2020

Analysis of recurrent events with associated informative censoring: application to HIV data.

Ejoku, Jonathan
Strathmore Institute of Mathematical Sciences
Strathmore University

Analysis of Recurrent Events with Associated Informative Censoring: Application to HIV Data

Ejoku Jonathan

Submitted in partial fulfillment of the requirements for the Degree of Master of Science in Statistical Sciences at Strathmore University

Strathmore Institute of Mathematical Sciences
Strathmore University Nairobi, Kenya

June, 2020

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

.... [Name of Candidate]

.... [Signature]

.... [Date]

Approval

The thesis of Ejoku Jonathan was reviewed and approved by the following:

Dr. Collins Odhiambo
Lecturer, Institute of Mathematical Sciences
Strathmore University

Ferdinand Otieno
Dean, Institute of Mathematical Sciences
Strathmore University

Dr. Bernard Shibwabo
Dean, School of Graduate Studies
Strathmore University

Abstract

In this study, we adapt a commonly used Cox-based model for recurrent events; the Prentice, Williams and Peterson Total -Time (PWP-TT) that has been largely used under the assumption of non-informative censoring and evaluate it under an informative censoring setting. Empirical evaluation was undertaken with the aid of the semi-parametric framework for recurrent events suggested by (Huang and Wang, 2004) where a subject specific latent variable is used to model the association between the recurrent event and hazard of the failure time. All implementations were made in R Studio software, using the reReg package (Chiou and Huang, 2019) and the method in the reReg function set to 'cox.HW'. For validation we used HIV data from a typical HIV care setting in Kenya. Results show that the PWP-TT model generally fit the data well, with a comparison to the Andersen-Gill method showing similar estimates, while the ordinary Cox model estimates were too unreliable

Keywords: Recurrent events, Loss to follow-up, HIV, Prentice, Williams and Peterson Gap-Time, Informative censoring

Contents

List of Figures

List of Tables

List of abbreviations

AG. Anderson and Gill

AIDS. Acquired Immunodeficiency Syndrome

ART. Antiretroviral Therapy

BBS. Block-Borges-Savits

GEE. Generalized Estimating Equations

HAART. Highly Active Antiretroviral Therapy

CCC. Comprehensive Care Center

DCM. Differentiated Care Model

HIV. Human Immunodeficiency Virus

HR. Hazard Ratio

LTFU. Lost to follow-up

LWA. Lee, Wei and Amato

PWP-GT. Prentice, Williams, and Peterson gap time

PLHIV. Persons Living with HIV

PWP-TT. Prentice, Williams, and Peterson total time

SD. Standard Deviation

SE. Standard Error

WLW. Wei, Lin and Weissfeld

INH. Isonicotinylhydrazide

WHO. World Health Organization

Acknowledgements

Firstly, I give thanks to God for the gift of life and understanding. He has been my constant source of inspiration through this long journey.

To my supervisors, Dr. Collins Odhiambo and Dr. Linda Chaba, I am highly indebted. This has been a learning curve for me, but your patience and guidance, in seeing this over the line has been constant and sure.

Finally, I thank my parents, and siblings for the support in more ways than I can list. The journey has had turns, valleys and hills, but your support has been unwavering.

Dedication

For the guidance, support, care and prayers, I dedicate this project to my parents, Mr. and Mrs. Omagor.

Chapter 1

INTRODUCTION

1.1 Background

Recurrent events occur in a variety of disciplines/areas of life such as recurrent opportunistic infections in HIV patients, and episodes of asthmatic attacks. Underlying processes that generate data from these events are called "recurrent event processes and the data they provide are called recurrent event data" (Cook and Lawless, 2007). There is a wide range of research on the analysis of this data including the analysis of recurrence of sports injuries (Ullah et al., 2014), multiple episodes of childhood infectious diseases (Kelly and Lim, 2000), and hospitalizations for chronic kidney disease (Yang et al., 2017). A key characteristic of recurrent events is that observations per individual are usually not independent and in many cases correlated, with current event incidences influenced by previous incidences.

There are a number of approaches that have been proposed to analyze recurrent events data, including both parametric and non-parametric methods such as the use of the Poisson and Negative binomial models (Lawless, 1987). Both of these usually accommodate only time-independent covariates, with the Poisson further pre-supposing a constant events rate per individual.

While these and a number of other models have been suggested, the models most predominant in recurrent event studies for ordered outcomes are the Anderson and Gill (Andersen and Gill, 1982), Prentice, Williams, and Petersen Total Time (PWP-TT) and gap time (PWP-GT) (Prentice et al., 1981), Wei, Lin and Weissfeld (Wei et al., 1989); and Lee, Wei and Amato (Lee et al., 1992). These approaches are essentially extensions of the widely popular semi-parametric Cox Proportional Hazards Model and based on extensive research been applied in many researches under an underlying assumption of non-informative censoring. To the best of my knowledge, there have been no noted applications with informative censoring.

Informative censoring in HIV related-studies may be noted in situations, where there are informed drop-outs; such as sicker persons or those in either WHO staging III or IV being withdrawn from the study or persons getting lost to the study due to stigma-related factors.

1.2 Statement of the Problem

In a typical HIV care clinic, the risk posed by instances of loss to follow-up (LTFU) is undesired and has the potential of undoing antiretroviral treatment benefits. Specifically, patients' retention in HIV care is critical to ensuring better health outcomes especially in reduced viral load suppression, mitigating mortality and averting possible drug resistance caused by non-adherence.

In spite of the multiple interventions in place to address incidences of LTFU among HIV patients like enhanced adherence counselling, within-person cases of LTFU are usually common and recurrent in nature, with the present likelihood of a person getting lost to follow-up influenced by previous occurrences.

There is a growing body of evidence that Cox-based models typically model recurrent data quite well. While application of these methods has been evidently widespread including in the medical field like cancer tumor recurrence and repeat asthma attacks among others, a review of literature indicates that it has been under the assumption of non-informative censoring.

Censoring is said to be non-informative if participants drop out for reasons unrelated to the study or recurrent process. Huang et al., 2010 note that a key characteristic of this is that the collection of individuals under observation at any time is a random sample of the study population as at the origin. On the other hand, censoring can be described as informative if study drop-outs or the incidences of lost-to-follow-up are due to reasons related to the event process.

1.2.1 Clinical significance

This study aims at adopting one of the five Cox-based models[1] and applying it to HIV data, under the assumption of informative censoring. This study adds to the body of HIV related literature an evaluation of factors that may lead to LTFU among HIV patients for instances where such episodes are recurrent and a terminal event is informative. For this study, we shall be focusing particularly on the Prentice, Williams and Peterson Total-Time (PWP-TT).

1.2.2 Statistical Significance

A review of literature reveals that the models for recurrent events have been reviewed, under the assumption of non-informative censoring. This study meets a gap by providing an assessment of the model performance of one of the models, PWP-TT, in the event that informative censoring is present.

1.3 Objectives

1.3.1 General Objective

To study and review the application of Cox-based recurrent models largely employed under the assumption of non-informative censoring and apply one of them (PWP-TT) to HIV data under the assumption of informative censoring. The incorporation of informative censoring is achieved using Huang and Wang, 2004 proposed joint model for recurrent events and failure time.

[1]The five common Cox-based models include the Andersen and Gill (AG); Wei, Lin and Weissfeld (WLW); Prentice, Williams and Peterson, total time (PWP-CP) and gap time (PWP-GT); and Lee, Wei and Amato (LWA)

1.3.2 Specific Objectives

1. To review Huang's proposed joint model.
2. To fit PWP-TT model using Huang's proposed joint model based on HIV data.
3. To compare estimates of the fitted PWP-TT model to those of the Andersen-Gill (AG) and ordinary Cox model

Chapter 2

LITERATURE REVIEW

2.1 Limitations of the Cox PH Model

Since its introduction by Cox, 1972, the Cox proportional hazards (Cox PH) model has by far become the most popular method for modelling the relationship of covariates to a survival outcome. Its application has spanned a number of fields, such as the biomedical in the study of infection following bone marrow transplant (Farewell, 1979), assessing survival times for breast cancer patients (Abadi et al., 2014), the identification of factors that affect the survival of tuberculosis patients (Tolosie and Sharma, 2014) through to finance such as the stock exchange market (Ni, 2009).

The Cox PH model is however not without its limitations. One key assumption it postulates is the proportionality of hazards (Collett, 2015), violation of which may lead to "overestimating the relative risk of covariates with hazard ratios that increase over time" (Schemper, 1992). Another limitation and of interest to this study is the application to recurrent events - "events that occur more than once over the follow-up time for a given subject" (David and Kleinbaum, 2016, p. 332). Thenmozhi et al., 2019 as well as (Therneau and Grambsch, 2000, p. 169) note that intra-subject correlation, a major trait of recurrent events data is not usually incorporated in the Cox PH models. Additionally, Villegas et al., 2013 notes that "it is not suitable for recurrent event data because survival times in the standard model terminate at the time of the event". Applying the standard

Cox model to recurrent data will thus lead to wastage of information as it only considers time to first event (Castañeda and Gerritse, 2010). Other limitations brought by onset of recurrent events data are related to the multiple time scales and the general structure of risk sets (Therneau and Grambsch, 2000, p. 169).

