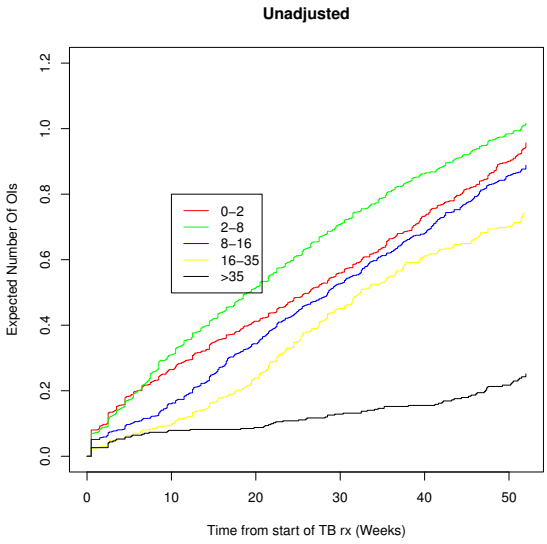
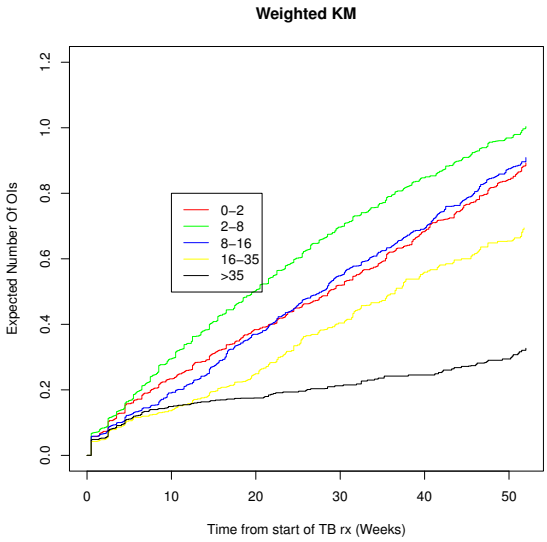


Figure 3.1: Possible patterns of observed data, where dN denotes occurrence of an event, A is time to treatment initiation (exposure), C is loss to follow up, and D is death.



(a)



(b)

Figure 3.2: Estimates of $E\{N(t) \mid U > t\}$, stratified by cART timing. Top panel uses only observed data. Bottom panel is weighted to adjust for censoring and confounding, and includes information on those with missing exposure. Curves are computed using (weighted) Nelson estimator.

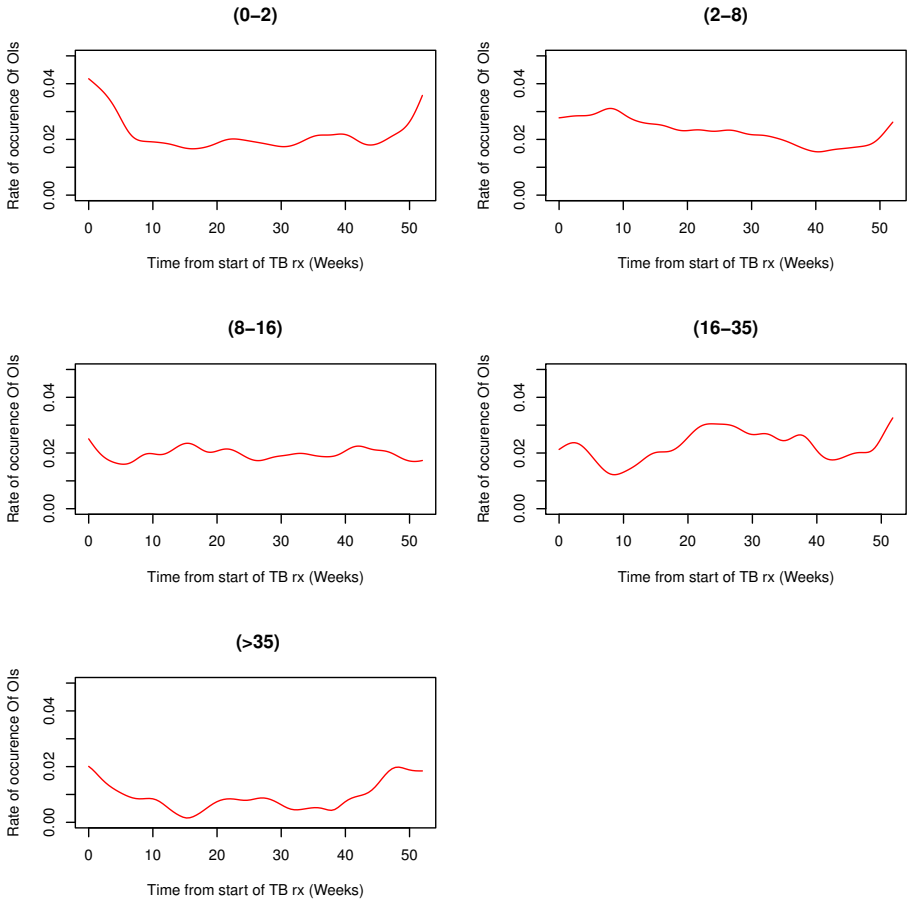


Figure 3.3: Estimated rate functions $d\mu^*(t)$, stratified by cART timing. No adjustments made for confounding, censoring, or missing exposure.

