DISCRETE TIME-TO-EVENT CONSTRUCTION FOR MULTIPLE RECURRENT STATE TRANSITIONS

UNIVERSITY OF
KWAZULU - NATAL

INYUVESI
YAKWAZULU-NATALI

Jesca Mercy Batidzirai

November, 2023

Discrete Time-to-Event Construction for Multiple Recurrent state Transitions

by

Jesca Mercy Batidzirai

A thesis submitted to the
University of KwaZulu-Natal
in fulfilment of the requirements for the degree
of
DOCTOR OF PHILOSOPHY
in
STATISTICS

Thesis Supervisors:
Professor Samuel O. M. Manda & Professor Henry G. Mwambi

UNIVERSITY OF KWAZULU-NATAL
SCHOOL OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
PIETERMARITZBURG CAMPUS, SOUTH AFRICA

Declaration - Plagiarism

I, Jesca Mercy Batidzirai, declare that

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowlegded as being sourced from other persons.
4. This thesis 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, or
   (b) where their exact words have been used, then their writing has been placed in italics and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the reference sections.

Jesca Mercy Batidzirai (Student)	Date
Professor Samuel O. M. Manda (Supervisor)	Date
Professor Henry G. Mwambi (Co-supervisor)	Date

Disclaimer

This document describes work undertaken as a PhD programme of study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institution.

ABSTRACT

Recent developments in multi-state models have considered discrete time rather than continuous time in the modeling of transition intensities, whose major drawback lies in the possibility of resulting in biased parameter estimates that arise from issues of handling ties. Discrete-time models have included univariate multilevel models to account for possible dependence among specific pairwise recurrent transitions within the same subject. However, in most cases, there would be several specific pairwise transitions of interest. In such cases, there is a need to model the transitions with the aim of identifying those transitions that are correlated. This provides insight into how the transitions are related to each other. In order to investigate the interdependencies between transitions, the unique contribution of this thesis is to propose a multivariate discrete-time multi-state model with multiple state transitions. In this model, each specific recurrent transition is associated with a random effect to capture possible dependence in the transitions of the same type or different types. The random effects themselves were then modeled by a multivariate normal distribution and model parameters were estimated using maximum likelihood methods with Gaussian quadratures numerical integration.

A simulation study was done to evaluate the performance of the proposed model. The model yielded satisfactory results for most fixed effects and random effects estimates. This is noticed by near-zero biases and mean square errors of the average estimates as well as high 95% coverage probabilities of the 95% confidence intervals from 1000 replications..

The proposed methodology was applied to marriage formation and dissolution data from KwaZulu-Natal province, South Africa. Five transitions were considered, namely: *Never Married* to *Married*, *Married* to *Separated*, *Married* to *Widowed*, *Separated* to *Married* and *Widowed* to *Married*. The presence of very small unobserved subject-to-subject heterogeneity for each transition and a weak positive correlation between transitions were produced. Statistically, the model produced smaller standard errors compared to those from univariate models, hence it is more precise on estimates. The multivariate modeling of discrete time-to-event models provides a better understanding of the evolution of all transitions simultaneously, thus in addition to covariate effects, giving an assessment of how one transition is associated with the other.

Empirical results confirmed well known important socio-demographic predictors of entering and exiting a marriage. Age at sexual debut played a positive critical role in most of the transitions. More educated subjects were associated with a lower likelihood of entering a first marriage, experiencing a marital dissolution as well as remarrying after widowhood. Subjects who had a sexual debut at younger ages were more likely to experience a marital dissolution than those who started late. Age at first marriage had a negative association with marital dissolution. We may, therefore, postulate that existing programs that encourage delay in onset of sexual activity for HIV risk reduction for example, may also have a positive impact on lowering rates of marital dissolution, thus ultimately improving psychological and physical health.

DEDICATION

This thesis is dedicated to God, my parents: Mr (late) and Mrs Sibanda, my daughter: Sheunesu, my twin sons: Xavier and Nicholas, and the Sibanda-Batidzirai family.

ACKNOWLEDGMENTS

I would like to thank God the almighty Father who made this work possible for the greater glory of His name, through Jesus Christ our Lord by the power of the Holy Spirit.

I would like to express my deepest gratitude and appreciation to my supervisors, Professor Samuel Manda and Professor Henry Mwambi, for such tremendous and tireless guidance throughout my study. They have always being available and supportive in my PhD studies.

My beloved family always gave me unconditional love, support and encouragement, I thank and appreciate them. I am grateful to my friends, Patience, Memory, Lucy, Shingi and Tinashe who have been supportive throughout my studies, even in the toughest times. I thank my husband, Andrew, for motivating me and bearing with me during my busy times.