These limitations have led to a number of suggested alternate approaches including the use of parametric methods. Lawless, 1987 notably proposed using Poisson and Negative binomial models, which come with their own drawbacks such as a fixed rate of event occurrence over time for the Poisson (Yang et al., 2017). Also, while parametric methods may offer some advantages over non-parametric approaches such as greater efficiency in that less parameters are estimated and that interpretations are also more relevant interpretations of the model if the parametric model matches the underlying data (Wheatley-Price et al., 2012), a model misspecification can lead to invalid inferences. Other authors such as (Therneau and Grambsch, 2000, p. 170) have however suggested modelling "a subject's correlation directly within the Cox framework" hence the extensions of the semi-parametric Cox PH model, discussed next.

2.2 Application of the five extensions of the Cox PH Models

We consider the applications of the five extended Cox models that include the Andersen-Gill (AG) Model, Wei, Lin and Weissfeld (WLW) Model, the Prentice, Williams, and Peterson (PWP) Total Time Model and the Prentice, Williams, and Peterson (PWP) - Gap Time Model as well as the Lee, Wei and Amato (LWA) Models.

In examining the recurrence of sports injuries, Ullah et al., 2014, compare three of the aforementioned Cox model extensions (excluding the LWA and PWP-Gap Time) while including the frailty and ordinary Cox model. A subsequent direct application of these methods to Australian National Rugby League recurrent injury data for the playing period 2008, revealed the poorest fit for the Cox PH model (unsurprising given its only efficient for the first event), while the AG and frailty models improved the fit.

On the other hand, Kelly and Lim, 2000 propose a method of systematically identifying the “components of the models that are appropriate for recurrent event data, that is risk intervals, baseline hazard, risk set, and correlation adjustment.”. From fits made to simulated data as well as a dataset of childhood recurrent infectious diseases, they deduce that the LWA model is inappropriate for recurrent data as it makes an individual to be “at risk several times for the same event”.

Yang et al., 2017 also compared three of the Cox based models (the AG, PWP-Gap Time and PWP- Total time) to the fully parametric Poisson regression model, with application to data on cardiovascular disease events. Generally, their assessment of these models indicated that some Poisson estimates were quantitatively different from those of the three Cox-type survival models. Deductively, they encourage caution in “using the Poisson regression to analyze recurrent events, due to its strong parametric assumptions”.

Together with the AG, and PWP-GT models, Sagara et al., 2014 additionally applied a shared gamma frailty model, and Generalized Estimating Equations model (GEE) using Poisson distribution to a clinical trial where recurrent malaria episodes from a clinical trial assessed the effectiveness of three malaria treatments – [artesunate+amodiaquine (AS+AQ), artesunate+ sulphadoxine-pyrimethamine (AS+SP) or artemether-lumefantrine (AL)]. They note similar results for the AG and shared frailty model, while the GEE models didn’t detect the effect of some confounders and advise for the use of methods that account for correlation within subjects with multiple events.

2.3 Rationale for the inclusion of informative censoring

Ghosh and Lin, 2003 note that informative censoring occurs in situations where the censoring times “depend on the observed or non-observed recurrent times”. An illustration of this in the HIV setting, is the informative drop-outs such as the withdrawal of sicker patients from a study way before study end.

Ghosh et al., (2003) specify two possible sources of informative censoring in practical settings:- the voluntary withdrawal of subjects for reasons related to the event process as well as death.

The exclusion of the censoring assumption from the study may generally lead to biased estimates. Castelli et al., 2006 in adapting the Inverse Probability of Censoring Weighted (IPCW) to study the survival times of asthma patients while including informative censoring (patients felt they were okay and did not need to consult a doctor) show that information coming from censoring process improves the survival estimate. Another approach is that by Ghosh and Lin, 2003 where they introduce a semi-parametric approach for recurrent events in the presence of dependent censoring and apply it ALIVE cohort study (Vlahov et al., 1991). A comparison with the accelerated failure times model as proposed by Lin et al., 1998, which assumes non-informative censoring found that their proposed method yielded "a much larger estimate for the effect of baseline HIV status on hospitalizations than the method proposed by Lin".

There is a body of literature to show that drop-outs in HIV programs are informative. (Berheto et al., 2014) in assessing the predictors of LTFU in Patients Living with HIV/AIDS after Initiation of Antiretroviral Therapy found that sicker patients (baseline CD4 $<$ 200 cells/mm3 – HR 1.7, 95% CIs 1.3-2.2, regimen substitution – HR 5.2; 95% CIs 3.6-7.3) and receiving non-isoniazid (INH) prophylaxis (HR 3.7; 95% CIs 2.3-6.2) as accelerators for LTFU. According to Assemie et al., 2018, being on WHO clinical stage IV as well as receiving isoniazid preventive therapy were significant predictors of LTFU.

On the premise of this, informative censoring is thus incorporated in this study.

2.4 Advances in the incorporation of informative censoring

A number of advances have been made to incorporate informative censoring through accounting for the dependence between the recurrent event and censoring times. Chang and Wang, 1999, suggest a joint scale model for both the dependent time and recurrent event gap times, while Ghosh and Lin, 2003 uses accelerated failure models to model the relationship between recurrent events and failure time. Ghosh and Lin, 2003 suggested the use of scale-change models to specify the distributions of recurrent event and dependent censoring. Fully parametric models were also studied by Lancaster and Intrator, 1998 for the recurrent event process and the failure time, where using a non-observed frailty terms, the recurrent event is allowed to be correlated with the failure time.

In this study, we adopt the approach suggested by Huang and Wang, 2004 where an unobserved frailty is used to account for the relationship between the recurrent event and censoring times. This is achieved using a joint model where both the failure time and recurrent event process are of interest.

Chapter 3

METHODOLOGY

3.1 Method and setting

Consider the rate of occurrence of recurrent events in a given time interval i.e. (0, Γ_O), where Γ_O>0 is set with the information of possible epochs of recurrences observable up to Γ_O. Suppose N(t) is the number of recurrent events that occur on or before t, t>0.

The functional rate of a recurrent event at t, t $\in$ [0,Γ_O], is expressed as

$$\lambda(t) = \lim_{\Delta \to 0^+} \frac{Pr[N[t+\Delta] - N[t]>0]}{\Delta}. \tag{3.1}$$

The rate is considered theoretically different from the intensity function. Specifically, we define the functional rate as the occurrence rate of recurrent events and is not conditional to the event history. The intensity function on the other hand is the occurrence rate and is conditional on the event history.
Let the cumulative rate function be described as

$$\Lambda(t) = \int_0^t \lambda[\mu] d\lambda. \tag{3.2}$$

Also, suppose γ be the censoring time for observation of the recurrent event process at termination. The interest remains the occurrence rate in the time interval $(0, \Gamma_O)$. We define $\tau = min[\gamma, \Gamma_0]$ as the new censoring time used in the proposed models.

3.2 Prentice, Williams, and Peterson (PWP) Model

The PWP model as proposed by Prentice, Williams, and Peterson (Prentice et al., 1981) is a 'conditional model' where the individuals are at risk for an event if and only if they were at risk for a previous event. To achieve this, each event occurrence is put into a different stratum with all participants at risk in the first stratum. Under this model, only participants that experienced the previous event would then be at risk for the next event (Amorim and Cai, 2015). The model estimates both overall and event-specific effects and uses robust standard errors to account for correlation.

When time since entry is of interest, this model condenses to total time model (PWP-TT), while when time since last event is of interest the model becomes a gap time model commonly abbreviated as PWP-GT. See Figure **3.1**. The key difference between this model and the Andersen-Gill (AG) is in terms of the effects of the covariates in different strata. In the AG model, effects of covariates are constant across all strata, while this varies in different strata for the PWP (Amorim and Cai, 2015).

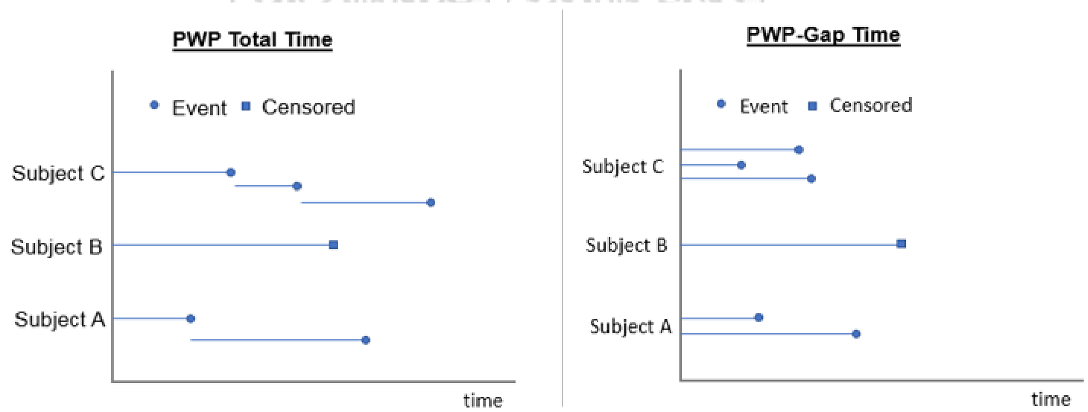

Figure 3.1: Representation of the PWP Models

The motivation for the choice of PWP model is informed by its underlying assumptions relative to the nature of data under consideration. The recurrent event of interest are the recurrences of LTFU. In an ideal HIV clinic set-up, LTFU clients are expected to be followed up by expert clients. Once found, enhanced adherence counseling is provided by clinicians. This affects the likelihood of the next LTFU incidence, as the next LTFU incidence may be affected by the previous incidence, hence a conditional relation between events.

