Adjusting the effect of integrating antiretroviral therapy and tuberculosis treatment on mortality for non-compliance: an instrumental variables analysis using a time-varying exposure

by

Fortunate Nonhlanhla Yende-Zuma

Submitted in partial fulfillment for the degree of

Doctor of Philosophy

in the Discipline of Statistics

School of Mathematics, Statistics and Computer Science

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

2018

Contents

Preface

The research contained in this thesis was completed by the candidate while based in the Discipline of Statistics, School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa. The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

Nonhlanhla Yende-Zuma Signed:  Date:13/03/2019

Supervisors:

Prof. Henry Mwambi Signed:  Date:14/03/2019

Prof. Stijn Vansteelandt Signed:_____ Date:_____

Declaration 1: Plagiarism

I, Nonhlanhla Yende-Zuma, declare that:

i. The research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

ii. This dissertation has not been submitted in full or in part for any degree or examination to any other university;

iii. This dissertation does not contain other persons data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

iv. This dissertation does not contain other persons writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

   (a) their words have been re-written, but the general information attributed to them has been referenced;

   (b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

v. Where I have used material for which publications followed, I have indicated in detail my role in the work;

vi. This dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

vii. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

Signed: _____ Date:_____

Declaration 2: Publications and Awards

Publication 1

This manuscript did not focus on instrumental variables analysis but it was published during my PhD study period. The dataset used for this manuscript is the same as the one used for this thesis.

Yende-Zuma, N. and Naidoo, K. (2016). The effect of timing of initiation of antiretroviral therapy on loss to follow-up in HIV-tuberculosis coinfected patients in South Africa: An open-label, randomised, controlled trial, *Journal of Acquired Immune Deficiency Syndromes* **72**(4):430-436.

Publication 2

This manuscript is the product of the current research. In Appendix 1, we provided the copy of the published manuscript.

Yende-Zuma, N., Mwambi, H. and Vansteelandt, S. (2019) Adjusting the effect of integrating antiretroviral therapy and tuberculosis treatment on mortality for noncompliance: a time-varying instrumental variables analysis, *Epidemiology* **30**(2):197-203

Oral presentations

1. Adjusting the effect of integrating antiretroviral therapy and tuberculosis treatment on mortality for non-compliance: an instrumental variables analysis.
   South African Statistical Association (SASA) conference, Cape Town, South Africa, 2016.

2. Adjusting the effect of integrating antiretroviral therapy and tuberculosis treatment on mortality for non-compliance: an instrumental variables analysis.
   International Society for Clinical Biostatistics (ISCB) conference, Vigo, Spain, 2017.

3. Adjusting the effect of integrating antiretroviral therapy and tuberculosis treatment on mortality for non-compliance: an instrumental variables analysis
   University of KwaZulu-Natal Statistics Research Seminar, Pietermaritzburg, South Africa, 2018.

Awards

1. 2017- International Society for Clinical Biostatistics (ISCB) conference award for scientist from developing countries.

Signed: _____ Date:_____

Abstract

In South Africa and elsewhere, research has shown that the integration of antiretroviral therapy (ART) and tuberculosis (TB) treatment saves lives. The randomised controlled trials (RCTs) which provided this compelling evidence used intent-to-treat (ITT) strategy as part of their primary analysis. As much as ITT is protected against selection bias caused by both measured and unmeasured confounders, but it is capable of drawing results towards the null and underestimate the effectiveness of treatment if there is too much non-compliance. To adjust for non-compliance, "as-treated" and "per-protocol" comparisons are commonly made. These contrast study participants according to their received treatment, regardless of the treatment arm to which they were assigned, or limit the analysis to participants who followed the protocol. Such analyses are generally biased because the subgroups which they compare often lack comparability.

In view of the shortcomings of the "as-treated" and "per-protocol" analyses, our objective was to account for non-compliance by using instrumental variables (IV) analysis to estimate the effect of ART initiation during TB treatment on mortality. Furthermore, to capture the full complexity of compliance behaviour outside the TB treatment duration, we developed a novel IV-methodology for a time-varying measure of compliance to ART. This is an important contribution to the IV literature since IV-methodology for the effect of a time-varying exposure on a time-to-event endpoint is currently lacking. In RCTs, IV analysis enable us to make use of the comparability offered by randomisation and thereby have the capability of adjusting for unmeasured and measured confounders; they have the further advantage of yielding results that are less sensitive to random measurement error in the exposure.

In order to carry out IV analysis, one needs to identify a variable called an instrument, which needs to satisfy three important assumptions. To apply the IV methodology, we used data from Starting Antiretroviral Therapy at Three Points in Tuberculosis (SAPiT) trial which was conducted by the Centre for the AIDS Programme of Research in South Africa. This trial enrolled HIV and TB co-infected patients who were assigned to start ART either early or late during TB treatment or after TB treatment completion. The results from IV analysis demonstrate that survival benefit of fully integrating TB treatment and ART is even higher than what has been reported in the ITT analysis since non-compliance has been accounted for.

Acknowledgements

This work would not have been possible without the assistance and guidance from my supervisors Professor, Henry Mwambi and Professor Stijn Vansteelandt. To Professor Vansteelandt, no amount of words can describe how grateful I am for your support and the time you spent guiding me through this journey. You agreed to be my co-supervisor even before meeting with me. Thank you for your unending patience and invaluable advice. I learnt a lot from you, amongst other things the importance of paying attention to finer details. To Professor Mwambi, thank you for allowing me to spread my wings and invite Professor Vansteelandt to be part of this project. Your guidance, generosity, words of encouragement and support is greatly appreciated.

To each and everyone at CAPRISA who worked on the SAPiT trial, the scientific leaders who formulated the research idea that resulted in the conduct of this clinical trial, to participants without whom none of this would have been possible, thank you very much.

I am immensely grateful for the funding that I received from different agencies in particular the support of the DST-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA). Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to SACEMA.

I am also grateful to the Department of Science and Technology (DST)-National Research Foundation (NRF) and TATA Africa for uplifting women in research.

This work was supported by grant number: 5R24TW008863 from the Office of Global AIDS Coordinator and the U. S. Department of Health and Human Services, National Institutes of Health (NIH OAR and NIH OWAR). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the government.

Finally, I am indebted to my children, Lungile and Minenhle whom I at times neglected while I was busy with this project.

List of Figures

List of Tables

Abbreviations

2SLS Two-stage Least Squares

2SPS Two-stage Predictor Substitution

2SRI Two-stage Residuals Inclusion

ARV Antiretrovirals

ART Antiretroviral therapy

ATE Average Treatment Effect

ATET Average Treatment Effect on the Treated

CI Confidence Interval

HIV Human Immunodeficiency Virus

HR Hazard Ratio

ITT Intent-to-Treat

IV Instrumental Variables

LATE Local Average Treatment Effect

LTFU Loss to Follow-up

PCZCDC Prince Cyril Zulu Communicable Disease Centre

RCT Randomised Controlled Trial

SAPiT Starting Antiretroviral Therapy at Three Points in Tuberculosis

SMC Data and Safety Monitoring Committee

TB Tuberculosis

WHO World Health Organization

Chapter 1

Introduction

1.1 Overview

Globally, there were an estimated 1.2 million people co-infected with tuberculosis (TB) and human immunodeficiency virus (HIV) in 2014 and around 74% of these people live in sub-Saharan Africa (World Health Organization, 2015). South Africa is one of the countries with the highest TB burden, and according to World Health Organization (2016b) there was an estimated incidence of 454,000 cases of active TB in 2015. Among these incident TB cases, 57% are also infected with HIV and 85% of those who are HIV infected are on antiretroviral therapy (ART) (World Health Organization, 2016b). In some parts of South Africa, it was estimated that almost 70% of TB patients are co-infected with HIV (Perumal et al., 2014). In 2012, approximately 6.4 million people (12.2%) were HIV positive in South Africa, with the province of KwaZulu-Natal having the highest HIV prevalence sitting at 16.2% (Shisana et al., 2012). South Africa has the largest ART roll-out programme with approximately just over 3 million people on ART in 2015 (Department of Health, 2015).

Despite this biggest ART-roll out and wide availability of TB treatment, TB and HIV are reported to be the leading causes of death in South Africa in age-groups 15 to 44 years (Statistics South Africa, 2014). In South Africa, ART used to be deferred until TB treatment completion

among patients with CD4+ cell count > 200 cells/mm^3 (Department of Health, 2004). Most low and middle-income countries were also following the same strategy. These guidelines were not based on clinical trials evidence and therefore Abdool Karim et al. (2010) conducted the Starting Antiretroviral Therapy at Three Points in Tuberculosis (SAPiT), an open label randomised controlled trial (RCT) to determine the optimal time to initiate ART in patients with HIV and TB co-infection who were receiving TB therapy. Results from the SAPiT trial showed that the integration of TB and ART reduced mortality by 56%. The integration may be at the same time or soon after the initiation of TB treatment. These results were adopted as part of clinical guidelines and treatment policy for HIV and TB co-infected patients (Department of Health, 2010, 2014, 2015, 2016; World Health Organization, 2009, 2016a).

Elsewhere, it was indeed confirmed that initiation of ART earlier or later during TB treatment improves survival (Blanc et al., 2011) more especially among patients with low CD4+ cell count (Abdool Karim et al., 2011; Havlir et al., 2011). Moreover, others showed that early initiation of ART in patients with high CD4+ cell count is beneficial over starting when CD4+ cell count reaches a certain threshold (Lundgren et al., 2015). The benefit of ART during TB treatment has not only been shown for drug susceptible TB but also for patients with multi-drug resistant tuberculosis (Padayatchi et al., 2014).

1.2 Research problem and objectives

The SAPiT trial and other open label RCTs provided remarkable evidence that the integration of TB treatment and ART reduces mortality (Abdool Karim et al., 2010, 2011; Blanc et al., 2011; Havlir et al., 2011). To preserve randomisation and obtain valid estimate, the ITT principle was used for their primary analysis. It is common knowledge that an ITT analysis provides estimates for the effect of treatment assignment rather than the effect of actually taking treatment. Furthermore, the analysis by treatment received suffer from selection bias where patients who take or comply with the assigned treatment versus those who do not may have different characteristics and therefore not be exchangeable.

ITT analysis are relevant for public policy because they provide realistic information about the effectiveness of implementing the intervention, keeping in mind that not everyone will adhere to it. However, such results might not be generalised to the whole population as trial participants are a selected group that met the trial's inclusion criteria. Robins and Tsiatis (1991) argued that once the treatment is proved to be efficacious in a trial, then the treatment uptake and compliance in the community might be higher. Therefore, the ITT effect may not represent the overall effect of treatment in the community.

As much as ITT is protected against selection bias caused by both measured and unmeasured confounders, it may shrink results towards the null and underestimate the effectiveness of treatment if there is too much non-compliance, especially in the active arm. This can potentially result in a patient who would like to comply with the treatment in the future doubting its benefit. According to Glymour (2006), if the ITT effect estimate is statistically significant, that indicates the existence of a causal relationship between the outcome and treatment, however it does not indicate the magnitude of the relationship.

Over and above the issue of non-compliance, there are other factors that may shrink results towards the null. For example, in an open label trial, it is possible that participants randomized to a control arm (which can be placebo) or treatment that is perceived to be inferior, may initiate effective treatment outside the study. When that alternative treatment is as effective as the intervention being studied (assuming everyone was adherent), then results will shrink towards the null. Even when participants comply with the treatment strategy they were assigned to (i.e. they are compliant), but if adherence is sub-optimal, then ITT results will be drawn towards the null. For example, in HIV prevention research, there are therapeutic levels that drugs need to reach in order to be protective against HIV acquisition. In these cases, participants can be compliant by taking the treatment they were assigned to, but if they do not fully adhere to the relevant dosing strategies and thus do not reach protective therapeutic levels, then ITT results can also be pushed towards the null.

To adjust for non-compliance in RCTs, "as-treated"and "per-protocol"comparisons are com-

monly made. These contrast study patients according to their received treatment, regardless of the treatment arm to which they were assigned, or limit the analysis to patients who followed the protocol. Such analyses are generally biased because the subgroups which they compare often lack comparability. However, in instances where non-compliance is unassociated with any measured or unmeasured confounders or prognostic factors, the "as-treated" analysis produces valid causal effects. In a scenario where non-compliance is associated with measured confounders, adjusting for them in a regression analyses will not render valid causal effects because that model will not be adjusted for unmeasured confounders which might not be balanced between the study groups. So far, instrumental variables is the only method that promises to adjust for both measured and unmeasured confounders.

The "as-treated" and "per-protocol" analyses are also likely to be underpowered because they involve either less people or fewer time-points than the ITT analysis. In the "per-protocol" analysis, patients who violated important protocol procedures get excluded from the analyses and this automatically reduces the sample size and may subsequently reduce the statistical power (the likelihood of showing an effect if it exists). However, when the variability is lower among the remaining patients who did not violate the protocol, there might be gains in the power largely due to a potential increase in effect size. Moreover, in the "per-protocol" analysis, time-points for data collected after protocol violations can be excluded resulting in fewer time-points being analysed than what was anticipated. In the "as-treated" analysis, participants are grouped according to their compliance status and depending on the level of non-compliance, that allocation can result in large imbalances between the groups. The imbalances can also arise in the "per-protocol" analysis since the patients who followed the protocol may also fail to be comparable between groups. Generally, deviations from equal allocation slightly reduces power. Also, heterogeneity between compliers and noncompliers in outcome distributions affects power (Jo, 2002).

In the SAPiT trial, 642 patients with TB and HIV were randomised to initiate ART either early or late during TB treatment or after the completion of TB treatment. Among the 362 patients who were randomised to start ART either early or late during TB treatment, 22.4% did not

initiate ART at the correct time (Abdool Karim et al., 2011). Moreover, following recommendations by the data and safety monitoring committee (SMC), some patients started ART earlier than the protocol specified time (Abdool Karim et al., 2010) and in this current project are regarded as non-compliant. More details about the SAPiT trial are given in Section 3.1.

In view of the shortcomings of the "as-treated" and "per-protocol" analyses, the objective of this study is to account for non-compliance in the SAPiT trial by using instrumental variables (IV) analysis to estimate the effect of ART initiation during TB treatment (exposure) on mortality. In the current analysis, non-compliance is defined as not starting ART at the correct time with respect to TB treatment, regardless of whether that was enforced by clinicians or by patients themselves. However, this definition of compliance does not take into account any temporary or permanent discontinuation of study drugs as well as adherence. In particular, we assumed that patients who were dispensed ART either early or late during TB treatment did not delay taking it until TB treatment completion.

Most IV analyses concentrate on baseline exposures or treatment, but the reality is that in RCTs and other biomedical studies, exposure to treatment changes over time. That part of research on longitudinal exposures in IV analysis is still largely lacking. Therefore, our second objective is to contribute to the IV literature by developing novel IV-methodology for time-varying measure of compliance to ART and assess its effect on mortality. IV analysis enable us to make use of the comparability offered by randomisation and thereby have the capability of adjusting for unmeasured and measured confounders; they have the further advantage of yielding results that are less sensitive to random measurement error in the exposure. Compared to ITT analysis which provide estimates of assigning patients to integrated treatment strategy, IV results produce estimates for the effect of receipt and complying to assigned treatment. As a result, IV analysis appeal to patients and clinicians who are always interested in the benefits of receiving treatment because they can provide an estimate of how receiving the assigned treatment affects outcome. Furthermore, IV analysis have been proposed as a technique to control for bias when using "as-treated" analysis.

1.3 Outline of the thesis

The rest of the thesis is organised as follows:

Chapter 2 introduces the concept of IV analysis and discusses how the IV methods works. We discuss different RCTs and other study designs where IV analyses were used. Most of the work done on the IV methods relates to continuous and binary outcomes. Even though remarkable work has been done on time-to-event outcomes, the field is still being developed. Therefore, in this chapter we also discuss some of the IV work done on time-to-event outcomes which is the outcome of interest in our analysis. We also give guidance on how to report results from the IV analysis. Lastly, we discuss the limitations and disadvantages of using the IV method.

Chapter 3 introduces the dataset from the SAPiT trial which will be used for application of the IV analysis. The trial enrolled patients who were co-infected with TB and HIV. The instrumental variable which is critical to this work, exposure variables (fixed and time-varying) and measured covariates that will be used for modelling purposes in different statistical models are explained in detail.

Chapter 4 presents background results from the SAPiT trial. Results of primary and secondary outcomes of the SAPiT trial are published in different journals (Abdool Karim et al., 2010, 2011; Naidoo et al., 2014, 2012). However, data relating to these outcomes were censored at 18 months of follow-up. In this thesis we report complete data up to 24 months of follow-up. Among others, we report baseline characteristics, mortality rates, longitudinal CD4+ count and viral load trajectories stratified by randomisation arm.

Chapter 5 outlines the additive and proportional hazards models used in the ITT analysis and their application using data from the SAPiT trial. Results from this Chapter will be compared with results from the IV analysis in Chapters 6 and 7.

In Chapter 6 we introduce the concept of two-stage modelling, namely two-stage predictor substitution (2SPS) and two-stage residuals inclusion (2SRI) methods, which forms the basis of most of our IV analysis. We performed the analyses based on additive and hazards regression models using the fixed exposure.

IV-methodology for the effect of a time-varying exposure on a time-to-event endpoint is currently lacking. In Chapter 7 we extend the IV approach introduced in Chapter 6 and develop novel IV-methodology for time-varying exposures. This was developed only for additive hazards model under the 2SPS approach.

In Chapter 8, we summarised results from Chapters 5, 6 and 7 for ease of comparison.

Chapter 9 shows a simulation study that was conducted to assess the validity of the 2SPS estimator developed in Chapter 7. The aim is to assess how much bias, if any, is introduced by the violation of some of the IV assumptions.

Chapter 10 gives the summary of the thesis, concluding remarks and future work.

Chapter 2

Review of instrumental variables (IV) method

2.1 What is an instrumental variable or instrument?

In order to carry out IV analysis, one needs to identify a variable called an instrument which needs to satisfy three important assumptions. Suppose Z, X, Y, C and U represent the instrument, exposure, outcome, measured and unmeasured confounders respectively.

An instrument or instrumental variable is an observed variable which:

1. must be associated with the exposure or treatment strategy being studied (relevance assumption): (i.e. $Z \not\perp\!\!\!\perp X|C$)
2. must have no direct effect on the outcome except through its association with the exposure (exclusion restriction assumption): (i.e. $Z \perp\!\!\!\perp Y|(X, U, C)$)
3. must not share common causes with the outcome as shown in Figure 2.1 (exchangeability assumption): (i.e. $Z \perp\!\!\!\perp U|C$)

In randomised studies, the randomisation arm can be used as an instrument. However, finding a suitable instrument in observational studies can be very challenging because it should mimic

the randomisation arm of an RCT such that all measured and unmeasured confounders should on average be equally distributed among different levels of a categorical instrument, leading to exchangeability.

The second and third assumptions are untestable, however (2) is somehow ensured by double-blind design and (3) by randomisation of the instrument. Even though double blinding can never completely guarantee exclusion restriction, it makes it more plausible because participants are not aware of the treatment arm they are assigned to and that does not enhance their expectation of success or failure. Therefore, participants might be hesitant to access alternative treatment outside the trial which could potentially introduce other pathways with which Z affects Y that are not through X. However, we are not oblivious to the fact that treatments with visible side effects can expose the treatment arm and thus undo the effects of double blinding. When these assumptions are met, the IV analysis can allow one to draw causal conclusions about the effect of exposure on the outcome in the presence of unmeasured confounders, provided that a monotonicity assumption, or alternatively, a treatment homogeneity assumption is fulfilled.

The monotonicity assumption means that increasing the level of the instrument cannot lead to a decrease in the exposure for some individuals and an increase for others. Glymour (2006) provides this example to illustrate monotonicity "In an RCT, the monotonicity assumption implies that there is nobody who would have refused the treatment if assigned to treatment but would have (perversely) sought out the treatment if assigned to control". On the other hand, treatment homogeneity assumption means that the effect of exposure on the outcome should be constant or similar across everybody, of which is biologically implausible.

In Figure 2.1, the strength of the correlation or association between X and Z determines whether the instrument is strong or weak. When the correlation is weak the instrument is regarded as weak and this affect assumption (1). The implications of a weak instrument, small sample size and bias produced when IV assumptions are violated are discussed in Section 2.7. In Figure 2.1, U and C represents all unobserved and observed confounders (of X and Y) that affect both the outcome and the decision to adhere to the assigned treatment.

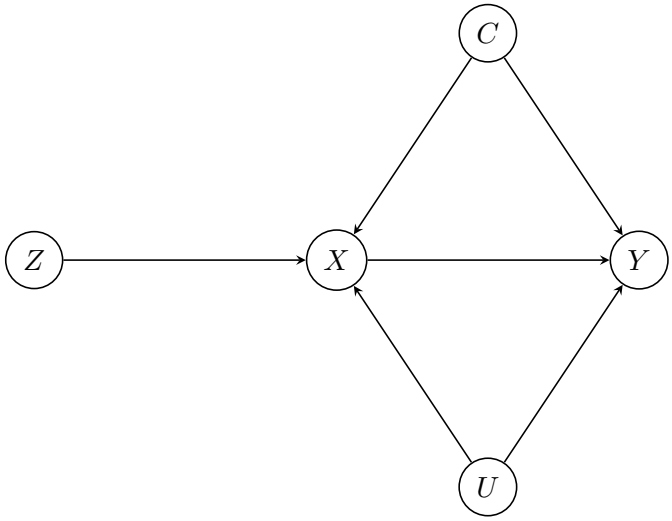


Figure 2.1: Causal diagram for an RCT with assignment Z, treatment X, outcome Y, unmeasured confounders U and measured confounders C

The IV method was developed in econometrics (Goldberger, 1972; Johnson, 1963) and has been increasingly used in medical and health research as well as in epidemiology (Greenland, 2000; Hernan and Robins, 2006; Newhouse and McClellan, 1998; Stukel et al., 2007). In economics and health studies, it has been used to estimate the causal effect of an exposure or treatment on the outcome in the presence of unmeasured confounding. In epidemiology, the IV methodology gained popularity with the special case of Mendelian randomisation (Didelez and Sheehan, 2007; Gray and Wheatley, 1991; Smith and Ebrahim, 2004). Mendelian randomisation studies utilise genetic factors or genotypes as instrument. The IV method also accounts for measurement error in covariates (Carroll and Stefanski, 1994; Newhouse and McClellan, 1998; Sheiner and Rubin, 1995; Stefanski, 1996; Wright, 1928). Even though the IV method has capabilities of identifying causal effects in the presence of unmeasured confounding, this comes at a cost. According to Jackson and Swanson (2015), "the IV methods shift the problem of knowing, measuring, and appropriately adjusting for confounders of the treatment-outcome relationship to confounders of the instrument-outcome relationship (i.e. to satisfy the exchangeability assumption)".

2.2 How does instrumental variable method work?

IV analysis enable us to make use of the comparability offered by randomisation and thereby have the capability of adjusting for unmeasured and measured confounders; they have the further advantage of yielding results that are less sensitive to random measurement error in the exposure (Vansteelandt et al., 2009). In this section we discuss how the IV method works in order to estimate the causal effect of nonrandomised exposure (X) on the outcome (Y), where X can be influenced by both measured and unmeasured confounders as shown in Figure 2.1. Consider using the model

$$Y = \beta_0 + \beta_x X + \varepsilon, \tag{2.1}$$

where X is nonrandomised and therefore correlated with ε (noise). Thus $\hat{\beta_x}$ will not have a causal interpretation. To find the causal estimate, one can use the comparability offered by the instrumental variable (Z). Here, Z and ε are uncorrelated since we are assuming Z is not affected or influenced by measured and unmeasured confounders. To further understand how Z can assist in finding a causal estimate ($\hat{\beta_x}$), we regress X on Z such that

$$X = \alpha_0 + \alpha_z Z + \omega. \tag{2.2}$$

Then we calculate the fitted values $\hat{X}$=$E[X|Z]$=$X - \omega$, where $\hat{X}$ represents variation in X that is independent of the unexplained variation ε given that Z is not correlated with ε. The instrument Z, assists in filtering out or removing the part of X that is correlated with the disturbances. Now $\hat{X}$ represents the values of X that are no longer correlated with both measured and unmeasured confounders, and it can be used in Equation 2.1 to estimate the causal effect of X on Y. This method is referred to as the two-stage least squares and its nonlinear extension will be presented in Chapters 6 and 7 and mainly form basis of the analytical component of this thesis.

When Z and X are both dichotomous variables, the usual IV estimand, also known as the Wald

estimator is given by

$$\frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]}, \tag{2.3}$$

where $Z=1$ and $Z=0$ represents assignment to the intervention and control arm respectively. The numerator of this estimator is the average causal effect of Z on Y which is the ITT effect. The denominator is the average causal effect of Z on X which is the measure of compliance with the assigned treatment. When there is minimal non-compliance, then $E[X|Z=1]-E[X|Z=0]=1$ resulting in the Wald estimator being identical to the ITT estimator.

2.3 Instrumental variable method in randomised controlled trials

In medical research, RCTs remain the gold standard in assessing the safety and effectiveness of new or existing treatment because they provide strong basis for causation. RCTs are the only clinical studies that are capable of removing bias due to confounding because treatment is randomly assigned (Grimes and Schulz, 2002). They provide reliable information about the efficacy of an intervention when the randomisation process is done perfectly. Among other things, the problem with some of the RCTs, specifically open label trials is that of non-compliance. Particularly because everyone who is involved in the trial is unblinded, therefore participants can decide to access the study treatment elsewhere if they feel that they have been assigned to either inferior or harmful treatment. Moreover, in case where the trial enrols sick people, the treating clinician or nurse can switch the patient's treatment any time based on their prognosis. All these issues lead to non-compliance where patients eventually receive the treatment they were not randomly assigned to. Noncompliance is not only an obstacle to fair statistical comparison between the intervention and the control group, but also a major threat to obtaining statistical power to detect a significant difference between intervention and control group, if it exists (Jo, 2002).

In blinded and unblinded RCTs where both the exposure and the instrument are binary, the IV method can be used to control for confounding due to non-compliance and participants can be

grouped into four possible compliance groups. According to Angrist et al. (1996), the groups are: compliers, always-takers, never-takers and defiers. Let us refer to Figure 2.1 and let Z be the randomisation arm, X the treatment received and Y the outcome of interest. We also let $X^{\underline{Z}}$ and $Y^{\underline{X}}$ denote the vector of potential treatments received under randomisation assignments $\underline{Z}$ and the vector of potential outcomes under treatment received $\underline{X}$, respectively.

Then the four groups are given by:

a) Compliers ($X^1 = 1,\ X^0 = 0$), are participants who follow the treatment strategy they are assigned to and would receive intervention only when assigned to.

b) Always-takers ($X^1 = 1,\ X^0 = 1$), are participants who always take the intervention even when assigned to the control treatment.

c) Never-takers ($X^1 = 0,\ X^0 = 0$), are participants who would not take the intervention even when assigned to it.

d) Defiers ($X^1 = 0,\ X^0 = 1$), are the participants who would take the opposite of what they are assigned to.

Always-takers, never-takers and defiers can be further grouped into non-compliers. Compliance groups can only be formed when treatment received is a dichotomous variable with the generalisation to continuous treatment somewhat artificial. Conducting a double blinded RCT removes the possibility of defiance because participants do not know whether they are getting intervention or control treatment.

In RCTs, whether open label or double blinded, the randomisation arm is mostly likely to serve as an instrument for the treatment received and this phenomenon provides causal estimates had every patient complied (Greenland, 2000). However, one has to keep in mind that the exclusion restriction assumption may be vulnerable especially in open label trials due to the fact that the knowledge of the randomisation arm may raise or dampen patient's expectations of success or failure. In RCT context, IV analysis can provide an estimate of how those who received treatment were affected by it, keeping in mind that the instrument has to satisfy three important assumptions described in Section 2.1, with the addition of monotonicity and treatment

homogeneity assumptions. It is worth mentioning that monotonicity and treatment homogeneity assumptions identifies the local average treatment effect (LATE) and the average treatment effect (ATE) estimands, respectively. Under the potential outcomes framework, we formally define the ATE and LATE estimands as $E[Y^1 - Y^0]$ and $E[Y^1 - Y^0|X^1 - X^0 = 1]$ respectively. ATE represents to the average casual effect in the entire population or community while LATE measures treatment effect among the subgroup of compliers.