I greatly appreciate the assistance I received in my PhD work from my colleagues, Halima, Forbes, Justine, Danielle, Farai, Retius, Knowledge, Mohanad, Innocent, Ashenafi, Alex, Faustin, Show, Kaombe, Alphonce, Christel, Bev, Prof Anita Heeren, the Pietermaritzburg community of Biostatisticians and IBS-SUSAN students members.

I gratefully acknowledge the support we received from the University of KwaZulu-Natal both in financial support and provision of research facilities. I acknowledge the South African Medical Research Council (SAMRC) for hosting me during my research visits.

I would also like to acknowledge Professor Frank Tanser and the Africa Health Research Institute (AHRI) for giving me permission to use the data for this research. Funding for the ARHI's Demographic Surveillance Information System and Population-based HIV Survey was received from the Wellcome Trust. Professor Frank Tanser received support from a UK Academy of Medical Sciences Newton Advanced Fellowship (NA150161).

Last but not least, I gratefully acknowledge those who funded me during my studies:

This work received some funding support for teaching relieve and conferences from the South African Department of Higher Education and Technology (DHET) through the University Capacity Development Programme (UCDP) and Teaching Development Grants [APP-TDG-096, APP-TDG-226, UCDP-455, UCDP-626 and UCDP-784].

The University Capacity Development Grant (UCDP) Staff Credentialing Funding by the University of KwaZulu-Natal gave me funding for teaching relief to work on completing my studies.

This work was partially supported through the DELTAS Africa Initiative [SSACAB], [grant 107754/Z/15/Z-DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme]

RESEARCH OUTPUTS

1. Discrete Survival Time Constructions for Studying Marital Formation and Dissolution in Rural South Africa. *Frontiers in Psychology 2020*.

   DOI:*10.3389/fpsyg.2020.00154*

CONFERENCE PRESENTATIONS FROM THIS THESIS

1. A Multivariate Discrete Time-to-Event Model for Multiple Recurring Events *Southern African Mathematical Association (SAMSA 2022:* **Mozambique**)
2. Multi-state Analysis of Discrete-Time Models: A South African Case Study of Family Formation & Dissolution: *International Biometric Conference- Sub-Saharan Africa (IBS-SUSAN 2019:* **CAPE TOWN**)
3. Discrete Time To Event Based Construction for Modeling Family Formation and Dissolution Studies in Rural South Africa: *Southern African Mathematical Association (SAMSA 2018:* **BOTSWANA**)
4. Multilevel Discrete Time-To-Event Modeling of Family Formation and Dissolution Data in rural South Africa: *International Biometric Conference (IBC 2018:* **BARCELONA**)
5. Multi-state transition modeling of family formation histories in rural South Africa: *International Biometric Conference- Sub-Saharan Africa (IBS-SUSAN 2017*: **MALAWI**)
6. Multi-state transition modeling of family formation and/or dissolution histories in South Africa: A comparison of Direct Likelihood method and Multiple Imputation of missing data: *Southern African Mathematical Association (SAMSA 2017:* **TANZANIA**)
7. Analysis of Family Formation and dissolution in rural South Africa using Multi-State Transition Models: *South African Statistical Association (SASA 2017:* **BLOEMFONTEIN**)
8. *South African Statistical Association (SASA 2016:* **CAPE TOWN**)
9. *South African Statistical Association (SASA 2015:* **PRETORIA**)

AWARDS FROM THIS THESIS

- 2^{nd} Prize winner for oral presentation during the 2017 SSACAB Meeting, Windhoek. Namibia. **Title of presentation:** Multilevel modeling of family formation and dissolution in rural South Africa using multi-state transition models with competing risks

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

CI	Confidence Interval
OR	Odds Ratio
AHRI	Africa Health Research Institute
LOCF	Last Observation Carried Forward
AIC	Akakie Information Criterion
BIC	Bayesian Information Criterion
GLM	Generalized Linear Models

CHAPTER 1

INTRODUCTION

1.1 Background

Multi-state models have gained popularity and have been widely used in modeling movements of subjects between the mutually exclusive stages (states) of a process such as the natural progression of a disease (Allison, 1982; Therneau et al., 2018). Willekens (2014) defines a multi-state process as a stochastic process with a finite state space where subjects transition from one state to the next (Therneau et al., 2018). As the process evolves, a history of observations is generated over time (Andersen & Keiding, 2002). In a multi-state process, certain states may exhibit distinct behaviors: some may only be visited once and never revisited (transient), others may subsequently be revisited (recurrent), and there are those that, once entered, cannot be exited (absorbing). Life histories analysis using multi-state models has been extensively used in biostatistics, demography and economics (Helbert, 2015; Willekens, 2014; Steele et al., 2004).