Additionally, due to enhanced adherence counseling, the baseline hazard for each event cannot be assumed to be similar for all events, a factor adjusted for by the PWP. These two factors rule out the AG. The WLW is primarily ruled out due to the fact that an individual is simultaneously at risk for all events concurrently. The frailty assumes that correlation among events is due to some individuals being more prone to develop recurrent event as compared to others because of some unobserved/unknown factors. Based on the data available, we have no rationale to indicate that individuals may be different.

The choice of PWP-TT over the PWP-GT is informed by the present capability of implementation within the reReg package. Currently the reReg package only works with total time specification.

3.3 Mathematical Representation of PWP Model

To include informative censoring, we built on the generic approach proposed by Huang and Wang, 2004 where they jointly model the recurrent event process and failure times.

The key to this approach is to model this relationship via a subject specific latent variable, Z_i, that models the association between the intensity of the recurrent event and the hazard of the failure time.

As relayed by Huang and Wang, 2004, the **intensity function** is

$$\lambda_i(t) = Z_i \lambda_O(t) exp\{\mathbf{x}\alpha\}, \tag{3.3}$$

And **Hazard of the failure time**, D, is given by:

$$h_i(t) = Z_i h_O(t) exp\{\mathbf{x}\beta\}. \quad (3.4)$$

Where:

λ_O and h_O are baseline intensity and hazard functions respectively.
α and β are coefficients.
$X_i(t)$ is the p-dimensional set of covariates

Based on the proposed intensity function by Huang and Wang, 2004, the PWP models are modified as below. A mathematical representation of each of the models is presented below, as well as the individual model specification with informative censoring incorporated (**Table 3.1**). Essentially, each model is multiplied by the unobserved frailty, Z_i.

Table 3.1: Mathematical Representation of the intensity functions of PWP models with informative censoring

Model Name	Mathematical presentation	Under Informative censoring
PWP-GT	$M_{ij}(t)\lambda_{Oj}(t - t_{j-1})exp\{X'\beta\}$	$Z_i M_{ij}(t)\lambda_{Oj}(t - t_{j-1})exp\{X'\beta\}$
PWP-TT	$M_{ij}(t)\lambda_{Oj}(t)exp\{X'\beta\}$	$Z_i M_{ij}(t)\lambda_{Oj}(t)exp\{X'\beta\}$

Where:-
PWP-GT is the Prentice, Williams, and Peterson Gap-Time Model,
PWP-TT is the Prentice, Williams, and Peterson Total-Time Model
And: -
$M_i(t)$ is the at risk indicator for the j-th event and ith person at time t. This is 1 when at risk for event j, and zero when not at risk for event j
λ_O is the baseline intensity function
$X(t)$ is the p-dimensional set of covariates

3.4 Evaluation of the PWP-GT model under informative censoring

3.4.1 Parameter Estimation

The model as proposed by Huang and Wang, 2004 jointly models the relationship between the intensity of the recurrent event process and hazard of the failure times, via a frailty. This then implies that we shall specifically estimate the cumulative hazard and intensity functions. For estimation of these in the PWP-TT, we maintain the approach suggested by Huang. A key difference in the PWP-TT and the generic model used by Huang is the definition of the baseline hazard. While a constant baseline hazard is assumed for the model by Huang, the PWP-TT assumes different baseline per person and event. Since this is treated as a nuisance parameter, the derivations as described by Huang also apply for the PWP-TT model.

We now estimate the cumulative hazard and intensity functions below:

Starting with notations:

Let $\mathbf{N(t)}$ represent events occurring at or before some time T_o.

D be failure time, in this case LTFU due to stigma.

C failure time due to other reasons than D.

x is a vector of 1xp covariates.

Let **Y** be the point at which the observation of recurrent events ceases, such that, **Y**=min(C, D, T_o). Note that C, D, and T_o are mutually independent.

Let m_i be the number of events occurring before censoring time Y_i We introduce a non-negative latent variable Z, such that the **intensity function**, $\lambda_i(t)$, is given by

$$\lambda_i(t) = Z_i \lambda_O(t) exp\{\mathbf{x}\alpha\}, \tag{3.5}$$

with

$$E[Z \mid x] = E[Z]. \tag{3.6}$$

Additionally, the Hazard function, $h(t)$ is given by:

$$h(t) = Z_i h_O(t) exp\{\mathbf{x}\beta\}. \tag{3.7}$$

Note that a large/small z increases or decreases the intensity and hazard respectively.

The rate function is defined as

$$\mu_z \lambda_0 exp\{\mathbf{x}\alpha\}. \tag{3.8}$$

Estimation of the cumulative intensity function

Assuming that for individual i, observations $(x_i, z_i, m_i(t_i, \ldots 0, t_{imi}), Y_i$ are made and are iid, then the cumulative intensity function, π_t, is given by

$$\pi_i(t) = \frac{Z_i M_i \lambda_O(t_i) exp\{\alpha \mathbf{x}_i\}}{Z_i M_i \Lambda(y_i) exp\{\alpha \mathbf{x}_i\}}. \tag{3.9}$$

However we assume that the density does not depend on Z_i, M_i or x, reducing the density to

$$\pi_i(t) = \frac{\lambda_O(t_i)}{\Lambda_O(y_i)}, 0 \leq t \leq y_i. \tag{3.10}$$

Where: $\Lambda_O y_i = \int_0^t \lambda_0(u) du = 1$. Since we assume iid, the conditional likelihood, L_c, is generated for n subjects, assuming m_i events per individual as:

$$L_c = \prod_{i=1}^{n} \cdot \prod_{j=1}^{m_i} \cdot \frac{\lambda_O(t_{ij})}{\Lambda_O(y_i)}. \tag{3.11}$$

An estimator for $\Lambda_O(y_i)$ is generated from the work of (Wang et al., 1986) and is given by:

$$\Lambda_O(t) = \prod (1 - \frac{d}{R}). \tag{3.12}$$

where d is the number of patients experiencing event at time t, and R is total persons at risk.

To estimate α, Huang proposes a class of unbiased estimating equations, shown below:

$$n^{-1}\sum_{i=1}^{n} w_i \mathbf{x}_i^{'}(m_i \Lambda_O(Y)^{-1} - \exp\{\mathbf{x}\gamma\}) = 0, \tag{3.13}$$

where $x_i^{'} = (1, \mathbf{X}_i),^{'} = In(\mu_z, \alpha^{'})$, w_i is a weighting function and $\Lambda_o(Y)^{-1}$ is defined in 1 above Huang proposes the estimation of α by replacing $\Lambda_O(Y)^{-1}$ by 3.12 above.

Estimation of the cumulative hazard function

A 'borrow-strength' method is proposed by Huang and Wang, 2004 for the estimation of the cumulative hazard function, $H_o(t)$. This involves first estimating the value of the latent variable from recurrent event data, then using the estimated value in the failure time model estimation.

1. Firstly the frailty for every individual, Z_i is computed using

$$Z_i = \frac{m_i}{(\Lambda_0(Y_i)\exp\{\alpha\mathbf{x}_i\}}, \tag{3.14}$$

with Λ_O and α computed as described above.

2. From 3.11, the score function of the likelihood function, L_c, is computed as follows:

$$U(\beta) = \frac{1}{n}\sum_{i=1}^{n}\Delta_i\{\mathbf{X}_i - \frac{\sum_{j=1}^{n}\mathbf{X}_i Z_j \exp\{\mathbf{X}_j\beta\} I(Y_j)}{\sum_{j=1}^{n} Z_j \exp\{\mathbf{X}_j\beta\} I(Y_j)}\}. \tag{3.15}$$

$$U(\beta) = \varepsilon\{\mathbf{X}\Delta_i I(Y)\} - \int_0^{T_o} \frac{\varepsilon\{\mathbf{X}_i Z_j \exp\{\mathbf{X}\beta\} I(Y_j)\}}{\varepsilon\{Z\exp\{\mathbf{X}_j\beta\} I(Y_j)\}} d\varepsilon\{\Delta I(Y)\}. \tag{3.16}$$

To estimate β,we equate $U(\beta) = 0$.

The estimator for $H_O(t)$ is given by

$$H_O(t) = \int_0^t \frac{d\varepsilon\{\Delta I(Y)\}}{\varepsilon\{Z\exp(\mathbf{X}\beta) I(Y)\}}. \tag{3.17}$$

3.5 Evaluation of the PWP-GT Model

Empirical Evaluation of Data

Patient-level data from four facilities in central Kenya, collected between 2013 and 2018 was used. The recurrent event of interest were the incidences of loss to follow-up. Censoring for a drop-out was considered informative, if on the basis of the reason provided for drop-out, a person cited 'stigma'. Additionally, time independent covariates were incorporated in the analysis. Data preparation was prepared to reflect the PWP-TT format as laid out by (Therneau and Grambsch, 2000, pp. 187–189). Primarily, we sought to establish if the incorporation of informative censoring improves estimation of the hazard and rate estimates. While this study acknowledges the possibility of competing risks, they were not incorporated in this study.