In this thesis, we will estimate the effect of treatment among patients who received treatment and thus our estimand of interest is known as the average treatment effect on the treated (ATET) formally defined as $E[Y^1 - Y^0|X = 1]$. It must be however highlighted that in RCTs, researchers often use ITT analysis for their primary analysis and seldom study the effect of the treatment that participants actually received. Throughout the thesis, the notation in Figure 2.1 will be used in different statistical models. This figure shows the IV framework in an RCT context. When the level of non-compliance is high, X may not be the same as Z and therefore studying the causal effect of X on Y without the use of Z will produce biased estimates as participants who chose to comply are different from non-compliers.

2.4 Some of the instruments used in healthcare research

In this section, some of the instruments that have been used in observational studies and also in RCTs pertaining to healthcare research will be discussed. The aim is to give readers an idea of what variables can be potentially used as instruments. As mentioned before, the most important aspect of the IV analysis is finding a valid instrument of high quality. Fortunately, in the current research we have a valid instrument since we are analysing data from an RCT, where randomisation arm will be used as an instrument. In practice, more especially in observational studies, the strong IV assumptions make it difficult to find a powerful instrument that mimics randomisation.

Sexton and Hebel (1984) designed a randomised encouragement clinical trial to study the effect of smoking on infant's birth weight. One group of women were assigned to an encouragement to stop smoking and others were not encouraged. Permutt and Hebel (1989) re-analysed that data using the IV method as opposed to ITT and the instrumental variable was the randomisation procedure. The treatment received was the number of cigarettes smoked per day. Since randomisation took place, it was assumed that the intervention has no direct effect on birth weight other than through the number of cigarettes smoked per day.

McClellan et al. (1994) studied the effect of more intensive treatments on mortality in elderly patients with acute myocardial infarction (AMI). They used distance to alternative types of hospitals as an instrument because differential distances roughly randomise patients to different likelihoods of receiving intensive treatments. The assumption was that distance to hospital is associated with receiving care as AMI patients should be taken to the nearest hospital due to the seriousness of their critical condition. Patients' differential distances to alternative types of hospitals seemed like a strong independent predictor of how an AMI patient will be treated and appears to be uncorrelated with their health status, thus satisfying the IV assumptions.

Angrist et al. (1996) evaluated the effect of serving in the military on health outcomes. In their study, random assignment was done through draft lottery. Those with low lottery numbers were assigned to serve in military. Whereas, those with high lottery numbers did not have to serve in the military. The instrument was randomised. The authors thought it was reasonable to assume that the draft lottery numbers have no direct effect on health outcome except through veteran status.

Korn and Baumrind (1998) utilized the treating clinician as an instrument because each clinician had preference for the choice of treatment that was given to patients. Such instruments are called preference-based instruments. However, in this particular study, authors were sceptical of using the IV approach because some of the assumptions in their analysis were not satisfied. The proposed instrument was later used by Brookhart et al. (2006), to estimate the effect of

COX-2 inhibitors on gastrointestinal toxicity, where they used a time-varying estimate of a physician's relative preference for a COX-2 inhibitor relative to a non-selective non-steroidal anti-inflammatory drugs (NSAIDs) on the effect of gastrointestinal complications. If the last prescription written by patient's physician was a COX-2 inhibitor then the clinician is classified as a "COX-2 prescriber", otherwise classified as "non-selective NSAID prescriber".

Austin et al. (2016) used an IV method to determine whether racial residential segregation lead to lower birth weight. Railroad division index, which measures the extent to which a metropolitan area was divided into subplots by railroad tracks was used an instrument. This division made it easy to segregate blacks into racially homogeneous enclaves. This instrument has been used before in economics research (Ananat, 2011). Austin et al. (2016) used metropolitan statistical area level segregation, quantified via the 2000 dissimilarity index as an exposure variable.

A summary of different types of instrumental variables that have been used in clinical research can be found in the systematic review which also discusses issues in the reporting and conduct of instrumental variables analysis (Davies et al., 2013).

2.5 Instrumental variable method in survival outcomes

In countries with high burden of HIV and TB, most RCTs that enrol HIV positive patients, with or without TB, use time to death or time to AIDS defining illness as the main study outcome, whereas, in HIV prevention trials time to HIV infection is usually the primary outcome. These outcomes are preferred because they are objectively measured and they are clinically meaningful. In this thesis, we will use the application of IV methods for time-to-event outcomes. However, one drawback which complicates the application of IV analysis in time-to-event outcomes is the issue of right censoring due loss to follow-up, voluntary withdrawal and many others. Right censoring also occurs when the study ends before a participant experience an event of interest (i.e. administrative censoring).

The Cox proportional hazards model is widely used and most popular regression in biomedical research. However, the IV assumptions in time-to-event outcomes can be met at the beginning of follow-up but can possibly be violated within risk sets as censoring will force some patients out of future risk sets. For these reasons, the Aalen's additive hazards model (Aalen, 1989) is a preferred regression in the context of IV analysis (Tchetgen Tchetgen et al., 2015) because its mathematical form resembles that of linear models. Compared to the hazard ratios, another useful property of additive hazards model is that the hazard difference is a collapsible effect measure, such that adjusting for a variable that is neither associated with the exposure nor the outcome will not change the magnitude of the hazard difference.

Compared to the two-stage least squares method, here the second stage regression is carried out using additive hazards regression. IV methods are well developed for continuous (Wooldridge, 2002; Wright, 1928) and to some extent binary outcomes (Robins and Rotnitzky, 2004; Vansteelandt et al., 2011). Robins and Tsiatis (1991) were the first to introduce IV analysis in the context of time-to-event outcomes. They developed G-estimation methods under a class of structural accelerated failure time models. However, application of these methods in applied research have been relatively low because of their complexity and often poor performance in the presence of censoring.

Even though the field is still being developed, a lot of researchers have produced interesting work, for example (Bijwaard and Ridder, 2005; Carslake et al., 2013; Gore et al., 2010; Hadley et al., 2010; Li et al., 2015; Martinussen et al., 2017a,b; Stukel et al., 2007; Tan et al., 2012; Tchetgen Tchetgen et al., 2015). In Section 2.1, we discussed three important assumptions that the instrumental variable needs to satisfy to produce valid estimates. When the analysis involves time-to-event outcomes, a fourth assumption needs to be added where it is assumed that censoring is independent of failure processes (i.e. censoring is non-informative). Informative censoring occurs when drop-out (censored) subjects are either more or less likely to experience the event of interest than remaining individuals in the future (Collett, 1994). In simple terms, informative censoring takes place when the reasons for censoring are related to the outcome of the study. In this IV analysis, we further assume that censoring is independent of failure processes given the

instrument, exposure and measured confounders.

Next, we highlight some of the work that has been done to estimate the causal effect of treatment on time-to-event outcome. As mentioned already, Robins and Tsiatis (1991) were the first to introduce IV methodology in time-to-event outcomes. They developed rank-preserving structural failure time models (RPSFTM) to estimate the effect of treatment on survival outcome using the G-estimation method. Robins and Tsiatis (1991) regarded RPSFTM as the strong version of accelerate failure time models. The limitation about their method was that there should be no censoring prior to the end of follow-up and no other missing data. This is an impossible requirement when dealing with longitudinal data.

Stukel et al. (2007) used IV analysis to determine the causal effect of invasive cardiac treatment on long-term mortality. They used regional cardiac catheterization rate as an instrument because prognostic factors related to mortality were similar across regions that have different cardiac catheterization rates. Mortality was considered as a binary variable and multiple linear regression was used. They acknowledged that their relative mortality rates, were comparable but not identical to those from proportional hazards models.

Similar to McClellan et al. (1994) who used the distance as an instrumental variable, Gore et al. (2010) also used distance from the center residence zip code of each study subject to the center of the zip code of the nearest cystectomy provider as an instrument. Their objective was to estimate the effect of undergoing radical cystectomy for bladder cancer versus receiving chemotherapy, radiation or surveillance on survival. The two-stage residual inclusion method was used where proportional hazards models were used in the second stage to study the effect of undergoing radical cystectomy for bladder cancer on survival. The two-stage residual inclusion method (Terza et al., 2008) is discussed further in Chapter 6.

Carslake et al. (2013) used a two-sample IV method to estimate the effect of father's height on mortality by using the son's height as an instrument. This two-sample method was initially

developed by Angrist and Krueger (1992). This approach makes maximum use of the data available on fathers mortality (even when the fathers height is not available, but his son is). Proportional hazards models were used in the second stage model to estimate father's all-cause and cause-specific mortality.

Tan et al. (2012) used differential distance to a partial nephrectomy provider as an instrumental variable to determine long-term survival among patients treated with partial versus radical nephrectomy. They calculated distance from the patients' residence to the nearest provider performing at least one partial nephrectomy in the year of treatment minus the distance from the patients' residence to the nearest surgeon performing any kidney cancer surgery. Using a two-stage residual inclusion approach, they calculated hazard ratios using the Weibull distribution to estimate the effect of treatment on survival. Models were adjusted for patient-level covariates, surgical approach, and post-operative complications.

Hadley et al. (2010) used IV methods to estimate the causal effect of radical prostatectomy versus conservative management on both prostate cancer-specific and all-cause mortality among patients with early-stage prostate cancer. They selected the lagged (i.e., previous years) local area treatment pattern for conservative management as the primary instrumental variable, because it varied substantially across geographic areas. It satisfied the key plausibility criterion of being independent of a current patients' health and other characteristics because it reflects provider treatment decisions from a previous time period (i.e., the year before the patient was diagnosed). Proportional hazards models with two-stage residual inclusion method was used.

Tchetgen Tchetgen et al. (2015) developed an IV approach for time-to-event outcomes under the additive hazards model where two-stage regression was used for continuous exposure or endogenous variables; that is, variables that share common causes with the outcomes as shown in Figure 2.1. They also developed a control-function approach for binary exposure. The control-function approach can also be seen as an extension of the two-stage residual inclusion approach of Terza et al. (2008). The control-function approach works like the two-stage least squares

method, however in the second stage of the control-function method, the outcome is regressed on the treatment and the residuals of the first stage regression. With this work, they confirmed the validity of the two-stage modelling under additive hazard models.

We concur with MacKenzie et al. (2014) in saying that a lot of statistical analysis in medical research uses proportional hazards models but most IV approaches are not designed for this regression technique. Therefore, MacKenzie et al. (2014) proposed a causal estimator within the proportional hazards model framework. They assumed that any omitted confounders have an additive rather than multiplicative effect on the baseline hazard. Simulation was used to justify the use of this estimator under the proportional hazards models. However, Tchetgen Tchetgen et al. (2015) have reservations about the validity of their estimator.

Li et al. (2015) developed a closed-form, two-stage estimator under the additive hazards models. Compared to Tchetgen Tchetgen et al. (2015), their estimator is restrictive and only suitable for both continuous and discrete exposures, but binary endogenous variables were not considered. Recently, Martinussen et al. (2017b) developed IV estimator under semiparametric structural cumulative survival models. Their models are closely related to but less restrictive than those of Tchetgen Tchetgen et al. (2015). They can handle arbitrary exposures and instrumental variables, and can accommodate adjustment for baseline covariates. Their approach does not require modelling the exposure distribution nor the association between covariates and outcome, and it deals with administrative censoring and certain forms of dependent censoring. However, it requires a correct model for the conditional mean of the instrumental variable, given covariates, for consistency of the estimated causal effect.

2.6 How to report instrumental variable analysis

Davies et al. (2013), Swanson and Hernan (2013) and Baiocchi et al. (2014) provided useful guidelines on how to report results from the IV analysis. In summary, Baiocchi et al. (2014) gave these six sub-topics on how to report IV analysis.

- Describe the theoretical basis for the choice of an instrument: The most convincing instrumental variable comes from a randomised experiment or clinical trial. One has to discuss why the IV is expected to be associated with treatment, and whether it is independent of unmeasured confounders, and does not have a direct effect on the outcome other than through its effect on the treatment.

- Report the strength of an instrument by showing a partial F-statistic obtained from regressing treatment received (X) on the instrument (Z). Also, the proportion of compliers and measured covariates should be reported. Other measures such as the concordance (C-statistic) or risk difference of exposure by level of instrument can be reported.

- Report the distribution of measured covariates across levels of the instrument and received treatment: Ideally an instrument should be unrelated to measured covariates.

- Explore concomitant treatments: One has to check whether the instrument is associated with concomitant medications other than those in the study. Possible violations of the exclusion restriction can be assessed by examining whether the instrument is associated with concomitant treatments. If the instrument is associated with concomitant treatments resulting in the violation of exclusion restriction assumption, then the IV approach may be biased for the effect of the exposure, especially if the concomitant treatments affect the outcome.

- Discuss the interpretation of the treatment effect estimated by the IV method.

- Report a sensitivity analysis: When the IV assumptions do not hold, one has to report a sensitivity analysis to show how sensitive inferences are to various violations of the IV assumption.

Baiocchi et al. (2014) gave a stern warning that newcomers to the IV methods may think that the validity of the IV can be easily tested by regressing the outcome on the treatment received and the instrument, and then testing whether the coefficient of the instrument is significantly different from zero.

2.7 Limitations and disadvantages of instrumental variable method

According to Greenland (2000), the major limitation for the IV method is that it relies mostly on the three assumptions as discussed in Section 2.1. Two of these three assumptions are not verifiable and as a result cannot always be tested. Weak correlation or association between the instrument and exposure leads to less precise or biased estimates with little power (Bound et al., 1995). When the instrument is weak enough (i.e. partial first-stage F statistics less than 10), it is important to consider the use of other causal inference methods. This issue is escalated when there are more than one weak instruments because that can increase finite sample bias in an IV estimator. The combination of weak correlation between Z and X and smaller sample size can be a threat to the validity of the IV method (Martens et al., 2006) and can have sizeable bias.

Chapter 3

Methods

3.1 Dataset used in the thesis

We used datasets from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) for the application of the IV method. CAPRISA conducted the Starting Antiretroviral Therapy at Three Points in Tuberculosis (SAPiT) study, an open-label, three armed randomised, controlled trial between 28 June 2005 and 04 July 2010. This study was conducted in the province of KwaZulu-Natal which had and still has the highest HIV prevalence in South Africa; 16.2% in 2012 (Shisana et al., 2012) and 27.0% in 2017 (Human Sciences Research Council, 2018). The trial was designed to determine the optimal time to initiate ART in patients with HIV and TB co-infection who were receiving TB therapy (Abdool Karim et al., 2010, 2011). Each patient was expected to be followed up for a maximum of 2 years. Prior to this study, there was no formal guidance on the timing of ART during TB treatment in South Africa. Ambulatory male and female patients aged 18 years or older, were enrolled. Pulmonary TB was confirmed by acid fast bacilli smear positivity. HIV infection was confirmed by two rapid HIV tests. All patients with a screening CD4+ count <500 cells/mm^3 were initiated on a standard TB treatment regimen.

The study was conducted at the CAPRISA's eThekwini Clinical Research site located adjacent

to the largest out-patient TB facility in South Africa, the Prince Cyril Zulu Communicable Disease Centre (PCZCDC) in the city of Durban. At the PCZCDC, TB patients were treated with a fixed drug combination of rifampicin, isoniazid, ethambutol and pyrazinamide for 2 months (intensive phase) with subsequent fixed-drug combination of isoniazid and rifampicin for 4 months (continuation phase). Patients with re-treatment TB received a 60-day intensive phase which included streptomycin, followed by a 100-day continuation phase (Department of Health, 2004). All patients received a standard package of care which included adherence counselling and cotrimoxazole prophylaxis. After providing written informed consent, patients were randomly assigned in a 1:1:1 ratio (with the use of sealed opaque envelopes) to one of the three study groups in permuted random blocks of six and nine with no stratification. Randomisation took place at the initiation of TB treatment. A total of 642 patients were randomised to initiate ART in the following three different points of their TB therapy:

- In the first group, antiretroviral therapy was to be initiated within 4 weeks after the start of TB therapy (early integrated arm).
- In the second group, antiretroviral therapy was to be initiated within 4 weeks after the completion of the intensive phase of TB therapy (i.e. within 3 months after TB treatment initiation) (late integrated arm).
- In the third group, antiretroviral therapy was to be initiated within 4 weeks after the completion of TB therapy (i.e. within 6-7 months after TB treatment initiation) (sequential arm).

Patients were initiated on a once daily antiretroviral therapy regimen that contained enteric-coated didanosine (250mg if weight <60kg and 400mg if weight $\geq$ 60kg), lamivudine (300mg) and efavirenz (600mg). Notwithstanding their study group assignment, patients could be initiated on antiretroviral therapy at any time at the discretion of the study clinicians or their personal physician.

After a planned interim analysis, on 01 September 2008, almost two months after completion of enrolment, the data and safety monitoring committee (SMC) made a recommendation (which

was not pre-specified in the protocol) that all patients in the sequential arm be initiated on ART as soon as possible but continue in follow-up until study completion. The SMC also recommended continuation of the early and late integrated arms without any alterations. The patients in the sequential arm were contacted within a week after the SMC meeting, and almost all those that were still retained started ART within a month. All patients had demographic, behavioural, physical, clinical and laboratory assessments done. Following randomisation, patients returned monthly for a period of 24 months for pill collection, physical and clinical examination. CD4+ cell counts, HIV-1 RNA viral load and laboratory assessments for safety were conducted at six monthly intervals or any time if clinically indicated.

The primary outcome was all-cause mortality. Information on deaths was ascertained through hospital chart notes, death certificates and oral reports from family or relatives. Primary cause of death was unknown for some patients. Patients who missed four consecutive visits were considered to be loss to follow-up. In the results published previously (Abdool Karim et al., 2010, 2011; Naidoo et al., 2012), administrative censoring took place at 18 months post randomisation but for this analysis we used data up to 24 months post randomisation.

The study was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal (E101/05) and the Medicines Control Council of South Africa (20060157).

3.2 Instrument

We used randomisation arm, denoted by Z as an instrument, which was modelled categorically (using dummy coding). We assume that patients in the three study arms are exchangeable, which is guaranteed by randomisation. However, we acknowledge that since our instrument comes from an open-label RCT, it might, to some extent, violate the exclusion restriction assumption.

3.3 Fixed exposure

SAPiT trial was mainly about TB and ART integration, however the study did not have an objective measure of compliance involving ART drug levels. The fixed value denoted by X, was defined as the fraction of time on ART during TB treatment (i.e. months on ART/months on TB therapy). We felt that this exposure measures non-compliance due to patients not starting ART at the correct time with respect to the TB treatment. All patients started TB treatment which is not usually a problem among co-infected patients because they know that its duration is approximately 6 months and cure or recovery is guaranteed given high adherence. This exposure is only defined during TB treatment period regardless of how long patients were on TB treatment for (as some took more than the expected 6 months to complete TB treatment).

We did not define compliance as a binary variable based on whether ART was initiated at the correct time with respect to TB treatment or not, because that would have left out patients who did not start ART. Currently, patients who were terminated before initiating ART were assigned an exposure of zero. The rationale for choosing this exposure is that in South Africa, TB and HIV are the leading causes of death in adults (Statistics South Africa, 2014), and TB-HIV co-infected patients are supposed to be co-treated by the same healthcare worker for both diseases. In a study conducted in Durban, they found that almost 70% of TB patients were co-infected with HIV (Perumal et al., 2014). However, in resource poor countries like South Africa, the integration has not been fully implemented. Among other things, one constraint is the low uptake of HIV testing among TB patients which deprives patients of the treatment integration (World Health Organization, 2015). For these reasons, this research focuses on strengthening the evidence of the benefit of integrated therapy on survival. The application of IV method using this exposure is shown in Chapter 6

3.4 Time-varying exposure

Among others, the limitation of using a fixed exposure over time, is that it cannot capture the full complexity of compliance behaviour. The occurrence of death, moreover influences the

magnitude of the exposure, in the sense that high compliance is more easily achieved amongst early deaths, which increases the opportunity for reverse causality. In view of this, we have moreover developed IV-methodology for a time-varying measure of compliance $X(t)$, which we define as 1 at time t when the considered patient was on ART at or prior to time t, and 0 otherwise. The time-varying exposure measures how much ART patients were exposed to during the entire study participation. Mathematical formulation and results pertaining to this exposure are shown in Chapter 7.

3.5 Outcome

We will focus on all-cause mortality. The survival time was defined as years from randomisation until the date of death. It was censored at the withdrawal date, last visit date for those who were loss to follow-up (LTFU) or date of the 24-month visit for those who completed the study.

3.6 Covariates

The association between X, the fraction of time on ART during TB treatment, and all-cause mortality is confounded because patients in poorer conditions (e.g. with lower CD4+ count) were at higher risk of death and thus more likely to initiate ART early irrespective of the study arm. Although the considered IV analysis do not require adjustment for measured confounders, we considered adjustment for CD4+ cell count, gender (0=male; 1=female) and employment status (0=unemployed; 1=employed), denoted by vector $\mathbf{C}$, to improve precision.

Statistical analyses were done using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) and `TIMEREG` package in R version 3.2.4.

Chapter 4

Background results

In this Chapter we provide SAPiT background results so that readers can have a better understanding of patient disposition and characteristics before we move to IV analysis. Among the 642 patients enrolled in the SAPiT trial, 214 (33.3%) were in the early integrated arm, 215 (33.5%) in the late integrated arm and 213 (33.2%) in the sequential arm. Out of the 642, only 501 (78.0%) were initiated on ART (Figure 4.1). Patient characteristics were similar across the three study arms (Table 4.1). Patients were followed for a median (IQR) time of 24 (16.0 - 24), 24 (10.4 - 24) and 24 (8.3 - 24) months in the early integrated, late integrated and sequential arms respectively. They were initiated on ART at a median time of 8 (IQR: 7 to 14), 85.5 (IQR: 63.5 to 118) and 239 (IQR: 195 to 271) days post randomisation in the early integrated, late integrated and sequential arms respectively.

Table 4.1: Baseline characteristics of patients in the SAPiT trial

Variable	Early integrated (N=214)	Late integrated (N=215)	Sequential (N=213)	p-value[c]
Mean age (SD), years	34.3 (8.0)	34.5 (8.7)	33.9 (8.2)	0.746
Number of males, n (%)	97 (45.3)	112 (52.1)	110 (51.6)	0.296
Employed, n (%)	135 (63.1)	117 (54.4)	117 (54.9)	0.123
Past history of TB, (%)	80 (37.4)	68 (31.6)	66 (31.0)	0.307
Extra pulmonary tuberculosis, n (%)[a]	10 (4.7)	9 (4.2)	9 (4.3)	1.000
WHO stage 4, n (%)	14 (6.5)	11 (5.1)	13 (6.1)	0.797
Median CD4+ count (IQR), cells/mm^3	155 (75-261)	149 (77-244)	140 (69-247)	0.605
Mean $\log_{10}$ viral load (SD), copies/ml[b]	5.0 (0.9)	5.0 (0.9)	5.1 (0.7)	0.250

IQR: Interquartile Range SD: Standard Deviation

[a] 1 patient in the late integrated arm and 3 patients in the sequential arm had missing extra pulmonary and missing data is not included in percentage calculation

[b] Viral load was not available for 16 patients in each of the early integrated and late integrated arms and 12 in the sequential arm

[c] p-values calculated using one-way analysis of variance or Kruskal Wallis test and Fisher's exact test

During follow-up, a total of 69 (10.7%) patients died (early integrated arm (n=17), late integrated arm (n=17) and sequential arm (n=35)); 417 (65.0%) completed the study (early integrated arm (n=151), late integrated arm (n=139) and sequential arm (n=127)). A total of 96 (15.0%) were loss to follow-up (early integrated arm (n=26), late integrated arm (n=36) and sequential arm (n=34)); while 60 (9.3%) either withdrew consent or relocated to other areas ((early integrated arm (n=20), late integrated arm (n=23) and sequential arm (n=17)). In addition, a total of 16 (7.5%), 51 (23.7%) and 74 (34.7%) patients never started ART in the early, late integrated and sequential arms respectively due to reasons such as death, loss to follow-up, relocation and voluntarily withdrawal (Figure 4.1).