Traditional approaches to modeling the transition intensities have often used models such as the Cox proportional hazard (Jackson et al., 2011) or accelerated failure time (Kridahl & Silverstein, 2017). These have been intensively done while the time to event is considered to be continuous. However, in some cases, observations are done in a truly discrete manner. In event history analysis, time- to-event models like survival analysis are quite common (Cox, 1972b; Ali et al., 2018; Allison, 1982; Tutz et al., 2016). These models have been developed and widely used where a process consists of only two states (one is usually absorbing). Interest is in modeling time until an event of interest occurs. The time to an event can be modeled using cross-sectional or longitudinal data. Longitudinal data has an added advantage over cross-sectional data in that it enables the researcher to detect changes or patterns in the outcome over time, both at the group or individual level. However, longitudinal studies may give rise to some complications in the analysis, such as a possible dependence in the data for the same subject. In situations where the two states are recurrent (such as the *no job* $\leftrightarrow$ *job* states used by Fahrmeir & Knorr-Held (1997)), the researcher is faced with yet more complex and demanding statistical problems in handling the data, because the within- and between- subject correlation need to be quantified. Hence, mixed effects survival models are applied to account for inter-

subject variability (Tutz et al., 2016; Austin, 2017; Crowther, 2019).

In cases where there exist more than two possible end states, competing risks models have been developed to model hazards of failure due to one of the causes that compete for failure (Fine & Gray, 1999; Beyersmann et al., 2011; Tutz et al., 2016; Berger et al., 2020; Schmid & Berger, 2020; Han, 1987; Han & Hausman, 1990; Allignol et al., 2011; Putter et al., 2011, 2007). A competing risks model may be regarded as a special case of a multi-state model (Andersen et al., 2002). For example, in the employment data, unemployment can end with either getting a new job or a recall to the previous job (Katz, 1986). Diamond & Hausman (1984) gave another example of a spell of unemployment that can end with either a new job or withdrawal from the labor force. Steele et al. (2004) used data on history of contraceptive use for women where a woman on a contraceptive can either switch contraceptive method or stop contraceptives altogether. Willekens (2014) modeled data on marital dissolution where he considered divorce or death of a spouse as the two competing causes of marital dissolution.

Competing risks models only consider transitions out of a single origin state. However, extensions have been made to processes where various origin states are considered and states are possibly recurrent and classes communicate (i.e. $i \longleftrightarrow j$) (Therneau et al., 2018; Putter et al., 2011). In such cases, the researcher may choose an option of considering separate two-way transitions, for all permissible transitions in the process, using a series of basic survival analysis models. However, this usually assumes independence of transitions and that each transition is occurring for the first time in an individual. Previous transitions of the same kind are ignored and the resulting estimates may be biased. In addition, correlation between the different types of transitions will be ignored. In continuous-time analysis, various multi-state models were developed, although they do not account for correlation. One example is the multi-state Markov model proposed by Jackson et al. (2011), which assumes that transition into the future state only depends on the current state with no influence from the previously occupied states. Recently, Asanjarani et al. (2021) also modeled a semi-Markov model to compare the sojourn times and transition intensities approaches in a continuous-time scale.

Various cases of multi-state models also include sequential or hierarchical transition models, which we will not cover in this work. These are common in terminal disease modeling where patients move through states progressively. An example is found in HIV studies where the states are considered to be the various CD4 count stages (Reddy et al., 2011).

What is common in all the highlighted multi-state models (Markov, 1971; Therneau et al., 2018; Putter et al., 2011; Meira-Machado et al., 2009; Craig & Sendi, 2002; Jackson et al., 2011; de Wreede et al., 2011; Putter et al., 2011; Spedicato et al., 2016) is their assumption that time is measured on a continuous scale. This, however, is not always the case in real life. The next section describes an alternative approach to multi-state modeling: discrete-time models.

1.2 Discrete Time-to-Event models

Most statistical methods of analyzing event history data are based on a continuous time variable (Cox & Oakes, 1984; Allison, 2014; Andersen et al., 2012; Cox, 1972a) such that the exact time of event occurrence is known. Allison (2014, 1982) describes time as always observed in discrete units, however small, but where the time intervals are very small relative to the rate of occurrence of the event, it is acceptable to ignore this discreteness. When time units are very large, observations are done at intervals of time. In such cases, continuous time-variates become tricky as they result in ties, since time intervals can be large enough that an individual may experience more than one transitions within the same time interval. Some discrete-time models of analyzing such data are, therefore recommended and have been reviewed and widely used in demography, social science and biostatistics (Allison, 1982; Andersen & Keiding, 2002; Fahrmeir & Knorr-Held, 1997; Bijwaard, 2014; Hox et al., 2010; Steele, 2011).