Computing environment

The models were assessed under the Rstudio computing environment. To account for informative censoring, the 'method' function within 'reReg function' in the reReg package (Chiou and Huang, 2019) was set to "cox.HW". Fitting the models was achieved by using the cluster and strata functions in the base survival package required by the reReg package. The Hmisc package was used to provide basic descriptive statistics.

3.5.1 Application To HIV Data

Data Description

Determination of the incidence of LTFU was computed using the cumulative baseline hazard technique with start of ART as time 0 and LTFU at the time a particular patient failed to return to CCC clinic for 38 weeks since the scheduled appointment date. The recurrent event of interest are the recurrences of LTFU, within the period of data abstraction (January 2013 to December 2018).

In a typical HIV set-up, after determination of LTFU, several consulted efforts by expert clients to retrace clients who are lost are implemented. The efforts include calling back clients and visiting them at their homes with the aim of returning them back to clinic. Patients who return back to care are exposed to recurrence of the event (LTFU). Censoring was considered informative if and only if at the terminal LTFU occurrence, LTFU was due to stigma. To gain this information, expert clients who got in touch with patients were informed by the patients that stigma and from highly active antiretroviral therapy (HAART) was the reason as to why they had ceased coming to the facility. On the other hand, it was non-informative for patients who were still in care as at the end of data abstraction period (December 2018) or for any other exists such as death as there was no additional information to indicate if these were linked to the event process. All these were right censored.

Among the sample of LTFU clients, additional parameters in the dataset included - Patient ID: patient identification, which may be repeated due to recurrence of LTFU. Time to LTFU: time to the event of interest, i.e. LTFU, which for some patients may be recurrent. Event: which is 1 for any instance of LTFU recurrence and 0 if no LTFU recurrence was noted. Status: for any who did not experience event (LTFU) recurrence (i.e. event=0). This variable records 1 for instances of informative censoring (i.e. LTFU due to stigma) and 0 for instances of non-informative censoring described above. Differentiated HIV care for patients who are either under differentiated care model or not, age of the patient, gender of the patient and ARV regimen line as defined by WHO. All PLHIV (Persons Living with HIV) adults on HAART who enrolled at four facilities in central Kenya in January 2013 to December 2018 were considered for analysis. Those who had at-least one follow-up visit prior to January 2013 were eligible to be included. Children below 15 years and confirmed pregnant mothers were excluded. Additionally, we also excluded PLHIV who had unknown ART start date, unknown outcome, and transferred in with incomplete baseline information. Data was pulled from point of care electronic medical database, consolidated in MS Excel, and exported to R Studio for further analysis.

The main event variable in the analysis was Time to LTFU (in months) with other exits treated as a competing event. Additionally, we employ a joint model for event recurrence and the hazard of failure, with both LTFU recurrence and the hazard due to stigma of interest.

Chapter 4

RESULTS AND DISCUSSIONS

4.1 Descriptive Statistics

Table A.1 gives a description of the dataset containing 58 patients. Majority of the patients were female 35(59.3%). The average age was 36.4 years (SD=6.28), and most patients were aged between 31 to 40 years of age. Most of the patients (89.8%) were on a differentiated care model and on a first line regimen (57.6%).

In this study, we jointly model the recurrent event process and a failure time, with the joint model described in **section 3.3** adopted. We define the recurrent event process as the LTFU recurrences and the failure event as LTFU due to stigma. Using the PWP-TT model, we seek to assess the effect of select explanatory variables (described next) on the rate of LTFU recurrence and the risk of LTFU due to stigma. The explanatory variables of interest were gender set to 1 for male and 2 for female, 1 for first line regimen and 2 for second line regimen, 1 if patient has ever been on a differentiated care model and 0 otherwise and age disaggregated into three categories of less than 30, 31 to 40, and greater than 40 years.

Cumulatively, there were 256 incidences of patients loss to follow up spread among 51 patients. Overall, 6 in 10 patients (61%) experienced between 1 (inclusive) to 6 (inclusive) LTFU incidences. The median time to lost to follow-up was 8.6 months (range: 3.0-82.9 months). Sixteen instances of informative censoring were reported. Event plots (see **Appendix**) give a visual depiction of the incidence of recurrent events of interest (LTFU) and failure event (LTFU due to stigma) disaggregated by variables of interest. Crude comparisons indicated an equal incidence of failure event by regimen (first line and second line), and by gender (male vs female). A higher incidence of both the recurrent event and failure events was noted for those aged between 31 to 40 years of age.

Additionally, cumulative sample mean plots are provided to provide the cumulative number of LTFU occurrences during the study period - **Figure A.6**. The number of incidences is shown to increase consistently across the monitoring window.

4.2 Application of the PWP-TT Model to HIV data

To empirically assess the performance of the proposed PWP-TT model under joint modelling of both the failure and event time, we apply it to the patient level data described in section 3.6.1. For this analysis, sex, age, regimen, and differentiated care were explanatory variables included. The event was LTFU recurrence (1 for LTFU occurrence, 0 for no occurrence). A 'composite' censoring variable is used. By this, first an event variable is included in the dataset (1 for LTFU incidence, 0 otherwise). For any person not experiencing the event (event=0), censoring (via 'status' variable), with 1 for informative censoring (LTFU due to stigma) and 0 for non-informative censoring is used. Specifically, we empirically investigate the effect of covariates on rate of LTFU and risk of LTFU arising from stigma using the PWP-TT model. Parameter estimates, standard errors (SE), and corresponding p-values are shown in **Table 4.1**. Standard errors were estimated by resampling 100 times from patient data. Given that there are currently no model checking methods for reReg objects, we compare PWP-TT results to those from AG and the Cox model.

Table 4.1: Estimates for risk factors with standard errors, and p-values under the PWP-TT model

LTFU Recurrence					
		coef	**exp (coef)**	**StdErr**	**p.value**
Gender (base - Male)	Female	0.042	1.043	0.177	0.814
Age (base - $\geq$ 30)	31-40	-0.004	0.996	0.326	0.991
	>40	-0.045	0.956	0.311	0.884
Regimen (base - First Line)	Second Line	0.061	1.063	0.211	0.773
Differentiated Care? (base - No)	Yes	-0.337	0.714	0.187	0.072 .
Hazard of LTFU due to stigma					
Gender (base - Male)	Female	-2.11	0.121	0.874	0.016 *
Age (base - $\geq$ 30)	31-40	-3.635	0.026	1.826	0.046 *
	>40	-0.213	0.808	0.824	0.796
Regimen (base - First Line)	Second Line	-0.394	0.674	0.776	0.612
Differentiated Care? (base - No)	Yes	-3.741	0.024	1.774	0.035 *

Codes: **0** '***' **0.001** '**' **0.01** '*' **0.05** '.' **0.1** ' ' **1**

From **Table 4.1**, it can be noted that patients on differentiated care had lower likelihoods for both LTFU recurrence (29% lower, p=0.072) as well as for the hazard of LTFU arising from stigma (98% less likely, p = 0.035). Gender was also a significant predictor of LTFU due to stigma. The instantaneous risk for LTFU from stigma is lower for females (88% less risk, p = 0.016) compared to their male counterparts, while also lower for those aged 31-40 when compared to those under 30 years.

Presently, the performance of model fit cannot be assessed using routine statistical approaches. In fact, Huang et al., 2010 note that "under informative censoring, model checking is expected to be a difficult task in general". Consequently, we make general comparisons with the Cox and AG. The Cox model used here assumed non-informative censoring assumed and only data from the first event included. The 'survival' package and 'coxph' function in R were used to fit these data. To check the proportionality assumption, the Schoenfeld residuals against the transformed time were computed with a non-significant relationship noted (p=0.316), implying the proportionality assumption was not violated.

Results from the Cox model **Table A.3** indicated that patients aged 31-40 had a significant (p=0.009) fourfold risk of getting LTFU, compared to those under 30 years, a finding generally inconsistent with the known fact that younger persons generally at a higher risk of getting LTFU - see (Teshale et al., 2020), Meloni et al., 2014). The wide confidence intervals for this variable also indicate that this findings may be volatile in the presence of different data. It should also be noted that since only the first event was considered, the Cox model used only 23% (58/256) of available information, hence a lot of information wastage.

On the other hand, the AG model indicated a 92% reduced risk (p = 0.068) of LTFU hazard from stigma for persons on a differentiated care program (**Table A.4**), with this effect similarly significant for the PWP-TT. The direction of effect for LTFU recurrence for differentiated care was also consistent with that of the PWP-TT, albeit a more pronounced for the PWP-TT. Absolute differences between the standard error estimates for the other variables between the PWP-TT and AG are also comparable, indicating that they might have fit the data well. Crude comparisons of the absolute difference in standard error estimates between the two models for LTFU recurrence estimates (range: 0.004, 0.149) and LTFU due to stigma/drug reaction (range: 0.060, 0.418) are comparable.

4.3 Discussion

The results from the PWP-TT model, indicate a reduced likelihood for LTFU recurrence and hazard arising from stigma. These are generally unsurprising given that one well documented benefit of differentiated care models is to fight stigma - Kalichman et al., 2019, Grimsrud et al., 2016, as well as to generally improve retention in care - Bacha et al., 2018, Wringe et al., 2018, Bemelmans et al., 2014, Mutasa-Apollo et al., 2017, Avong et al., 2018. Additionally, and consistent with known literature, the hazard of LTFU arising from stigma is relatively high for among the youth - Wolf et al., 2014, Megerso et al., 2016.