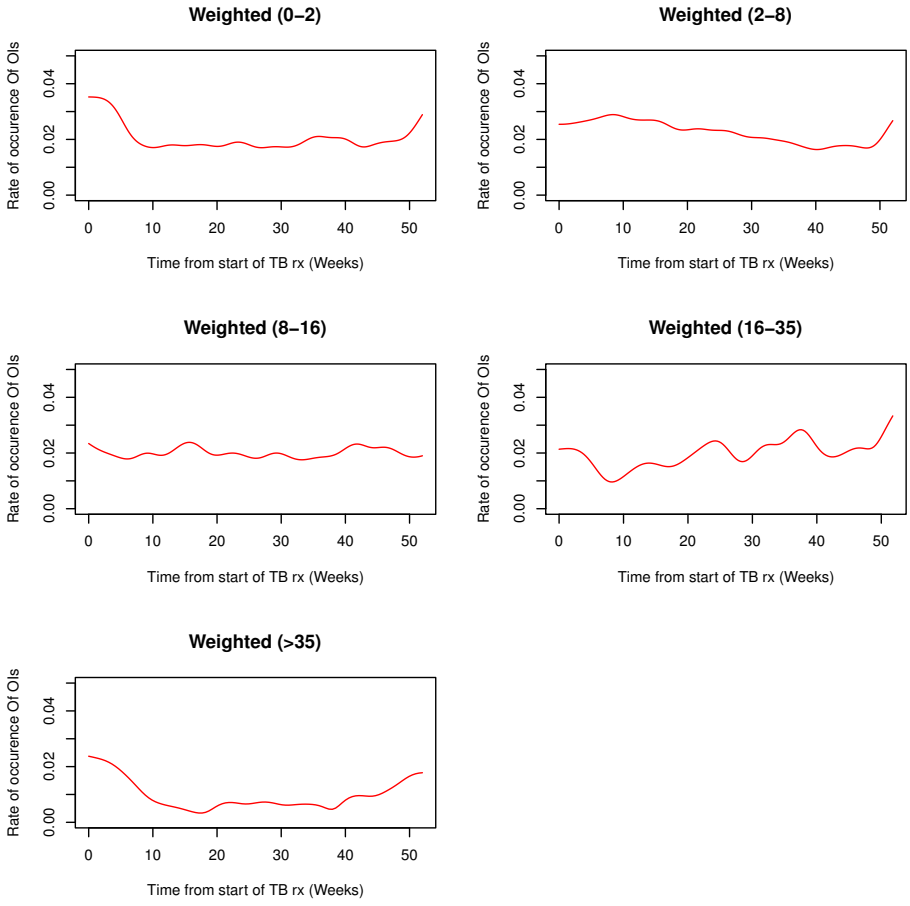


Figure 3.4: Estimated rate functions $d\mu^*(t)$, stratified by cART timing, with adjustments made for confounding, censoring, and missing exposure.

Chapter 4

Estimating survival distribution from double-sampled data by modeling the informative censoring process

Abstract

Estimates of a survival distribution can be biased when event times are informatively censored. In many HIV care programs in the developing world, mortality rate is a key indicator of program effectiveness, but estimates based on observed data can suffer significant bias because large fractions of patients are lost to follow up. The bias derives from dependence between the censoring process and the survival time.

To alleviate this problem, some programs implement what is sometimes known as 'double sampling', where a subset of those with missing mortality status are followed up in an effort to ascertain their status (Glynn et al., 1993; Frangakis & Rubin, 2001). In HIV programs, recent papers by Yiannoutsos et al. (2008) and An et al. (2008) have shown that losses to follow up tend to have very high mortality rates, and that substantial corrections are possible when incorporating information on double-sampled data.

Censoring is informative if the loss to follow up time L is related to the event time T. In most

settings this relationship is not known because we only observe $\min(T, L)$. Usually we assume that T is independent of L (Klein & Moeschenberger, 2003), or that T is independent of L conditional on some measured covariates V. In double sampling, we observe T and L for some individuals, affording opportunity to model their dependence via a hazard model $\lambda^L(L|T)$.

With double-sampled data, the survival distribution can be estimated with inverse probability weighting (IPW) (Frangakis & Rubin, 2001; Robins et al., 2001), or with a parametric model that accounts for double sampling. In this paper, we show how to use the double sampled data to estimate a model for the hazard of informative censoring as a function of the actual survival time, and in turn how to use the censoring model to construct an IPW estimator of the survival distribution.

It is possible to estimate the survival distribution without explicitly modeling this dependence. A key motivation for our approach is that the informative censoring model, once estimated, can be adapted to other analyses of data from the same program, which is advantageous because double-sampled data are difficult and expensive to collect.

4.1 Introduction

Survival analysis is an important tool for research and program evaluation in clinical care programs in developing countries. For example, The President's Emergency Plan for AIDS Relief (PEPFAR), a response by the US government to the fight against HIV in the developing world, depend on regular estimates of mortality in order to evaluate treatment delivery programs, monitor progress, and assess public health impact on HIV infection. A key impediment to constructing accurate mortality estimates and models is censoring caused by patient loss to follow up (LTFU). In a review of 33 African cohorts, Rosen et al. (2007) calculated a weighted mean attrition rate of 1.8% to 3.3% per month, which they attributed largely to follow-up losses, with 40% of patients being lost from care by 2 years after initiation of ART in low income countries. Another study of 13 African cohorts (Braitstein et al., 2006) noted an average of 15% loss to follow-up at 12 months following ART initiation, with variability ranging from 0% to 44% across programs.

Informative right censoring in general is an important issue in survival analysis, and if not properly adjusted can cause bias in event rate estimates, regression models and hypothesis tests. In HIV it is widely assumed that patients lost to follow up (LTFU) have higher mortality rates. Morgan

et al. (2002) reported that patients lost to follow up not due to transfer to other care providers are likely to die within 1 year of loss to follow up. In a study carried out in Malawi, 3,203 patients were traced and vital status were ascertained for 94%; of these, 345 (12%) were confirmed to be deaths (Mukhuna et al., 2009). In another study in South Africa, Maskew et al. (2007) identified 182 lost to follow up patients and death accounted for 27% of loss to follow up. Results from a systematic review and meta analysis by Martin WG Brinkhof (July 2008) showed that 20-60% of patients lost to follow up by subsequently tracked had died.

Correction of censoring bias relies on knowing the relationship between L and T. Two commonly used approaches to address informative censoring are to use covariates for modeling the informative censoring process, whereby survival time is considered independent of censoring conditional on covariates; and sensitivity analysis (Scharfstein et al., 2001), whereby the effect of departures from the conditional independence can be assessed.

Double sampling gives the opportunity to learn the relationship between L and T. It involves aggressively pursuing information on those patients lost to follow-up. A double sampling design has two phases (Frangakis & Rubin, 2001), as shown in Figure 4.1. In phase I the follow up effort is based on regularly scheduled clinic visit. In this phase an individual is either observed to die (Case 1), or observed to survive to the end of the follow up period (Case 2), or is observed to be lost to follow up. (Case 3, 4, 5 and 6) . In phase II a subset of LTFU's from phase I are selected and pursued intensively (Case 3, 4 and 5); a subset has vital status ascertained (Case 3 and 4). The information from the two phases is then combined to estimate the survival function.

To estimate the survival function in the presence of right censoring, it is sometimes assumed that LTFU is ignorable after conditioning on covariates. In this paper, we show how to use the double sampled data to estimate a model for the informative censoring distribution, and in turn how to use the censoring model to construct an IPW estimator of the survival distribution. This method can be ported over to other inferential approaches such as log rank test, Cox regression etc. It is also possible to use the informative censoring model as prior information for other studies, especially if it is possible to adjust for important differences in covariate distributions.

Analyses by Yiannoutsos et al. (2008) and An et al. (2008) using data from a cohort of HIV-infected individuals initiating combination antiretroviral therapy in East Africa show the degree to which double sampled data can be used to correct potential biases in estimation of overall mortality.

In their study, the goal is to estimate one-year mortality from date of treatment initiation. Assuming non-informative censoring and confining the analysis to phase I data, the observed mortality rates were 1.7% (95% CI 1.3%-2.0%). Including the double sampled data but assuming non-informative censoring, the results were revised to 2.8% (95% CI 2.3%-3.1%); in this analysis, data from the two phases were combined and no distinction was made between patients not lost to follow up and double sampled LTFU's. Using methods based on properly incorporating phase II data (described below), the mortality rates were revised to 9.9% (95% CI 8.4%-11.5%).