Discrete time data essentially arises in two ways: derived or intrinsic discrete. Derived discrete data comes as a result of continuously recorded underlying data which is discretized through grouping or rounding into a discrete time (Wolkewitz et al., 2008). Because of some assumptions made about the data, it may be acceptable to discretize it, resulting in discrete event times due to grouping effects (Tutz et al., 2016; Berger et al., 2020; Schmid & Berger, 2020). Intrinsic discrete data is obtained when observations are made on truly discrete time points such as yearly, monthly, at hospital visits or menstrual cycles (Scheike & Keiding, 2006). In such cases, analyzing the data using continuous-time models may lead to biased estimates (Allison, 1982). Many longitudinal studies have their data collected in this way, so using continuous-time methods would not yield satisfactory results. Hence, focus in this study will be on the assumption that time is measured on a discrete scale. In continuous-time analysis, the state at any given time, however small, is observed and known. In the discrete-time analysis, observations are only made at

pre-determined time points. If a transition is to occur, the exact time will not be known; only the interval is known. Discrete-time methods may be used to approximate results of a continuous-time survival analysis (Jenkins, 1995; Vermunt, 1996) and are conceptually and computationally straightforward (Dean et al., 2014). Singer & Willett (1993) and Mills (2011a) highlighted many advantages of using a discrete-time approach including easy implementation of time varying covariates as well as straight forward testing of the proportional odds (hazards) assumption, hence its flexibility when modeling time-varying covariates and their interactions with time (Allison, 1982; Clark et al., 2013). Although the inclusion of time-varying covariates is apparent when modeling time as discrete, it is important to note that it is possible for a covariate to be time-varying, yet its effect over time does not significantly vary (Singer & Willett, 1993) and vice-versa. Furthermore, discrete time methods allow for interim analysis. Lastly, it produces reasonable predicted probabilities that can be interpreted independently and therefore not affected by an inability to estimate a reliable baseline hazard (often the case with Cox proportional hazards models).

In discrete-time multi-state modeling, there are various constructions that may be considered. Firstly, multinomial models may be used to model state occupation of a subject. Secondly, for modeling the hazards of transition from one state to the other, pairwise binary transition models may be used. Lastly, competing risks models are used to model the hazard of exiting a state.

All the models are executed while adjusting for covariates, some of which may be time-dependent. Furthermore, as repeated occurrences of the same event may be observed in a subject, the event times can be correlated. Multilevel models were developed to account for these correlations and have been widely used in social sciences (Steele, 2011; Raudenbush & Bryk, 1986, 2002; Hox et al., 2010), medical statistics (Ali et al., 2018) and other areas of research (Steele et al., 2004). Moreover, there may be different transitions (events) of interest, each of which may reoccur. The different transitions could be correlated too. Thus, we have intra- and inter- dependencies in the event histories in the same subject. Ignoring this correlation in the analysis might lead to underestimation of standard errors (Rice & Leyland, 1996). A robust model which takes into consideration all the above-mentioned structure of data is required.

1.3 Aims and Objectives

This study aimed at deriving a multivariate discrete-time survival model with multiple state transitions where each transition has its own separate random effect. The joint model of these transition-specific random effects is given a multivariate normal distribution and its variance-covariance matrix gives estimates of dependence between and within transitions, which is an added advantage over univariate models that only produce estimates of covariates effects. The highly complex multi-state structure arises because several types of transition occur repeatedly over time, with interdependences between the different transitions. In this thesis, model parameters are estimated using a maximum likelihood approach with Gaussian quadratures numerical integration. The proposed model is then applied to a routine data set from an ongoing longitudinal study based on life course history on marriage formation and dissolution events in rural KwaZulu-Natal of South Africa.

This study will contribute to the body of statistical knowledge on multi-state modeling of discrete-time longitudinal data that is hierarchical and retrospective through this advanced statistical method. It offers a better understanding of the processes and dynamics of family formations and dissolutions and will demonstrate how multilevel discrete-time event history models can be used to jointly model family formation and dissolution data to efficiently and robustly estimate various effects. Statistical methods contained in this study will afford researchers tools with which to analyze such types of data which is gaining popularity in statistics.

1.4 Structure of the Thesis

Chapter 2 gives a thorough review of a univariate discrete-time survival model as the simplest multi-state model with only two states and its extension to multilevel data. Chapter 3 is dedicated to the application of the discrete-time survival model to routine data set to model age at first marriage. Other existing alternatives of dealing with univariate multi-state data are presented in Chapter 4 via three models namely: multinomial logit model for state occupations, a series of discrete survival models for the one-step transitions between two states and discrete-time competing risks models for marital dissolution. Preliminary work from this chapter forms part of the work that was published in the Journal of Frontiers in Psychology (abstract attached in Appendix A).