On the other hand, estimates from the Cox model proved too unreliable (very wide CIs), potentially arising from the fact they used only 23% of the data. This 'wastage of information' remains a downside for the Cox model - Ullah et al., 2014, among other limitations. Overall, these results underline what is generally known as regards the Cox model; that it may not be suitable in observational recurrent studies.

The differences in estimates noted between the AG and PWP-TT are majorly down to the assumptions of these two models. A major assumption in AG model is that the inter-event time increments are conditionally uncorrelated with given covariates. It also assumes the same baseline hazard for all persons, which may not be the case for LTFU in an HIV setting, given that intensive adherence counseling may alter the subsequent likelihood of an individual getting LTFU. On the other hand, it is best suited to cases of independent increments across observation units, and is also the easiest extension to the Cox model to replicate. On the other hand, either of the PWP models (GT or TT) adjusts for varying baseline hazards across observation units which may be efficient in the case of LTFU, where due to adherence counseling, the baseline hazard may change for a subsequent event may change.

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

In this study, we extended PWP-TT model to cover informative censoring when making inferences on recurrent events under HIV retention setting. The proposed methodology intrinsically extends the existing method in literature on recurrent events under typical HIV resource limited setting. A well-defined truncation time T, with Z_i overall i.e. $Z_i M_{ij}(t)\lambda_{oj}(t - t_{j-1}) exp(X_i(t)'\beta$ has been assumed. Here, we have concentrated on the LTFU as the recurrent event, and LTFU due to stigma as the hazard and fitted real data. It can also be developed further to other regression/Cox methods. The concept can be extended to other approaches by modeling $\lambda_{oj}(t)$ and $\lambda_{oj}(t - t_j - 1)$ with proportional hazards models.

The main purpose of this work was to provide an overview, applicable statistical techniques when analyzing recurrent event data under informative censoring in HIV retention setting. The typical real data used in this work is from a routine well established HIV care clinic. The longitudinal approach to analyze recurrent-event data applied here can also be applicable to other observational cohort studies. Because the technical approach employed here are extensions of Cox proportional hazards regression, explicit issues that affect model modification can also be handled in the same manner as the classic applied techniques.

Generally, the choice of recurrent event data analysis technique is determined by several factors, i.e. events; relationship between events; varying effects across recurrences; the medical/biological process; and independence/dependence structure. Usually the stratified models, as PWP (total or gap times) or multi-state models, are useful whenever there are relatively few recurrent-events per individual and the risk of recurrences.

A recurrent events model will ideally help to provide insights into the program/disease structure and process. Hence, it is critical to consider the censoring mechanism and perform analysis that enhances comprehension of the risk factors.

Finally, the proposed method is not without limitations, particularly as regards model checking. As highlighted by Huang et al., 2010, this is expected to be a generally difficult task for recurrent events with informative censoring. There are no current implementations in the reReg package in R.

Bibliography

Abadi, A., Yavari, P., Dehghani-Arani, M., Alavi-Majd, H., Ghasemi, E., Amanpour, F., & Bajdik, C. (2014). Cox models survival analysis based on breast cancer treatments. *Iranian journal of cancer prevention*, *7*(3), 124.

Amorim, L. D., & Cai, J. (2015). Modelling recurrent events: A tutorial for analysis in epidemiology. *International journal of epidemiology*, *44*(1), 324–333.

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The annals of statistics*, 1100–1120.

Avong, Y. K., Aliyu, G. G., Jatau, B., Gurumnaan, R., Danat, N., Kayode, G. A., Adekanmbi, V., & Dakum, P. (2018). Integrating community pharmacy into community based anti-retroviral therapy program: A pilot implementation in abuja, nigeria. *PloS one*, *13*(1).

Bacha, J. M., Aririguzo, L. C., Malingoti, B., Wanless, R. S., Ngo, K., Campbell, L. R., Schutze, G. E., Et al. (2018). The standardized pediatric expedited encounters for art drugs initiative (speedi): Description and evaluation of an innovative pediatric, adolescent, and young adult antiretroviral service delivery model in tanzania. *BMC infectious diseases*, *18*(1), 1–9.

Bemelmans, M., Baert, S., Goemaere, E., Wilkinson, L., Vandendyck, M., van Cutsem, G., Silva, C., Perry, S., Szumilin, E., Gerstenhaber, R., Et al. (2014). Community-supported models of care for people on hiv treatment in sub-s aharan a frica. *Tropical Medicine & International Health*, *19*(8), 968–977.

Berheto, T. M., Haile, D. B., & Mohammed, S. (2014). Predictors of loss to follow-up in patients living with hiv/aids after initiation of antiretroviral therapy. *North American journal of medical sciences*, *6*(9), 453.

Castañeda, J., & Gerritse, B. (2010). Appraisal of several methods to model time to multiple events per subject: Modelling time to hospitalizations and death. *Revista Colombiana de Estadıstica*, *33*(1), 43–61.

Castelli, C., Saint-Pierre, P., & Daures, J. (2006). Informative censoring in survival analysis and application to asthma. *Far East Journal of Theoretical Statistics*, *19*(2), 203.

Chang, S.-H., & Wang, M.-C. (1999). Conditional regression analysis for recurrence time data. *Journal of the American Statistical Association*, *94*(448), 1221–1230.

Chiou, S. H. (, & Huang, C.-Y. (2019). *Rereg: Recurrent event regression* [R package version 1.2.1]. R package version 1.2.1. https://CRAN.R-project.org/package=reReg

Collett, D. (2015). *Modelling survival data in medical research*. CRC press.

Cook, R. J., & Lawless, J. (2007). *The statistical analysis of recurrent events*. Springer Science & Business Media.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

David, G., & Kleinbaum, K. (2016). *Survival analysis: A self-learning text*. Springer-Verlag New York.

Farewell, V. (1979). An application of cox's proportional hazard model to multiple infection data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *28*(2), 136–143.

Ghosh, D., & Lin, D. (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*, *59*(4), 877–885.

Grimsrud, A., Bygrave, H., Doherty, M., Ehrenkranz, P., Ellman, T., Ferris, R., Ford, N., Killingo, B., Mabote, L., Mansell, T., Et al. (2016). Reimagining hiv service delivery: The role of differentiated care from prevention to suppression. *Journal of the International AIDS Society*, *19*(1), 21484.

Huang, C.-Y., & Wang, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*, *99*(468), 1153–1165.

Huang, C.-Y., Qin, J., & Wang, M.-C. (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics*, *66*(1), 39–49.

Kalichman, S. C., Mathews, C., Banas, E., & Kalichman, M. O. (2019). Stigma management intervention to improve antiretroviral therapy adherence: Phase-i test of concept trial, cape town south africa. *Global public health*, *14*(8), 1059–1074.

Kelly, P. J., & Lim, L. L.-Y. (2000). Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in medicine*, *19*(1), 13–33.

Lancaster, T., & Intrator, O. (1998). Panel data with survival: Hospitalization of hiv-positive patients. *Journal of the American statistical association*, *93*(441), 46–53.

Lawless, J. F. (1987). Regression methods for poisson process data. *Journal of the American Statistical Association*, *82*(399), 808–815.

Lee, E. W., Wei, L., Amato, D. A., & Leurgans, S. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations, In *Survival analysis: State of the art*. Springer.

Lin, D., Wei, L., & Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika*, *85*(3), 605–618.

Megerso, A., Garoma, S., Tolosa Eticha, T. W., Daba, S., Tarekegn, M., & Habtamu, Z. (2016). Predictors of loss to follow-up in antiretroviral treatment for adult patients in the oromia region, ethiopia. *HIV/AIDS (Auckland, NZ)*, *8*, 83.

Meloni, S. T., Chang, C., Chaplin, B., Rawizza, H., Jolayemi, O., Banigbe, B., Okonkwo, P., & Kanki, P. (2014). Time-dependent predictors of loss to follow-up in a large hiv treatment cohort in nigeria, In *Open forum infectious diseases*. Oxford University Press.

Mutasa-Apollo, T., Ford, N., Wiens, M., Socias, M. E., Negussie, E., Wu, P., Popoff, E., Park, J., Mills, E. J., & Kanters, S. (2017). Effect of frequency of clinic visits and medication pick-up on antiretroviral treatment outcomes: A systematic literature review and meta-analysis. *Journal of the International AIDS Society*, *20*, 21647.

Ni, J. (2009). Application of cox proportional hazard model to the stock exchange market. *BS Undergraduate Mathematics Exchange*, *6*(1), 12–18.

Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, *68*(2), 373–379.

Sagara, I., Giorgi, R., Doumbo, O. K., Piarroux, R., & Gaudart, J. (2014). Modelling recurrent events: Comparison of statistical models with continuous and discontinuous risk intervals on recurrent malaria episodes data. *Malaria journal*, *13*(1), 293.

Schemper, M. (1992). Cox analysis of survival data with non-proportional hazard functions. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *41*(4), 455–465.

Teshale, A. B., Tsegaye, A. T., & Wolde, H. F. (2020). Incidence and predictors of loss to follow up among adult hiv patients on antiretroviral therapy in university of gondar comprehensive specialized hospital: A competing risk regression modeling. *PloS one*, *15*(1), e0227473.