4.2 Data source: Academic model providing access to health care (AMPATH)

AMPATH is a PEPFAR funded site in western Kenya, currently treating over 70,000 patients seeking HIV care at AMPATH clinics. AMPATH has an outreach program that traces patients who fail to adhere to their clinic visits. At the initial encounter with the program, outreach staff collects and records information in a locator form from patients about how they can be traced if they miss appointments. Outreach staff support patients who miss appointments to improve retention by following them up by phone or home visit and encouraging them to return to clinic. They use the AMPATH medical records system (AMRS), a computerized database designed for clinical data management (Tierney et al., 2007) to generate lists of patients who have missed scheduled visits. For patients newly initiated on combination antiretroviral therapy (cART) the tracing starts on the second day of their defaulted scheduled appointments. However, for those already on cART and stable this might take 2-4 weeks. The main goals of outreach are to determine the patients vital status and encourage return to the treatment program. The program is described in detail by Wools-Kaloustian et al. (2006), Einterz et al. (2007) and Ochieng-Ooko et al. (2010).

In this paper we use data from a double sampling study carried out in AMPATH for a cohort of 8977 adults who were enrolled into two AMPATH clinics between January 2005 and January 2007. In this study a patient was considered to be LTFU if on cART and has not kept a visit for more than 3 months or if not receiving cART has not come to clinic for more than 6 months. At AMPATH the decision to track patients is based on their cART status and tracking is done by first

trying to reach the patients by telephone and where necessary a home visit in remote areas. It is hard to follow some patients due to several reasons such as: distance, stigma surrounding HIV, return to clinic before attempts to double sample and lack of information in the locator forms due to patients refusal. Among the 8,977 adult clients 3,624 were LTFU; of these, 1,143 were targeted for tracking and 621 had their vital status ascertained (Yiannoutsos et al., 2008). The outreach data has been used to adjust mortality rate estimation (An et al., 2008; Yiannoutsos et al., 2008) and to investigate the effect of gender on loss to follow up (Ochieng-Ooko et al., 2010).

4.3 Current statistical approaches for double sampling designs

4.3.1 Terminology and notation

Terminology is important because we are dealing with multiple types of censoring. A patient is *lost to follow up (LTFU)* if, while actively taking cART, he or she has not kept a visit for more than 3 months. For those not actively taking cART, a patient is LTFU if they have not come to clinic for more than 6 months. A patient is *administratively censored* if the study period ends before we observe the event. Finally, a patient is *double sampled* if, after being lost to follow up, the patient is targeted for ascertainment of vital status. Notably, not every patient who is targeted for double sampling has their vital status ascertained.

In the absence of LTFU a subject is observed to either fail at time T or to be administratively censored at time C. Thus for each individual we observe $X = \min(T, C)$ and $\Delta = I(X = T)$ which is an event indicator. If, in addition to administrative censoring we have LTFU, we let L be the loss to follow up time, which can occur prior to both T and C.

In a double sampling design, a subset of LTFU are targeted for double sampling. Among LTFU, let S be an indicator for whether an individual is targeted for double sampling, and let A be an indicator of whether X and Δ are subsequently ascertained. From Figure 4.1, X and Δ are ascertained for Cases 1 and 2 in phase I and for Cases 3 and 4 in phase II. Some targeted subjects are not found, and not all losses to follow up are targeted for double sampling; for these individuals, we do not observe X and Δ (Cases 5 and 6 in Figure 4.1).

Following Robins et al. (2001), to take into account the double sampling design, we define a

new censoring time

$$Q = \begin{cases} L & \text{if } S = 1, A = 0, L < X \\ L & \text{if } S = 0, A = 0, L < X \\ C & \text{otherwise} \end{cases}$$

Thus for each subject in the aggregated sample we can define $\widetilde{X} = \min(Q, X)$ and $\widetilde{\Delta} = I(\widetilde{X} = T)$. In addition, we observe a vector of baseline covariates V such as baseline CD4 count, WHO staging, gender and clinic type (urban vs rural).

4.3.2 Statistical Approaches

For an event time T, let $N(u)$ denote the counting process, $Y(u)$ the at risk indicator, and $\Lambda^T(u)$ the cumulative hazard. Our goal is to estimate features of the hazard $\Lambda^T(u)$ using the double sampling data. We can obtain estimates of $S^T(u)$ using the relation between the cumulative hazard and the survival function. Before describing our approach, we review methods that have been applied to this problem by others. With the exception of parametric models, the available methods can be characterized as weighted Nelson-Aalen estimators.

4.3.2.0.5 Use Phase I data only. It is instructive to review the standard estimator that ignores double sampled data, which is the Nelson-Aalen estimator applied to the Phase I data, treating administratively censored observations and losses to follow up as non-informatively censored. Using notation from above, this estimator can be written as

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\sum_i dN_i(u) I\{u \leq \min(X_i, L_i)\}}{\sum_i Y_i(u) I\{u \leq \min(X_i, L_i)\}}$$

This estimator is biased if LTFU is associated with survival; if for example much sicker patients are more likely to be lost to follow up.

4.3.2.0.6 Combine Phase I and Phase II data without adjusting for sampling probabilities.

This approach involves combining data from phase I and double sampled data from phase II, making no distinction between failures from phase I and those from phase II. The Nelson Aalen estimator is of the form

$$\widehat{\Lambda}(t) = \int_0^t \frac{\sum_i d\widetilde{N}_i(u)}{\sum_i \widetilde{Y}_i(u)}$$

where now $d\widetilde{N}(u) = I(\widetilde{\Delta} = 1, \widetilde{X} = u)$ and $\widetilde{Y}(u) = I(\widetilde{X} \geq u)$. As discussed by Frangakis & Rubin (2001), failing to account for differential sampling probabilities between Phase I and II observations will generally lead to bias, even if LTFU is a noninformative censoring process.

4.3.2.0.7 Frangakis-Rubin (F-R) approach.

The basic FR approach does not use covariates. In Section 5 we show how to incorporate covariates. Frangakis & Rubin (2001) proposed a means to adjust survival estimates when censoring is dependent on the event of interest using double sampling data. The method relies on a number of key assumptions:

(a) All LTFU subjects have same probability of being of double sampled

(b) All double sampled LTFU have X and Δ ascertained

(c) Administrative censoring time is independent of both failure and loss to follow up time.

The F-R method involves using data for non drop outs in phase I and the double sampled observations whose status was ascertained. Robins et al. (2001) showed that the F-R estimator for the cumulative hazard has the form of an IPW estimator

$$\widehat{\Lambda}_2(t) = \int_0^t \frac{\sum_i dN_i(u)K_i(u)}{\sum_i Y_i(u)K_i(u)}$$

where $K_i(u) = R_i/\pi_i$. The variable R_i is an indicator of not being lost in phase I or being lost in phase I and targeted and ascertained in phase II; in terms of the event history variables,

$$R_i = I\{\text{min}(T_i, C_i) < L_i \text{ or } (S_i = 1, A_i = 1)\}.$$

The probability π_i is defined as

$$\pi_i = \begin{cases} 1 & \text{if } \min(T_i, C_i) < L_i \\ \\ \dfrac{P(S=1, A=1)}{P\{\min(T, C) > L\}} & \text{otherwise} \end{cases}$$

The F-R method uses only the data for those not lost in phase I or those lost in phase I and targeted and ascertained in phase II. However, the excluded observations $\{i : X_i > L_i, S_i = 0\}$ still contains information about the survival time distribution because we know that survival time is greater than drop-out time. Hence efficiency can be potentially improved by extending the F-R method to incorporate this portion of the data.