In Chapter 5, the theory on joint modeling of discrete-time survival models is ex-

tended to multi-state processes. Hence, a multivariate discrete-time multi-state model is proposed for possibly recurrent transitions while accounting for correlation within- and between- transitions. To demonstrate this model, an example data set on transitions between marital states is used from rural South Africa. Work from this chapter forms a manuscript which will be submitted to the Journal of Applied Statistics (attached in Appendix B).

A multivariate competing risks model in discrete time is further constructed in Chapter 6 for joint multi-state modeling through exiting each state. Chapter 7 concludes the work and gives recommendations and suggests areas of future research.

CHAPTER 2

DISCRETE-TIME-TO-EVENT MODELS FOR TRANSITION INTENSITIES

2.1 Introduction

A survival model is considered as a special case of multi-state models which has been widely used to consider time to a single event such as time until a marriage (for those who are single), time to infection with (or relapse of) a disease, or time to death (Allison, 1982; Cleves, 2008; Cox, 1972b). It is one of the simplest methods used to analyze a 2-state process (Allison, 1982) where the state space is assumed to consists of only two states and there is only one possible transition into an absorbing destination state. In cases where the two states are recurrent, special modification is done to the model to account for the correlation within the subjects. In this chapter, we review the existing survival methods for discrete event times (for transition into a single absorbing state) together with their extension to account for the multilevel nature of data (for recurrent events) by including a subject-specific random effect which measures the correlation within observations of subjects. It differs from Manda & Meyer (2005), for example, who used cross sectional data for analyzing first marriages among Malawian women. Longitudinal data in the current study, has an added advantage of capturing outcome effects over time. Section 2.2 reviews the standard univariate survival model and its likelihood function and discusses the approaches of modeling the baseline hazard. Section 2.3 introduces a frailty model to account for the recurring nature of the transitions and reviews the numerical approximation methods for the integrals. Section 2.4 deals with a multilevel survival model with more than 2-levels.

2.2 Standard Survival Model with Discrete Transition Times

For discrete-time survival analysis, the time scale is subdivided into τ unit intervals, denoted by $I_t = [a_{t-1}, a_t)$, such that $0 = a_0 < a_1 < ... < a_\tau = \infty$ which do not

necessarily have to be of equal length. The observed discrete time is $T = t$ where $t \in 1, 2, ..., \tau$ denotes the interval I_t. If censoring exists, it is assumed to occur at the end of the intervals (Narendranathan & Stewart, 1991, 1993). Suppose the discrete survival time for subject $k(k = 1, ..., K)$ is represented by T_k. Then T_k is a random variable which takes on τ discrete values such that $T_k = \{1, ..., \tau\}$, where $\tau \in \mathbb{N}$. The discrete-time hazard function which is defined as the conditional probability of failure given that the subject has survived up to the beginning of the time (age) interval, t, is given by

$$\lambda\left(t|\boldsymbol{Z}_t\right) = P\left(T_k = t|T_k \geq t; \boldsymbol{Z}_t\right), \tag{2.1}$$

where $\boldsymbol{Z}_t$ is a $p-$ dimensional vector of known covariates, possibly time-varying. Consequently, the survivor function is the probability that the subject has not yet experienced an event by time (age) interval t and is denoted by

$$S\left(t|\boldsymbol{Z}_t\right) = P\left(T_k > t|\boldsymbol{Z}_t\right) = \prod_{b=1}^{t}\left(1 - \lambda\left(b|\boldsymbol{Z}_t\right)\right). \tag{2.2}$$

Age may be used as the time scale,rather than calendar time, since it can be used to represent stages of life (Andersen et al., 2002). The discrete hazard function in equation (2.1) may be linked to a time-varying predictor η_{kt}, according to (Fahrmeir & Knorr-Held, 1997; Lee et al., 2018), by

$$\lambda(t|\boldsymbol{Z}_t) = h(\eta_{kt}), \tag{2.3}$$

through a suitable link function, where $h(.)$ is a fixed response function and η_{kt} may then be modeled as a function of t and $\boldsymbol{Z}_t$ either parametrically or non-parametrically. If we consider the class of parametric regression models the predictor may commonly be modeled with a linear form following generalized linear models (GLM) (Molenberghs & Verbeke, 2005; Fitzmaurice et al., 2008) as:

$$\eta_{kt} = \beta_{0t} + \boldsymbol{Z}_t'\boldsymbol{\beta}, \tag{2.4}$$

where $\boldsymbol{\beta}_{0t}$ is a real-valued baseline hazard which is allowed to be time-dependent and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients independent of t.