Over 984.79 person-years of follow-up, the mortality rates were 4.9 per 100 person-years (py) (95% confidence interval (CI): 2.9, 7.9) in the early integrated arm; 5.2 per 100 py (95% CI: 3.0, 8.2) in the late integrated arm and 11.3 per 100 py (95% CI: 7.9, 15.8) in the sequential arm (Table 4.2). Mortality rates were significantly different across the three study arms at 12, 18 and 24 months of follow-up. Even though the overall mortality rate for early integrated arm is slightly lower than that of late integrated arm, we noted that in the first six months of follow-up,

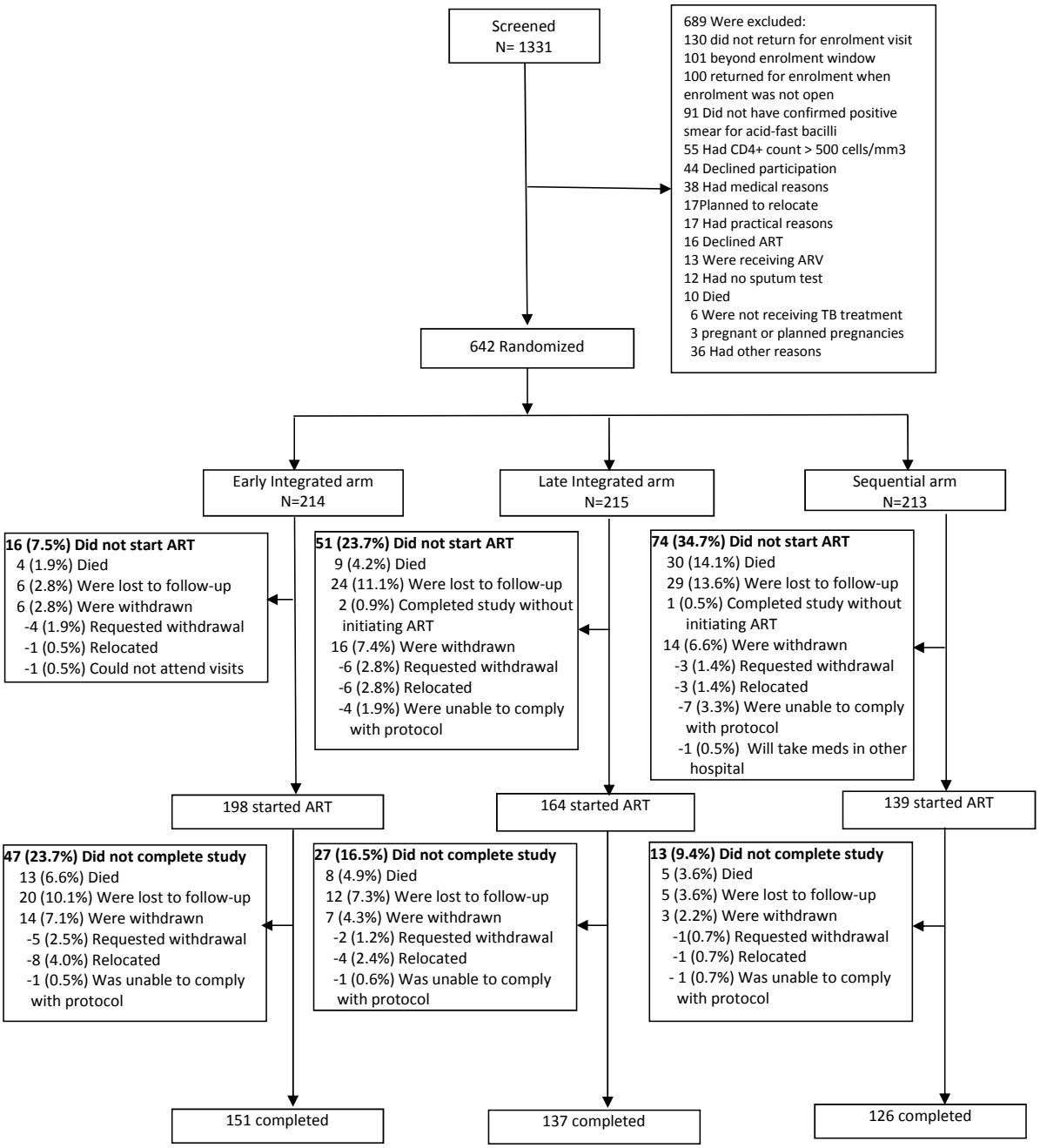


Figure 4.1: Screening, randomisation and follow-up of patients

Table 4.2: Mortality rates per arm over time

Follow-up (months)	Early integrated arm			Late integrated arm			Sequential arm			p-value
	Events	Person-yr	Rate per 100 person-yr (95% CI)	Events	Person-yr	Rate per 100 person-yr (95% CI)	Events	Person-yr	Rate per 100 person-yr (95% CI)	
6	10	98.63	10.11 (4.9, 18.6)	5	97.59	5.12 (1.7, 12.0)	13	98.16	13.2 (7.05, 22.65)	0.1773
12	12	186.80	6.4 (3.3, 11.2)	12	181.15	6.6 (3.4, 11.6)	32	176.66	18.1 (12.4, 25.6)	0.0004
18	15	268.95	5.6 (3.1, 9.2)	15	258.45	5.8 (3.2, 9.6)	35	244.672	14.3 (10.0, 19.9)	0.0008
24	17	345.82	4.9 (2.9, 7.9)	17	330.01	5.2 (3.0, 8.2)	35	308.96	11.3 (7.9, 15.8)	0.0037

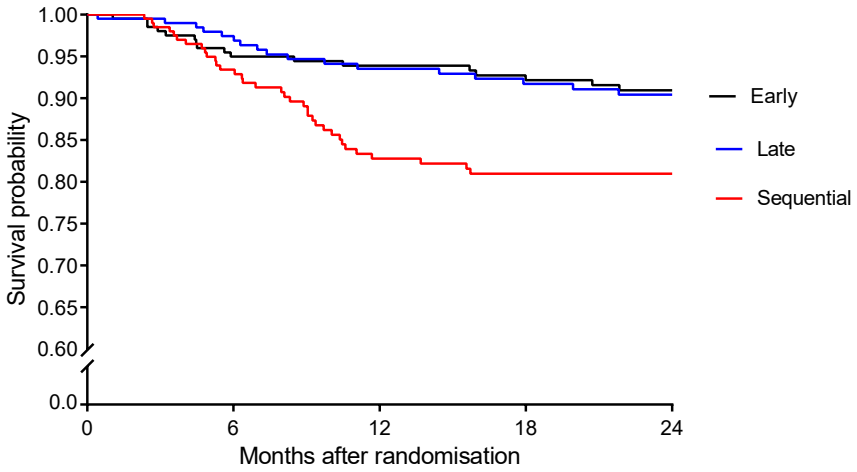
yr: years

mortality rate in the early integrated arm was double that of late integrated arm. Most of the deaths occurred in the first 12 months after randomisation in all three arms (Figure 4.2, Table 4.2). Other than the study arm, gender, employment status and baseline CD4+ count were associated with mortality. Males, unemployed patients and obviously those with low baseline CD4+ count had a significantly higher risk of death (Table 4.3). Other studies in our setting have linked male gender and low CD4+ count with high mortality rates (Cornell et al., 2012; Naidoo et al., 2017).

Table 4.3: Analysis of factors associated with mortality using multivariable proportional hazards regression

Variable	Deaths/ person-years	Mortality rate/100 p-y	95% CI	aHR	95%CI	p-value
Sequential	35/308.96	11.3	7.9, 15.8	ref.		
Early	17/345.82	4.9	2.9, 7.9	0.51	0.28, 0.91	0.024
Late	17/330.01	5.2	3.0, 8.2	0.47	0.26, 0.83	0.010
Male	44/468.11	9.4	6.8, 12.6	ref.		
Female	25/516.68	4.8	3.1, 7.1	0.42	0.25, 0.71	0.001
Unemployed	41/410.87	10.0	7.2, 13.5	ref.		
Employed	28/573.91	4.9	3.2, 7.1	0.39	0.24, 0.64	<0.001
Age (years)				0.97	0.83, 1.12	0.662
CD4+ count (per 50 cells/mm^3 increase)				0.73	0.64, 0.84	<0.001

aHR: adjusted hazard ratio CI: confidence interval p-y: person-years



Number at risk (number died)					
Early integrated	214 (0)	183 (10)	171 (12)	157 (15)	151 (17)
Late integrated	215 (0)	182 (5)	159 (12)	149 (15)	139 (17)
Sequential	213 (0)	179 (13)	143 (32)	132 (35)	127 (35)

Figure 4.2: Kaplan-Meier curve for survival

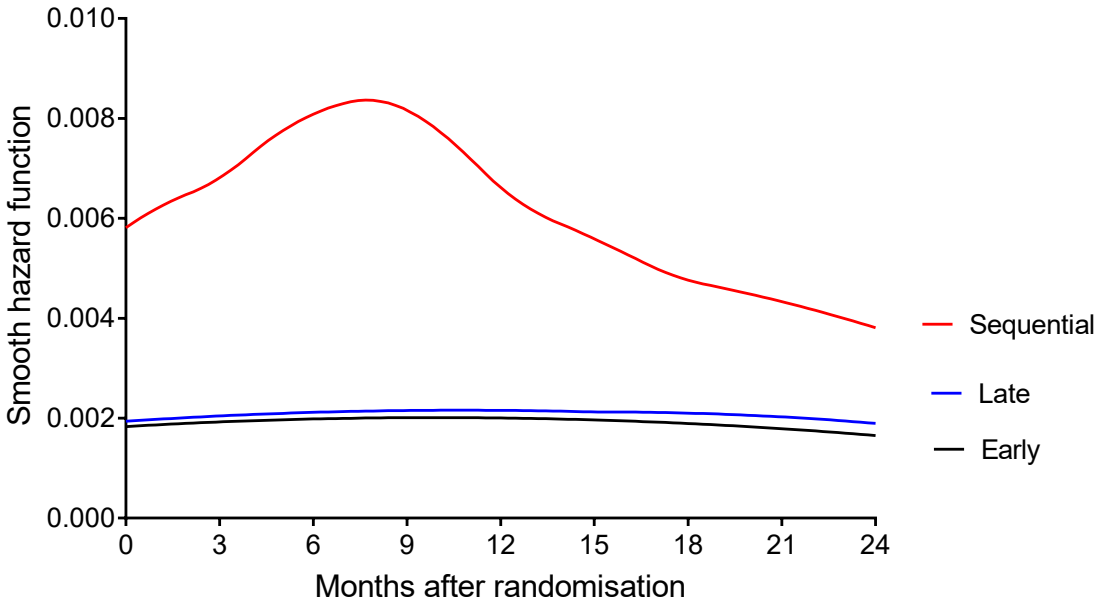
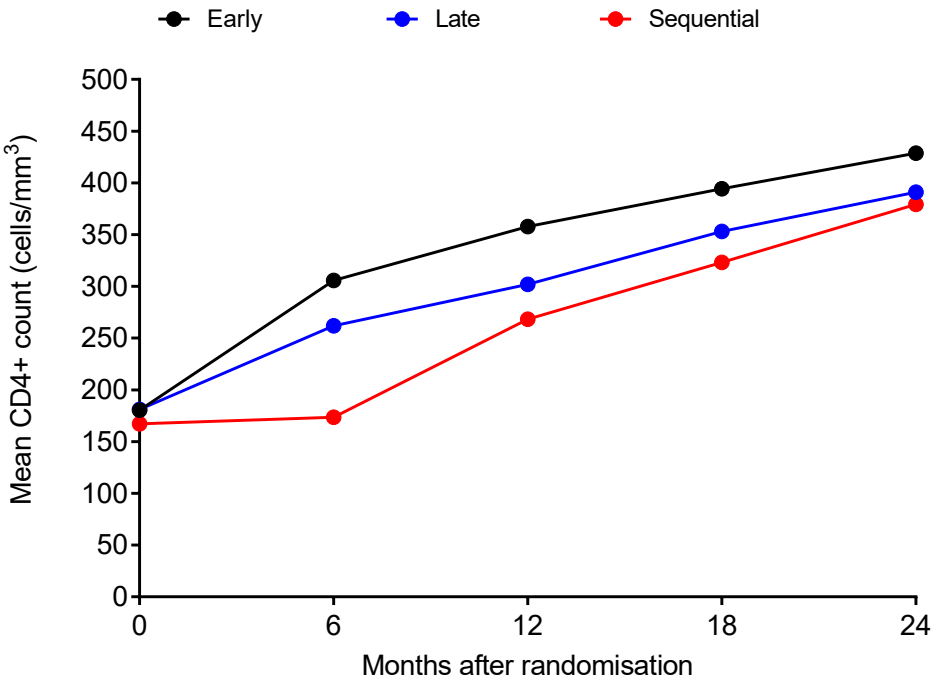


Figure 4.3: Instantaneous probability of death during follow-up

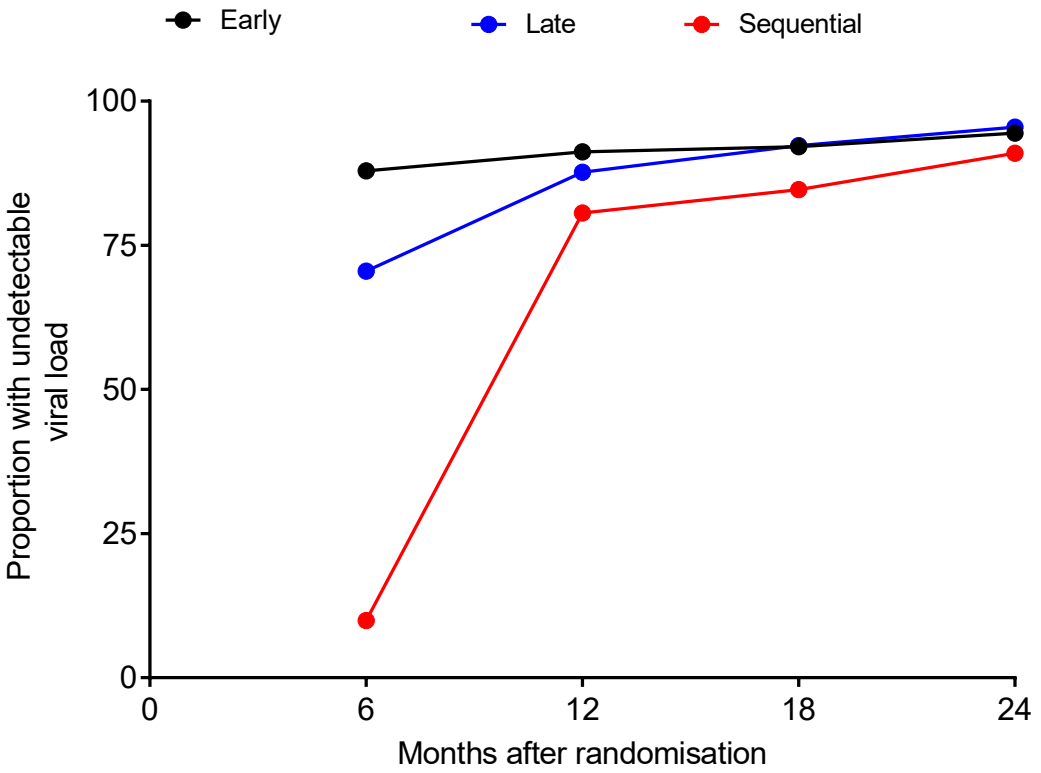
Figure 4.3 shows that the risk of death per unit time for both the early and late integrated arms was fairly constant over the duration of study follow-up. However, there was a steady

incline in the risk of death from baseline to eight months and a sharp decline thereafter in the sequential arm. This decline did not reach the levels seen in the early and late integrated arms. Even though the virologic response at the end of the study was impressive for patients in the sequential arm (Figure 4.5), their immunological response was slightly lower when compared to patients in the early integrated arm (Figure 4.4).



Number of patients with results at each time point					
Early integrated	214	173	160	152	145
Late integrated	215	159	149	143	132
Sequential	213	154	129	124	123

Figure 4.4: Mean CD4+ count (cells/mm^3) over time



Patients with results at each time point				
Early integrated	174	159	152	145
Late integrated	156	146	143	133
Sequential	151	129	124	122

Figure 4.5: Proportion of patients with undetectable viral load (<400 copies/ml)

Treatment-limiting toxicities, drug interactions (Lalloo, 2009), high pill burden and immune reconstitution inflammatory syndrome (IRIS) (Breen et al., 2004; Burman et al., 2007) were shown to be associated with TB and HIV co-treatment. Despite high rates of IRIS observed in the SAPiT trial among patients randomised to early integrated arm (Naidoo et al., 2012), Figures 4.2 and 4.4 paints a clear picture of survival benefits and better immunological outcomes respectively when the treatments are integrated. ARV drug changes were uncommon and not significantly different across the three study arms (Naidoo et al., 2014). Moreover, it has been shown that the incidence rate of loss to follow-up was lower among TB patients initiated on ART in the SAPiT study (Yende-Zuma and Naidoo, 2016).

Chapter 5

Intent-to-treat analysis

5.1 Introduction

Results from the IV analysis will be compared with the results from the ITT analyses to understand the extent of non-compliance in the SAPiT trial. Therefore, we started with presenting ITT results before moving to IV analysis. These analyses were based on semiparametric additive hazards models as well as proportional hazards models. All multivariable models were adjusted for gender, baseline CD4+ cell count and employment status.

5.2 Model formulation

We started by fitting a nonparametric additive hazards model, where the effect of Z, gender, baseline CD4+ cell count and employment status were allowed to change over time. This way, the regression coefficients are allowed to depend on time. The slope the cumulative regression function plot against time gives information on whether the particular covariate has any effect on the hazard function and whether the effect is constant or time dependent. Positive slopes occur during time when an increase in covariate values are associated with an increase in the hazard function. However, negative slopes occur during the times when increasing covariate values are associated with a decrease in the hazard function. Sometimes one can observe a slope

closer to zero which basically shows that the covariate has no effect on the hazard. When the follow-up time is longer, one can observe all three types of slopes within the same plot.

The nonparametric model is given by:

$$\lambda(t|Z, \mathbf{C}) = \lambda_0(t) + \beta_z(t)Z + \beta_c(t)^T\mathbf{C}, \tag{5.1}$$

where $\lambda(t|Z, \mathbf{C})$ is the hazard function conditional on Z and $\mathbf{C}$, $\lambda_0(t)$ is the unknown baseline hazard function, $\beta_z(t)$ and $\beta_c(t)^T$ are time-dependent coefficients for the instrumental variable and the measured confounders respectively. Even though results from model (5.1) gives us an important information regarding whether the covariate lose or maintain their effect over time, but we also fitted semiparametric additive and Cox proportional hazards models shown in equations (5.2) and (5.3) respectively. We will obtain constant hazard difference and hazard ratios respectively, from these two models. These estimates have easy public health interpretation compared to the time-varying hazards differences.

$$\lambda(t|Z, \mathbf{C}) = \lambda_0(t) + \beta_z Z + {\beta_c}^T\mathbf{C}, \tag{5.2}$$

$$\lambda(t|Z, \mathbf{C}) = \lambda_0(t)\exp(\beta_z Z + {\beta_c}^T\mathbf{C}). \tag{5.3}$$

In these models we assumed that the randomisation arm and the baseline variables have constant effects. Moreover, in the Cox proportional hazards models we assume that that the ratio of the hazards for any two participants is constant over time (i.e. they are proportional). In both equations, β_z and ${\beta_c}^T$ represents the effect of randomisation arm and baseline variables (CD4+ cell count, gender and employment status) on time to death.

5.3 Results

The cumulative regression plots in Figures 5.1 and 5.2 from the nonparametric additive hazards model show that all variables have an effect on mortality. However, study arm and employment status have a time-varying effect on the hazard of death where we observe strong initial effect

that seems to disappear after 12 months after randomisation. On the other hand, CD4+ count and gender seem to have a persistent influence.

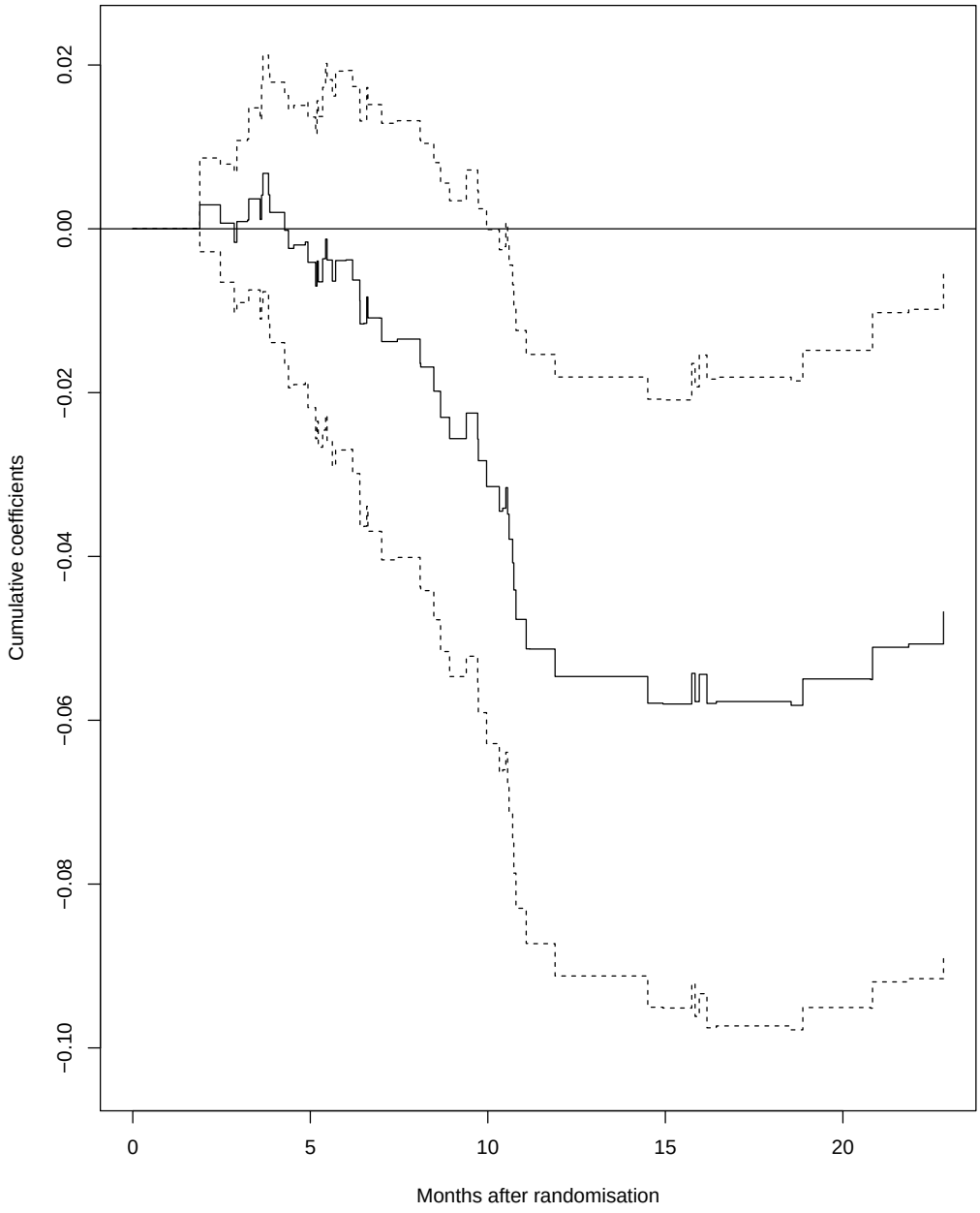
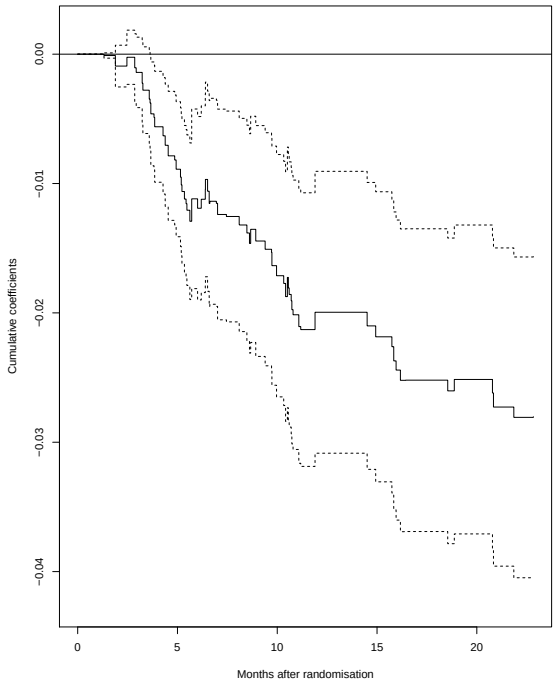
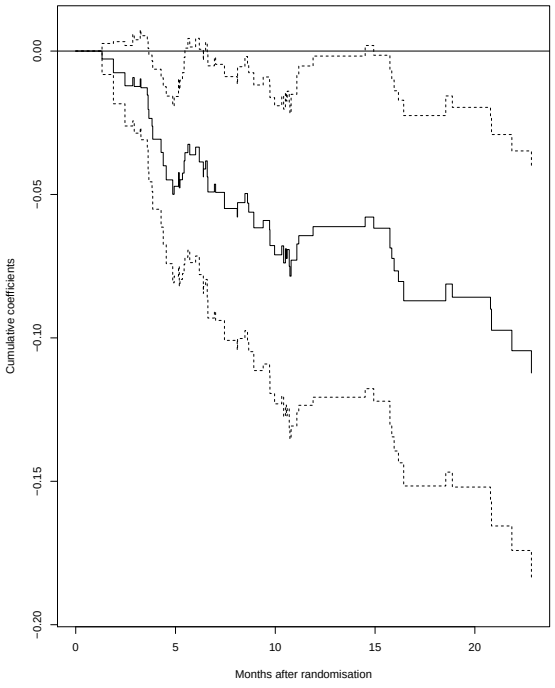


Figure 5.1: Cumulative regression for randomisation arm

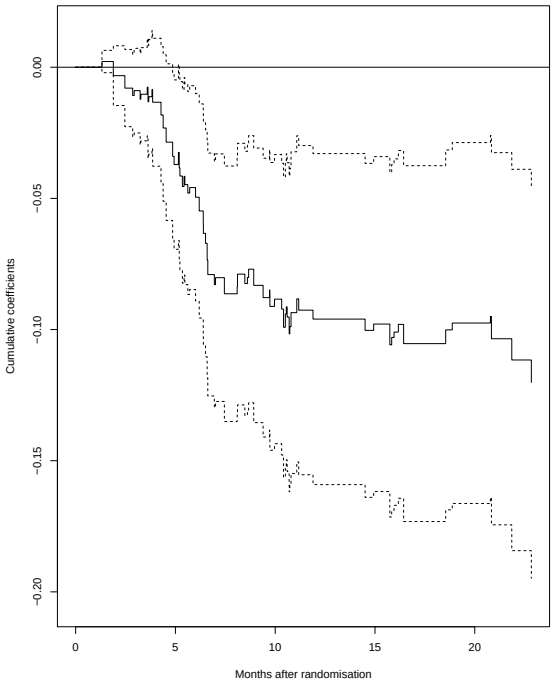
Interestingly, all variables seem to have a stronger effect on mortality during the first 12 months.



(a) CD4+ count



(b) Gender



(c) Employment

Figure 5.2: Cumulative regression for measured confounders

It is not surprising that the effect of the arm on the hazards of death diminishes after 12 months as there were fewer deaths observed after that period. Of the 69 deaths observed in the trial, 81% (56/69) occurred in the first 12 months of follow-up.

Table 5.1: ITT estimates for the effect of study arm on mortality using additive hazards models

	Univariable		Multivariable[a]	
Study arm	β	95% CI	β	95% CI
Early integrated	-0.06	-0.11, -0.02	-0.05	-0.09, -0.01
Late integrated	-0.06	-0.11, -0.02	-0.06	-0.11, -0.02
Sequential	0		0	

[a] adjusted for gender, baseline CD4+ count and employment status

Multivariable results from the ITT analysis show that patients in the early and late integrated arms had a hazard difference of -0.05 (95% CI: -0.09 to -0.01) and -0.06 (95% CI: -0.11 to -0.02) respectively, when compared to the sequential arm (Table 5.1). This indicates that on average, five and six deaths were averted for each year of follow-up in each 100 patients randomised to early and late integrated arms compared with each 100 patients in the sequential arm.

Table 5.2: ITT estimates for the effect of study arm on mortality using proportional hazards models

	Univariable		Multivariable[a]	
Study arm	HR	95% CI	HR	95% CI
Early integrated	0.44	0.25, 0.80	0.51	0.28, 0.92
Late integrated	0.46	0.26, 0.83	0.46	0.26, 0.83
Sequential	1.0		1.0	

[a] adjusted for gender, baseline CD4+ count and employment status
HR: hazard ratio

ITT results based on proportional hazards models showed that the hazard of death was reduced by 49% (hazard ratio (HR): 0.51; 95% CI: 0.28 to 0.92) in the early and 54% (HR: 0.46; 95% CI: 0.26 to 0.83) in the late integrated arm when compared to the sequential arm (Table 5.2).

5.4 Summary

As previously shown by Abdool Karim et al. (2010), the early and late integration of TB and ART reduced the hazard of death. Results from additive hazards models give an indication of absolute number of deaths averted with treatment integration, whereas, hazard ratios provide relative hazard of death which speaks to the magnitude of the association between the study arm and time to death.

Chapter 6

Instrumental variables analysis using fixed exposure

6.1 Introduction

One of the objectives of this study is to account for non-compliance by using instrumental variables analysis. In Chapter 3 we introduced variables that we are going to use in IV analysis. Here, we introduce some of them again for ease of reference. We used randomisation arm (Z) as an instrument and by design it is independent of measured and unmeasured confounders. The measure of compliance (X) is defined as a fraction of time on ART during TB treatment and referred to here as fixed exposure since it was defined cross sectionally per patient. We aim to assess the effect of X on time to death and expect the effect size from this analysis to be higher than that of ITT analysis.

IV methods have been shown to work well under additive hazards models because its mathematical form and the interpretation of the estimates resembles that of linear models. Under the proportional hazards models, IV methods work well when the outcome is rare over the entire follow-up (Tchetgen Tchetgen et al., 2015), which was not well satisfied in the SAPiT trial. Therefore, the main analysis in this thesis were based on the additive hazards model while the

proportional hazards model was used for comparison purposes.

We will use the two-stage modelling, a concept that was introduced in Section 2.2 which will form the basis of our IV analysis. We utilised two-stage predictor substitution (2SPS) and two-stage residuals inclusion (2SRI) methods. 2SPS is a nonlinear extension of the linear two-stage least squares (2SLS) (Rassen et al., 2009). 2SRI was first introduced by Hausman (1978) and later proposed for the analysis of time-to-event endpoints by Terza et al. (2008). 2SRI has been the method of choice in clinical research involving time-to-event outcomes (Gore et al., 2010; Hadley et al., 2010; Tan et al., 2012). Both these approaches work by fitting a first-stage, univariable or multivariable regression models of the association between X and Z. The second-stage models are then used for studying the effect of X on time to death as explained in Section 2.2.