4.3.2.0.8 Inverse probability weighting using covariate information. Robins et al. (2001), in the discussion of Frangakis & Rubin (2001) paper, showed that the F-R estimator can be generalized to include information about covariates. They proposed the following IPCW estimator for $\Lambda(t)$ when there is data $\widetilde{X}, \widetilde{\Delta}$ and (potentially time-varying) covariates $V(u)$:

$$\widehat{\Lambda}_3(t) = \int_0^t \frac{\sum_i d\widetilde{N}_i(u) K_i(u)}{\sum_i \widetilde{Y}_i(u) K_i(u)},$$

where $K_i(u) = 1/\{P(Q > u|\overline{V}_i(u))\}$ is the conditional probability of remaining uncensored up to time u based on the models for the following:

$$\begin{aligned} \lambda^Q(u|V(u)) = \lambda_0^Q(u)\exp\{\theta V(u)\}, &\qquad \text{when } u \neq L \\ P\{A = 0|S = 1, V(u)\}, &\qquad \text{when } u = L \\ P\{S = 1|L = u, V(u)\}, &\qquad \text{when } u = L \end{aligned}$$

These models can be specified in terms of Cox regressions, logit regression etc and can make use of time varying covariates (Robins et al., 2001)

4.4 Hazard estimation using a model for informative censoring

We view the problem as an informative censoring problem where the double sampled data provides information about the association between T and L. We make use of the double sampling data to estimate the relation between informative censoring L and the event time T by modeling the hazard $\lambda^L(U|T)$ for loss to follow up given the failure time.

Scharfstein et al. (2001) provided a semi-parametric approach to estimate the distribution of survival with dependent censoring for discrete survival data using IPCW when double sampled data are not available; they also propose sensitivity analyses to assess the effect of potentially informative censoring. Individuals are weighted by inverse probability of not being lost to follow up. The weights are obtained from a model that captures the dependence of L on T

$$S^L(u) = P(L > u|T > u, T, V(u))$$

In the standard setting, one cannot estimate $S^L(u)$. Scharfstein et al. (2001) assumed that the relationship between the non identifiable distribution i.e $f(T|X = u, \Delta = 0)$ and the identifiable distribution, $f(T|X > u)$ can be represented by the exponential tilt model

$$f(T|X = u, \Delta = 0) = f(T|X > u)\frac{\exp\{q_u(T)\}}{E(\exp\{q_u(T)|X > u)\}} \tag{4.1}$$

where q_u is a known function of u and T that is generally not identifiable from observed data, referred to as a *censoring bias function*. Relation (4.1) implies $\text{logit}\{\lambda^L(u|T \geq u, T)\} = h_u + q_u(T)$, where

$$\lambda^L(u|T \geq u, T) = P(L = u|T \geq u, T)$$

is the hazard for L.

We treat L as an informative censoring process, where double sampling provides information to actually estimate the censoring model $\lambda^L(u|T)$. Scharfstein et al. (2001) show that when the

exponential tilt assumption holds, $\Lambda(t)$ can be consistently estimated by the IPCW estimator

$$\widehat{\Lambda}_4(t) = \int_0^t \frac{\sum_i dN_i(u)K_i(u)}{\sum_i Y_i(u)K_i(u)} \tag{4.2}$$

where now $R_i = I(T_i = X_i)$ and $K_i(u) = R_i/P\{L > u|\overline{V}_i(u), T\}$. The discrete-time hazard $P\{L = u|T > u, T, V(u)\}$, needed to compute the denominator of $K(u)$, can be estimated as the root of the function

$$U = \sum_{i=1}^{n} \sum_{u=0}^{M-1} W_i I(L_i \geq u) I(T_i > u) \times \{I(L_i = u) - P(L = u|T_i > u, T_i, V_i(u))\} \tag{4.3}$$

Note that the estimating equation (4.3) is itself based on IPW because T is not observed on everyone. It is observed only among the failures in phase I; i.e., if $\{T < C, T < L\}$, and among failures in phase II; i.e., if $\{T < C, T > L, S = 1, A = 1\}$. To estimate the weights W in (4.3) in the presence of covariates V, we need to estimate the following:

$$\begin{aligned} P(T < C, T < L|V) &= P(T < C|T < L, V)P(T < L|V) \\ &= \underbrace{P(\Delta = 1|X < L, V)}_{\pi_1} \underbrace{P(X < L, V)}_{\pi_2} \end{aligned}$$

and

$$\begin{aligned} P(T < C, T > L, S = 1, A = 1|V) = &\underbrace{P(T < C|T > L, S = 1, A = 1, V)}_{\pi_3} \\ &\times \underbrace{P(S = 1|A = 1, T > L, V_1(t))}_{\pi_4} \underbrace{P(A = 1|T > L, V_2(t))}_{\pi_5} \\ &\times \underbrace{P(T > L|V)}_{\pi_6} \end{aligned}$$

Thus for each individual for whom T is observed, the weight is

$$W = \frac{I(X < L)\Delta}{\pi_2\pi_1} + \frac{I(X > L)SA\Delta}{\pi_3\pi_4\pi_5\pi_6}$$

The probabilities $\pi_1, \ldots, \pi_6$ can be estimated by fitting logistic regression models with covariates V.

4.5 Application to AMPATH Double sampling Data

4.5.1 Data summaries

There were 8977 individuals enrolled in the study between January 2005 and January 2007. Figure 4.2 shows a flow chart indicating the different paths in the double sampling study. There were 106 deaths discovered through routine follow up and an extra 124 discovered through following a sample of those LTFU. We discretize the continuous time into two month intervals, and define $M = 12$; this is equivalent to 24 months, or the highest possible administrative censoring time in the dataset. In the first phase there were 5379 administratively censored individuals with 462 being administratively censored at or after 24 months. Of the 621 double sampled there were 126 confirmed deaths.

In our dataset there was no indicator of being targeted for double sampling; we only have an indicator for those that were targeted and had their status ascertained. Figure 4.3 shows the time to LTFU stratified by double sampling status. Among those lost to follow up, we observe that the double sampled individuals have better survival compared to the never double sampled. The median LTFU time among the double sampled individuals is 133 days, whereas for the non double sampled it is 83.