Depending on the nature of the data at hand in the case of a 2-state model $(i \rightarrow j)$, the most common choices of link functions are the logit link (Singer & Willett, 1993; Muenz & Rubinstein, 1985; Manda & Meyer, 2005; Tutz, 2011; Tutz et al., 2016; Dean

et al., 2014), discrete proportional hazards (also known as the complementary log-log (clog-log) link) and the probit link (Mills, 2011a). A clog-log specification (Scheike & Jensen, 1997; Jenkins, 1997; Mills, 2011b; Kridahl & Silverstein, 2017), which is commonly used when data is derived-discrete, would yield a hazard of the form

$$\lambda\left(t|\boldsymbol{Z}_t\right) = 1 - \exp\left[-\exp\left(\beta_{0t} + \boldsymbol{Z}_t'\boldsymbol{\beta}\right)\right]. \tag{2.5}$$

It is an analogue of the Cox proportional hazard model in the continuous time methods. A probit hazard would be of the form

$$\lambda\left(t|\boldsymbol{Z}_t\right) = \boldsymbol{\Phi}\left(\beta_{0t} + \boldsymbol{Z}_t'\boldsymbol{\beta}\right), \tag{2.6}$$

where $\boldsymbol{\Phi}\left(\beta_{0t} + \boldsymbol{Z}_t'\boldsymbol{\beta}\right)$ is the standard normal cumulative distribution function as used by Kandala & Ghilagaber (2006). It is often used if the underlying distribution of the process is known to be approximately normal. A logit link gives a hazard of the form

$$\lambda\left(t|\boldsymbol{Z}_t\right) = \frac{\exp\left(\beta_{0t} + \boldsymbol{Z}_t'\boldsymbol{\beta}\right)}{1 + \exp\left(\beta_{0t} + \boldsymbol{Z}_t'\boldsymbol{\beta}\right)}, \tag{2.7}$$

where it defines the proportional continuation ratio model. It is preferred because the exponents of its parameter estimates give an odds ratio interpretation which has practical importance in most cases (Manda, 1998). For that reason, it will be used in this study although the choice of specification of the link function has little impact on result (Steele et al., 2009) and is based on convenience of interpretation (Mills, 2011a).

2.2.1 The Likelihood

For subject k, let T_k and C_k be the survival time and censoring time, respectively. Assuming random censoring, we observe time as $t_k := \min\left(T_k, C_k\right)$. An indicator variable is also defined as

$$\delta_k = \begin{cases} 0 & \text{if } T_k > C_k \\ 1 & \text{if } T_k \leq C_k. \end{cases}$$

The contribution to the likelihood function for subject k is therefore given by

$$L_k = \lambda\left(t|\boldsymbol{Z}_t\right)^{\delta_k}\left(1 - \lambda\left(t|\boldsymbol{Z}_t\right)\right)^{1-\delta_k}\prod_{b=1}^{t_k-1}\left(1 - \lambda\left(b|\boldsymbol{Z}_t\right)\right), \tag{2.8}$$

(see Tutz et al., 2016, pages 52 and 53 for standard inferential text in survival analysis for a description of likelihood function).. Further define a binary variable, $\varepsilon_{kt} \in [0, 1]$ as

$$\varepsilon_{kt} = \begin{cases} 0 & \text{if subject } k \text{ is still in the origin state} \\ 1 & \text{if subject } k \text{ has made a transition into the destination state,} \end{cases}$$

which is only defined as long as subject k is at risk, i.e. $t \leq t_k$. It is clear that for subject k who is censored $(\delta_k = 0)$, the binary sequence will then be

$$(\varepsilon_{k1}, ..., \varepsilon_{kt_k}) = (0, 0, ..., 0),$$

while for subject k who experiences the event $(\delta_k = 1)$, the binary sequence will be

$$(\varepsilon_{k1}, \ldots, \varepsilon_{kt_k}) = (0, 0, ..., 1).$$

The likelihood 2.8 is equivalent to that for a binary response model as used in GLM's so that

$$L_k = \prod_{b=1}^{t_k} \lambda\left(b|\boldsymbol{Z}_t\right)^{\varepsilon_{kb}} \left(1 - \lambda\left(b|\boldsymbol{Z}_t\right)\right)^{1-\varepsilon_{kb}}. \tag{2.9}$$

The complete log-likelihood over all subjects is then given by

$$\ell \propto \sum_{k=1}^{K} \sum_{b=1}^{t_k} \varepsilon_{kb} \log \lambda\left(b|\boldsymbol{Z}_t\right) + (1 - \varepsilon_{kb}) \log\left(1 - \lambda\left(b|\boldsymbol{Z}_t\right)\right), \tag{2.10}$$

which is maximized with respect to $\boldsymbol{\beta}$ to obtain the maximum likelihood estimates (MLE's) of $\boldsymbol{\beta}$. This is easily done using readily available standard software that is designed for binary regression models (Allison, 1982; Mills, 2011b; Singer & Willett, 1993; Jenkins, 1995).