6.2 Model formulation

The first-stage multivariable linear model is

$$X = \alpha_0 + \alpha_z Z + \alpha_c^T \mathbf{C} + \varepsilon, \tag{6.1}$$

where α_0, α_z and α_c^T are coefficients for the intercept, the instrumental variable and measured confounders respectively. Z was modelled categorically using dummy coding. The 2SPS approach then proceeds by regressing the survival time on the fitted values $\hat{X}$ from the first-stage regression model (6.1) and on the measured confounders ($\mathbf{C}$) either using additive hazard or proportional hazards regression models. The 2SRI approach proceeds likewise, but regressing survival time on the residuals $X - \hat{X}$ from the first-stage regression model (6.1). Tchetgen Tchetgen et al. (2015) showed that under certain conditions specified next, the coefficient of $\hat{X}$ in the resulting additive hazard model can be interpreted as the exposure effect β_x in the additive hazard model

$$\lambda(t|X, Z, U, \mathbf{C}) = \lambda_0(t) + \beta_x X + \beta_c^T \mathbf{C} + \beta_u(t) U \tag{6.2}$$

which also involves adjustment for possible unmeasured confounders U.

Deducing from Equation 6.2, $\exp(-\beta_x t)$ can be interpreted as the relative chance of surviving time t with exposure 1 versus 0; note that it takes the length of the exposure period into account via the value of t. For the 2SPS approach, the condition is that the error term ε in the exposure model (6.1) is independent of randomisation arm (given the confounders). For the 2SRI approach, a more subtle additional assumption is needed, which is satisfied when the error term (ε) equals the unmeasured confounder (U) apart from (additive) random noise. Because tests of the null hypothesis of no exposure effect are robust against model misspecification in the 2SPS approach (unlike the 2SRI approach), we generally recommend the 2SPS approach (Tchetgen Tchetgen et al., 2015; Vansteelandt et al., 2011; Wooldridge, 2002).

In the 2SPS approach, the predicted exposure is a function of the IV, and thus under the null hypothesis of no exposure effect, one would not detect an association at the 5% significance level more than 5% of times, regardless of whether the first stage model is correct or not (Vansteelandt and Didelez, 2018). Indeed, even when that model is incorrect, the predicted exposure is still a function of the IV, and thus not associated with outcome under the null hypothesis. This is not the case with the 2SRI method where one models the residual, and where model misspecification may induce bias (Vansteelandt et al., 2011; Vansteelandt and Didelez, 2018).

Both 2SPS and 2SRI approaches were extended to proportional hazards models even though the outcome is not rare. 2SPS under proportional hazards models is not consistent, however Li et al. (2015) argued that it may provide reasonable inferences in at least some empirical applications. Under the rare event assumption, then it follows that the coefficient of $\hat{X}$ can be interpreted as the exposure effect β_x in the proportional hazards regression model

$$\lambda(t|X, Z, U, \mathbf{C}) = \lambda_0(t)\exp(\beta_x X + \beta_c^T \mathbf{C} + \beta_u U) \tag{6.3}$$

but not otherwise.

Two-stage estimation has its own disadvantages, especially outside linear models. If the first-

stage model is wrong, the standard errors in the second stage will be incorrect leading to biased effect size. To get the correct standard errors and confidence intervals, bootstrapping methods can be used (Efron, 1979; Efron and Tibshirani, 1993). Alternatively, the jackknife method introduced by Angrist et al. (1999) can also be used. We generated 1000 nonparametric bootstrap samples with replacement to calculate the standard errors and subsequently used the bias-corrected and accelerated (BCa) method (DiCiccio and Efron, 1996) to calculate 95% confidence intervals (CIs) for β_x in the second-stage of both 2SPS and 2SRI. This is due to the fact that even when the instrument is valid, IV method produces standard errors that are higher than non-IV estimates (Little et al., 2009).

For ease of comparison with the results from the ITT analysis, we also fitted a nonparametric additive hazards model

$$\lambda(t|X, Z, U, \mathbf{C}) = \lambda_0(t) + \beta_x(t)X + \beta_c^T(t)\mathbf{C} + \beta_u(t)U. \tag{6.4}$$

Compared to the semiparametric additive hazards model (6.2), the nonparametric model (6.4) allows the effect of the exposure ($\beta_x(t)$) to change over time. Moreover, we fitted semiparametric additive hazards and proportional hazards models under the "as treated"analysis given by

$$\lambda(t|X, \mathbf{C}) = \lambda_0(t) + \beta_x X + \beta_c^T \mathbf{C} \tag{6.5}$$

and

$$\lambda(t|X, \mathbf{C}) = \lambda_0(t)\exp(\beta_x X + \beta_c^T \mathbf{C}) \tag{6.6}$$

respectively. As mentioned earlier, results from the "as treated"analyses are likely biased because they disregard the randomisation arm. Compared to the IV analyses, they do not make use of the comparability offered by randomisation. Results from the "as treated"analyses will be compared with the results from the ITT and IV analyses (both 2SPS and 2SRI approaches), but they will obviously be interpreted with caution.

6.3 Results

Even though the fixed exposure was continuous and we preferred keeping it that way, we were able to identify compliant and non-compliant patients. Among those who started ART in all three study arms, compliance was defined as starting ART within 4 weeks after the start of TB therapy (early integrated arm); starting ART within 4 weeks after the completion of the intensive phase of TB therapy (late integrated arm); starting ART within 4 weeks after the completion of TB therapy (sequential arm). Non-compliance was defined as not starting ART at the correct time with respect to TB treatment.

Following guidance from Baiocchi et al. (2014) on how to report IV analysis, we reported the distribution of measured covariates for compliers and non-compliers (Table 6.1). The distribution of measured covariates across levels of the instrument was presented in Table 4.1, where we showed that the instrument was independent of measured confounders. The 16 patients who never started ART in the early integrated arm had the highest median baseline CD4+ count (Table 6.1). All 24 patients from the sequential arm who were non-compliant due to starting ART before completing TB treatment had the lowest median baseline CD4+ count. This is not surprising as these patients were immuno-compromised and hence the decision from clinicians to initiate them on ART earlier than expected.

The mean exposure to ART during TB treatment was 0.78, 0.39 and 0.04 in the early integrated, late integrated and sequential arms respectively (Figure 6.1). This means that during TB treatment, patients in the early integrated, late integrated and sequential arm spent on average 78%, 39% and 4% of the time on both ART and TB treatment respectively. Had all patients complied with their arm assignment and also remained in the study until TB treatment is completed, one would expect all patients in the early integrated arm to have $X \geq 0.8$. Whereas, in the late integrated arm it would have been $0.6 \leq X < 0.8$ and $X = 0$ in the sequential arm. The median (IQR) duration on TB treatment was 6.7 (6.4, 8.3) months in each of the three study arms.

Table 6.1: Baseline characteristics of patients stratified by compliance status in the SAPiT trial

Variable	**Compliant**			**Non-compliant**			**No ART initiation**		
	Early integrated (N=153)	Late integrated (N=106)	Sequential (N=115)	Early integrated (N=45)	Late integrated (N=58)	Sequential (N=24)	Early integrated (N=16)	Late integrated (N=51)	Sequential (N=74)
Mean age (SD), year	34.8 (7.8)	34.7 (7.7)	34.1 (8.1)	33.8 (9.3)	35.8 (11.3)	38.1 (9.6)	30.7 (5.4)	32.8 (7.3)	32.2 (7.4)
Male, n (%)	71 (46.4)	54 (50.9)	57 (49.6)	19 (42.2)	25 (43.1)	11 (45.8)	7 (43.8)	33 (64.7)	42 (56.8)
Employed patients, n (%)	99 (64.7)	58 (54.7)	71 (61.7)	26 (57.8)	30 (51.7)	12 (50.0)	10 (62.5)	29 (56.9)	34 (45.9)
Median CD4+ count (IQR), cells/mm^3	150 (75-256)	141 (70-228)	146 (84-245)	127 (88-273)	154 (52-247)	72 (41-179)	208 (83-305)	154 (123-350)	156 (55-304)
Mean $\log_{10}$ viral load (SD), copies/ml[a]	5.1 (1.0)	5.1 (0.8)	5.1 (0.8)	4.8 (0.9)	5.1 (0.9)	5.1 (0.9)	4.8 (1.0)	4.9 (1.0)	5.1 (0.7)
WHO stage 4, n (%)	10 (6.5)	7 (6.6)	8 (7.0)	3 (6.7)	4 (6.9)	0	1 (6.3)	0	5 (6.8)

[a] Viral load was not available for 16 patients in each of the early and late integrated arms and 12 in the sequential arm

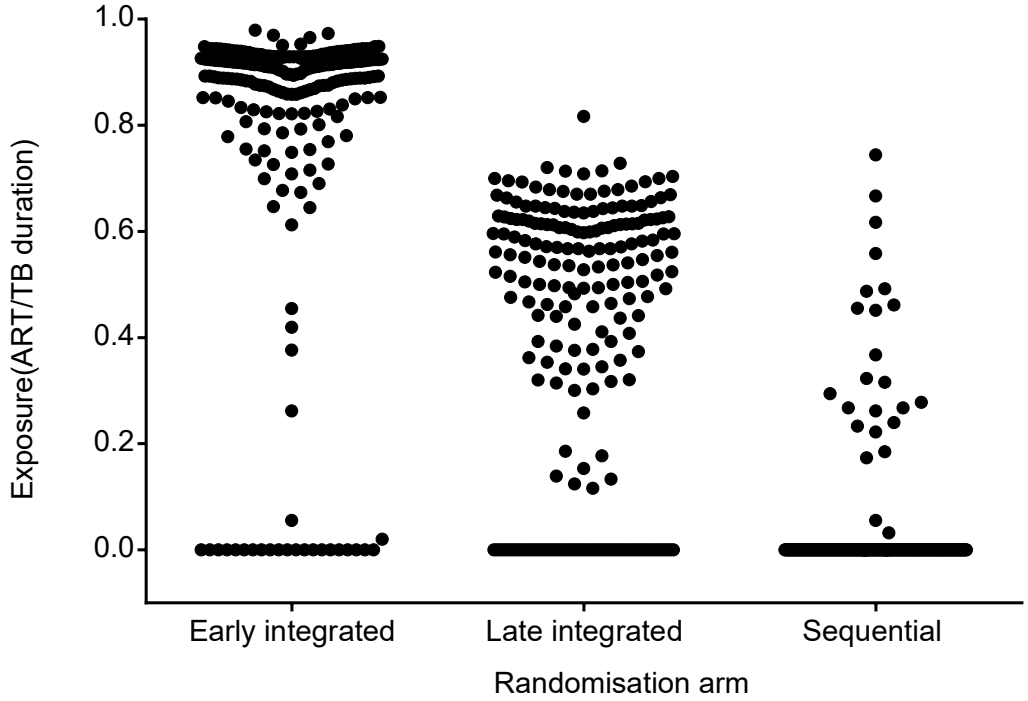


Figure 6.1: Exposure (fraction of time on ART during TB treatment) in the three randomised arms

Non-compliance is evident in Figure 6.1, as some of the patients in the sequential arm had high exposure because they started ART before the completion of TB therapy. Also, some patients in the early and late integrated arms had very low exposure suggesting that they delayed starting ART.

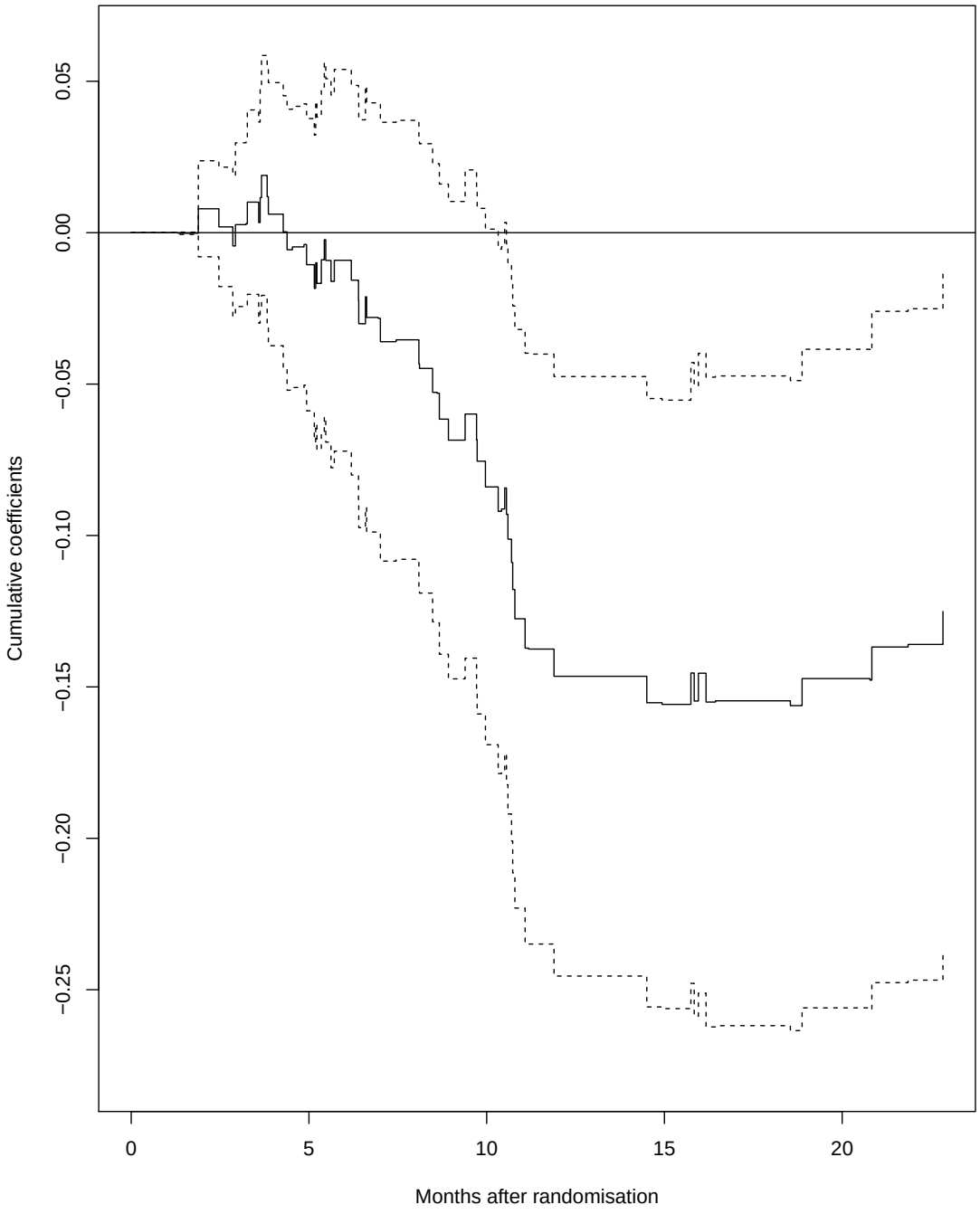


Figure 6.2: Cumulative regression for fixed exposure under model (6.4)

Similar to the study arm, fixed exposure had time-varying effect on the hazard of to death (Figure 6.2). This has been attributed to fewer deaths observed after 12 months of follow-up.

We did not present the plots for measured confounders because we are mostly interested in the results for the exposure variable.

Table 6.2: First-stage linear regression model predicting the fixed exposure

	Univariable			Multivariable[a]		
Effect	Estimate	Std Err.	Pr> $\|t\|$	Estimate	Std Err.	Pr> $\|t\|$
Intercept	0.04	0.02	0.015	0.06	0.03	0.019
Early integrated	0.74	0.02	<.001	0.74	0.02	<.001
Late integrated	0.35	0.02	<.001	0.35	0.02	<.001
Sequential	0			0		
F-value; Partial R^2	514.68; 0.62			211.66; 0.62		

[a] adjusted for gender, baseline CD4+ count and employment status

Our instrument was strong (F-value= 514.68, R^2=0.62) and also associated with the exposure (Table 6.2). Roughly, 62% of the variation in the instrument was explained by the exposure.

Table 6.3: IV estimates for the effect of exposure on mortality using additive hazards models

	Univariable		Multivariable[a]	
Method/exposure	β	95% CI	β	95% CI
As-treated	-0.12	-0.17, -0.08	-0.12	-0.17, -0.07
2SPS	-0.08	-0.14, -0.03	-0.07	-0.12, -0.01
2SRI	-0.06	-0.13, -0.01	-0.05	-0.11, 0.01
First-stage residuals	-0.19	-0.35, -0.06	-0.23	-0.43, -0.09

[a] adjusted for gender, baseline CD4+ count and employment status

Results from 2SPS analysis in Table 6.3 showed that on average, seven deaths (hazard difference= -0.07; 95% CI: -0.12 to -0.01) were prevented for each year of follow-up in each 100 patients with full exposure to ART during TB treatment (as would be the case under perfect compliance in the early integrated arm) as opposed to 100 patients with no ART exposure during TB treatment (as would be the case under perfect compliance in the sequential arm). The 2SRI method resulted in slightly weaker effects (hazard difference = -0.05; 95% CI: -0.11; 0.01). The strong association found between the first-stage residuals and time to death in this analysis provides

strong evidence of unmeasured confounding, which the IV analysis accounted for. The findings from the "as-treated"analysis (hazard difference = -0.12; 95% CI: -0.17; -0.07), showed higher effect but likely biased.

Table 6.4: IV estimates for the effect of exposure on mortality using proportional hazards models

	Univariable		Multivariable[a]	
Method/exposure	HR	95% CI	HR	95% CI
As-treated	0.15	0.07, 0.31	0.14	0.06, 0.29
2SPS	0.30	0.12, 0.70	0.35	0.14, 0.85
2SRI	0.25	0.09, 0.64	0.26	0.09, 0.70
First-stage residuals	0.24	0.08, 0.71	0.18	0.05, 0.64

[a] adjusted for gender, baseline CD4+ count and employment status

Results from the Cox proportional hazards models are shown in Table 6.4. The 2SPS and 2SRI methods showed that full exposure to ART during TB treatment reduced the hazard of death by 65% (HR: 0.35; 95% CI: 0.14 to 0.85) and 74% (HR: 0.26; 95% CI: 0.09 to 0.70) respectively. However, according to Tchetgen Tchetgen et al. (2015), results from Cox proportional hazards models are valid when the outcome is rare over the entire follow-up duration. This also explains why time-varying exposure analysis was not used under proportional hazards models. Results from the "as-treated"analysis showed a stronger reduction of 86% (HR: 0.14; 95% CI: 0.06 to 0.29) in hazard of death due to full ART exposure during TB treatment (as opposed to no exposure).

6.4 Model-based predicted survival probabilities

We calculated model-based survival probabilities for ITT and IV analysis under the 2SPS approach of the additive and proportional hazards models. In the IV analysis, the survival probabilities were estimated for three fixed exposure levels of 0 (no exposure), 0.6 (partial exposure) and 1.0 (full exposure) (Figure 6.3). These correspond to how the survival probabilities would have looked like in the sequential, late integrated and early integrated arms, respectively, had there been perfect compliance.

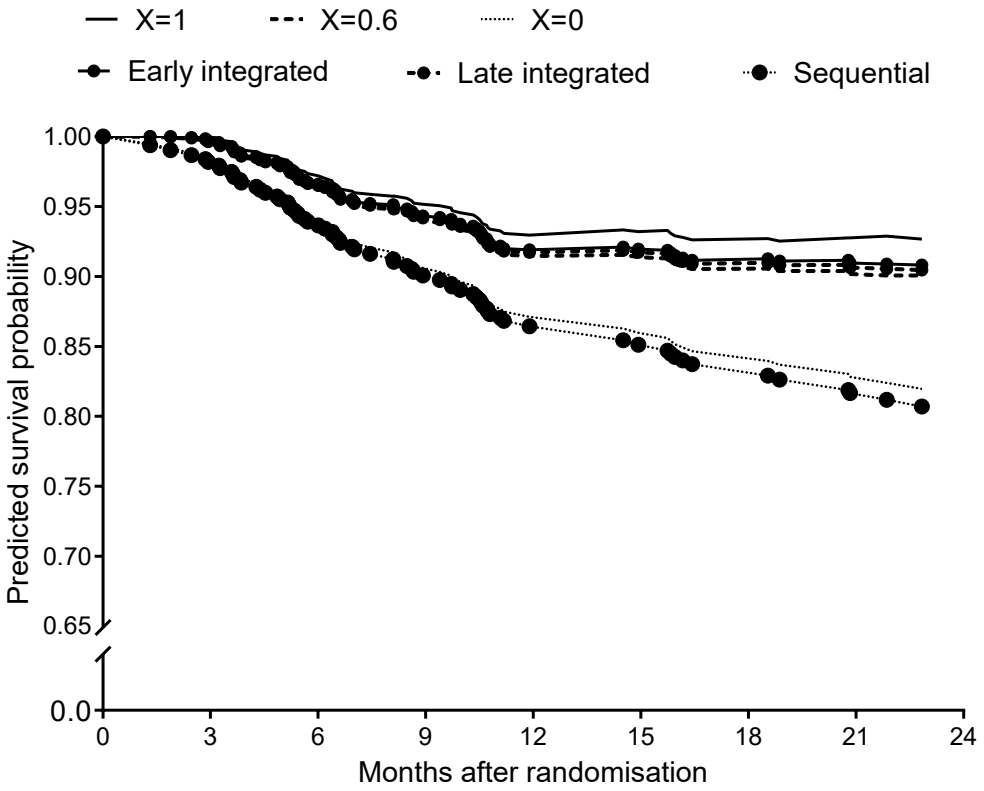


Figure 6.3: Predicted survival curves comparing ITT and IV estimates from additive hazards model with fixed exposure

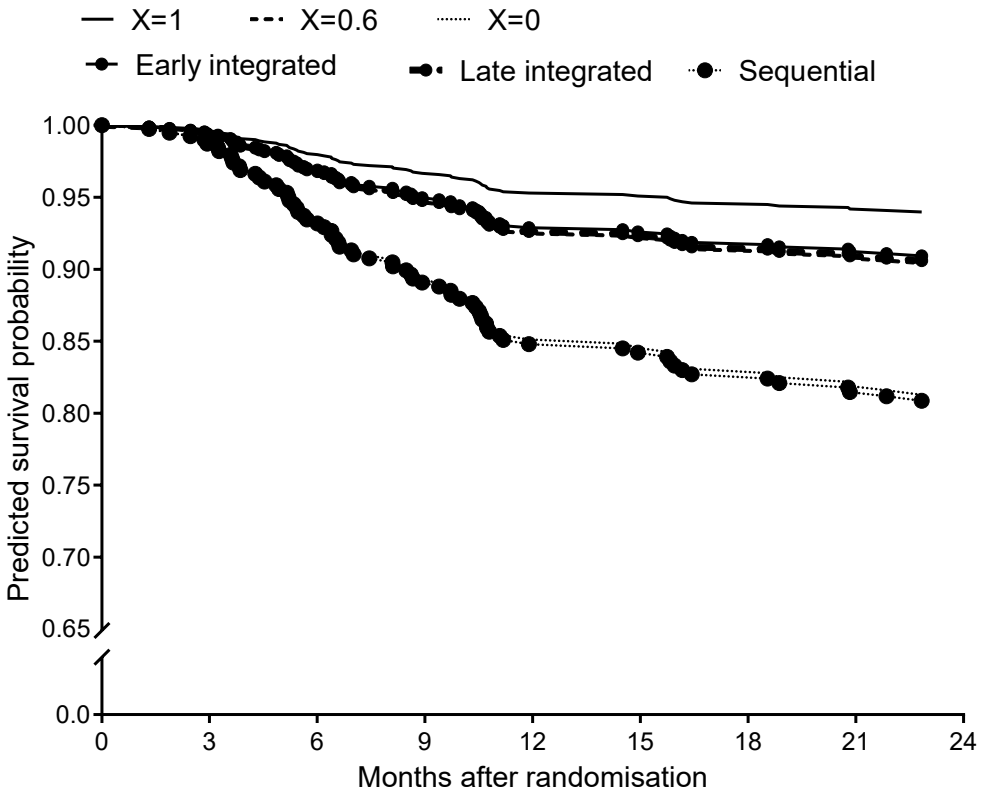


Figure 6.4: Predicted survival curves comparing ITT and IV estimates from proportional hazards model with fixed exposure

In both Figures 6.3 and 6.4 the predicted survival probabilities for X=1 are slightly higher than that of the early integrated arm. However, survival probabilities for X=0.6 and X=0 are closer to those for late integrated and sequential arms respectively.

6.5 Sensitivity analysis 1

We saw in Figures 5.1 and 6.2 that the effect of both exposure and randomisation arm on mortality was only stronger during the first 12 months of follow-up. The accumulation of person-years with fewer deaths after 12 months of follow-up potentially underestimated the mortality rate. The aim of this sensitivity analysis is to understand how the results would have looked like had the study ran for 12 rather than 24 months. We let T^*= min(T,12), where T is the follow-up time defined as the minimum of time to death and time to censoring and T^* is the shortened

follow-up time. For patients who died after 12 months, we let them become censored.

Table 6.5: Sensitivity: ITT and IV estimates for the effect of study arm and exposure on mortality using additive hazards models

Method/exposure		Univariable		Multivariable[a]	
		β	95% CI	β	95% CI
ITT	Early integrated	-0.11	-0.19, -0.04	-0.11	-0.19, -0.04
	Late integrated	-0.12	-0.19, -0.04	-0.10	-0.17, -0.03
	Sequential	0		0	
As-treated	X	-0.20	-0.28, -0.13	-0.20	-0.28, -0.12
2SPS	X	-0.16	-0.27, -0.06	-0.14	-0.25, -0.05
2SRI	X	-0.14	-0.26, -0.04	-0.12	-0.23, -0.03
First-stage residuals		-0.20	-0.43, -0.01	-0.26	-0.52, -0.06

[a]adjusted for gender, baseline CD4+ count and employment status

As expected, sensitivity analysis produced higher effect estimates (Table 6.5) than those from the original analysis presented in Tables 5.1 and 6.3. On average, 11 and 10 deaths were averted for each year of follow-up in each 100 patients randomised to early and late integrated arms compared with each 100 patients in the sequential arm. These estimates are doubled, when compared to the original estimates of -0.05 and -0.06. Likewise, the same trend was observed where the hazard difference changed from -0.07 and -0.05 (Table 6.3) to -0.14 and -0.12 under the 2SPS and 2SRI approach respectively.

6.6 Sensitivity analysis 2

Earlier on, we mentioned that the initiation of ART during TB treatment has been shown to improve survival, especially among patients with low CD4+ cell count. This may imply that the effect of ART initiation during TB treatment differed according to baseline CD4+ cell count. We used baseline median CD4+ cell count of 148 cells/mm^3 as a cut-off point to categorise patients to low and high CD4+ cell count levels.

Results for patients with low CD4+ cell count expressed higher effect than for those with high

Table 6.6: First-stage linear regression model predicting the exposure stratified by baseline CD4+ cell count

	CD4+ cell count $\leq$ 148 cells/mm^3				**CD4+ cell count $>$ 148 cells/mm^3**			
Variable	Univariable		Multivariable[a]		Univariable		Multivariable[a]	
Effect	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept	0.05	0.02	0.03	0.03	0.03	0.02	0.001	0.03
Early integrated arm	0.74	0.03	0.73	0.03	0.74	0.02	0.73	0.03
Late integrated arm	0.36	0.03	0.36	0.03	0.33	0.03	0.33	0.03
Sequential arm	0	0	0	0	0	0	0	
F-value; Partial R^2	267.05; 0.62		265.45; 0.63		243.83; 0.61		235.80; 0.61	

S.E.: standard error [a] adjusted for gender and employment status

Table 6.7: ITT and IV results from additive hazards models stratified by baseline CD4+ cell count

Method/exposure		**CD4+ cell count $\leq$ 148 cells/mm^3**		**CD4+ cell count $>$ 148 cells/mm^3**	
		Univariable β (95% CI)	Multivariable[a] β (95% CI)	Univariable β (95% CI)	Multivariable[a] β (95% CI)
ITT	Early integrated	-0.09 (-0.17, -0.02)	-0.08 (-0.16, -0.01)	-0.03 (-0.07, 0.02)	-0.02 (-0.06, 0.02)
	Late integrated	-0.10 (-0.17, -0.02)	-0.11 (-0.18, -0.03)	-0.02 (-0.06, 0.03)	-0.02 (-0.06, 0.03)
	Sequential	0	0	0	0
As-treated	X	-0.18 (-0.26, -0.09)	-0.17 (-0.25, -0.09)	-0.08 (-0.12, -0.03)	-0.07 (-0.11, -0.03)
2SPS	X	-0.12 (-0.22, -0.02)	-0.11 (-0.23, -0.02)	-0.03 (-0.08, 0.02)	-0.03 (-0.07, 0.03)
2SRI	X	-0.10 (-0.21, 0.01)	-0.09 (-0.21, 0.02)	-0.02 (-0.08, 0.04)	-0.01 (-0.07, 0.05)
	First-stage residuals	-0.25 (-0.65, -0.03)	-0.26 (-0.66, -0.03)	-0.16 (-0.37, -0.04)	-0.16 (-0.40, -0.04)

[a] adjusted for gender and employment status

CD4+ count. This effect is even higher than that for the combined sample (see Table 6.3). It is not surprising that the effect of ART initiation during TB treatment had no impact on patients with high CD4+ count because most deaths [76.8% (53/69)] occurred among patients with CD4+ count below the median and therefore that analysis lacked statistical power.