4.5.2 Summary of statistical models

To estimate the censoring bias function (4.1) we used the model

$$\text{logit } h^L(u|V,T) = h_u + \alpha(T_i - u)I(T_i < M+1) + \alpha_1 I(T_i = M+1) + \theta V,$$

where h_u is a baseline hazard (intercept), α is the conditional log odds ratio of being lost to follow up in the interval between times u and $u+1$ per unit increase in T for subjects who have not yet experienced death at time u, and θ is the effect of covariates V. As mentioned in Section 4, to estimate the censoring bias parameters α and α_1 using (4.3) we need to estimate the weights W. In the dataset we only had baseline covariates and therefore to compute the weight W, we considered two logistic regression models to estimate the probability of observing T, the first being the probability of observing T in phase I conditional on baseline covariates (case 1, Figure 4.1) and the other being the probability of observing T in phase II conditional on baseline covariates (case 4, Figure 4.1). The weight W is thus given by

$$W = \frac{I(\text{case} = 1)}{P(\text{case} = 1|V)} + \frac{I(\text{case} = 4)}{P(\text{case} = 4|V)}$$

As a form of sensitivity analysis in the estimation of $\Lambda(t)$ in (4.2), we considered different models to estimate the probability of observing T and also for estimating the censoring bias function. We fitted the following models:

- Model 1: An unweighted pooled logistic regression model for estimating the censoring bias function without controlling for covariates.
- Model 2: Same as Model 1, but controlling for covariates V
- Model 3: A weighted pooled logistic regression model for estimating the censoring bias function without controlling for covariates, with weight models for W that did not control for V.
- Model 4: Same as Model 3, but with a censoring bias function that controls for covariates V.
- Model 5: Same as Model 3, but with weight model for W that controls for covariates V.
- Model 6: Same as Model 4, but with weight model for W that controls for covariates V.

To summarize, these models are the various combinations formed when the censoring bias function and the weight function do and do not include V.

The estimated α parameter and associated estimated mortality rate at 1 year corresponding to each of the above specifictions are shown in Table 4.5. The (stabilized) weights W for the model

with covariates ranged from 0.39 to 7.79 with 25th, 50th and 75th percentiles 0.82, 0.99 and 1.09 respectively. The Hosmer and Lemeshow goodness of fit was used to assess model adequacy for each of the models, and we used the test to select subsets of covariates where necessary to ensure adequate fit to the data.

To correctly account for uncertainty in the estimation of the confidence intervals for one-year mortality, we used bootstrap re-sampling with 500 replicates. We report 95% confidence intervals based on the lower and upper 2.5 percentiles from the ordered 500 replicates.

4.5.3 Results

We report results for Models 2 and 6, using data on the 773 subjects who had T observed. We fitted a pooled logistic regression model to estimate the parameters in the censoring model. Table 4.1 and Table 4.2 show the unadjusted and weighted estimates of α. The weights were obtained by estimating for each subject the probability of observing their failure time T controlling for the following covariates in V: baseline CD4 count, WHO stage, gender and clinic type (urban vs rural). The model was unstable when we allowed separate intercepts by time, so we used a common intercept. The same covariates were used for the censoring model as for the weight models. From Table 4.2 we observe that $\widehat{\alpha}$ is negative which indicates that individuals dying later are less likely to be censored between u and $u+1$ compared to those dying sooner.

Table 4.3 shows the estimated mortality rates at 1 year based on the different methods. The results for the naive method ignoring LTFU are based on using the data for the 8977 under the assumption that there was no double sampling. The naive combined is based on including the deaths from double sampling but making no statistical adjustments. The FR method uses data for the 5455 non drop outs and the 621 double sampled to compute the survival function. Our approach uses double sampling data to estimate the censoring process as shown in Table 4.2 and then estimate the survival function using data for those that had T observed in phase I. From the Table we observe that there is an increase in mortality rate with the naive method, giving a mortality rate of 1.7% and the FR method a mortality of 9.7%. Our approach gives a mortality rate of 18%. Our results were fairly comparable when we included a weight for the probability of observing T conditional on the observed covariates and when we did not. Both lead to an estimated mortality rate of 16.6%

at one year. As shown in Table 4.5 the estimated coefficient for the censoring bias function were comparable for the different model specifications.

4.6 Discussion

The results suggest that methods that ignore data from double sampling yield lower mortality estimates, as noted by An et al. (2008) and Frangakis & Rubin (2001). This is expected if much sicker patients are more likely to drop out compared to the healthier ones. From our method we observe that early drop outs are more likely to be deaths compared to later ones, based on the estimated censoring bias parameter α. We observe that our method leads to a higher mortality estimate compared to the other methods. One key difference between our approach and the FR approach is that we make use of time varying weights whereas the FR method considers only the sampling fraction. This can lead to different results because, as shown in Figure 4.3, the drop out times are different among the double sampled compared to the non double sampled (median drop out time 133 vs. 83). The double sampled individuals therefore might not be a random sample of the drop outs.

Our approach can be easily ported to other informative censoring problem where there is information on the covariates considered in our analysis but no double sampling data. We can also construct priors for the censoring bias function for other studies using the estimated censoring bias function.

A significant limitation to our approach is its dependence on correct model specification for the censoring bias function and for the associated weight models. Estimating mortality under a variety of specifications is intended to serve as a sensitivity analysis, which is highly recommended in practice. In our application, mortality rate seems not to be sensitive to the model specification. Possible extension are incorporating time varying weights in estimating the censoring bias function when we have time varying covariates.

Table 4.1: Regression estimates for the censoring bias function without adjusting for differential probability of observing failure time T.

	Estimate	Std. Error	z value	p-value
(Intercept)	-3.006	0.200	-14.988	$<$ 2e-16
T-(u+1)	-0.352	0.033	-10.601	$<$ 2e-16
WHO stage 3 and 4	0.523	0.179	2.927	0.003
Male	0.320	0.165	1.940	0.052
Urban clinic	0.009	0.176	0.049	0.961
Mean centered CD4	-0.105	0.029	-3.617	0.0002

Table 4.2: Weighted regression estimates for censoring bias function, adjusting for differential probability of observing failure time T.

	Estimate	Std. Error	t value	p-value
(Intercept)	-2.740	0.188	-14.589	$<$2e-16
T-(u+1)	-0.289	0.026	-10.890	$<$2e-16
WHO stage 3 and 4	0.237	0.162	1.462	0.1438
Male	0.018	0.162	0.114	0.9091
Urban clinic	0.071	0.180	0.395	0.6927
Mean centered CD4	0.03128	0.018	1.726	0.0844

Table 4.3: Estimated mortality rates at 1 year from enrollment and 95% bootstrap confidence intervals based on different methods of estimating hazard function.

Method	Mortality Rate
Naive (Ignore LTFU)	1.6 (1.3, 2.0)
Naive (Combine phase I and II)	2.7 (2.4, 3.1)
Frangakis and Rubin	9.7 (8.2, 11.2)
Estimating censoring function	16.6 (14.3, 18.8)

Table 4.4: Estimates for logistic regression models for constructing weights W; standard errors are listed in parentheses.

	$P(\text{case} = 1\|V)$	$P(\text{case} = 4\|V)$
Intercept	-2.50 (0.09)	-4.37 (0.19)
WHO stage 3 or 4	-0.04 (0.09)	0.51 (0.17)
Male	-0.19 (0.05)	0.16 (0.16)
Urban clinic	-0.17 (0.09)	-0.17 (0.03)
Mean centered CD4	-0.03 (0.01)	-0.14(0.03)

Table 4.5: Sensitivity analysis: estimated selection bias coefficient α and mortality rates at 1 year from enrollment using different models for estimating the censoring function.