2.2.2 Modeling the Functional form of the Discrete Baseline Hazard

The linear predictor in 2.4 may be rewritten to include a design vector, $\mathbf{Y}_t$, as

$$\eta_{kt} = \mathbf{Y}_t' \boldsymbol{\beta}_{0t} + \boldsymbol{Z}_t' \boldsymbol{\beta}, \tag{2.11}$$

where $\mathbf{Y}_t$ denotes the duration dependence. Several approaches in the literature have been proposed regarding the flexibility in modeling $\boldsymbol{\beta}_{0t}$, which may be specified either parametrically (such as piece-wise constant (Steele et al., 2005) or a polynomial function of time (age)) or non-parametrically (such as splines (Berger et al.,

2018)). In cases where the number of event times is large relative to the sample size, the baseline coefficients may be represented by a smooth function of an unspecified form, such as P-splines or smoothing splines (Berger et al., 2020; Tutz et al., 2016). However, the choice of the shape of the hazard function is usually up to the researcher (Jenkins, 1995, 2005). We expound upon some of these options for modeling the baseline hazard.

Constant Hazard This assumes that the hazard is constant throughout the observation period. It requires only one parameter to be estimated, but produces a very high deviance, hence not preferable.

Piece-wise Constant Hazard This is also referred to as the general specification. Here, dummy variables are created which correspond to each (possibly grouped) interval of time, say yearly or monthly. The hazard rate in each interval is constant, but differs between these intervals. The $\mathbf{Y}_t$ in equation 2.11 will be a vector of dummy variables for intervals where Y_t is a binary variable which is equal to 1 for the corresponding interval and zero otherwise (Jenkins, 2005). However, when a piece-wise constant specification is used, the intercept term within the vector will not be used, or else it would be co-linear. If one insists to include it, then one of the dummy variables will have to be dropped. In the analysis, the time variable is now treated as a categorical variable and assuming that the hazard in each interval is different. One advantage of the piece-wise constant hazard lies in its ease to interpret the model estimates, whose shape generally depicts that produced by life table estimates.

On the downside, this approach lacks parsimony as it requires a large number of parameters where observation periods are long. In addition, due to sampling variation, the fitted hazard functions are often erratic due to sampling variation. In cases where a large number of discrete time periods are observed, a large number of dummy variables will be required. These might result in the hazards nearing zero in some time periods, which may yield difficulties in generating maximum likelihood estimates. It is also unable to discern differences in situations where some time periods have small risk sets. For these reasons and others, more parsimonious specifications of the baseline hazard have been suggested in literature, which equally produce goodness of fit of models.

Polynomial Function In a polynomial specification of order r, the baseline log odds is given by $\mathbf{Y}_t'\boldsymbol{\beta}_{0t} = \beta_0 + \beta_1 t + ... + \beta_r t^r$. For example, the duration vector, $\boldsymbol{Y}_t$, may be linear ($\boldsymbol{Y}_t = (1, t)$, with a straight line shape), quadratic ($\boldsymbol{Y}_t = (1, t, t^2)$,

where the hazard is U-shaped or inverse-U-shaped) or any polynomial function of time (age) where ($\boldsymbol{Y}_t = (1, t, t^2, ..., t^r)$.

Logarithmic Specification In a logarithmic function of time, the baseline log odds is given by $\boldsymbol{Y}_t'\boldsymbol{\beta}_{0t} = r\text{log}t$, where the parameter r is to be estimated together with intercept and slope parameters within the vector $\boldsymbol{\beta}_{0t}$. The logarithmic specification in the discrete-time specification is analogous to the Weibull specification of the continuous time methods (Jenkins, 2005). This is mainly because the shape of the hazard remains constant if $r = 0$, monotonically decreases when $r < 0$ or monotonically increases when $r > 0$. Note that in cases where the number of time periods are small, this specification is indistinguishable from linear specification. (Jenkins, 1995).

Non-parametric Parametric specification of the baseline hazard has a major drawback of lacking flexibility as a result of the restrictions that are put on the function, based on the different assumptions that are made by the different models. One example is the normality assumption of the data, which is required by many statistical tests. Another example is the linearity assumption of η_{kt} on the β's which implies that each covariate has a linear effect on the transformed hazard. These assumptions are too restrictive and in cases where they are violated, serious bias may arise in the interpretation of results as well as in the application of the statistical tests themselves. This problem is circumvented by non-parametric modeling of the baseline hazard. Models that allow for a more flexible predictor which is not necessarily linear are required (Tutz et al., 2016). These include use of kernels (similar to moving averages) and splines (smoothing functions). Additive models may be used to avoid numerical problems associated with estimation of the baseline hazard where $\mathbf{Y}_t$ in equation 2.11 is a smooth function of time.