Thenmozhi, M., Jeyaseelan, V., Jeyaseelan, L., Isaac, R., & Vedantam, R. (2019). Survival analysis in longitudinal studies for recurrent events: Applications and challenges. *Clinical Epidemiology and Global Health*, *7*(2), 253–260.

Therneau, T. M., & Grambsch, P. M. (2000). The cox model, In *Modeling survival data: Extending the cox model*. Springer.

Tolosie, K., & Sharma, M. (2014). Application of cox proportional hazards model in case of tuberculosis patients in selected addis ababa health centres, ethiopia. *Tuberculosis research and treatment*, *2014*.

Ullah, S., Gabbett, T. J., & Finch, C. F. (2014). Statistical modelling for recurrent events: An application to sports injuries. *Br J Sports Med*, *48*(17), 1287–1293.

Villegas, R., Julià, O., & Ocaña, J. (2013). Empirical study of correlated survival times for recurrent events with proportional hazards margins and the effect of correlation and censoring. *BMC medical research methodology*, *13*(1), 95.

Vlahov, D., Anthony, J. C., Muñoz, A., Margolick, J., Nelson, K. E., Celentano, D. D., Solomon, L., & Polk, B. F. (1991). The alive study: A longitudinal study of hiv-1 infection in intravenous drug users: Description of methods. *Journal of Drug Issues*, *21*(4), 759–776.

Wang, M.-C., Jewell, N. P., & Tsai, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *the Annals of Statistics*, 1597–1605.

Wei, L.-J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American statistical association*, *84*(408), 1065–1073.

Wheatley-Price, P., Hutton, B., & Clemons, M. (2012). The mayan doomsday's effect on survival outcomes in clinical trials. *CMAJ*, *184*(18), 2021–2022.

Wolf, H. T., Halpern-Felsher, B. L., Bukusi, E. A., Agot, K. E., Cohen, C. R., & Auerswald, C. L. (2014). it is all about the fear of being discriminated [against]... the person suffering from hiv will not be accepted": A qualitative study exploring the reasons for loss to follow-up among hiv-positive youth in kisumu, kenya. *BMC public health*, *14*(1), 1154.

Wringe, A., Cawley, C., Szumilin, E., Salumu, L., Amoros Quiles, I., Pasquier, E., Masiku, C., & Nicholas, S. (2018). Retention in care among clinically stable antiretroviral therapy patients following a six-monthly clinical consultation schedule: Findings from a cohort study in rural malawi. *Journal of the International AIDS Society*, *21*(11), e25207.

Yang, W., Jepson, C., Xie, D., Roy, J. A., Shou, H., Hsu, J. Y., Anderson, A. H., Landis, J. R., He, J., Feldman, H. I., Et al. (2017). Statistical methods for recurrent event analysis in cohort studies of ckd. *Clinical Journal of the American Society of Nephrology*, *12*(12), 2066–2073.

Appendix A

APPENDIX

A.1 Tables

Table A.1: A description of the dataset

Variable	**Disaggregation**	**Frequency (%)**
Gender	Male	24 (40.7%)
	Female	35 (59.3%)
Age	≥ 30	8 (13.6%)
	31–40	31 (52.5%)
	>40	20 (33.9%)
On differentiated Care?	No	6 (10.2%)
	Yes	53 (89.8%)
Regimen	First Line	34 (57.6%)
	Second Line	25 (42.4%)
Viral Load Result	Suppressed	51 (86.4%)
	Unsuppressed	8 (13.6%)

Table A.2: LTFU recurrences and LTFU due to stigma disaggregated by variables of interest

		Terminal LTFU due to stigma	Number of LTFU Recurrences		
			0	**1-4**	**5-9**
Sex	Male	8 (50%)	2 (28.6%)	11 (47.8%)	10 (35.7%)
	Female	8 (50%)	5 (71.4%)	12 (52.2%)	18 (64.3%)
Age	≦ 30	3 (18.8%)	2 (28.6%)	3 (13%)	3 (10.7%)
	31 - 40	5 (31.3%)	3 (42.9%)	11 (47.8%)	17 (60.7%)
	> 40	8 (50%)	2 (28.6%)	9 (39.1%)	8 (28.6%)
On differentiated Care?	No	2 (12.5%)	1 (14.3%)	2 (8.7%)	3 (10.7%)
	Yes	14 (87.5%)	6 (85.7%)	21 (91.3%)	25 (89.3%)
Regimen	First Line	8 (50%)	4 (57.1%)	12 (52.2%)	17 (60.7%)
	Second Line	8 (50%)	3 (42.9%)	11 (47.8%)	11 (39.3%)
Viral Load Result	Suppressed	12 (75%)	5 (71.4%)	20 (87%)	26 (92.9%)
	Unsuppressed	4 (25%)	2 (28.6%)	3 (13%)	2 (7.1%)
Total		16	7	23	28

Table A.3: Summary of Cox Results

		coef	**exp(coef)**	**se(coef)**	**95% CI (lower)**	**95% CI (upper)**
Gender (base - Male)	Female	-0.525	0.591	0.314	0.32	1.094 .
Age (base - $\geq$ 30)	31-40	1.481	4.397	0.568	1.446	13.374 **
	> 40	0.875	2.399	0.605	0.733	7.854
Regimen (base - First Line)	Second Line	-0.154	0.857	0.331	0.448	1.64
On Differentiated Care? (base - No)	Yes	-0.107	0.898	0.525	0.321	2.511

Codes: **0** '***' **0.001** '**' **0.01** '*' **0.05** '.' **0.1** ' ' **1**

Table A.4: Estimates for risk factors with standard errors, and p-values under the AG model

LTFU Recurrence					
		coef	**exp (coef)**	**StdErr**	**p.value**
Gender (base - Male)	Female	0.001	1.001	0.08	0.992
Age (base - $\geq$ 30)	31-40	0.264	1.302	0.177	0.137
	>40	0.163	1.177	0.253	0.519
Regimen (base - First Line)	Second Line	0.13	1.139	0.227	0.568
Differentiated Care? (base - No)	Yes	-0.267	0.766	10.191	0.164 .
Hazard of LTFU due to stigma					
Gender (base - Male)	Female	-0.642	0.526	0.534	0.229
Age (base - $\geq$ 30)	31-40	-1.033	0.356	1.283	0.421
	>40	0.809	2.246	0.746	0.29
Regimen (base - First Line)	Second Line	0.936	2.55	0.667	0.161
Differentiated Care? (base - No)	Yes	-2.478	0.084	1.356	0.068

Codes: **0** '***' **0.001** '**' **0.01** '*' **0.05** '.' **0.1** ' ' **1**

A.2 Event and cumulative Hazard Plots

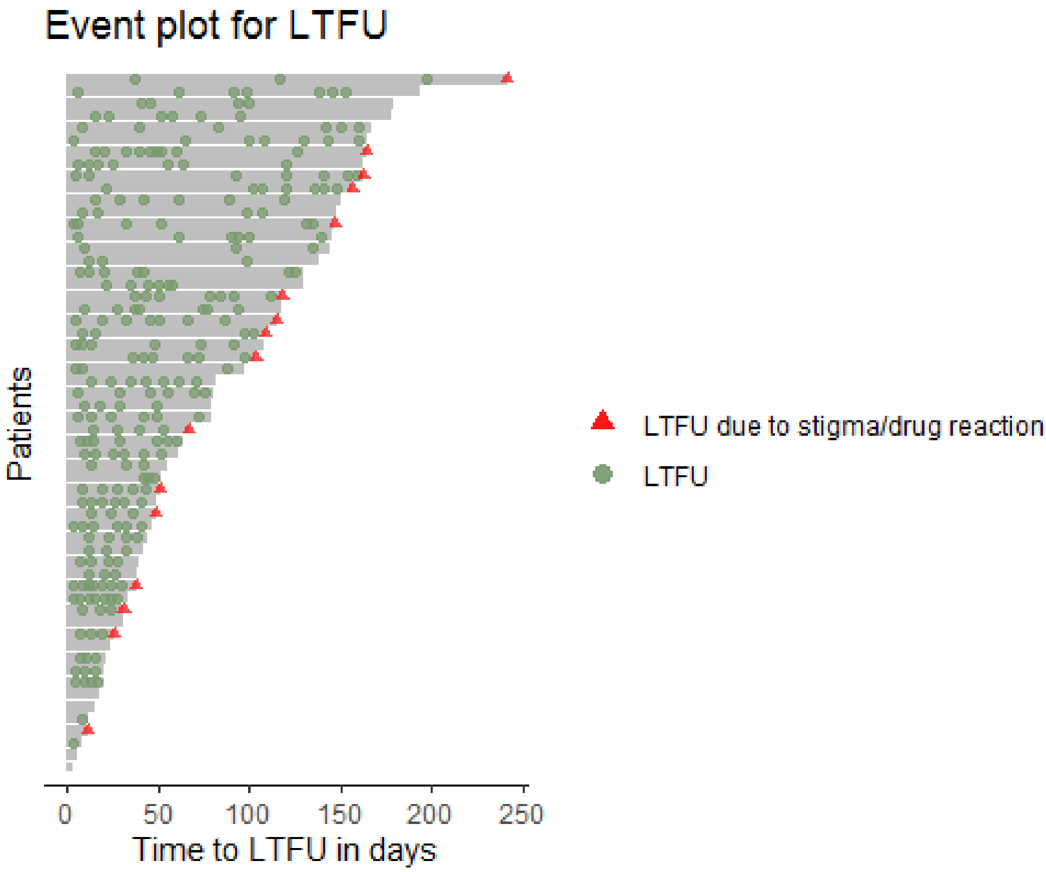

Figure A.1: Event Plot for LTFU recurrence and LTFU due to stigma (Terminal Event).