Method	$\widehat{\alpha}$ (s.e.)	Mortality Rate (%)
No covariates in estimating $K(u)$		
Model 1	-0.36 (0.03)	17
Model 3	-0.38 (0.04)	19
Model 5	-0.29 (0.04)	17
Incorporating covariates in estimating $K(u)$		
Model 2	-0.35 (0.03)	17
Model 4	-0.37 (0.04)	17
Model 6	-0.29(0.03)	16

float placement: (h)ere, page (t)op, page (b)ottom, other (p)age

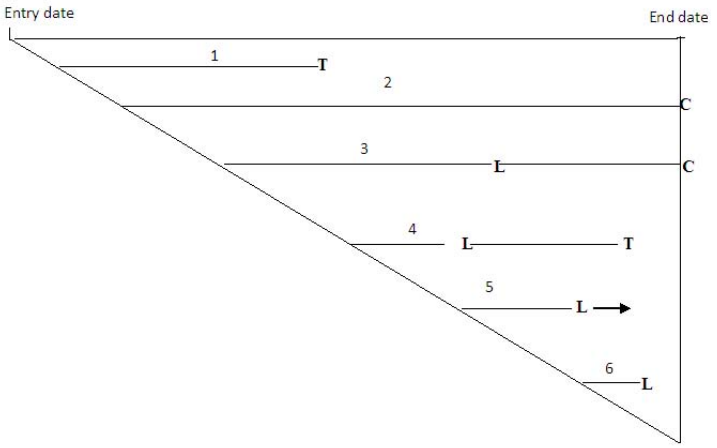


Figure 4.1: Double sampling design T is death time, L drop out time and C is administrative censoring time

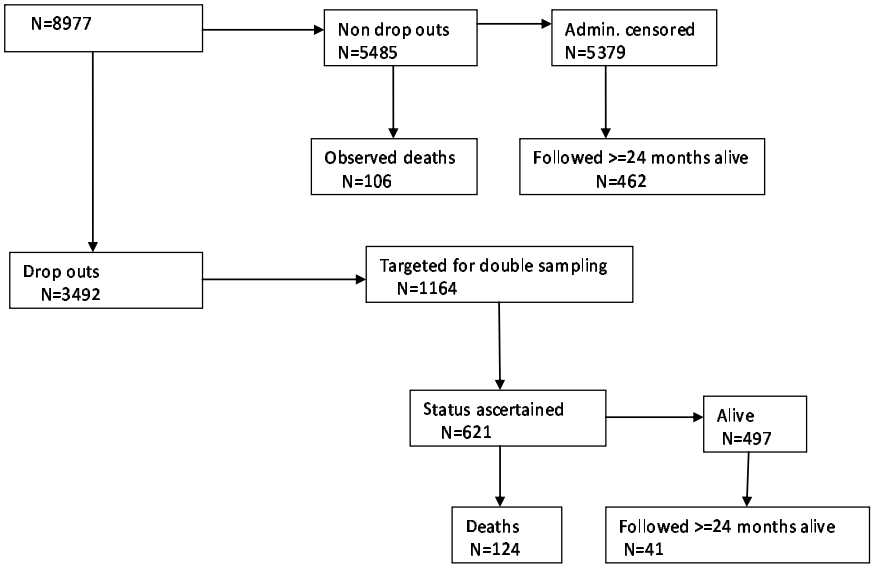


Figure 4.2: Flow chart of AMPATH double sampling study

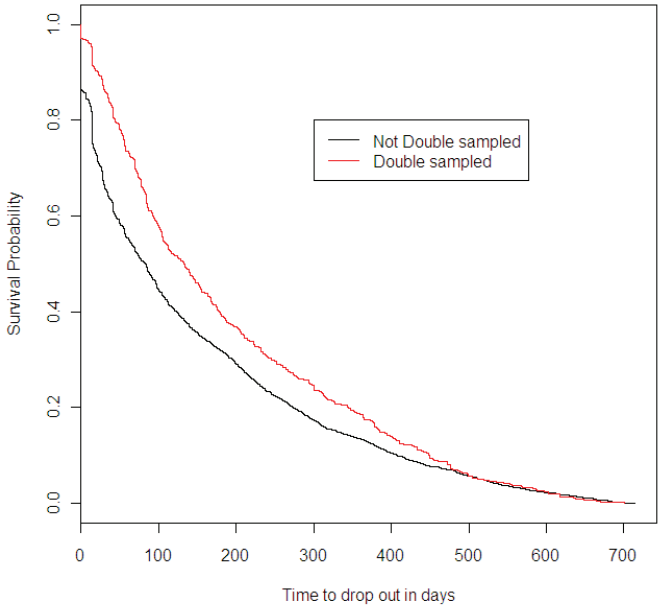


Figure 4.3: Observed survival curves among drop outs, stratified by double sampling status

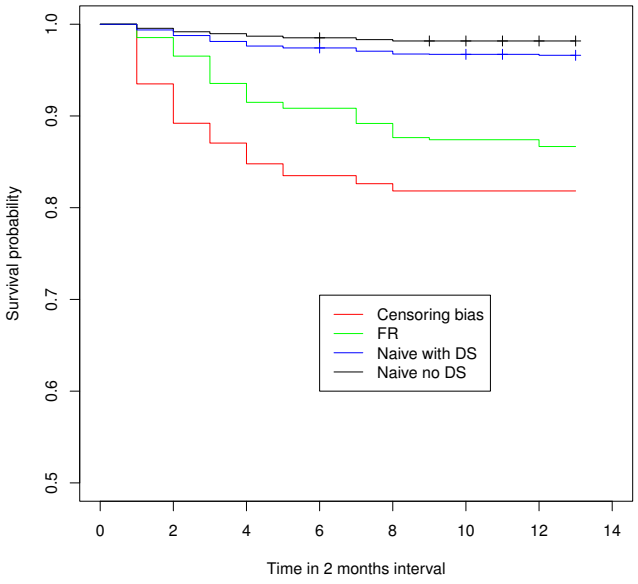


Figure 4.4: Estimated survival curves based on the different methods described in section 3 and 4

Bibliography

AN, M.-W., FRANGAKIS, C. E., MUSICK, B. S. & YIANNOUTSOS, C. T. (2008). The Need for Double-Sampling Designs in Survival Studies: An Application to monitor PEPFAR. *Biometrics* **65**, 301–306.

ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Annals of Statistics* **10**, 1100–1120.

BALETA, A. (2008). Trial finds simultaneous HIV/tuberculosis treatment beneficial. *Lancet Infectious Diseases* **8**, 669–669.

BIJWAARD, G. E. & RIDDER, G. (2005). Correcting for selective compliance in a re-employment bonus experiment. *Journal of Econometrics* **125**, 77–111.

BLANC, F.-X., HAVLIR, D. V., ONYEBUJOH, P. C., THIM, S., GOLDFELD, A. E. & DELFRAISSY, J.-F. (2007). Treatment strategies for HIV-infected patients with tuberculosis: ongoing and planned clinical trials. *Journal of Infectious Disease* **196 Suppl 1**, S46–S51.