6.7 Summary

In this Chapter we have provided IV analysis of the causal effect of exposure to ART during TB treatment on time to death. This analysis accounted for non-compliance as a result of not all patients sticking to randomisation and starting ART at the correct time with respect to TB treatment. The SAPiT trial data has been analysed using ITT methods and showed that integration of ART and TB treatment saves lives (Abdool Karim et al., 2010). Our results express more precisely how many lives could be saved under perfect compliance, of which appeals to patients and clinicians who are interested in the benefits of initiating and adhering to received treatment.

The analyses were carried out using the semiparametric additive hazards models under 2SPS and 2SRI methods. Cox proportional hazards models were used for comparison purposes. Results from the IV analysis, which accounted for non-compliance produces slightly larger estimates compared to the ITT results. These results demonstrate that the survival benefit of fully integrating TB treatment and ART is even higher than what has been reported in the ITT analysis since non-compliance has been accounted for. Results from the "as-treated"analysis produced the largest estimates and one could potentially be tempted to draw conclusions from them, but they are likely to be biased.

Large estimates observed when follow-up was censored at 12 months are an indication that the benefit of ART and TB integration was seen as early as 12 months while revealing the extent of early mortality. After 12 months, only 9 deaths were observed and the study continued accumulating person-years and thus reducing mortality rate across all study arms. In our settings, it has been shown that early mortality among adult patients accessing ART programmes is

high (Lawn et al., 2008), due to patients presenting late for care with advanced clinical diseases (Kigozi et al., 2009; Lawn et al., 2005). Studies done elsewhere showed that late presentation to care is still rife even when the eligibility threshold was increased or when ART was recommended irrespective of CD4+ cell count (Darcis et al., 2018; Fomundam et al., 2017; Larsen et al., 2018; Nyika et al., 2016).

The validity and interpretation of IV estimates is built on three key strong assumptions. There is also a fourth and also strong assumption about treatment homogeneity. The first, that patients on the different arms of the study are exchangeable, is guaranteed by randomisation. Secondly, that the instrument is associated with exposure is shown in Table 6.2. The third, so-called exclusion restriction, that randomised assignment may only influence all-cause-mortality by changing ART exposure could be violated. One possible cause of violation concerns our definition of exposure, which may not fully capture all relevant components such as adherence which could have an effect on mortality. A second possible reason is that, in the open-label SAPiT trial, being assigned to either of the integrated arms may have enhanced patients expectation of success, and in contrast, assignment to the sequential arm might have reduced such an expectation. Moreover, those randomised to integrated arms who did not start ART soon after TB treatment initiation might have deliberately delayed ART initiation because they were still feeling well and did not see the need to integrate TB treatment and ART. Our analysis moreover ignored differential ART exposure outside the TB treatment window. All of this in turn violates the exclusion restriction assumption, which underlies our analysis. The violation of this untestable assumption can lead to biased IV estimates. Lastly, through treatment homogeneity we assume that TB treatment and ART has the same effect for compliers and defiers.

Tchetgen Tchetgen et al. (2015) and Li et al. (2015) validated the two-stage estimation method under the additive hazards models. In linear models, the two-stage estimation method, namely 2SLS is robust against first-stage regression model misspecification (Wooldridge, 2002). However, the 2SPS and 2SRI requires the first-stage regression to be correctly specified in order for the IV estimates to be valid (Tchetgen Tchetgen et al., 2015). Furthermore, 2SPS requires the error term (ε) in model (6.1) to be independent of randomisation arm given, the measured

confounders. When it comes to Cox proportional hazards models, the 2SPS and 2SRI methods are not guaranteed to produce valid estimates unless the incidence proportion is relatively low during follow-up.

We acknowledge several additional limitations in our analyses. Our fixed exposure did not differentiate between patients who were on TB treatment for six months and those who were on TB treatment for a longer period. A patient who was on TB treatment for six months and only took ART for three months had similar exposure level to a patient who was on TB treatment for 12 months and took ART for 6 months. The latter patient is more likely to have drug resistant TB and thus more likely to die. Also, for patients who died soon after enrolment and those who did not start ART, the exposure is not well defined. Moreover, this exposure can potentially introduce reverse causality especially among those who either died or were terminated early. We believe this somehow explains why the ITT and IV estimates under the additive hazards models are not that dramatically different.

Chapter 7

Instrumental variables analysis using time-varying exposure

7.1 Introduction

IV-methodology for the effect of a time-varying exposure on a time-to-event endpoint is currently lacking, with the exception of G-estimation for structural accelerated failure time models (Robins and Tsiatis, 1991). While G-estimation strategies have been proposed to infer the effect of a time-varying exposure on a time-to-event endpoint in the presence of an instrumental variable under an alternative class of structural accelerated failure time models (Robins and Tsiatis, 1991), application of these methods in applied research has been relatively infrequent because of their complexity and often poor performance in the presence of censoring (Vansteelandt and Joffe, 2014a).

In this Chapter, we developed a novel IV-methodology for time-varying exposure under additive hazard models. Our proposal overcomes above-mentioned concerns because it is applicable in standard software for additive hazard models, which naturally accommodates non-informative censoring without requiring further adjustments. Over and above the lack of IV methodology for time-varying exposure, we developed this method due to the limitations of our fixed exposure

discussed in Chapter 6. Among others, it cannot capture the full complexity of compliance behaviour and may, moreover, be indirectly influenced by censoring or death. The time-varying measure of compliance ($X(t)$) to ART is defined as 1 at time t when the considered patient was on ART at or prior to time t, and 0 otherwise. We will assess the effect of $X(t)$ on time to death using under the 2SPS approach. Our IV analysis based on this exposure will express the effect of continuous ART use versus no ART use (regardless of TB treatment).

7.2 Model formulation

Let $\lambda(t)$ be the conditional hazard at time t given $X(t)$, Z and U, where $X(t)$ is the exposure at time t, Z is the instrument and U represents unmeasured confounders of the association between the exposure ($X(t)$) and the time-to-event endpoint (T). Let $dN(t)$ be the increment in the counting process related to the event at time t. Then $\lambda(t)$ is more precisely defined as $E\big\{dN(t)|T \geq t, \bar{X}(t), Z, U\big\}$, where $\bar{X}$ is the history of exposure up to time t. Consider the model

$$\lambda(t) = \omega(t, U) + \beta X(t) \tag{7.1}$$

and for subjects with $T \geq t$, let

$$X(t) = E\big\{X(t)|T \geq t, Z\big\} + \Delta(t), \tag{7.2}$$

where we assume $\Delta(t)$ is independent of Z among subjects alive at time t. Therefore, the conditional hazard at time t is given by

$$\lambda^*(t) \equiv E\big\{dN(t)|T \geq t, Z, U\big\} = \omega^*(t, U) + \beta E\big\{X(t)|T \geq t, Z\big\},$$

where $E\big\{X(t)|T \geq t, Z\big\} \equiv M(t)$ such that

$$\lambda^*(t) \equiv E\big\{dN(t)|T \geq t, Z, U\big\} = \omega^*(t, U) + \beta M(t),$$

and

$$\begin{aligned}\omega^*(t,U) &= \omega(t,U) - \frac{d}{dt}\log E\big[\exp\big\{-\beta\Delta(t)|T \geq t, Z, U\big\}\big] \\ &= \omega(t,U) - \frac{d}{dt}\log E\big[\exp\big\{-\beta\Delta(t)|T \geq t, U\big\}\big].\end{aligned}$$

It further follows that

$$P(T \geq t|Z,U) = \exp\Big[-\int_0^t \big\{\omega^*(s,U) + \beta M(s)\big\}ds\Big]$$

from which, by Bayes' rule

$$f(Z,U|T \geq t) = \exp\Big[-\int_0^t \omega^*(s,U)ds\Big]\exp\Big[-\int_0^t \beta M(s)ds\Big]\frac{f(Z)f(U)}{P(T \geq t)}.$$

We note that $M(s)$ is a deterministic function of time and the baseline Z and not a time-varying covariate. Factorisation of this joint density into terms involving either Z or U demonstrate that $Z \perp\!\!\!\perp U|T \geq t$ for all $t > 0$. Using this, we can now demonstrate that the following estimating equation for β,

$\big\{M(t) - E\big[M(t)|T \geq t\big]\big\}R(t)\big\{dN(t) - \beta M(t)dt\big\}$, where $M(t) \equiv E\big\{X(t)|T \geq t, Z\big\}$ is unbiased, where $R(t)$ is the at risk indicator.

Indeed, its expectation equals

$$\begin{aligned}&E\big(\big\{M(t) - E[M(t)|T \geq t]\big\}R(t)\big\{dN(t) - \beta M(t)dt\big\}\big) \\ &= E\big(\big\{M(t) - E[M(t)|T \geq t]\big\}R(t)\big\{\omega(t,U)dt + \beta\Delta(t)dt\big\}\big)\end{aligned}$$

Since $M(t)$ is a function of Z only, and $Z \perp\!\!\!\perp U|T \geq t$, we have that

$$E\big(\big\{M(t) - E[M(t)|T \geq t]\big\}R(t)\big\{\omega(t,U)dt + \beta\Delta(t)dt\big\}\big) = 0.$$

Since $\Delta(t)$ is independent of Z among subjects alive at time t, we further have that

$$E\big(\big\{M(t) - E[M(t)|T \geq t]\big\}R(t)\beta\Delta(t)dt\big) = 0.$$

A consistent estimator of β can thus be obtained by solving an estimating equation based on the above estimating function such that

$$\hat{\beta} = \Bigg[\sum_{i=1}^{n}\big\{M_i(t) - E[M_i(t)|T_i \geq t]\big\}dN_i(t)\Bigg]\Bigg[\sum_{i=1}^{n}\big\{M_i(t) - E[M_i(t)|T_i \geq t]\big\}M_i(t)R_i(t)\Bigg]^{-1}$$

for $i = 1, \ldots, n$ individuals.

It follows from Vansteelandt et al. (2014b) that the Aalen least squares estimator under the model

$$\tilde{\lambda}(t) = \psi_0(t) + \beta M(t) \tag{7.3}$$

solves precisely such an equation, thus confirming the validity of the 2SPS estimator.

Using the 2SPS approach, the first-stage model for $X(t)$ is given by

$$X(t) = \alpha_0(t) + \alpha_z(t)Z + \alpha_c^T(t)\mathbf{C} + \varepsilon(t) \tag{7.4}$$

defined for patients who are alive at time t, which we consider at each observed event time t. Under the assumption that $\varepsilon(t)$ is uncorrelated with Z, conditional on $\mathbf{C}$ (vector of measured confounders) for patients who are alive at time t, then show that the 2SPS approach can be extended to an additive hazard regression of the survival time on the fitted values $\hat{X}(t)$ from the first-stage regression (model 7.4). In particular, we show that the resulting effect β_x of $\hat{X}(t)$ can be interpreted as the effect of $X(t)$ in the additive hazard model

$$\lambda(t|\bar{X}(t), Z, U, \mathbf{C}) = \lambda_0(t) + \beta_x X(t) + \beta_c^T\mathbf{C} + \beta_u(t)U. \tag{7.5}$$

However, since $X(t)$ is a binary variable, model (7.4) is substituted by the logit model.

7.2.1 Model formulation with time-varying unmeasured confounders

Next, we show that the proposed IV approach for time-varying exposure is valid in the presence of time-varying unmeasured confounders. Consider, without loss of generality, a study design which intends to collect data on an instrumental variable Z at baseline (time $t = 0$), on exposure measurements $X(t)$ taken at discrete time points $t = 0, 1, ..., K$ and on time to death T. The exposure measurements are only recorded up to the time of death, the time L to censoring or the end K of the measurement period, whichever comes first. For each individual, we thus observe the following data: $Z, T^* \equiv \min(T, L), \Delta \equiv I(T \leq L), \bar{X} \equiv \min(T^*, K)$, where $\bar{X}(t)$ refers to the history of the exposure measurements $X(s)$ at the discrete times $s = 0, 1, ...$ up to time t. Further, let $U(t)$ be a vector of unmeasured variables such that for each time $t = 0, 1, ..., K$, the history $\bar{U}(t)$ along with $\bar{X}(t-1)$ is sufficient to adjust for confounding of the effect of $X(t)$ on T among individuals who are alive at that time (i.e., for whom $T \geq t$).

Throughout, we will assume that $U(t)$ at each time t is not influenced by the previous exposure measurements $\bar{X}(t-1)$. While this is a potentially strong assumption, it is one that is difficult to relax in IV analysis. Indeed, without this assumption, there may be pathways from the instrumental variable Z via $\bar{X}(t-1)$ and $U(t)$ towards the time-to-event endpoint, thereby inducing a dependence between the instrumental variable and the unmeasured confounders, and thus violating the instrumental variables assumptions. This assumption is therefore often implicitly implied by common IV analysis which ignore the time-varying nature of the exposure. More formally, we will impose the generalised IV assumption

$$U(\lfloor t \rfloor) \perp\!\!\!\perp Z | T \geq t, \bar{U}(\lfloor t \rfloor - 1), \tag{7.6}$$

at each measurement time $t = 0, 1, ..., K$, where we define $\bar{U}(-1) \equiv \emptyset$ and $\lfloor t \rfloor$ to be the integer value just below (or equal to) t. It is indeed seen from the causal diagram in Figure 7.1 that this assumption is justified when $U(t)$ at each time t is not influenced by the previous exposure measurements $\bar{X}(t-1)$.

Let $\lambda(t)$ be the conditional hazard at time t, given $\bar{X}(\lfloor t \rfloor), \bar{U}(\lfloor t \rfloor)$ and Z. Let $dN(t)$ be the

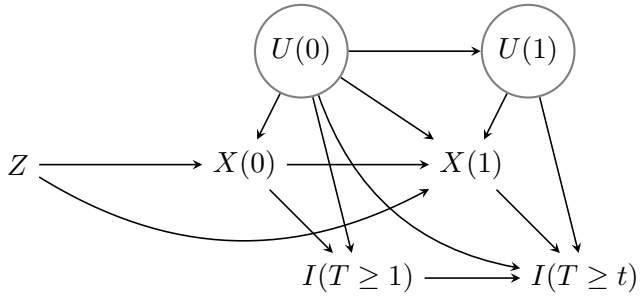


Figure 7.1: Causal diagram with time-varying exposure and unmeasured confounders

increment in the counting process related to the event time at time t. Then $\lambda(t)$ is more precisely defined as $E\left\{dN(t)|T \geq t, \bar{X}(\lfloor t \rfloor), \bar{U}(\lfloor t \rfloor), Z\right\}$. Consider the model

$$\lambda(t) = \omega\left\{t, \bar{U}(\lfloor t \rfloor)\right\} + \beta X(\lfloor t \rfloor). \tag{7.7}$$

The fact that the right hand side does not involve $\bar{X}(\lfloor t \rfloor - 1)$ embodies the assumption that early exposures $\bar{X}(\lfloor t \rfloor - 1)$ can only influence the hazard at time t by influencing the later exposure $\bar{X}(\lfloor t \rfloor)$. Making this assumption is partly a necessity, as the instrumental variable Z, which only takes 3 levels in our motivating application, does not carry sufficient information to be able to distinguish short-term from long-term exposure effects. Under the above assumptions (7.6) and (7.7), along with specific conditions to be specified later, we will demonstrate that the following estimating function for β

$$\left\{M(t) - E\left[M(t)|T \geq t\right]\right\} R(t) \left\{dN(t) - \beta M(t)dt\right\},$$

where $M(t) \equiv E\left\{X(\lfloor t \rfloor)|T \geq t, Z\right\}$, is unbiased.

We will first investigate under what conditions, in addition to assumptions (7.6) and (7.7), the generalised IV assumption

$$U(\lfloor t \rfloor) \perp\!\!\!\perp Z | T \geq s, \bar{U}(\lfloor t \rfloor - 1), \tag{7.8}$$

holds for all $s \geq \lfloor t \rfloor$. For this, note that, by Bayes' rule, for each t and $s \geq t$,

$$\begin{aligned} f(Z, U(\lfloor t \rfloor) | T \geq s, \bar{U}(\lfloor t \rfloor - 1)) &= f(Z | T \geq \lfloor t \rfloor, \bar{U}(\lfloor t \rfloor - 1)) f(U(\lfloor t \rfloor) | T \geq \lfloor t \rfloor, \bar{U}(\lfloor t \rfloor - 1)) \\ &\quad \times \frac{P(T \geq s | T \geq \lfloor t \rfloor, Z, \bar{U}(\lfloor t \rfloor))}{P(T \geq s | T \geq \lfloor t \rfloor, \bar{U}(\lfloor t \rfloor - 1))}, \end{aligned}$$

by assumption (7.6). Further, by assumption (7.7)

$$\begin{aligned} P(T \geq t | T \geq \lfloor t \rfloor, Z, \bar{U}(\lfloor t \rfloor)) &= \exp\left[- \int_{\lfloor t \rfloor}^{t} \omega \left\{ s, \bar{U}(\lfloor t \rfloor) \right\} ds \right] \exp\left[- \int_{\lfloor t \rfloor}^{t} \beta M(s) ds \right] \\ &\quad \times E\left\{ \exp\left[- \int_{\lfloor t \rfloor}^{t} \beta \Delta(s) ds \right] | T \geq \lfloor t \rfloor, Z, \bar{U}(\lfloor t \rfloor) \right\}, \end{aligned}$$

which factorises as a product $\phi_z(t, Z)\phi_u(t, \bar{U}(\lfloor t \rfloor))$ when

$$\Delta(t) \perp\!\!\!\perp Z | T \geq \lfloor t \rfloor, \bar{U}(\lfloor t \rfloor). \tag{7.9}$$

We conclude that (7.8) holds for $s < \lfloor t \rfloor + 1$. By repeated averaging, it is moreover seen that for $\lfloor t \rfloor + 1 \leq s < \lfloor t \rfloor + 2$:

$$\begin{aligned} P(T \geq s | T \geq \lfloor t \rfloor, Z, \bar{U}(\lfloor t \rfloor)) &= P(T \geq s | T \geq \lfloor t \rfloor + 1, Z, \bar{U}(\lfloor t \rfloor)) P(T \geq \lfloor t \rfloor + 1 | T \geq \lfloor t \rfloor, Z, \bar{U}(\lfloor t \rfloor)) \\ &= E\left\{ P(T \geq s | T \geq \lfloor t \rfloor + 1, Z, \bar{U}(\lfloor t \rfloor + 1)) | T \geq \lfloor t \rfloor + 1, Z, \bar{U}(\lfloor t \rfloor) \right\} \\ &\quad \times P(T \geq \lfloor t \rfloor + 1 | T \geq \lfloor t \rfloor, Z, \bar{U}(\lfloor t \rfloor)). \end{aligned}$$

Here, both terms $P(T \geq s | T \geq \lfloor t \rfloor + 1, Z, \bar{U}(\lfloor t \rfloor + 1)$ and $P(T \geq \lfloor t \rfloor + 1 | T \geq \lfloor t \rfloor, Z, \bar{U}(\lfloor t \rfloor))$ can be written as a product of a function of t and Z, and a function of t and $\bar{U}(\lfloor t \rfloor + 1)$. Since furthermore $U(\lfloor t \rfloor + 1) \perp\!\!\!\perp Z | T \geq \lfloor t \rfloor + 1, \bar{U}(\lfloor t \rfloor)$ by (7.6), this factorisation is maintained after averaging, so that (7.8) also holds for $\lfloor t \rfloor + 1 \leq s < \lfloor t \rfloor + 2$. Next, for $\lfloor t \rfloor + 2 \leq s < \lfloor t \rfloor + 3$:

$$\begin{aligned} P(T \geq s | T \geq \lfloor t \rfloor, Z, \bar{U}(\lfloor t \rfloor)) &= P(T \geq s | T \geq \lfloor t \rfloor + 2, Z, \bar{U}(\lfloor t \rfloor)) \\ &\quad \times P(T \geq \lfloor t \rfloor + 2 | T \geq \lfloor t \rfloor + 1, Z, \bar{U}(\lfloor t \rfloor)) \\ &\quad \times P(T \geq \lfloor t \rfloor + 1 | T \geq \lfloor t \rfloor, Z, \bar{U}(\lfloor t \rfloor)). \end{aligned}$$

Let us concentrate on the first term; the other two are similar to earlier terms and enjoy the

earlier factorisation. We have that

$$P(T \geq s|T \geq \lfloor t \rfloor + 2, Z, \bar{U}(\lfloor t \rfloor)) \quad = \quad E\left\{P(T \geq s|T \geq \lfloor t \rfloor + 2, Z, \bar{U}(\lfloor t \rfloor + 2))|T \geq \lfloor t \rfloor + 2, Z, \bar{U}(\lfloor t \rfloor\right\}.$$

Here, the term $P(T \geq s|T \geq \lfloor t \rfloor + 2, Z, \bar{U}(\lfloor t \rfloor + 2))$ can be written as a product of a function of t and Z, and a function of t and $\bar{U}(\lfloor t \rfloor + 2)$. Since furthermore $U(\lfloor t \rfloor + 2) \perp\!\!\!\perp Z|T \geq \lfloor t \rfloor + 2, \bar{U}(\lfloor t \rfloor + 1)$ by (7.6) and $U(\lfloor t \rfloor + 1) \perp\!\!\!\perp Z|T \geq \lfloor t \rfloor + 2, \bar{U}(\lfloor t \rfloor)$ by (7.8), this factorisation is maintained after averaging, so that (7.8) also holds for $\lfloor t \rfloor + 2 \leq s < \lfloor t \rfloor + 3$. Repeating along these lines, it is seen that (7.8) holds for all $s \geq t$.

Let us now revisit the unbiasedness of the estimating function under assumptions (7.7) and (7.8) (which, by the above result, are implied by assumptions (7.6), (7.7) and (7.9)). This function has expectation

$$\begin{aligned}
&E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t) \{dN(t) - \beta M(t)dt\}\right) \\
&= E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t) \left[\omega\left\{t, \bar{U}(\lfloor t \rfloor)\right\} dt + \beta\Delta(t)dt\right]\right) \\
&= E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t)\omega\left\{t, \bar{U}(\lfloor t \rfloor)\right\} dt\right),
\end{aligned}$$

where we make the testable assumption that $\Delta(t) \perp\!\!\!\perp Z|T \geq t$ (which is implied by (7.9)).

We further have that

$$\begin{aligned}
&E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t)\omega\left\{t, \bar{U}(\lfloor t \rfloor)\right\} dt\right) \\
&= E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t) E\left[\omega\left\{t, \bar{U}(\lfloor t \rfloor)\right\}|\bar{U}(\lfloor t \rfloor - 1), T \geq t, Z\right]\right) \\
&= E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t) E\left[\omega\left\{t, \bar{U}(\lfloor t \rfloor)\right\}|\bar{U}(\lfloor t \rfloor - 1), T \geq t\right]\right) \\
&= E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t)\omega^*\left\{t, \bar{U}(\lfloor t \rfloor - 1)\right\}\right),
\end{aligned}$$

for some function $\omega^*(.)$, where we use that

$$U(\lfloor t \rfloor) \perp\!\!\!\perp Z|T \geq t, \bar{U}(\lfloor t \rfloor - 1).$$

Continuing, working from the final display, we obtain that

$$\begin{aligned}
&E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t)\omega^*\left\{t, \bar{U}(\lfloor t \rfloor - 1)\right\}\right) \\
&= E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t) E\left[\omega^*\left\{t, \bar{U}(\lfloor t \rfloor - 1)\right\} | \bar{U}(\lfloor t \rfloor - 2), T \geq t, Z\right]\right) \\
&= E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t) E\left[\omega\left\{t, \bar{U}(\lfloor t \rfloor - 1)\right\} | \bar{U}(\lfloor t \rfloor - 2), T \geq t\right]\right) \\
&= E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t)\omega^{**}\left\{t, \bar{U}(\lfloor t \rfloor - 2)\right\}\right),
\end{aligned}$$

for some function $\omega^{**}(.)$, where we use that

$$U(\lfloor t \rfloor - 1) \perp\!\!\!\perp Z | T \geq t, \bar{U}(\lfloor t \rfloor - 2).$$

Further continuing, we eventually obtain that the estimating function has expectation

$$E\left(\{M(t) - E\left[M(t)|T \geq t\right]\} R(t)\tilde{\omega}\left\{t, U(0)\right\}\right),$$

for some function $\tilde{\omega}(.)$. Using that

$$U(0) \perp\!\!\!\perp Z | T \geq t,$$

it is immediate that the latter expectation equals zero. A consistent estimator of β can thus be obtained by solving an estimating equation based on the above estimating function. It follows from Vansteelandt et al. (2014b) that the Aalen least squares estimator under model

$$\tilde{\lambda}(t) = \psi_0(t) + \beta M(t)$$

solves precisely such equation, thus confirming the validity of the 2SPS estimator described in Section 7.2.

7.3 Results

A total of 13080 observations or study visits were used to describe the time-varying exposure, where 74% of the visits had exposure to ART (88%, 75% and 57% in the early, late integrated

and sequential arms respectively).

Table 7.1: IV estimates for the effect of time-varying exposure on mortality using additive hazards models

	Univariable		Multivariable[a]	
Method/exposure[b]	β	95% CI	β	95% CI
As-treated	-0.22	-0.30, -0.13	-0.21	-0.31, -0.13
2SPS	-0.41	-0.72, -0.09	-0.29	-0.54, -0.03

[a] adjusted for gender, baseline CD4+ count and employment status
[b] measure of compliance defined as 1 at time t when the considered patient was on ART at or prior to time t, and 0 otherwise

The IV analysis of time-varying exposure expresses a much higher effect of continuous ART exposure versus no exposure (hazard difference of -0.29; 95% CI: -0.54 to -0.03) (Table 7.1). It indicates that an average of 29 deaths were prevented for each year of follow-up in each 100 patients on (continuous) ART, compared with 100 patients not on ART, conditional on gender, employment status and CD4+ cell count.

The large differences found between the estimated effects of fixed and time-varying exposures are largely attributable to the different exposure definitions: compliance during TB treatment was relatively good, resulting in clear differences in exposure distribution between the different arms; however, since ART exposure was generally high for all patients outside the TB treatment window, differences between the arms were much less pronounced in terms of the time-varying exposure. A major drawback of our IV-methodology for time-varying exposures is that it relies on a location-shift assumption, which is unlikely to hold for dichotomous exposures.

7.4 Model-based predicted survival probabilities

The model-based survival probabilities were estimated for values, 1 (on ART) and 0 (not on ART) (Figure 7.2).