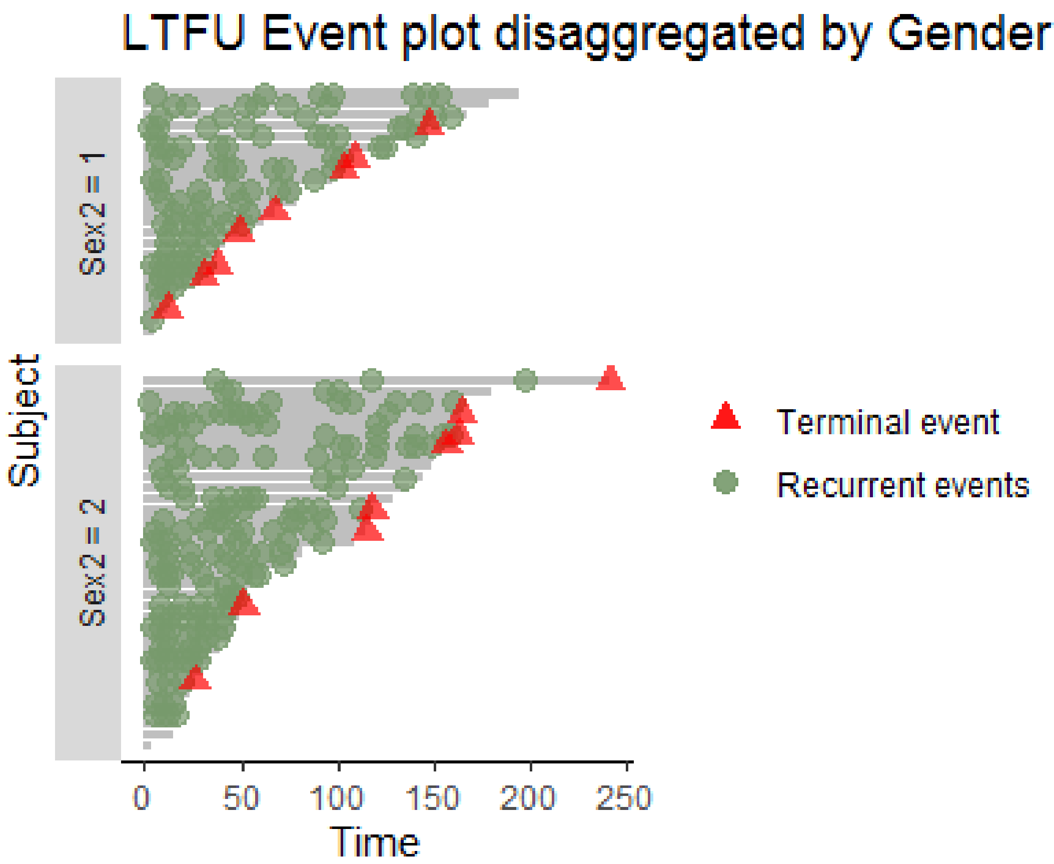

Figure A.2: Event plot by Gender

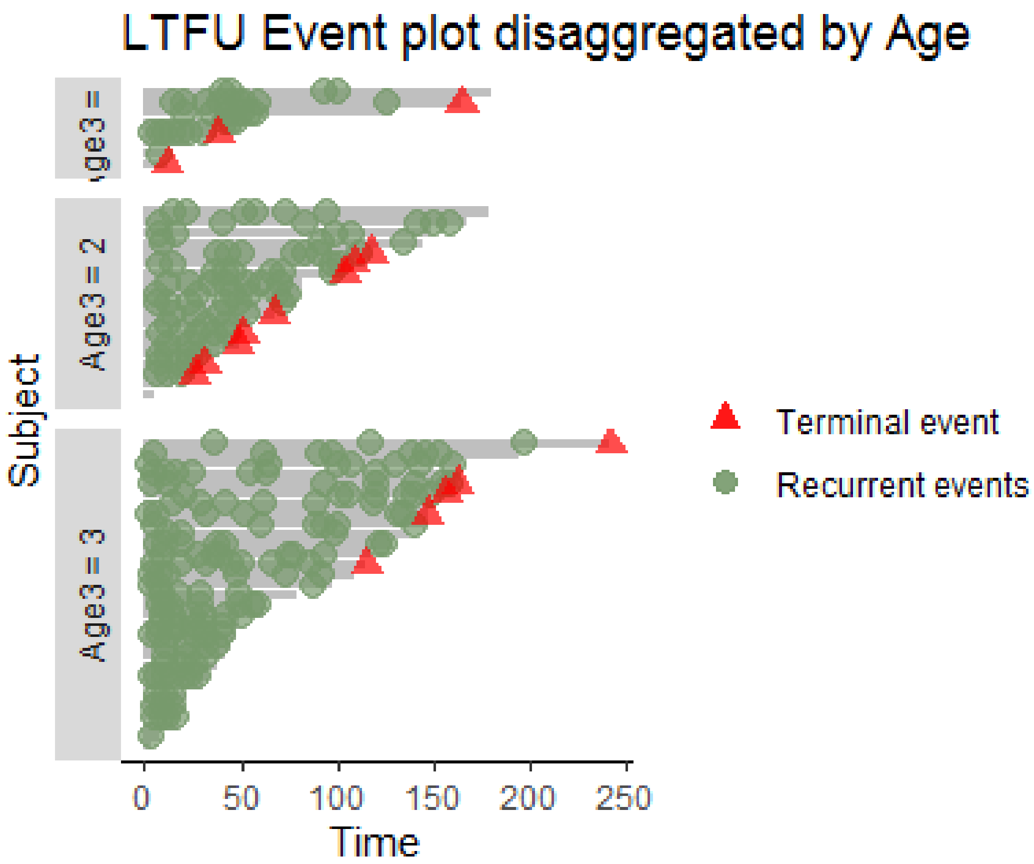

Figure A.3: Event plot by Age

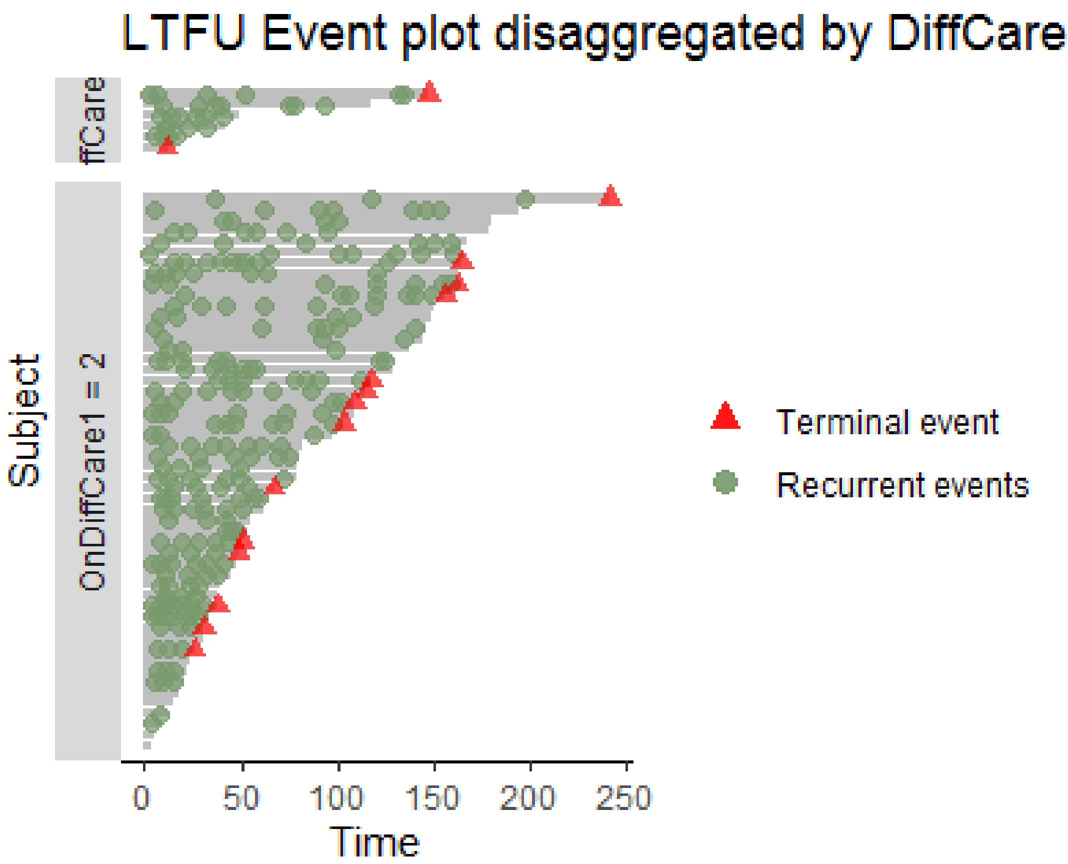

Figure A.4: Event plot by Differentiated Care

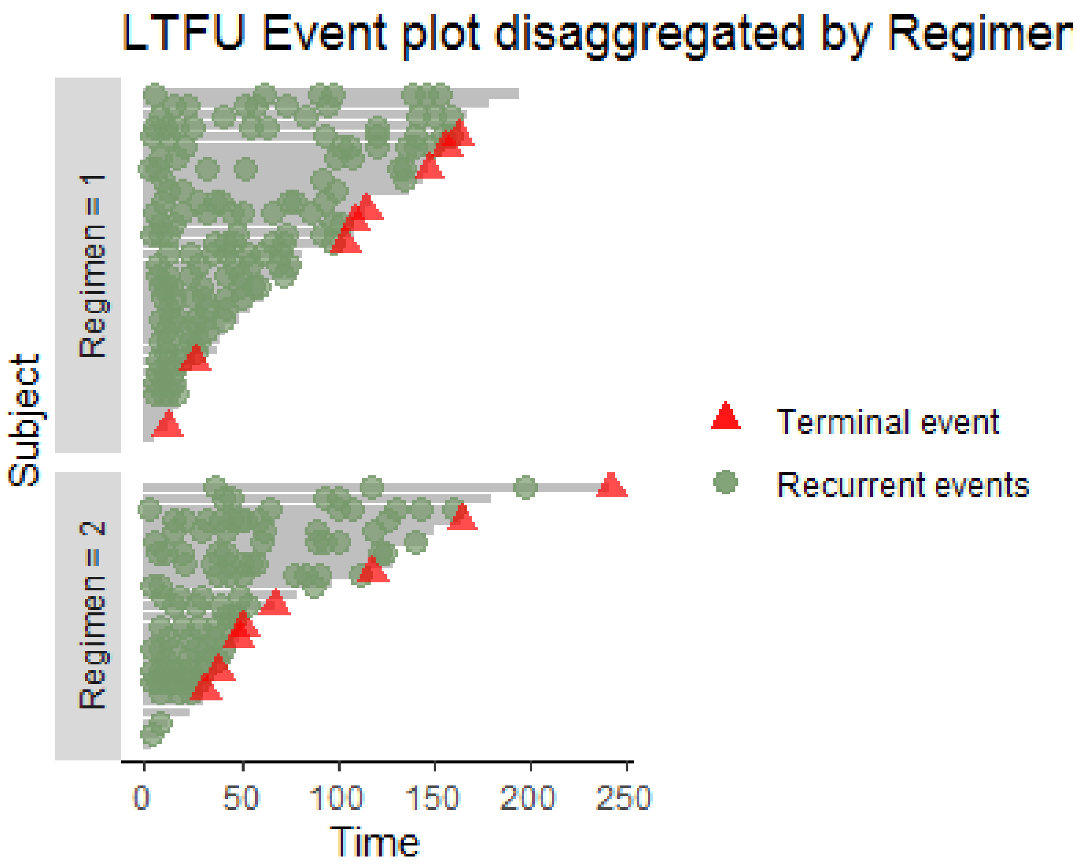

Figure A.5: Event plot by Regimen

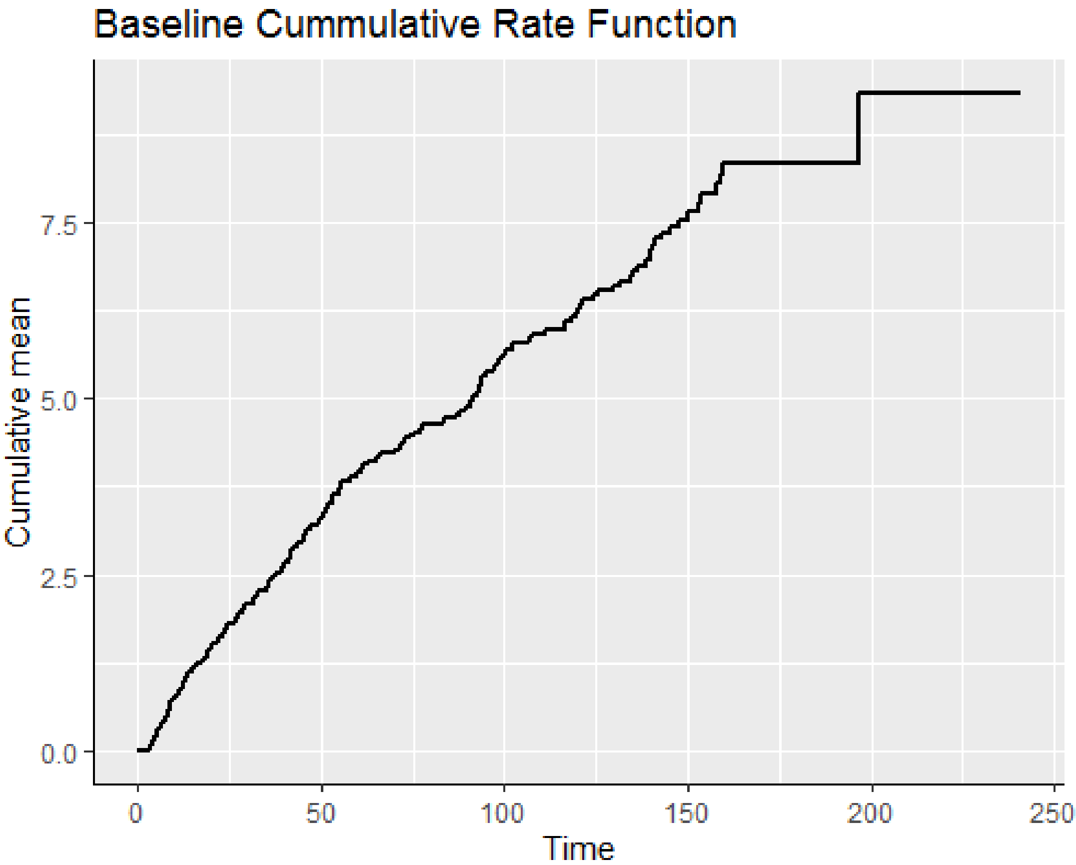

Figure A.6: Cumulative number of incidences over time

A.3 R Code

```
# Recurrent Events Analysis

# Ejoku Jonathan. 111385

# ==========================================================

# INITIALIZE——————————————

setwd("D:/Google Drive/Masters/MSc/Research/Final Datasets")

rm(list=ls()) #clear workspace

DataRecurrent_Event ¡- read.csv("DataRecurrent_Event2.csv")

set.seed(53240)

# PACKAGES——————————————

library(Hmisc) # for descriptive stats, preloads survival package needed for Cox model

library(reReg) # for recurrent events

attach(DataRecurrent_Event)

# CONVERTING VARIABLE TYPES——————————

DataRecurrent_Event$Sex3 < − as.factor(Sex2)# Gender

DataRecurrent_Event$Age2 < − as.factor(Age2) # Age

DataRecurrent_Event$OnDiffCare1 < − as.factor(OnDiffCare) # On Diff Care

DataRecurrent_Event$Regimen < − as.factor(Endpoint_Regimen) # Regimen Line

str(DataRecurrent_Event)

attach(DataRecurrent_Event)

str(DataRecurrent_Event)

summary(DataRecurrent_Event)

names(DataRecurrent_Event)
```

```
describe(DataRecurrent_Event)

sd(Age)