BRAITSTEIN, P., AYUO, P., MWANGI, A., WOOLS-KALOUSTIAN, K., MUSICK, B., SIIKA, A. & KIMAIYO, S. N. (2010). Sustainability of First-Line Antiretroviral Regimens: Findings From a Large HIV Treatment Program in Western Kenya. *Journal of Acquired Immune Deficiency Syndrome* **53**, 254–259.

BRAITSTEIN, P., BRINKHOF, M. W. G., DABIS, F., SCHECHTER, M., BOULLE, A., MIOTTI, P., WOOD, R., LAURENT, C., SPRINZ, E., SEYLER, C., BANGSBERG, D. R., BALESTRE, E., STERNE, J. A., MAY, M., EGGER, M., IN LOWER INCOME COUNTRIES (ART-LINC) COLLABORATION, A. T. & GROUPS, A. C. C. (2006). Mortality of HIV-1-infected patients in the

first year of antiretroviral therapy: comparison between low- income and high-income countries. *Lancet* **367**, 817–824.

BRAITSTEIN, P., NYANDIKO, W. M., VREEMAN, R. C., WOOLS-KALOUSTIAN, K., SANG, E., S., M. B., SIDLE, J., YIANNOUTSOS, C. T., AYAYA, S. & CARTER, E. (2009). The clinical burden of tuberculosis among HIV-infected children in western Kenya and the impact of combination antiretroviral treatment. *Pediatric Infectious Disease* **28**, 1–7.

BURMAN, W. (2005). Issues in the management of HIV-related tuberculosis. *Clinics in Chest Medicine* **26**, 283–294.

CAIN, L. E., ROBINS, J. M., LANOY, E., LOGAN, R., COSTAGLIOLA, D. & HERNAN, M. A. (2010). When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *The International Journal of Biostatistics* **6**, Article 18, Issue 2.

CDC (2009). Guidelines for prevention and treatment of opportunistic infections in HIV-infected adults and adolescents. Tech. rep., Center for Disease Control.

COOK, R. & LAWLESS, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer, New York.

COOK, R., LAWLESS, J., LAKHAL-CHAIEB, L. & LEE, K. (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: Application of skeletal complications in cancer metastatic to bone. *Journal of the American Statistical Association* **104**, 60–75.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

D'AGOSTINO, R. B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* **17**, 2265–2281.

EINTERZ, R. M., KIMAIYO, S., MENGECH, H. N. K., KHWA-OTSYULA, B. O., ESAMAI, F., QUIGLEY, F. & MAMLIN, J. J. (2007). Responding to the HIV pandemic: The power of an academic medical partnership. *Academic Medicine* **82**, 812–818.

FORSTER, M., BAILEY, C., BRINKHOF, M. W. G., GRABER, C., BOULLE, A., SPOHR, M., BALESTRE, E., MAY, M., KEISER, O., JAHN, A., EGGER, M., ART-LINC & INTERNATIONAL EPIDEMIOLOGICAL DATABASES TO EVALUATE AIDS (2008). Electronic medical record systems, data quality and loss to follow-up: Survey of antiretroviral therapy programmes in resource-limited settings. *Bulletin of the World Health Organization* **86**, 939–947.

FRANGAKIS, C. & RUBIN, D. (2001). Addressing an idiosyncrasy in estimating survival curves using double sampling in the presence of self-selected right censoring. *Biometrics* **57**, 333–342.

GHOSH, D. & LIN, D. (2002). Marginal Regression Models for Recurrent and Terminal Events. *Statistica Sinica* **12**, 663–688.

GLYNN, R. J., LAIRD, N. M. & RUBIN, D. B. (1993). Multiple imputation in mixture models for nonignorable nonresponse with follow-ups. *Journal of the American Statistical Association* **88**, 984–983.

GOLUB, J. E., DUROVNI, B., KING, B. S., CAVALACANTE, S. C., PACHECO, A. G., MOULTON, L. H., MOORE, R. D., CHAISSON, R. E. & SARACENI, V. (2008). Recurrent tuberculosis in HIV-infected patients in Rio de Janeiro, Brazil. *AIDS* **22**, 2527–2533.

HARRIES, A. D., ZACHARIAH, R. & LAWN, S. D. (2009). Providing HIV care for co-infected tuberculosis patients:a perspective from sub-Saharan Africa. *International Journal of Tuberculosis and Lung Disease* **13**, 6–16.

HERNAN, M. A., BRUMBACK, B. & ROBINS, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* **11**, 561–570.

HERNAN, M. A., BRUMBACK, B. & ROBINS, J. M. (2001). Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments. *Journal of the American Statistical Association* **96**, 440–448.

HERNAN, M. A., LANOY, E., COSTAGLIOLA, D. & ROBINS, J. M. (2006). Comparison of dynamic treatment regimes via inverse probability weighting. *Basic & Clinical Pharmacology & Toxicology* **98**, 237–242.

HOGAN, J. W. & LANCASTER, T. (2004). Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. *Statistical Methods in Medical Research* **13**, 17–48.

KARIM, S., NAIDOO, K., GROBLER, A., PADAYATCHI, N., NAIR, G., BAMBER, S., PIENAAR, J., FRIEDLAND, G., EL-SADR, W. & KARIM, Q. A. (2009). Initiating ART during TB treatment significantly increases survival: Results of a randomized controlled clinical trial in TB/HIV co-infected patients in South Africa. In *16th Conference on Retroviruses and Opportunistic Infections. Montreal, Canada*.

KLEIN, J. P. & MOESCHENBERGER, M. L. (2003). *Survival Analysis: Techniques for censored and truncated Data*. Springer.

KURLAND, B. F., JOHNSON, L. L., EGLESTON, B. L. & DIEHR, P. H. (2009). Longitudinal data with follow-up truncated by death: Match the analysis method to research aims. *Statistical Science* **24**, 211–222.

LI, Q. & LAGAKOS, S. (1997). Use of the Wei-Lin Weissfeld method for the analysis of a recurring and a terminating event. *Statistics in Medicine* **16**, 925–940.

MARTIN WG BRINKHOF, FRANÇOIS DABIS, L. M. D. R. B. A. B. D. N. M. S. C. L. O. K. M. M. E. S. M. E. X. A. F. T. A.-L. O. I. C. (July 2008). Early loss of HIV-infected patients on potent antiretroviral therapy programmes in lower-income countries. *Bulletin of the World Health Organization* **Volume 86, Number 7**, 497–576.

MASKEW, M., MACPHAIL, P. & C MENEZES, D. R. (2007). Lost to follow up contributing factors and challenges in South African patients on antiretroviral therapy. *South African Medical Journal* **97(9)**, 853–857.

MELLORS, J. W., MUÑOZ, A., GIORGI, J. V., MARGOLICK, J. B., TASSONI, C. J., GUPTA, P., KINGSLEY, L. A., TODD, J. A., SAAH, A. J., DETELS, R., PHAIR, J. P. & RINALDO, C. R. (1997). Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Annals of Internal Medicine* **126**, 946–954.