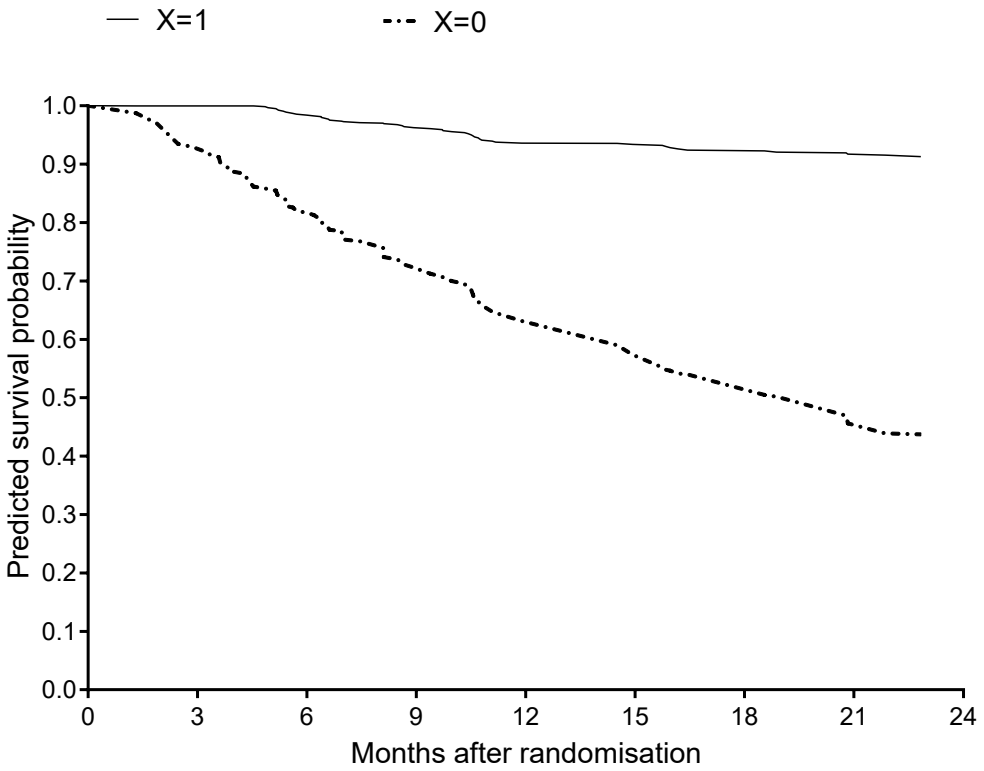


Figure 7.2: Predicted survival curves from additive hazards model with time-varying exposure

The predicted survival probabilities for $X = 1$ in Figure 7.2 are higher than that of the early integrated as seen in Figure 6.3. On the other hand, the predicted survival probabilities for a patient without ART exposure, that is, $X(t)$=0 for all t, were very low compared to those for $X = 0$ in Figure 6.3. $X = 0$ is obviously better than $X(t)$=0 because the former is an indication of delayed ART (i.e. ART initiated after TB treatment completion), whereas the latter indicates a patient who did not initiate ART at all.

7.5 Summary

In this Chapter, we extended the IV-methodology to time-varying measures of compliance. Our analysis of time-varying exposure is based on novel IV-methodology which returns effects on the additive hazard scale. We found that continuous use of ART averted on average of 29 deaths for each year of follow-up in each 100 patients on (continuous) ART, compared with 100 patients not

on ART, conditional on gender, employment status and CD4+ cell count. The large differences found between the estimated effects of fixed and time-varying exposures are largely attributable to the different exposure definitions: while the analysis of a fixed exposure focus on the effect on integration of ART and TB treatment (as compared to a regimen with delayed ART), the analysis of a time-varying exposure focuses on the overall effect of continuous use of ART as opposed to no ART at all. In contrast to G-estimation, it does have the drawback of making assumptions on the distribution of the exposure (as shown in Equation 7.4). However, the proposed methods can be applied using standard software and do not require specific corrections for non-informative censoring.

We have already mentioned that the analysis of fixed exposure is less reliable because of the previously mentioned concern for reverse causality, and the fact that differential ART exposure outside the period of TB treatment may have induced a violation of the exclusion restriction in this analysis. In contrast, also the time-varying IV analysis is subject to a violation of the assumption that the residual in Equation (7.4) is (conditionally) uncorrelated with the instrumental variable. Even though this could be fixed by using the control function approach (Tchetgen Tchetgen et al., 2015), it was shown to work for continuous exposures and unlikely to hold for binary variables, more especially time-varying binary exposures.

We showed that the proposed IV approach for time-varying exposure is valid in the presence of unmeasured time-varying confounders $U(t)$, so long as these are not influenced by previous exposure measurement $\bar{X}(t-1)$, under the data-generating mechanism visualised in Figure 7.1. While this is a potentially strong assumption, it is one that is implicit in all standard IV analyses that reduce time-varying exposure to a fixed exposure, and one that is difficult or impossible to relax. Indeed, without this assumption, there may be pathways from the instrumental variable Z via $\bar{X}(t-1)$ and $U(t)$ towards the time-to-event endpoint, thereby inducing a dependence between the instrumental variable and the unmeasured confounders, and thus violating the instrumental variables assumptions.

Chapter 8

Summary of results across different analytical approaches

In order to ease a comparison between different analytical methods or approaches and regressions models, it is helpful to restate different comparisons and also present a summary of the results. This way it is easy to make sense of all the results from different chapters. Here, we present results from semiparametric additive and Cox proportional hazards models under the ITT, "as treated"and IV analyses. The following comparisons were made:

1. Results from semiparametric additive hazards models (equations 5.2, 6.2 and 7.5) are presented in Table 8.1, where we compared hazard differences of
   fixed exposure
   a) randomisation arm (Z) under the ITT analyses
   b) exposure (X) under the "as treated"analyses
   c) exposure ($\hat{X}$) under the IV analyses using both 2SPS and 2SRI approaches
      time-varying exposure
   d) exposure ($X(t)$) under the "as treated"analyses
   e) exposure ($\hat{X}(t)$) under the IV analyses using the 2SPS approach.
2. Results from semiparametric Cox proportional hazards models (equations 5.3 and 6.3) are

presented in Table 8.2, where we compared hazard ratios of

a) randomisation arm (Z) under the ITT analyses

b) exposure (X) under the "as treated"analyses

c) exposure ($\hat{X}$) under the IV analyses using both 2SPS and 2SRI approaches.

Table 8.1: ITT and IV estimates for the effect of study arm and exposure on mortality using additive hazards models

		Univariate		Multivariate[a]	
Method	Arm/exposure type	β	95% CI	β	95% CI
ITT	Early integrated	-0.06	-0.11, -0.02	-0.05	-0.09, -0.01
	Late integrated	-0.06	-0.11, -0.02	-0.06	-0.11, -0.02
	Sequential	0		0	
As-treated	X	-0.12	-0.17, -0.08	-0.12	-0.17, -0.07
IV: 2SPS	X	-0.08	-0.14, -0.03	-0.07	-0.12, -0.01
IV: 2SRI	X	-0.06	-0.13, -0.01	-0.05	-0.11, 0.01
	First-stage residuals	-0.19	-0.35, -0.06	-0.23	-0.43, -0.09
As-treated	X(t)	-0.22	-0.30, -0.13	-0.21	-0.31, -0.13
IV: 2SPS	X(t)	-0.41	-0.72, -0.09	-0.29	-0.54, -0.03

[a] adjusted for gender, baseline CD4+ count and employment status

Table 8.2: ITT and IV estimates for the effect of study arm and exposure on mortality using proportional hazards models

		Univariable		Multivariable[a]	
Method	Arm/exposure type	HR	95% CI	HR	95% CI
ITT	Early integrated	0.44	0.25, 0.80	0.51	0.28, 0.92
	Late integrated	0.46	0.26, 0.83	0.46	0.26, 0.83
	Sequential	1.0		1.0	
As-treated	X	0.15	0.07, 0.31	0.14	0.06, 0.29
IV: 2SPS	X	0.30	0.12, 0.70	0.35	0.14, 0.85
IV: 2SRI	X	0.25	0.09, 0.64	0.26	0.09, 0.70
	First-stage residuals	0.24	0.08, 0.71	0.18	0.05, 0.64

HR:hazard ratio

[a] adjusted for gender, baseline CD4+ count and employment status

Chapter 9

Simulation study

We conducted a simulation study to investigate the potential bias from 2SPS and 2SRI approaches based on additive hazards models, where the instrument is a categorical variable with three levels. The aim was to assess the robustness of the IV models and quantify the amount of bias (if any) when using linear and binary exposures. In particular, when using time-varying binary exposure, the two assumptions might not hold:

1. $X(t) = E\{X(t) \mid T \geq t, Z\} + \Delta(t)$
2. $\Delta(t)$ is independent of Z given $T \geq t$

For fixed exposures, Tchetgen Tchetgen et al. (2015) were able to circumvent this problem when the instrumental variable is dichotomous; see also Martinussen et al. (2017b) for general instrumental variables. We used time-varying exposure in our analysis but only simulated datasets with fixed exposures because the simulation of time-varying exposures is complex and therefore was not performed. The R code for the simulation study is provided in Appendix 2.

Data was generated to mimic the SAPiT clinical trial structure. The variables were generated in such way that exposure (X) was influenced by the study arm (Z) and failure times were influenced by X alone or X and unmeasured confounders (U) as shown in Figure 2.1. Measured confounders were not included in the simulation study. Data were simulated for arbitrary

strengths of the exposure effect on time-to-event outcome, such as low (β_x between 0 and 0.3), moderate (β_x between 0.4 and 0.6) and large ($\beta_x > 0.7$) effect. The aim was to determine whether the amount of bias changes as the exposure effect becomes larger. The effect of unmeasured confounders on failure times and the sample size were kept constant as (β_u=0.1) and (N=1000) respectively. Sample size was kept constant because it has been shown elsewhere that an increase in sample size does not have major impact in the magnitude of the bias (Li et al., 2015). We simulated 1000 datasets for each scenario presented in Table 9.1.

Datasets for both linear and binary exposures were generated as follows:

1. Generate the instrumental variable Z (dummy coded), where each of the three study arm forms one third of the total sample size.

2. Generate linear exposure $X|Z$=$0.04 + 0.74Z_1 + 0.34Z_2+\varepsilon$, where Z_1 and Z_2 represents early and late integrated arms respectively and ε is an error term from normal with mean 0 and standard deviation 1. The values 0.74, 0.34 and 0.04 represent the mean exposure in the early, late integrated and sequential arms respectively.

3. Generate binary exposure $P(X = 1|Z) = \text{expit}(0.8Z_1 + 0.5Z_2)$, where 0.8 and 0.5 measures an increase in log-odds of exposure for early and late integrated arms when compared to the sequential arm.

4. Generate failure times using $\lambda(t|X, Z) = \lambda_0(t) + \beta_x X$, where the baseline hazard ($\lambda_0(t)$) was allowed to vary under different exposure effects to avoid getting negative failure times. The censoring took place at $t = 2.0$.

Using simulated data, we fitted additive hazards models under the 2SPS and 2SRI approaches. We further assessed whether the inclusion of unmeasured confounders either exacerbates or minimise bias (if any) in the 2SPS approach. Therefore, we generated data this way:

1. Generate $U \sim N(0, 1)$

2. Linear exposure was generated as $X|Z, U$=$0.04 + 0.74Z_1 + 0.34Z_2 + U$

3. Binary exposure was generated as $P(X = 1|Z, U) = \text{expit}(0.8Z_1 + 0.5Z_2 + U)$

4. Failure times were generated using $\lambda(t|X, Z, U) = \lambda_0(t) + \beta_x X + \beta_u U$. The baseline hazard ($\lambda_0(t)$) was allowed to vary under different exposure effects to avoid getting negative failure times and the censoring took place at $t = 2.0$. In the first stage, linear and logistic regression were used for linear and binary exposure respectively.

Bias under each scenario was calculated as $\frac{1}{1000}\sum_{j=1}^{1000}\left(\hat{\theta}_j - \theta\right)$.

Table 9.1: Average bias from 1000 simulations

		Linear exposure			Binary exposure		
Sample size	Exposure effect (θ)	Bias: 2SPS	Bias: 2SRI	Bias: 2SPS[a]	Bias: 2SPS	Bias: 2SRI	Bias: 2SPS[a]
1000	0	0.0012	-0.0007	0.0043	0.0084	-0.0426	-0.0120
1000	0.05	0.0033	-0.0259	0.0235	-0.0036	-0.0214	0.0487
1000	0.1	-0.0023	0.0049	-0.0034	-0.0404	-0.0091	0.0538
1000	0.3	-0.0274	0.0960	-0.0200	-0.0449	0.0968	-0.0606
1000	0.5	0.0061	0.1975	-0.0159	0.0401	0.1961	-0.0515
1000	0.8	0.0214	0.3608	0.0308	-0.0339	0.3420	-0.0579

[a] 2SPS approach adjusted for unmeasured confounding

Results in Table 9.1 show that both linear and binary exposure performed well under 2SPS approach even in the presence of unmeasured confounding. On the other hand, results from 2SRI approach produced biased estimate especially when the exposure effect is moderate or larger. These results suggest that the 2SPS approach is fairly robust when dichotomous exposures are used. Even though our results show that the degree of bias for binary exposure is small under these considered settings. However, we must warn the reader that no guarantees can be given that the proposed approach for time-varying exposures is (nearly) unbiased when the exposure is dichotomous, unless when the exposure effect is low.

Chapter 10

Concluding remarks

In this thesis we focused on instrumental variables analysis to account for non-compliance in an open label RCT which was designed to determine the optimal time to initiate ART in patients on tuberculosis treatment. We acknowledge that IV method can produce biased estimates unless the instrumental variable meets the three crucial assumptions as well as the treatment homogeneity assumption. Two of the three assumptions are untestable. The bias will be exacerbated when the instrument is weak. We showed that the testable IV assumption was met in our analysis since the instrument was strong and evidently associated with the exposure (i.e. partial F-statistics less $\geq$ 10) as shown in Table 6.2. The instrumental variable was randomised and therefore exchangeability is guaranteed. We also showed that the instrumental variable was not associated with measured confounders (Table 4.1) and therefore not expected to share common causes with the time-to-event outcome.

The measure of compliance used in Chapter 6, focused on the level of TB and ART integration. Non-compliance was as a result of patients not starting ART at correct time after TB treatment initiation. Results from ITT and IV analyses were not very different and this was either due to the fact that the level of non-compliance in the SAPiT trial was not very high or the issues of reverse causality which affected the fixed exposure definition. Producing estimates that take non-compliance to account is relevant in South African setting where TB and ART integration remains an important topic due to a larger number of patients who are co-infected.

In Chapter 7 we contributed to the IV literature by developing novel IV-methodology for time-varying measure of compliance, which is currently lacking in IV research. As much as this was applied using dichotomous exposure variable, it can be easily adapted to other exposure types. One caveat about our novel IV approach is that it puts restriction on the exposure distribution which should be correctly specified, and therefore has to be modelled using the two-stage method. In future, we plan to develop or apply (in case it is developed by then) an instrumental variables approach that incorporates time-varying exposure but does not require two-stage modelling and thus avoid restrictions on the exposure distribution. While more reliable, the analysis of time-varying exposure did not immediately allow us to assess the effectiveness of ART treatment during the TB treatment window because the average TB duration was short (i.e. on average six months). Furthermore, since the location-shift assumption is unlikely to hold for this particular exposure, therefore we cannot exclude the possibility of large biases when the exposure effect is large. It would be of interest to evaluate how the effect of time-varying ART exposure is modified by TB treatment. This will require novel methodology to include interactions between time-varying TB treatment and ART exposure in the model.

Some advances have been made in IV methodology for time-to-event outcomes including ours, however most of those advances are built around additive hazards models. The Cox proportional hazards model is the most popular and widely used regression for time-to-event data in health research. However, the IV developments around this famous time-to-event regression are very slow and we hope that recent work from Martinussen et al. (2017a) for IV under a structural Cox model and that of MacKenzie et al. (2014) will stimulate more developments. In the same breath, we hope this work will persuade statisticians in health research to utilise additive hazards model in their analysis especially when the proportional hazards assumption is violated. The advantage of using additive hazards models is that the hazard difference has collapsibility properties compared to the hazard ratios. Basically, adjusting for variables that are not associated with either the exposure or outcome should not change the hazard difference but it can potentially change the magnitude of the hazard ratios. The interpretation of the hazard difference is also straightforward because it gives an indication of how many cases will

either be averted or detected depending on the nature of the time-to-event outcome, which can be useful for public health planning and community interventions. The underutilisation of additive hazards models was probably due the fact that for the longest time it could not be easily applied in many statistical softwares other than in SAS using macros developed by Howell and Kein (1998). Currently, additive hazards models including the application of our method can be easily applied using the `TIMEREG` package in R, and hopefully its awareness will increase even outside the IV topic.

In RCTs, "as-treated", "safety"and "per-protocol"analyses are often used in conjunction with the ITT analysis. Even though results from IV analysis should be interpreted with caution, keeping in mind all the assumptions needed to render causal estimands, but we encourage RCTs with non-compliance to also add IV analysis as part of the sensitivity or secondary analysis. This will offer readers, patients and healthcare providers flexibility to choose an estimand that resonate with them or their situations. The advantage of using IV analysis in this thesis was that the randomisation arm was used as an instrument, whereas in non-randomisation studies it is difficult to find an instrument. However, it was difficult to find a reliable exposure since we did not have drug level data to confirm whether patients were actually taking their treatment and moreover, adherent to it.

IV methods sometimes produce estimated treatment effect that may not be generalised to the entire population when there is treatment effect heterogeneity. We note that our analysis infer treatment effect in the treated, and thereby assume treatment effect homogeneity (i.e. patients with a given ART exposure level on the different arms on average experience the same benefit of it). We do not infer the effect in compliers, as this would be more difficult to define with a time-varying and/or continuous measure of compliance. We are moreover concerned about the usefulness of inferring treatment effects for a subgroup of individuals that we cannot identify.

References

Aalen, O. O. (1989). A linear regression model for the analysis of life times, *Stat Med* **8**(8): 907–25.

Abdool Karim, S. S., Naidoo, K., Grobler, A., Padayatchi, N., Baxter, C., Gray, A., Gengiah, T., Nair, G., Bamber, S., Singh, A., Khan, M., Pienaar, J., El-Sadr, W., Friedland, G. and Abdool Karim, Q. (2010). Timing of initiation of antiretroviral drugs during tuberculosis therapy, *N Engl J Med* **362**(8): 697–706.

Abdool Karim, S. S., Naidoo, K., Grobler, A., Padayatchi, N., Baxter, C., Gray, A. L., Gengiah, T., Gengiah, S., Naidoo, A., Jithoo, N., Nair, G., El-Sadr, W. M., Friedland, G. and Abdool Karim, Q. (2011). Integration of antiretroviral therapy with tuberculosis treatment, *N Engl J Med* **365**(16): 1492–501.

Ananat, E. O. (2011). The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality, *American Economic Journal: Applied Economics* **3**(2): 34–66.

Angrist, J. D., Imbens, G. W. and Krueger, A. B. (1999). Jackknife instrumental variables estimation, *Journal of Applied Econometrics* **14**(1): 57–67.

Angrist, J. D. and Krueger, A. B. (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples, *Journal of the American Statistical Association* **87**(418): 328–336.

Angrist, J., Imbens, G. and Rubin, D. B. (1996). Identification of causal effects using instrumental variables, *Journal of the American Statistical Association* **91**(434): 444–455.

Austin, N., Harper, S. and Strumpf, E. (2016). Does segregation lead to lower birth weight?: An instrumental variable approach, *Epidemiology* **27**(5): 682–9.

Baiocchi, M., Cheng, J. and Small, D. S. (2014). Instrumental variable methods for causal inference, *Stat Med* **33**(13): 2297–340.

Bijwaard, G. E. and Ridder, G. (2005). Correcting for selective compliance in a re-employment bonus experiment, *Journal of Econometrics* **125**(1-2): 77–111.

Blanc, F. X., Sok, T., Laureillard, D., Borand, L., Rekacewicz, C., Nerrienet, E., Madec, Y., Marcy, O., Chan, S., Prak, N., Kim, C., Lak, K. K., Hak, C., Dim, B., Sin, C. I., Sun, S., Guillard, B., Sar, B., Vong, S., Fernandez, M., Fox, L., Delfraissy, J. F. and Goldfeld, A. E. (2011). Earlier versus later start of antiretroviral therapy in HIV-infected adults with tuberculosis, *N Engl J Med* **365**(16): 1471–81.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak, *Journal of the American Statistical Association* **90**(430): 443–450.

Breen, R. A., Smith, C. J., Bettinson, H., Dart, S., Bannister, B., Johnson, M. A. and Lipman, M. C. (2004). Paradoxical reactions during tuberculosis treatment in patients with and without HIV co-infection, *Thorax* **59**(8): 704–7.

Brookhart, M. A., Wang, P. S., Solomon, D. H. and Schneeweiss, S. (2006). Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable, *Epidemiology* **17**(3): 268–75.

Burman, W., Weis, S., Vernon, A., Khan, A., Benator, D., Jones, B., Silva, C., King, B., LaHart, C., Mangura, B., Weiner, M. and El-Sadr, W. (2007). Frequency, severity and duration of immune reconstitution events in HIV-related tuberculosis, *Int J Tuberc Lung Dis* **11**(12): 1282–9.

Carroll, R. J. and Stefanski, L. A. (1994). Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses, *Statistics in Medicine* **13**(12): 1265–1282.

Carslake, D., Fraser, A., Davey Smith, G., May, M., Palmer, T., Sterne, J., Silventoinen, K., Tynelius, P., Lawlor, D. A. and Rasmussen, F. (2013). Associations of mortality with own height using son's height as an instrumental variable, *Econ Hum Biol* **11**(3): 351–9.

Collett, D. (1994). *Modelling Survival Data in Medical Research*, London: Chapman & Hall.

Cornell, M., Schomaker, M., Garone, D. B., Giddy, J., Hoffmann, C. J., Lessells, R., Maskew, M., Prozesky, H., Wood, R., Johnson, L. F., Egger, M., Boulle, A. and Myer, L. (2012). Gender differences in survival among adult patients starting antiretroviral therapy in South Africa: a multicentre cohort study, *PLoS Med* **9**(9): e1001304.

Darcis, G., Lambert, I., Sauvage, A. S., Frippiat, F., Meuris, C., Uurlings, F., Lecomte, M., Leonard, P., Giot, J. B., Fombellida, K., Vaira, D. and Moutschen, M. (2018). Factors associated with late presentation for HIV care in a single Belgian reference center: 2006-2017, *Sci Rep* **8**(1): 8594.

Davies, N. M., Smith, G. D., Windmeijer, F. and Martin, R. M. (2013). Issues in the reporting and conduct of instrumental variable studies: a systematic review, *Epidemiology* **24**(3): 363–9.

Department of Health (2004). National antiretroviral treatment guidelines. Pretoria, South Africa.

Department of Health (2010). The South African Antiretroviral Treatment Guidelines.

Department of Health (2014). National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults. 2014. http://www.kznhealth.gov.za/family/hiv-guidelines-jan2015.pdf. Accessed April 4, 2016.

Department of Health (2015). National Department of Health. National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults. 2015. http://www.sahivsoc.org/files/art%20guidelines%2015052015.pdf. Accessed April 31, 2016.

Department of Health (2016). Department of Health, South Africa. Universal test and treat. http://www.sahivsoc.org/files/22%208%2016%20circular%20utt%20%20%20decongestion%20ccmt%20di
Accessed April 4, 2018.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals, *Statistical Science* **11**(3): 189–228.

Didelez, V. and Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference, *Stat Methods Med Res* **16**(4): 309–30.

Efron, B. (1979). Bootstrap methods: another look at the jackknife, *The Annals of Statistics* **7**(1): 1–26.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman and Hall.

Fomundam, H. N., Tesfay, A. R., Mushipe, S. A., Mosina, M. B., Boshielo, C. T., Nyambi, H. T., Larsen, A., Cheyip, M., Getahun, A. and Pillay, Y. (2017). Prevalence and predictors of late presentation for HIV care in South Africa, *S Afr Med J* **107**(12): 1058–1064.

Glymour, M. (2006). *Natural experiments and instrumental variable analyses in social epidemiology. In JMichael Oakes, Jay S Kaufman (ed) Methods in social epidemiology.*, San Francisco, Jossey-Bass.

Goldberger, A. (1972). Structural equation methods in social-sciences, *Econometrica* **40**(6): 979–1001.

Gore, J. L., Litwin, M. S., Lai, J., Yano, E. M., Madison, R., Setodji, C., Adams, J. L., Saigal, C. S. and Urologic Diseases in America, P. (2010). Use of radical cystectomy for patients with invasive bladder cancer, *J Natl Cancer Inst* **102**(11): 802–11.

Gray, R. and Wheatley, K. (1991). How to avoid bias when comparing bone marrow transplantation with chemotherapy, *Bone Marrow Transplant* **7 Suppl 3**: 9–12.

Greenland, S. (2000). An introduction to instrumental variables for epidemiologists, *Int J Epidemiol* **29**(4): 722–9.

Grimes, D. A. and Schulz, K. F. (2002). An overview of clinical research: the lay of the land, *Lancet* **359**(9300): 57–61.

Hadley, J., Yabroff, K. R., Barrett, M. J., Penson, D. F., Saigal, C. S. and Potosky, A. L. (2010). Comparative effectiveness of prostate cancer treatments: evaluating statistical adjustments for confounding in observational data, *J Natl Cancer Inst* **102**(23): 1780–93.

Hausman, J. A. (1978). Specification tests in econometrics, *Econometrica* **46**(6): 1251–1271.

Havlir, D. V., Kendall, M. A., Ive, P., Kumwenda, J., Swindells, S., Qasba, S. S., Luetkemeyer, A. F., Hogg, E., Rooney, J. F., Wu, X., Hosseinipour, M. C., Lalloo, U., Veloso, V. G., Some, F. F., Kumarasamy, N., Padayatchi, N., Santos, B. R., Reid, S., Hakim, J., Mohapi, L., Mugyenyi, P., Sanchez, J., Lama, J. R., Pape, J. W., Sanchez, A., Asmelash, A., Moko, E., Sawe, F., Andersen, J. and Sanne, I. (2011). Timing of antiretroviral therapy for HIV-1 infection and tuberculosis, *N Engl J Med* **365**(16): 1482–91.

Hernan, M. A. and Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream?, *Epidemiology* **17**(4): 360–72.

Howell, A. M. and Kein, J. (1998). A SAS Macro for the Additive Hazards Regression Model.

Human Sciences Research Council (2018). The Fifth South African National HIV Prevalence, Incidence, Behaviour and Communication Survey, 2017: HIV Impact Assessment Summary Report. Cape Town, HSRC Press. http://www.hsrc.ac.za/uploads/pagecontent/9225/ sabssmv_impact_assessment_summary_za_ads_cleared1%20(002).pdf. Accessed October 20, 2018.

Jackson, J. W. and Swanson, S. A. (2015). Toward a clearer portrayal of confounding bias in instrumental variable applications, *Epidemiology* **26**(4): 498–504.

Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance, *Psychol Methods* **7**(2): 178–93.

Johnson, J. (1963). *Econometric Methods*, McGraw-Hill: New York.

Kigozi, I. M., Dobkin, L. M., Martin, J. N., Geng, E. H., Muyindike, W., Emenyonu, N. I., Bangsberg, D. R. and Hahn, J. A. (2009). Late-disease stage at presentation to an HIV clinic in the era of free antiretroviral therapy in Sub-Saharan Africa, *J Acquir Immune Defic Syndr* **52**(2): 280–9.

Korn, E. L. and Baumrind, S. (1998). Clinician preferences and the estimation of causal treatment differences, *Statistical Science* **13**(3): 209–227.

Lalloo, U. G. (2009). Efavirenz and nevirapine interactions with rifampicin: resolving the dilemmas?, *Clin Infect Dis* **48**(12): 1760–2.