# SOME GRAPHS————————————————

# Plots of recurrent events, and corresponding LTFU incidence as well as informative censoring instances

reObj < − Recur(Start %to% Stop, PatientID, Event2, status2)

plot(reObj, cex = 1.5, xlab = ”Time to LTFU in days”, ylab = ”Patients”,

main = ”Event plot for LTFU”,

terminal.name = ”LTFU due to stigma”,

recurrent.name = ”LTFU”)

plotEvents(reObj   Sex2, main = ”LTFU Event plot disaggregated by Gender”)

plotEvents(reObj   Age3, main = ”LTFU Event plot disaggregated by Age”)

plotEvents(reObj   OnDiffCare1, main = ”LTFU Event plot disaggregated by DiffCare”)

plotEvents(reObj   Regimen, main = ”LTFU Event plot disaggregated by Regimen”)

# Cummulative Sample Mean Plots

library(gridExtra)

plotCSM(reObj, data = DataRecurrent_Event, main = ”Baseline Cummulative Rate Function”)

# To assess if there is any visual trend

# Taking Sex + Age + Diff Care + Regimen Line + Viral Load

——————————————-

# A. Cox Model

surv_object < − Surv(time = Stop, event = Event2)
```

fit_Cox2 < − coxph(surv_object Sex2 + Age3 + Regimen + OnDiffCare1, subset=(Stratum_PWP==1), data = DataRecurrent_Event)

cox.zph(fit_Cox2) # Proportionality assumption not violated. cox.zph global value = .316 ¿ 0.05

summary(fit_Cox2)

B. PWP TT

fit_PWP2 < − reReg(reObj Sex2 + Age3 + Regimen + OnDiffCare1+ strata(Stratum_PWP),

data = DataRecurrent_Event, method = ”cox.HW”, se = ”resampling”, B=100)

summary(fit_PWP2)

C. AG

fit_AG2 < − reReg(reObj Sex2 + Age3 + Regimen + OnDiffCare1, data = DataRecurrent_Event, method = ”cox.HW”, se = ”resampling”, B=100)

summary(fit_AG2)