MORGAN, D., MAHE, C., MAYANJA, B., OKONGO, J. M., LUBEGA, R. & WHITWORTH, J. A. (2002). HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries. *AIDS* **16**, 597–603.

MUKHUNA, B., MCGUIRE, M., TAREMBA, K., YASIN, K., PAIH, M. L., HOLMGREN, C., SZUMILIN, E., SAID, A., GOUMEIZEL, S., HEINZELMANN, A., MBALE, E. & PUJADES-RODRÍGUEZ, M. (2009). Outcomes of patient active tracing by lay community workers in a decentralized HIV care program of rural Malawi. In *5th IAS Conference on HIV Pathogenesis and Treatment*.

MWANGI, A., HOGAN, J., CARTER, E. J., KANTOR, R. & SIIKA, A. (2010). Causal Inference for Survival Times with Informative Censoring and Missing Exposure, with Application to Treatment of TB/HIV-co-infection in Western Kenya. Thesis Paper 1.

NYANDIKO, W. M., AYAYA, S., NABAKWE, E., TENGE, C., SIDLE, J. E., YIANNOUTSOS, C. T., MUSICK, B. S., WOOLS-KALOUSTIAN, K. & TIERNEY, W. M. (2006). Outcomes of HIV-Infected Orphaned and Non-Orphaned Children on Antiretroviral Therapy in Western Kenya. *Journal of Acquired Immune Deficiency Syndrome* **43**, 418–425.

OCHIENG-OOKO, V., OCHIENG, D., SIDLE, J. E., HOLDSWORTH, M., WOOLS-KALOUSTIAN, K., SIIKA, A. M., YIANNOUTSOS, C. T., OWITI, M., KIMAIYO, S. & BRAITSTEIN, P. (2010). Influence of gender on loss to follow-up in a large HIV treatment programme in western Kenya. *Bulletin of the World Health Organization* **88**, 681–688.

PALELLA, F. J., KM, D., AC, M., MO, L., J, F., GA, S., DJ, A. & SD., H. (1998). Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV outpatient study investigators. *New England Medical Journal* **338**, 853–860.

PRENTICE, R. L., WILLIAMS, B. J. & PETERSON, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **18**, 373–379.

ROBERTSON, J., MEIER, M., WALL, J., YING, J. & FICHTENBAUM, C. J. (2006). Immune reconstitution syndrome in HIV: Validating a case definition and identifying clinical predictors in persons initiating antiretroviral therapy. *Clinical Infectious Diseases* **42**, 1639–1646.

ROBINS, J., ORELLANA, L. & ROTNITZKY, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. *Statistics in Medicine* **27**, 4678–4721.

ROBINS, J., ROTNITZKY, A. & BONETTI, M. (2001). Discussion of the Frangakis and Rubin Article. *Biometrics* **57**, 343–347.

ROBINS, J. M. (1999a). Association, causation and marginal structural models. *Synthese* **121**, 151–179.

ROBINS, J. M. (1999b). Marginal structural models versus structural nested models as tools for causal inference. *Statistical Models in Epidemiology: The Environment and Clinical Trials* , 95–134.

ROBINS, J. M. & FINKELSTEIN, D. M. (2000). Correcting for noncompliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. *Biometrics* **56**, 779–788.

ROBINS, J. M., HERNÁN, M. A. & BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560.

ROSEN, S., FOX, M. & GILL, C. (2007). Patient retention in antiretroviral therapy programs in sub-Saharan Africa: a systematic review. *PLoS Medicine* **4**, e298.

ROSENBAUM, P. & RUBIN, D. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* **79**, 516–524.

SCHARFSTEIN, D., ROBINS, J. M., EDDINGS, W. & ROTNITZKY, A. (2001). Inference in randomized studies with informative censoring and discrete time-to-event endpoints. *Biometrics* **57**, 404–413.

SHORTREED, S. M. & FORBES, A. B. (2010). Missing data in the exposure of interest and marginal structural models: A simulation study based on the Framingham Heart Study. *Statistics in Medicine* **29**, 431–443.

SIIKA, A. M., AYUO, P. O., MWANGI, A. W., SIDLE, J. E., WOOLS-KALOUSTIAN, K., KIMAIYO, S. N. & TIERNEY, W. M. (2008). Admission characteristics, diagnoses and outcomes

of HIV-infected patients registered in an ambulatory HIV-care programme in western Kenya. *East African Medical Journal* **85**, 523–528.

SMITH, D. K., WARREN, D. L., VLAHOV, D., SCHUMAN, P., STEIN, M. D., GREENBERG, B. L. & HOLMBERG, S. D. (1997). Design and baseline participant characteristics of the Human Immunodeficiency Virus Epidemiology Research (HER) Study: a prospective cohort study of human immunodeficiency virus infection in US women. *American Journal of Epidemiology* **146**, 459–469.

TIERNEY, W. M., ROTICH, J. K., HANNAN, T. J., SIIKA, A. M., BIONDICH, P. G., MAMLIN, B. W., NYANDIKO, W. M., KIMAIYO, S., WOOLS-KALOUSTIAN, K., SIDLE, J. E., SIMIYU, C., KIGOTHO, E., MUSICK, B., MAMLIN, J. J. & EINTERZ, R. M. (2007). The AMPATH medical record system: creating, implementing, and sustaining an electronic medical record system to support HIV/AIDS care in western Kenya. *Study Health Technology Informatics* **129**, 372–376.

UNAIDS/WHO (2009). Aids epidemic update 2009:nov 2009. Tech. rep., UNWHO.

VELASCO, M., CASTILLA, V., SANZ, J., GASPAR, G., CONDES, E., BARROS, C., CERVERO, M., TORRES, R., GUIJARRO, C. & THE COMESUM COHORT (2009). 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.

WEI, L. J., LIN, D. Y. & WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065–1073.

WHO (2006a). Antretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach (2006 revision). Tech. rep., WHO.

WHO (2006b). Guidelines on co-trimoxazole prophylaxis for HIV-related infections among children, adolescents and adults. Tech. rep., WHO.

WHO (2007). Tuberculosis care with TB-HIV co-management. Tech. rep., WHO.

WHO (2009). HIV-related tb deaths higher than past estimates. Tech. rep., WHO.

WOOLS-KALOUSTIAN, K., KIMAIYO, S., DIERO, L., SIIKA, A., SIDLE, J., YIANNOUTSOS, C. T., MUSICK, B., EINTERZ, R., FIFE, K. H. & TIERNEY, W. M. (2006). Viability and effectiveness of large-scale HIV treatment initiatives in sub-Saharan Africa: experience from western Kenya. *AIDS* **20**, 41–48.

YIANNOUTSOS, C. T., AN, M.-W., FRANGAKIS, C. E., MUSICK, B. S., BRAITSTEIN, P., WOOLS-KALOUSTIAN, K., OCHIENG, D., MARTIN, J. N., BACON, M. C., OCHIENG, V. & KIMAIYO, S. (2008). Sampling Based Approaches to Improve Estimation of Mortality among Patient Dropouts: Experience from a Large PEPFAR- Funded Program in Western Kenya. *PLos ONE* **3(12)**.