Larsen, A., Cheyip, M., Tesfay, A., Vranken, P., Fomundam, H., Wutoh, A. and Aynalem, G. (2018). Timing and predictors of initiation on antiretroviral therapy among newly-diagnosed HIV-infected persons in South Africa, *AIDS Behav* [**Epub ahead of print**].

Lawn, S. D., Harries, A. D., Anglaret, X., Myer, L. and Wood, R. (2008). Early mortality among adults accessing antiretroviral treatment programmes in sub-Saharan Africa, *AIDS* **22**(15): 1897–908.

Lawn, S. D., Myer, L., Orrell, C., Bekker, L. G. and Wood, R. (2005). Early mortality among adults accessing a community-based antiretroviral service in South Africa: implications for programme design, *AIDS* **19**(18): 2141–8.

Li, J., Fine, J. and Brookhart, A. (2015). Instrumental variable additive hazards models, *Biometrics* **71**(1): 122–30.

Little, R. J., Long, Q. and Lin, X. (2009). A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance, *Biometrics* **65**(2): 640–9.

Lundgren, J., Babiker, A. and Gordin, F. (2015). INSIGHT START Study Group. Initiation of antiretroviral therapy in early asymptomatic HIV infection, *New England Journal of Medicine* **373**(9): 795–807.

MacKenzie, T. A., Tosteson, T. D., Morden, N. E., Stukel, T. A. and A., O. J. (2014). Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding, *Health Serv Outcomes Res Method* **14**(1-2): 54–68.

Martens, E. P., Pestman, W. R., de Boer, A., Belitser, S. V. and Klungel, O. H. (2006). Instrumental variables: application and limitations, *Epidemiology* **17**(3): 260–7.

Martinussen, T., Norbo Sorensen, D. and Vansteelandt, S. (2017a). Instrumental variables estimation under a structural Cox model, *Biostatistics* .

Martinussen, T., Vansteelandt, S., Tchetgen Tchetgen, E. J. and Zucker, D. M. (2017b). Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models, *Biometrics* **73**(4): 1140–1149.

McClellan, M., McNeil, B. J. and Newhouse, J. P. (1994). Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? analysis using instrumental variables, *JAMA* **272**(11): 859–66.

Naidoo, A., Naidoo, K., Yende-Zuma, N., Gengiah, T. N., Padayatchi, N., Gray, A. L., Bamber, S., Nair, G. and Karim, S. S. (2014). Changes to antiretroviral drug regimens during integrated TB-HIV treatment: results of the SAPiT trial, *Antivir Ther* **19**(2): 161–9.

Naidoo, K., Hassan-Moosa, R., Yende-Zuma, N., Govender, D., Padayatchi, N., Dawood, H., Adams, R. N., Govender, A., Chinappa, T., Abdool-Karim, S. and Abdool-Karim, Q. (2017). High mortality rates in men initiated on anti-retroviral treatment in KwaZulu-Natal, South Africa, *PLoS One* **12**(9): e0184124.

Naidoo, K., Yende-Zuma, N., Padayatchi, N., Naidoo, K., Jithoo, N., Nair, G., Bamber, S., Gengiah, S., El-Sadr, W. M., Friedland, G. and Abdool Karim, S. (2012). The immune reconstitution inflammatory syndrome after antiretroviral therapy initiation in patients with tuberculosis: findings from the SAPiT trial, *Ann Intern Med* **157**(5): 313–24.

Newhouse, J. P. and McClellan, M. (1998). Econometrics in outcomes research: the use of instrumental variables, *Annu Rev Public Health* **19**: 17–34.

Nyika, H., Mugurungi, O., Shambira, G., Gombe, N. T., Bangure, D., Mungati, M. and Tshimanga, M. (2016). Factors associated with late presentation for HIV/AIDS care in Harare City, Zimbabwe, 2015, *BMC Public Health* **16**: 369.

Padayatchi, N., Abdool Karim, S. S., Naidoo, K., Grobler, A. and Friedland, G. (2014). Improved survival in multidrug-resistant tuberculosis patients receiving integrated tuberculosis and antiretroviral treatment in the SAPiT trial, *Int J Tuberc Lung Dis* **18**(2): 147–54.

Permutt, T. and Hebel, J. R. (1989). Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight, *Biometrics* **45**(2): 619–22.

Perumal, R., Padayatchi, N., Naidoo, K. and Knight, S. (2014). Understanding the profile of tuberculosis and human immunodeficiency virus coinfection: Insights from expanded HIV surveillance at a tuberculosis facility in Durban, South Africa, *ISRN AIDS* **2014**: 260329.

Rassen, J. A., Brookhart, M. A., Glynn, R. J., Mittleman, M. A. and Schneeweiss, S. (2009). Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships, *J Clin Epidemiol* **62**(12): 1226–32.

Robins, J. and Rotnitzky, A. (2004). Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models, *Biometrika* **91**(4): 763–783.

Robins, J. and Tsiatis, A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models, *Communications in Statistics, Theory and Methods* **20**(8): 2609–2631.

Sexton, M. and Hebel, J. R. (1984). A clinical trial of change in maternal smoking and its effect on birth weight, *JAMA* **251**(7): 911–5.

Sheiner, L. B. and Rubin, D. B. (1995). Intention-to-treat analysis and the goals of clinical trials, *Clin Pharmacol Ther* **57**(1): 6–15.

Shisana, O., Rehle, T., Simbayi, L. C., Zuma, K., Jooste, S., Zungu, N., Labadarios, D. and Onoya, D. (2012). South African National HIV Prevalence, Incidence and Behaviour Survey. Cape Town, HSRC Press. Accessed on 19 March 2015 at :http://www.hsrc.ac.za/uploads/pageContent/4565/SABSSM%20IV%20LEO%20final.

Smith, G. D. and Ebrahim, S. (2004). Mendelian randomization: prospects, potentials, and limitations, *Int J Epidemiol* **33**(1): 30–42.

Statistics South Africa (2014). Mortality and causes of death in South Africa, 2014: Findings from death notification. http://www.statssa.gov.za/publications/p03093/p030932014.pdf. Accessed March 14, 2016.

Stefanski, L. A. (1996). Instrumental variable estimation in generalized linear measurement error models au - buzas, jeffrey s, *Journal of the American Statistical Association* **91**(435): 999–1006.

Stukel, T. A., Fisher, E. S., Wennberg, D. E., Alter, D. A., Gottlieb, D. J. and Vermeulen, M. J. (2007). Analysis of observational studies in the presence of treatment selection bias: effects of

invasive cardiac management on ami survival using propensity score and instrumental variable methods, *JAMA* **297**(3): 278–85.

Swanson, S. A. and Hernan, M. A. (2013). How to report instrumental variable analyses (suggestions welcome), *Epidemiology* **24**(3): 370–374.

Tan, H. J., Norton, E. C., Ye, Z., Hafez, K. S., Gore, J. L. and Miller, D. C. (2012). Long-term survival following partial vs radical nephrectomy among older patients with early-stage kidney cancer, *JAMA* **307**(15): 1629–35.

Tchetgen Tchetgen, E. J., Walter, S., Vansteelandt, S., Martinussen, T. and Glymour, M. (2015). Instrumental variable estimation in a survival context, *Epidemiology* **26**(3): 402–10.

Terza, J. V., Basu, A. and Rathouz, P. J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling, *J Health Econ* **27**(3): 531–43.

Vansteelandt, S., Babanezhad, M. and Goetghebeur, E. (2009). Correcting instrumental variables estimators for systematic measurement error, *Stat Sin* **19**: 1223–1246.

Vansteelandt, S., Bowden, J., Babanezhad, M. and Goetghebeur, E. (2011). On instrumental variables estimation of causal odds ratios, *Statistical Science* **26**(3): 403–422.

Vansteelandt, S. and Didelez, V. (2018). Improving the robustness and efficiency of covariate-adjusted linear instrumental variable estimators, *Scandinavian Journal of Statistics* **45**(4): 941–961.

Vansteelandt, S. and Joffe, M. (2014a). Structural nested models and g-estimation: The partially realized promise, *Statistical Science* **29**(4): 707–731.

Vansteelandt, S., Martinussen, T. and Tchetgen Tchetgen, E. J. (2014b). On adjustment for auxiliary covariates in additive hazard models for the analysis of randomized experiments, *Biometrika* **101**(1): 237–244.

Wooldridge, J. (2002). Econometric analysis of cross section and panel data. Cambridge, Mass.: MIT Press.

World Health Organization (2009). Rapid advice: antiretroviral therapy for HIV infection in adults and adolescents. 2009. geneva. http://www.who.int/hiv/pub/arv/rapid_advice_art.pdf.

World Health Organization (2015). HIV-Associated Tuberculosis. TB/HIV fact sheets. 2015. http://www.who.int/tb/challenges/hiv/tbhiv_factsheet_2015.pdf. Accessed March 10, 2015.

World Health Organization (2016a). Consolidated guidelines on the use of antiretroviral drugs for treating and preventing hiv infection. geneva. http://www.who.int/hiv/pub/guidelines/en/. Accessed July 07, 2018.

World Health Organization (2016b). Global Tuberculosis Control 2016, Geneva. www.who.int/tb/en. Accessed Apri 4, 2018.

Wright, S. (1928). *Appendix to the Tariff on Animal and Vegetable Oils*, New York: MacMillan.

Yende-Zuma, N. and Naidoo, K. (2016). The effect of timing of initiation of antiretroviral therapy on loss to follow-up in HIV-tuberculosis coinfected patients in South Africa: An open-label, randomized, controlled trial, *J Acquir Immune Defic Syndr* **72**(4): 430–6.

Appendix 1: Manuscript

Adjusting the Effect of Integrating Antiretroviral Therapy and Tuberculosis Treatment on Mortality for Noncompliance

A Time-varying Instrumental Variables Analysis

Nonhlanhla Yende-Zuma,[a,b,c] *Henry Mwambi,*[d] *and Stijn Vansteelandt*[e,f]

Background: Using intent-to-treat comparisons, it has been shown that the integration of antiretroviral therapy (ART) and tuberculosis (TB) treatment improves survival. Because the magnitude of the effect of ART initiation during TB treatment on mortality is less well understood owing to noncompliance, we used instrumental variables (IV) analyses.

Methods: We studied 642 HIV-TB co-infected patients from the Starting Antiretroviral Therapy at Three Points in Tuberculosis trial. Patients were assigned to start ART either early or late during TB treatment or after TB treatment completion. We used 2-stage predictor substitution and 2-stage residuals inclusion methods under additive and proportional hazards regressions with a time-fixed measure of compliance defined as the fraction of time on ART during TB treatment. We moreover developed novel IV methods for additive hazards regression with a time-varying measure of compliance.

Results: Intent-to-treat results from additive hazards models showed that patients in the early integrated arms had a reduced hazard of -0.05 (95% confidence interval [CI]: -0.09, -0.01) when compared with the sequential arm. Adjustment for noncompliance changed this effect to -0.07 (95% CI: -0.12, -0.01). An additional time-varying IV analysis on the overall effect of ART exposure suggested an effect of -0.29 (95 % CI: -0.54, -0.03).

Conclusion: IV analyses enable assessment of the effectiveness of TB and ART integration, corrected for noncompliance, and thereby enable a better public health evaluation of the potential impact of this intervention.

Keywords: Additive hazards model; Instrumental variable; Noncompliance; Time-varying exposure; HIV; TB

(*Epidemiology* 2019;30: 197–203)

Submitted September 10, 2016; accepted September 17, 2018.

From the [a]Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, Durban, South Africa; [b]MRC-CAPRISA HIV-TB Pathogenesis and Treatment Research Unit, Durban, South Africa; [c]The South African DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), University of Stellenbosch, Stellenbosch, South Africa; [d]School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa; [e]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, Belgium; and [f]Medical Statistics Unit, the London School of Hygiene and Tropical Medicine, London, United Kingdom.

Nonhlanhla Yende-Zuma is supported by CAPRISA, which was established as part of the Comprehensive International Program of Research on AIDS (CIPRA) (grant # AI51794) from the US National Institutes of Health. The U.S. President's Emergency Plan for AIDS Relief (PEPfAR) funded the care of all the participants in the trial. The Global Fund to fight AIDS, Tuberculosis and Malaria funded the cost of the drugs used in the trial. The research infrastructure to conduct this trial, including the data management, laboratory and pharmacy cores were established through the CIPRA grant. Nonhlanhla Yende-Zuma is also supported by the South African Medical Research Council (SAMRC). This publication was made possible by grant number 5R24TW008863 from the Office of Global AIDS Coordinator and the U. S. Department of Health and Human Services, National Institutes of Health (NIH OAR and NIH OWAR). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the government. The support of the DST-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), University of Stellenbosch toward this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to SACEMA.

The authors report no conflicts of interest.

Availability of data and code: Dataset and code will be made available upon request but access will be controlled and each request will be considered on a case by case basis.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Nonhlanhla Yende-Zuma, Centre for the AIDS Programme of Research in South Africa (CAPRISA), 2nd floor K-RITH tower, University of KwaZulu-Natal, 719 Umbilo Road, Congella, Durban, South Africa, 4013. E-mail: nonhlanhla.yende@caprisa.org.

ISSN: 1044-3983/19/3002-0197
DOI: 10.1097/EDE.0000000000000923

In 2014, an estimated 1.2 million people were co-infected with tuberculosis (TB) and HIV, around 74% of them living in sub-Saharan Africa.[1] In some parts of South Africa, it is estimated that almost 70% of TB patients are co-infected with HIV.[2] South Africa has the largest antiretroviral therapy (ART) roll-out programme with approximately 3 million people on ART in 2015.[3] Despite the size of the ART roll-out and wide availability of TB treatment, TB and HIV are reported to be the leading causes of death in South Africa in the age group 15–44 years.[4]

Initiation of ART within 4 weeks after the start of TB therapy or within 4 weeks after the completion of the intensive phase of TB therapy has been shown to improve survival,[5,6] especially among patients with low CD4+ cell count.[7,8] In

South Africa, TB patients were the first ones to be initiated on ART irrespective of their CD4+ cell count.[9] Patients who develop TB while on ART should continue with ART throughout TB therapy duration. One of the studies that informed guidelines on treatment integration was the Starting Antiretroviral Therapy at Three Points in Tuberculosis (SAPiT) trial. SAPiT was an open-label randomized controlled trial where patients were randomly assigned to start ART either early or late during TB treatment or after the completion of TB therapy. Irrespective of the randomization arm, patients in the SAPiT study could be started on ART at any time by study clinicians or personal physicians at their discretion. In the current analyses, noncompliance is defined as not starting ART at the correct time with respect to TB treatment, regardless of whether that was enforced by clinicians or by patients themselves.

However, this definition of noncompliance does not take into account any temporary or permanent discontinuation of study drugs and adherence. Among the 362 patients who initiated ART in early and late integrated arms of the SAPiT trial, 22.4% did not initiate ART at the correct time.[8] Moreover, following recommendations by data and safety monitoring committee, some patients started ART earlier than the protocol specified time[5] and are regarded as noncompliant.

To preserve the balance brought about by randomization, intent-to-treat (ITT) comparison was used for the primary analyses of the trial. This provides valid estimates of the effect of randomized assignment, but likely underestimates the effectiveness of treatment integration in the presence of noncompliance. To adjust for noncompliance, "as-treated" and "per-protocol" comparisons are commonly made. These contrast study participants according to their received treatment, regardless of the treatment arm to which they were assigned, or limit the analysis to participants who followed the protocol. Such analyses are generally biased because the subgroups that they compare often lack comparability.

In view of the shortcomings of the "as-treated" and "per-protocol" analyses, our objective was to account for noncompliance by using instrumental variables (IV) analyses to estimate the effect of ART initiation during TB treatment (exposure) on mortality. This will also be referred to as effectiveness. IV analyses enable us to make use of the comparability offered by randomization and thereby have the capability of adjusting for unmeasured and measured confounders; they have the further advantage of yielding results that are less sensitive to random measurement error in the exposure.[10] The key challenge with IV analyses especially in nonrandomized studies is obtaining a valid instrument, which must (1) be associated with an exposure; (2) only affect the outcome through its association with an exposure; and (3) not share common causes with an exposure. We used randomization arm as an instrument, with the exposure being defined as the fraction of time on ART during TB treatment (i.e., months on ART/months on TB therapy).

A limitation of using such fixed exposure over time is that it cannot capture the full complexity of compliance behavior and may, moreover, be indirectly influenced by censoring or death. In view of this, we also provide analyses for a time-varying measure of compliance to ART. IV-methodology for the effect of a time-varying exposure on a time-to-event endpoint is currently lacking, with the exception of G-estimation for structural accelerated failure time models.[11] Because this is complex and often performs poorly in the presence of censoring, we developed novel methodology under so-called additive hazard models. The proposed methods can be applied using standard software and do not require specific corrections for noninformative censoring.

METHODS

Dataset

This analysis is based on 642 HIV-TB co-infected patients from the SAPiT open-label randomized trial that was conducted between June 2005 and July 2010 in South Africa. The primary objective of the trial was to determine the optimal timing of ART initiation in patients co-infected with HIV and TB. More details about the study and the results for primary and secondary outcomes have been published in detail elsewhere.[5,8,12,13]

Patients who had confirmed HIV infection and newly diagnosed pulmonary TB were randomly assigned to start ART at the following 3 different points of their TB therapy. In the first arm, ART was to be initiated within 4 weeks after the start of TB therapy (early integrated arm). In the second arm, ART was to be initiated within 4 weeks after the completion of the intensive phase of TB therapy (late integrated arm). In the third arm, ART was to be initiated within 4 weeks after the completion of TB therapy (sequential arm). All patients received prophylaxis to control opportunistic infections. After a planned interim analysis, on September 1, 2008, almost 2 months after completion of enrollment, the data and safety monitoring committee made a recommendation that all patients in the sequential arm be initiated on ART as soon as possible but stay in follow-up until study completion. The committee also recommended continuation of the early and late integrated arms without any modifications.

The SAPiT study was approved by the Biomedical Research Ethics Committee of the University of KwaZulu-Natal (E107/05) and the Medicines Control Council of South Africa (20060157).

Exposure

As mentioned earlier, the fixed exposure, denoted by X, is defined as the fraction of time on ART during TB treatment (i.e., months on ART/months on TB therapy) (Figure 1). This was defined regardless of how long patients were on TB treatment (as some took more than the expected 6 months to complete TB treatment). Patients who were terminated before initiating ART were assigned an exposure of zero.

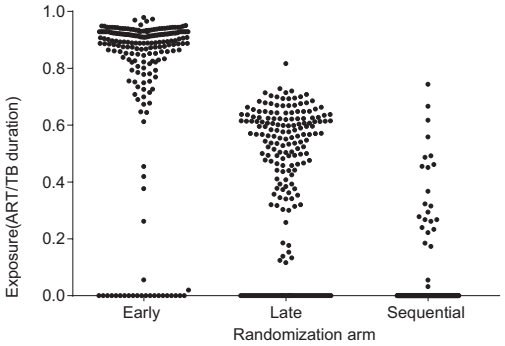


FIGURE 1. Exposure (fraction of time on ART during TB treatment) in the 3 randomized arms.

The rationale for choosing this exposure is as follows. In South Africa, TB and HIV are the leading causes of death in adults,[4] and TB-HIV co-infected patients are supposed to be co-treated by the same healthcare worker for both diseases. However, in resource-poor countries like South Africa, this integration has not been fully implemented. Among other things, one constraint is the low uptake of HIV testing among TB patients that deprive patients of treatment integration.[1] For these reasons, this article focuses on strengthening the evidence of the benefit of integrated therapy on mortality. Our IV analysis based on this exposure will express how, under perfect compliance, the ITT-analysis comparing the early integrated arm (where the exposure is then 1) and the sequential arm (where the exposure is then 0) would have fallen out.

Owing to the above-explained limitations of a fixed exposure over time, we have moreover developed IV-methodology for a time-varying measure of compliance to ART, which we defined as 1 at time t when the considered patient was on ART at or before time t, and 0 otherwise. Our IV analysis based on this exposure will express the effect of continuous ART use versus no ART use (regardless of TB treatment).

Outcome

We will focus on all-cause mortality. The survival time was defined as time from randomization until the date of death. It was censored at the withdrawal date, last visit date for those who were loss to follow-up or date of the 24-month visit for those who completed the study.

Covariates

The association between the fraction of time on ART during TB treatment and all-cause mortality is confounded because patients in poorer conditions (e.g., with lower CD4+ count) were at higher risk of death and thus more likely to initiate ART early irrespective of the study arm. Although the considered IV analysis do not require adjustment for measured confounders, we considered adjustment for CD4+ cell count, sex (0 = male; 1 = female) and employment status (0 = unemployed; 1 = employed) to improve precision. The instrument Z (0 if assigned to sequential arm, 1 if assigned to late integrated arm and 2 if assigned to early integrated arm). This was modeled categorically.

Statistical Analysis of Fixed Exposure

We first performed ITT analyses based on additive hazards models and Cox proportional hazards models. These assess the association between assignment Z to one of the arms and all-cause mortality. To investigate the association between the exposure X and all-cause mortality, we next performed IV analyses.

In particular, we used 2-stage predictor substitution (2SPS) and 2-stage residuals inclusion (2SRI) methods. Two-stage predictor substitution is a nonlinear extension of the linear 2-stage least squares. Two-stage residuals inclusion was first introduced by Hausmann[14] and recently proposed for the analysis of time-to-event endpoints by Terza et al.[15] Both these approaches work by fitting, in a first-stage, univariable or multivariable linear regression models of the association between X and Z. For instance, our results below are based on the multivariable model:

$$X = \alpha_0 + \alpha_z Z + \alpha_m^T \boldsymbol{M} + \varepsilon \quad (1)$$

where **M** is vector of measured covariates such as sex, employment status, and CD4+ cell count and where Z was modeled categorically (using dummy coding). The 2-stage predictor substitution approach then proceeds by regressing the survival time on the fitted values $\hat{X}$ from the first-stage regression (1) and on the measured covariates **M** either using additive hazard or Cox proportional hazards regression models. The 2-stage residuals inclusion approach proceeds likewise, but regressing additionally on the residuals X-$\hat{X}$ from the first-stage regression (1).

Tchetgen Tchetgen et al.[16] showed that, under certain conditions specified next, the coefficient of $\hat{X}$ in the resulting additive hazard model can be interpreted as the exposure effect β_x in the additive hazard model

$$\lambda(t|X, U, Z, \boldsymbol{M}) = \lambda_0(t) + \beta_x X + \beta_m^T \boldsymbol{M} + \beta_u(t) U \quad (2)$$

which involves adjustment for possible unmeasured confounders U. Here, exp ($-\beta_x t$) can be interpreted as the relative chance of surviving time t with exposure 1 vs 0; note that it takes the length of the exposure period into account via. the value of t. For the 2-stage predictor substitution approach, the condition is that the error term ε in the exposure model (1) is independent of randomization arm (given the covariates **M**). For the 2-stage residuals inclusion approach, a more subtle additional assumption is needed, which is satisfied when the error term ε equals the unmeasured confounder U apart from (additive) random noise. Because tests of the null hypothesis of no exposure effect are robust against model misspecification in the

2-stage predictor substitution approach (unlike the 2-stage residuals inclusion approach), we generally recommend the predictor substitution approach. Both these approaches can be extended to Cox proportional hazard models when the event (all-cause mortality) is rare, which is not well satisfied in the SAPiT trial. Under the rare assumption, then the coefficient of $\hat{X}$ can be interpreted as the exposure effect β_x in the Cox proportional hazards regression model

$$\lambda(t|X, Z, U, \boldsymbol{M}) = \lambda_0(t)\exp(\beta_x X + \beta_m^T \boldsymbol{M} + \beta_u U), \quad (3)$$

but not otherwise.

We calculated model-based survival probabilities for ITT and IV analyses under the 2SPS approach of the additive and proportional hazards models. In the IV analyses, the survival probabilities were estimated for 3 fixed exposure levels of 0 (no exposure), 0.6 (partial exposure), and 1.0 (full exposure). These correspond to what the survival probabilities would have looked like in the sequential, late, and early integrated arms, respectively, had there been perfect compliance.

Standard errors reported below are based on 1,000 nonparametric bootstrap samples with replacement, refitting both stages of the procedure each time. We used the bias-corrected and accelerated method[17] to calculate 95% confidence intervals (CIs) for β_x in the second-stage.

Statistical Analysis of Time-varying Exposure

In view of the aforementioned limitations of using a fixed exposure over time, we next extended the 2-stage predictor substitution approach to time-varying measures X(t) of compliance (see eAppendix 1; http://links.lww.com/EDE/B410). Our results are based on the multivariable model

$$X(t) = \alpha_0(t) + \alpha_z(t)Z + \alpha_m^T(t)\boldsymbol{M} + \varepsilon(t) \quad (4)$$

defined for patients who are alive at time *t,* which we consider at each observed event time *t*. Under the assumption that $\varepsilon(t)$ is uncorrelated with Z, conditional on $\boldsymbol{M}$ for patients who are alive at time *t*, we then show in the eAppendix 1; http://links.lww.com/EDE/B410 that the 2-stage predictor substitution approach can be extended to an additive hazard regression of the survival time on the fitted values $\hat{X}(t)$ from the first-stage regression (4) (which is limited to patients who are alive at time *t*) and on the measured covariates **M**. In particular, we show that the resulting effect β_x of $\hat{X}(t)$ can be interpreted as the effect of X(t) in the additive hazard model

$$\lambda\left(t|\bar{X}(t), U, Z, \boldsymbol{M}\right) = \lambda_0(t) + \beta_x X(t) + \beta_m^T \boldsymbol{M} + \beta_u(t)U \quad (5)$$

It follows from the eAppendix 2; http://links.lww.com/EDE/B410 that the proposed IV approach for time-varying exposure is valid in the presence of unmeasured time-varying confounders, U(t), so long as these are not influenced by previous exposure measurement $\bar{X}(t-1)$; that is, under the data-generating mechanism visualized in eFigure 1; http://links.lww.com/EDE/B410. Although this is a potentially strong assumption, it is one that is implicit in all standard IV analyses that reduce time-varying exposure to a fixed exposure, and one that is difficult or impossible to relax. Indeed, without this assumption, there may be pathways from the instrumental variable Z via. $\bar{X}(t-1)$ and U(*t*) toward the time-to-event endpoint, thereby inducing a dependence between the IV and the unmeasured confounders, and thus violating the IVs assumptions.

We calculated model-based survival probabilities under the proposed additive hazards models for values, 1 (on ART) and 0 (not on ART). Statistical analyses were done using SAS version 9.4 (SAS Institute Inc., Cary, NC) and R version 3.2.4.

RESULTS

Baseline and Follow-up

A total of 642 patients were enrolled: 214 in the early integrated arm, 215 in the late integrated arm, and 213 in the sequential arm (eFigure 2; http://links.lww.com/EDE/B410). Patients in the 3 study arms had similar baseline demographic and clinical characteristics (eTable 1; http://links.lww.com/EDE/B410).

During a median follow-up of 24 months (interquartile range, 10.5–24 months), 69 (10.7%) patients died (17 in each of the early and late integrated arms and 35 in the sequential arm). A total of 417 (65.0%) completed the study (early integrated arm [n = 151], late integrated arm [n = 139], and sequential arm [n = 127]). A total of 96 (15.0%) were lost to follow-up (early integrated arm [n = 26], late integrated arm [n = 36], and sequential arm [n = 34]); while 60 (9.3%) either withdrew consent or relocated to other areas (early integrated arm [n = 20], late-integrated arm [n = 23], and sequential arm [n=17]). A total of 16 (7.5%), 51 (23.7%), and 74 (34.7%) patients did not start ART in the early, late-integrated, and sequential arms, respectively, owing to reasons such as death, loss to follow-up, relocation, and voluntarily withdrawal (eFigure 2; http://links.lww.com/EDE/B410). The baseline and clinical characteristics by study arm are presented in eTable 1; http://links.lww.com/EDE/B410, whereas the characteristics of the compliant and noncompliant patients in all arms are shown in eTable 2; http://links.lww.com/EDE/B410.

Over 984.79 person–years of follow-up, the mortality rates were 4.9 per 100 person–years (py) (95% CI: 2.9, 7.9) in the early integrated, 5.2 per 100 py (95% CI: 3.0, 8.2) in the late integrated, and 11.3 per 100 py (95% CI: 7.9, 15.8) in the sequential arm.

The median (interquartile range) duration on TB treatment was 6.7 (6.4–8.3) months in each of the 3 study arms. The mean exposure to ART during TB treatment was 0.78 in the early integrated, 0.39 in the late integrated, and 0.04 in the sequential arm (Table 1, Figure 1).

Additive Hazards Analysis on the Effect of Integrating ART and TB Treatment

Multivariable results from ITT analyses showed that patients in the early and late integrated arms had a hazard

difference of -0.05 (95% CI: -0.09, -0.01) and -0.06 (95% CI: -0.11, -0.02), respectively, when compared with the sequential arm. This indicates that on average, 5 and 6 deaths were averted for each year of follow-up in each 100 patients randomly assigned to early and late integrated arms compared with each 100 patients in the sequential arm (Table 2).

The 2-stage predictor substitution analysis of the fixed exposure showed that on average, 7 deaths (hazard difference = -0.07; 95% CI: -0.12, -0.01) were prevented for each year of follow-up in each 100 patients with full exposure to ART during TB treatment (as would be the case under perfect compliance in the early integrated arm) as opposed to 100 patients with no ART exposure during TB treatment (as would be the case under perfect compliance in the sequential arm) (Table 2). The 2-stage residual inclusion method for the fixed exposure resulted in slightly weaker effects (hazard difference = -0.05; 95% CI: -0.11, 0.01). The strong association found between the first-stage residuals and time to death in this analysis provides strong evidence of unmeasured confounding, which the IV-analysis accounted for. The findings from the "as-treated" analyses of the fixed exposure (hazard difference = -0.12; 95% CI: -0.17, -0.07) are thus likely biased (Table 2). Results from proportional hazards regression are shown in the Web Appendix (eTable 6; http://links.lww.com/EDE/B410).

Figure 2A shows model-based predicted survival probabilities from univariable ITT and IV (fixed exposure) analyses under the 2-stage predictor substitution approach from additive hazards model. Figure 2B shows predicted survival probabilities obtained from the model with time-varying exposure. The survival probabilities for X = 1 in both figures are higher than that of the early integrated arm. However, survival probabilities for X = 0.6 and X = 0 are closer to those for late integrated and sequential arms, respectively (Figure 2A). Predicted survival probabilities under proportional hazards models are shown in eFigures 3A and 3B; http://links.lww.com/EDE/B410.

TABLE 1. First-stage Linear Regression Model Predicting the Fixed Exposure

	Univariable		Multivariable[a]	
Effect	**Estimate**	**S.E.**	**Estimate**	**S.E.**
Intercept	0.04	0.02	0.06	0.03
Early integrated arm	0.74	0.02	0.74	0.02
Late integrated arm	0.35	0.02	0.35	0.02
Sequential arm	0		0	
F-value; partial R^2	514.68; 0.62		211.66; 0.62	

[a]Adjusted for sex, baseline CD4+ count, and employment status.
S.E., standard error.

Time-varying Additive Hazards Analysis on the Effect of ART

The IV-analysis of time-varying ART exposure expresses the effect of continuous ART exposure versus no exposure (hazard difference of -0.29; 95% CI: -0.54, -0.03) (Table 2). It indicates that an average of 29 deaths were prevented for each year of follow-up in each 100 patients on (continuous) ART, compared with 100 patients not on ART, conditional on sex, employment status, and CD4+ cell count. Corresponding predicted survival probabilities for a patient without ART exposure, X(t) = 0 for all *t,* were very low (Figure 2B). The large differences found between the estimated effects of fixed and

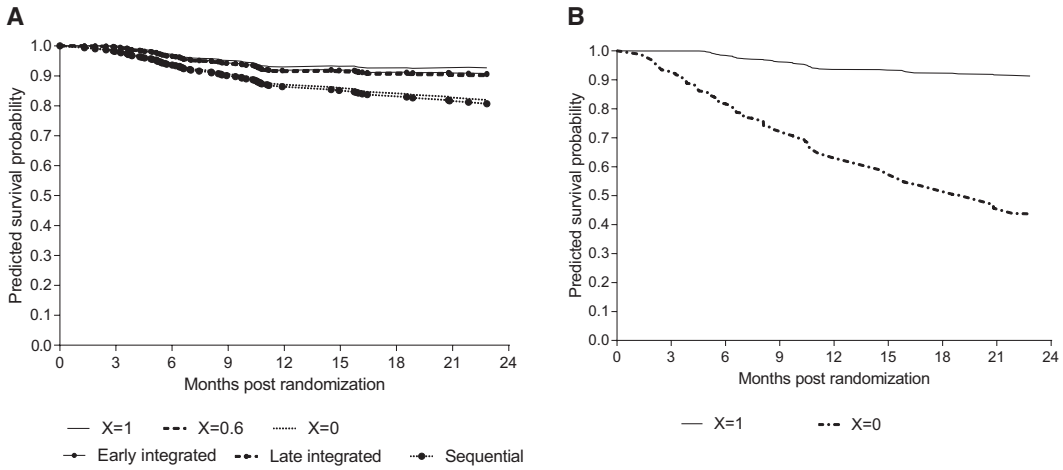


FIGURE 2. Predicted survival curves comparing ITT and IV estimates from additive hazards model (A: fixed exposure, B: time-varying exposure).

TABLE 2. ITT and IV Estimates for the Effect of Study Arm and Exposure on Mortality Using Additive Hazards Models

Method/Exposure	Arm/Exposure type	Univariable β (95% CI)	Multivariable[a] β (95% CI)
ITT	Early integrated arm	-0.06 (-0.11, -0.02)	-0.05 (-0.09, -0.01)
	Late integrated arm	-0.06 (-0.11, -0.02)	-0.06 (-0.11, -0.02)
	Sequential arm	0	0
Time-varying exposure			
As-treat	Exposure to ART during follow-up	-0.22 (-0.30, -0.13)	-0.21 (-0.31, -0.13)
2SPS	Exposure to ART during follow-up	-0.41 (-0.72, -0.09)	-0.29 (-0.54, -0.03)
Fixed exposure			
As-treated	Exposure to ART during TB treatment	-0.12 (-0.17, -0.08)	-0.12 (-0.17, -0.07)
2SPS	Exposure to ART during TB treatment	-0.08 (-0.14, -0.03)	-0.07 (-0.12, -0.01)
2SRI	Exposure to ART during TB treatment	-0.06 (-0.13, -0.01)	-0.05 (-0.11, 0.01)
	First-stage residuals	-0.19 (-0.35, -0.06)	-0.23 (-0.43, -0.09)

[a]Adjusted for sex, baseline CD4+ count, and employment
2SRI, 2-stage residuals inclusion.

time-varying exposures are largely attributable to the different exposure definitions: while the analyses of a fixed exposure focus on the effect of integration of ART and TB treatment (as compared with a background regimen with ART), the analysis of a time-varying exposure focuses on the overall effect of ART. Arguably, the analysis of fixed exposure is also less reliable because of the previously mentioned concern for reverse causality, and the fact that differential ART exposure outside the period of TB treatment may have induced a violation of the exclusion restriction in this analysis. In contrast, also the time-varying IV analysis is subject to a violation of the assumption that the residual in equation (4) is (conditionally) uncorrelated with the IV.

This location shift assumption could be plausible for continuous exposures, but is known to be violated for dichotomous exposures. In view of this, we report the results of limited simulation studies for time-fixed exposures in eTable 7; http://links.lww.com/EDE/B410. These show that the degree of bias under violation of this assumption is small under the considered settings. However, we must warn the reader that we cannot guarantee that the proposed approach for time-varying exposures is (nearly) unbiased when the exposure is dichotomous, unless when the exposure effect is close to, or equal to zero.

DISCUSSION

We have provided IV analyses of the causal effect of exposure to ART (during TB treatment) on time to death, to account for noncompliance as a result of not all patients adhering to randomization and starting ART at the correct time with respect to TB treatment. The SAPiT trial data have been analyzed using ITT methods and showed that integration of ART and TB treatment saves lives.[5] Our results express more precisely how many lives could be saved under perfect compliance. The IV results thus appeal to patients and clinicians who are interested in the benefits of initiating and adhering to received treatment. The results from the "as-treated" analyses showed even higher effectiveness, but these results are biased because patients who comply and those who do not comply with the randomized treatment are not always comparable.

The analyses were carried out using the semiparametric additive hazards models. Cox proportional hazards models were used for comparison. Our instrumental variables analysis relies on 2 key assumptions. The first, that patients on the different arms of the study are exchangeable, is guaranteed by randomization. The second, so-called exclusion restriction, that randomized assignment may only influence all-cause mortality by changing ART exposure, could be violated. One possible cause of violation concerns our definition of exposure, which may not fully capture all relevant components such as adherence, which could have an effect on mortality. A second possible reason is that, in the open-label SAPiT trial, being assigned to either of the integrated arms may have enhanced patient's expectation of success, and in contrast, assignment to the sequential arm might have reduced such an expectation. Moreover, those randomized to integrated arms who did not start ART soon after TB treatment initiation might have deliberately delayed ART initiation because they were still feeling well and did not see the need to integrate TB treatment and ART. Our analysis moreover ignored differential ART exposure outside the TB treatment window. All of this, in turn, violates the exclusion restriction assumption, which underlies our analysis. The violation of this untestable assumption can lead to biased IV estimates.

We acknowledge several additional limitations in our analyses. Our fixed exposure did not differentiate between patients who were on TB treatment for 6 months and those who were on TB treatment for a longer period. A patient who was on TB treatment for 6 months and only took ART for 3 months had similar exposure level to a patient who was on

TB treatment for 12 months and took ART for 6 months. The latter patient is more likely to have drug-resistant TB and thus be more likely to die. Also, for patients who died soon after enrollment and those who did not start ART, the exposure is not ideally defined. We have tried to counteract that limitation by developing an IV-methodology for a time-varying measure of compliance. This analysis answers a different scientific question, but better recognizes the complexity of the exposure. Indeed, the magnitude of a summary exposure over the observation period is likely influenced by censoring owing to loss to follow-up, and death, and therefore ill defined. In particular, while more reliable, the analysis of time-varying exposure did not immediately allow us to assess the effectiveness of ART treatment during the TB treatment window.

However, a drawback of our IV-methodology for time-varying exposures is that it relies on a location-shift assumption, which is unlikely to hold for dichotomous exposures. For such exposures, we cannot exclude the possibility of large biases when the exposure effect is large. For time-fixed exposures, Tchetgen Tchetgen et al.[16] were able to circumvent this problem when the instrument is dichotomous; see also Martinussen et al.[18] for general IVs. It is an open question whether their proposals can be extended to time-varying exposures.

Our analysis of time-varying exposure is based on novel IV-methodology that returns effects on the additive hazard scale. Although alternative G-estimation strategies have been proposed to infer the effect of a time-varying exposure on a time-to-event endpoint in the presence of an IV under an alternative class of structural accelerated failure time models,[11] application of these methods in applied research has been relatively infrequent because of their complexity and often poor performance in the presence of censoring.[19] Our proposal overcomes these concerns by being applicable in standard software for additive hazard models, which naturally accommodates noninformative censoring without requiring further adjustments. In contrast to G-estimation, it does have the drawback of making assumptions on the distribution of the exposure (as shown in model (4)).

Further, our analyses infer a treatment effect in the treated, and thereby assume treatment effect homogeneity (i.e., that patients with a given ART exposure level on the different arms on average experience the same benefit of it). We do not infer the effect in compliers, as this would be more difficult to define with a time-varying and/or continuous measure of compliance. Finally, we assumed that censoring is noninformative, given the covariates that we controlled for, and is moreover independent of the exposure and IV.

In conclusion, results from IV analyses demonstrate that survival benefit of fully integrating TB treatment and ART is even higher than what has been reported in the ITT analyses since noncompliance has been accounted for. IV estimates are clinically important because knowing the effectiveness of the TB and ART integration in the absence of noncompliance enables a much better public health evaluation of the potential impact of this intervention.

ACKNOWLEDGMENTS

We are thankful to the protocol investigator Professor Salim Abdool Karim and project director Dr. Kogieleum Naidoo for allowing us to use data from the SAPiT trial. We also thank the patients for their participation in the study and the entire SAPiT study team.

REFERENCES

1. World Health Organization. HIV-associated tuberculosis. TB/HIV fact sheets. 2015. Available at: http://www.who.int/tb/challenges/hiv/tbhiv_factsheet_2015.pdf. Accessed March 10, 2015.
2. Perumal R, Padayatchi N, Naidoo K, Knight S. Understanding the profile of tuberculosis and human immunodeficiency virus coinfection: insights from expanded HIV surveillance at a tuberculosis facility in Durban, South Africa. *ISRN AIDS*. 2014;2014:260329.
3. National Department of Health. Annual Report 2014/2015. Available at: http://www.health-e.org.za/wp-content/uploads/2015/10/Department-of-Health-Annual-Report-201415.pdf. Accessed April 23, 2016.
4. Statistics South Africa. Mortality and causes of death in South Africa, 2014: Findings from death notification. Available at: http://www.statssa.gov.za/publications/P03093/P030932014.pdf. Accessed March 14, 2016.
5. Abdool Karim SS, Naidoo K, Grobler A, et al. Timing of initiation of antiretroviral drugs during tuberculosis therapy. *N Engl J Med*. 2010;362:697–706.
6. Blanc FX, Sok T, Laureillard D, et al.; CAMELIA (ANRS 1295–CIPRA KH001) Study Team. Earlier versus later start of antiretroviral therapy in HIV-infected adults with tuberculosis. *N Engl J Med*. 2011;365:1471–1481.
7. Havlir DV, Kendall MA, Ive P, et al.; AIDS Clinical Trials Group Study A5221. Timing of antiretroviral therapy for HIV-1 infection and tuberculosis. *N Engl J Med*. 2011;365:1482–1491.
8. Abdool Karim SS, Naidoo K, Grobler A, et al. Integration of antiretroviral therapy with tuberculosis treatment. *N Engl J Med*. 2011;365:1492–1501.
9. Department of Health. National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults. 2014. Available at: http://www.kznhealth.gov.za/family/HIV-Guidelines-Jan2015.pdf. Accessed April 4, 2016.
10. Vansteelandt S, Babanezhad M, Goetghebeur E. Correcting instrumental variables estimators for systematic measurement error. *Stat Sin*. 2009;19:1223–1246.
11. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Commun Stat Theory Methods*. 1991;20:2609–2631.
12. Naidoo K, Yende-Zuma N, Padayatchi N, et al. The immune reconstitution inflammatory syndrome after antiretroviral therapy initiation in patients with tuberculosis: findings from the SAPiT trial. *Ann Intern Med*. 2012;157:313–324.
13. Naidoo A, Naidoo K, Yende-Zuma N, et al. Changes to antiretroviral drug regimens during integrated TB-HIV treatment: results of the SAPiT trial. *Antivir Ther*. 2014;19:161–169.
14. Hausman JA. Specification tests in econometrics. *Econometrica*. 1978;46:1251–1271.
15. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J Health Econ*. 2008;27:531–543.
16. Tchetgen Tchetgen EJ, Walter S, Vansteelandt S, Martinussen T, Glymour M. Instrumental variable estimation in a survival context. *Epidemiology*. 2015;26:402–410.
17. DiCiccio TJ, Efron B. Bootstrap confidence intervals. *Stat Sci*. 1996;11:189–228.
18. Martinussen T, Vansteelandt S, Tchetgen Tchetgen EJ, Zucker DM. Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. *Biometrics*. 2017;73:1140–1149.
19. Vansteelandt S, Joffe M. Structural nested models and G-estimation: the partially realized promise. *Stat Sci*. 2014;29:707–731.

Appendix 2: R code

2SPS and 2SRI approach using fixed continuous exposure

```
install.packages("timereg")
library(timereg)
# stage 1:  univariable linear model
unilin <- lm(exposure ~ factor(dummy1)+factor(dummy2), data=mydata)
summary (unilin)
#get predicted values
mydata$preduni <- predict(unilin)
#get residuals
mydata$resuni <- residuals(unilin)

# stage 2:  univariable additive hazards model under 2SPS
uni_sps <- aalen(Surv(Yearz, Died==1) ~ const(preduni), robust=0, data=mydata)
summary(uni_sps)
# stage 2:  univariable additive hazards model under 2SRI
uni_sri <- aalen(Surv(Yearz, Died==1) ~ const(exposure)+const(resuni), robust=0,
data=mydata)
summary(uni_sri)

#bootstrap get standard errors and confidence intervals
i<-0
```

```
nuke.fun <- function(data, indices)
{
dat<-data[indices,]
unilin <- lm(exposure ~ factor(dummy1)+factor(dummy2), data=dat)
dat$preduni<-NA
dat$resuni<-NA
dat$preduni <-predict(unilin)
dat$resuni <-residuals(unilin)
uni_sps <-aalen(Surv(Yearz, Died==1) ~ const(preduni), robust=0, data=dat)
uni_sri <-aalen(Surv(Yearz, Died==1) ~ const(exposure)+ const(resuni), robust=0,
data=dat)
c(coef.aalen(uni_sps)[,1],coef.aalen(uni_sri)[,1])
}
nuke.boot <- boot(mydata,nuke.fun, R = 1000)
# median
vv <-apply(nuke.boot$t,2, median)
# get confidence intervals
boot.ci(nuke.boot, type=c("bca"), index=1, conf=0.95) #2SPS
boot.ci(nuke.boot, type=c("bca"), index=2, conf=0.95) #2SRI
boot.ci(nuke.boot, type=c("bca"), index=3, conf=0.95) #2SRI

# stage1:  multivariable linear model
multilin <- lm(exposure ~ factor(dummy1)+factor(dummy2)+const(CD4count)+
const(Gender)+const(Employ), data=mydata)
summary (multilin)
#get predicted values
mydata$predmulti <- predict(multilin)
#get residuals
mydata$resmulti <- residuals(multilin)

#stage 2:  multivariable additive hazards model under 2SPS
```

```
multi_sps <- aalen(Surv(Yearz, Died==1) ~ const(predmulti)+const(CD4count)+
const(Gender)+const(Employ), robust=0, data=mydata)
summary(multi_sps)

#stage 2:  multivariable additive hazards model under 2SRI
multi_sri <- aalen(Surv(Yearz, Died==1) ~ const(exposure)+const(resmulti)+
const(CD4count)+ const(Gender)+const(Employ), robust=0, data=mydata)
summary(multi_sri)

#bootstrap get standard errors and confidence intervals
i<-0
nuke.fun <- function(data, indices)
{
dat<-data[indices,]
multilin <- lm(exposure ~ factor(dummy1)+factor(dummy2)+const(CD4count)+
const(Gender)+const(Employ), data=dat)
dat$predmulti<-NA
dat$resmulti<-NA
dat$predmulti <-predict(multilin)
dat$resmulti <-residuals(multilin)
multi_sps <-aalen(Surv(Yearz, Died==1) ~ const(predmulti)+const(CD4count)+
const(Gender)+const(Employ), robust=0, data=dat)
multi_sri <-aalen(Surv(Yearz, Died==1) ~ const(exposure)+ const(resmulti)+
const(CD4count)+ const(Gender)+const(Employ), robust=0, data=dat)
c(coef.aalen(multi_sps)[,1],coef.aalen(multi_sri)[,1])
}
# median
nuke.boot <- boot(mydata,nuke.fun, R = 1000)
vv <-apply(nuke.boot$t,2, median)
# get confidence intervals
boot.ci(nuke.boot, type=c("bca"), index=1, conf=0.95) #2SPS
```

```
boot.ci(nuke.boot, type=c("bca"), index=5, conf=0.95) #2SRI
boot.ci(nuke.boot, type=c("bca"), index=6, conf=0.95) #2SRI
```

2SPS approach using time-varying binary exposure

```
# univariable
# time is ordered time-points (t)
# yrlast is observed failure times (T)
# start_yr is the start of the time interval
# stop_yr is the stop of the time interval
# rx is the binary exposure

# Calculating E(X(t) | Z, T ≥ t)
time <- sort(unique(c(longdata$start_yr,longdata$stop_yr)))
longdata$xmean <- NULL
for (i in 1:length(time)){
mod <- glm(rx ~ dummy1+dummy2, family="binomial", data = longdata,
subset = (longdata$yrlast >= time[i]))
longdata$xmean[longdata$stop_yr == time[i]] <-
predict(mod,newdata=longdata[longdata$stop_yr == time[i],],type="response")
}
# second stage additive hazards model
unilong<-aalen(Surv(start_yr,stop_yr, fdied==1) ~ const(xmean),robust=0,
data=longdata)
summary(unilong)

#multivariable models
for (i in 1:length(time)){
modm <- glm(rx ~ dummy1+dummy2 + CD4count+ factor(Gender)+factor(Employ),
```

```
family="binomial", data = longdata, subset = (longdata$yrlast >= time[i]))
longdata$xmeanm[longdata$stop_yr == time[i]] <-
predict(modm,newdata=longdata[longdata$stop_yr == time[i],],
type="response")
}
#second stage additive hazards model under 2SPS
unilongm<-aalen(Surv(start_yr,stop_yr, fdied==1) ~ const(xmeanm)+const(CD4count)+
const(Gender)+const(Employ),robust=0, data=longdata)
summary(unilongm)
```

IV simulation for continuous exposure

```
#nsum is the sample size
#nsim is number of simulations per scenario
# x is an exposure
# d is a censor variable (1 if an event occurred, 0 if censoring occurred)
# u is unmeasured confounders
# ti is failure times
# beta0 is baseline hazard
#beta1 and betau are exposure and unmeasured confounder effect respectively

library(timereg)
nsum <- 1000
nsim<-1000
out.cross.iv<-numeric()
out.cross.iv_u<-numeric()
out.cross_u<-numeric()

#create study arms
```

```
arms <- sample(1:3,size=nsum,replace=TRUE,prob=c(0.333,0.333,0.333))
#create dummy variables
arm1<- ifelse(arms==1, 1, ifelse(arms==2,0,ifelse(arms==3,0,NA)))
arm2<- ifelse(arms==1, 0, ifelse(arms==2,1,ifelse(arms==3,0,NA)))
for(i in 1:nsim)
{
x<- rnorm(nsum, 0.04+0.74*arm1+0.34*arm2)
#change intercept when you change beta1 value and plot failure times
#to make sure they still follow exponential distribution
ti=rexp(nsum)/abs(lambda0+beta1*x)
d <- ifelse(ti<2,1,0)
ti <- pmin(ti,2)

data.cross<-data.frame(cbind(d,arm1,arm2,arms,ti,x))
lin <- lm(x ~ arm1+arm2,data=data.cross)
data.cross$preduni <- predict(lin)
# second stage additive hazards model under 2SPS
model.cross.iv<-aalen(Surv(ti,d)~ const(preduni),data=data.cross)
model.cross.iv.gamma=c(model.cross.iv$gamma)
out.cross.iv<-append(out.cross.iv,model.cross.iv.gamma)

u<-rnorm(nsum)
x_u<- rnorm(nsum, 0.04+0.74*arm1+0.34*arm2+u)
#change intercept when you change beta1 value and plot failure times
#to make sure they still follow exponential distribution
ti_u=rexp(nsum)/abs(lambda0+beta1*x_u+beta0*u)
d_u <- ifelse(ti_u<2.0,1,0)
ti_u <- pmin(ti_u,2)
data.cross_u<-data.frame(cbind(d_u,arm1,arm2,arms,ti_u,x_u))
#second stage additive hazards model under 2SRI
model.cross_u=aalen(Surv(ti_u,d_u)~ const(x_u)+const(u),data=data.cross_u)
```

```
model.cross_u.gamma=c(model.cross_u$gamma)
out.cross_u<-append(out.cross_u,model.cross_u.gamma)
# 2SPS approach adjusted for unmeasured confounding
lin_u <- lm(x_u∼ arms,data=data.cross_u)
data.cross_u$preduni_u <- predict(lin_u)
model.cross.iv_u=aalen(Surv(ti_u,d_u)∼ const(preduni_u),data=data.cross_u)
summary(model.cross.iv_u)
model.cross.gamma_u=c(model.cross.iv_u$gamma)
out.cross.iv_u<-append(out.cross.iv_u,model.cross.gamma_u)
}

est.cross.iv<-c(rep(beta1,length(out.cross.iv)))
bias.cross.iv<-est.cross.iv-(out.cross.iv)
av.bias.cross.iv<-mean(bias.cross.iv)

est.cross_u<-c(rep(beta1,length(out.cross_u)))
bias.cross_u<-est.cross_u-(out.cross_u)
av.bias.cross_u<-mean(bias.cross_u)

est.cross.iv_u<-c(rep(beta1,length(out.cross.iv_u)))
bias.cross.iv_u<-est.cross.iv_u-(out.cross.iv_u)
av.bias.cross.iv_u<-mean(bias.cross.iv_u)
```

IV simulation for binary exposure

```
# variables including the study arms are explained as shown under the simulation
for linear exposure

for(i in 1:nsim)
```

```
{
p =exp(0.8*arm1+0.5*arm2) / (1 + exp(0.8*arm1+0.5*arm2))
x =rbinom(nsum, size=1, prob=p)
#change intercept when you change beta1 value and plot failure times
#to make sure they still follow exponential distribution
ti=rexp(nsum)/abs(lambda0+beta1*x)
d <- ifelse(ti<2,1,0)
ti <- pmin(ti,2)

data.cross<-data.frame(cbind(d,arm1,arm2,arms,ti,x))
bin <- glm(a factor(arm1)+factor(arm2),binomial(link='logit'),data=data.cross
data.cross$preduni<-predict(bin,type="response")

# second stage additive hazards model under 2SPS
model.cross.iv<-aalen(Surv(ti,d)~ const(preduni),data=data.cross)
model.cross.iv.gamma=c(model.cross.iv$gamma)
out.cross.iv<-append(out.cross.iv,model.cross.iv.gamma)

u<-rnorm(nsum)
p_u =exp(0.8*arm1+0.5*arm2+u) / (1 + exp(0.8*arm1+0.5*arm2+u))
x_u =rbinom(nsum, size=1, prob=p_u)
#change intercept when you change beta1 value and plot failure times
#to make sure they still follow exponential distribution
ti_u=rexp(nsum)/abs(lambda0+beta1*x_u+beta0*u)
d_u <- ifelse(ti_u<2.0,1,0)
ti_u <- pmin(ti_u,2)
data.cross_u<-data.frame(cbind(d_u,arm1,arm2,arms,ti_u,x_u))

#second stage additive hazards model under 2SRI
model.cross_u=aalen(Surv(ti_u,d_u)~ const(x_u)+const(u),data=data.cross_u)
model.cross_u.gamma=c(model.cross_u$gamma)
```

```
out.cross_u<-append(out.cross_u,model.cross_u.gamma)

# 2SPS approach adjusted for unmeasured confounding
bin_u <- glm(a factor(arm1)+factor(arm2),binomial(link='logit'),data=data.cross)
data.cross$preduni_u<-predict(bin_u,type="response")
model.cross.iv_u=aalen(Surv(ti_u,d_u)~ const(preduni_u),data=data.cross_u)
model.cross.gamma_u=c(model.cross.iv_u$gamma)
out.cross.iv_u<-append(out.cross.iv_u,model.cross.gamma_u)
}

est.cross.iv<-c(rep(beta1,length(out.cross.iv)))
bias.cross.iv<-est.cross.iv-(out.cross.iv)
av.bias.cross.iv<-mean(bias.cross.iv)

est.cross_u<-c(rep(beta1,length(out.cross_u)))
bias.cross_u<-est.cross_u-(out.cross_u)
av.bias.cross_u<-mean(bias.cross_u)

est.cross.iv_u<-c(rep(beta1,length(out.cross.iv_u)))
bias.cross.iv_u<-est.cross.iv_u-(out.cross.iv_u)
av.bias.cross.iv_u<-mean(bias.cross.iv_u)
```