Splines The baseline in model 2.4 may be expressed as a function given by splines. A spline of order, say q, is a piece-wise polynomial function of degree q which is joined together at points by knots. Each segment is a polynomial but continuous and has continuous derivatives of orders $1, 2, ..., q-1$ at the knots. In general, the spline-based approach may be classified into regression splines, smoothing splines and penalized regression splines. In regression splines, the basis functions are joined at a predetermined set of knots. For example Harrell et al. (2001) placed the knots at quantiles of the data. Although this is relatively simple in practice, a shortcoming of this approach lies in its high sensitivity to the number and locations of the knots. To avoid issues associated with knot placement, smoothing splines have been suggested in literature, where a relatively large number of knots are placed at each data point, but at the same time, a penalty term that penalizes curvature is incorporated

in the model fit objective function (Wood & Augustin, 2002). This approach gives rise to as many parameters as the data are to be smoothed, which in turn results in high computational intensity when estimating the splines. A better alternative is the penalized regression splines which retain the good properties of smoothing splines at the same time with a reduced knot set to lower the computational expense (Gray, 1992; Roshani & Ghaderi, 2016; Mantel & Hankey, 1978; Fahrmeir & Wagenpfeil, 1996).

Let $g(z)$ represent a univariate function (used to model the log of the baseline, β_{0t}) with $\{c_k(z) : k = 1, \ldots, d\}$ being a set of functions. Then the univariate function can be defined as

$$g(z) = \sum_{k=1}^{d} \alpha_k c_k(z),$$

where α_k's and m are known parameters. The function $g(z)$ is comprised of a linear combination of the basis functions, $c_k(z)$. Suppose also $\{z^* : k = 1, \ldots, d\}$ is a set of points which are referred to as knots.

Regression splines yield more stable estimates compared to polynomial regression functions (mentioned above) since they introduce flexibility by increasing the number of knots but keep the degree of polynomial fixed, unlike the polynomial regression functions which must use a high degree polynomial to produce flexible fits. In the regression splines, log baseline is expanded into B-spline basis functions. Cubic splines are such examples, where $q = 3$ for each segment, which are piece-wise cubic functions that are continuous and have continuous first and second derivatives. Cubic splines are piece-wise polynomials whose curves are joined smoothly at knots as opposed to other polynomials which are not smooth at points. In addition to first derivatives, their curves are twice differentiable at all knots, hence flexible in interpolation. Letting $c_m(z) = |z - z_k^*|$ for $k = 1, \ldots, d, c_{k+1}(z) = 1, c_{k+2}(z) = z$,. The function $g(z)$ is defined as

$$g(z) = \sum_{k=1}^{d+2} \alpha_k c_k(z). \tag{2.12}$$

Some scholars believe that there is seldom any good reason to model beyond cubic splines, as the spline can be made smoother by simply increasing the number of knots. More approaches on interpolation using splines are detailed in De Boor & De Boor (1978) and Wang (2011).

Smoothing splines are designed to balance model fit with smoothness in the result-

ing function, thereby addressing the non-linearity in models. Unlike the regular splines which fit through the data points (interpolate the data), smoothed splines are not exact, so they do not run through all the data points because a constraint for some smoothness is placed. Their other major drawback lies in the fact that the smoothness of the graph depends on visually assessing the goodness of fit of the graph through trial and error. The choice of the number of knots as well as position of the knots is subjective and may give rise to inaccurate estimates; over- or under-smoothed lines. According to Wang (2011), the piece-wise polynomials help to assign an appropriate weight to the different data points so as to avoid local variations in data, which could possibly assign excessive influence to the regression line. Non-parametric bootstraping maybe used to construct confidence intervals for splines to quantify uncertainty.

$$\log\beta_{0t} = k + \sum_{d=1}^{m} k_d B_d(t, q), \tag{2.13}$$

where $k = (k_0, k_1, ..., k_m)'$ are the spline coefficients and $m = \bar{m} + q - 1$ denotes the number of interior knots while q denotes the degree of the B-splines basis function, $B(.)$. Increasing the number of knots also increases the flexibility in approximating $\beta_{0,ij}(t)$ (Rizopoulos, 2012). Where restricted cubic spline is used,, a continuous smooth function is obtained which is linear before the first knot, a piece-wise cubic polynomial between adjacent knots and linear again after the last knot.
Penalized cubic regression splines may be used to represent β_{0t} in model 2.4 as

$$\beta_{0t} = \sum_{d=1}^{m} \alpha_{0d} C_d(t), \tag{2.14}$$

where $C_d(t)$'s are the cubic spline basis and α_{0d}'s are the knot coefficients. Then penalized cubic regression splines are knot-based approximations on the knots $t_1 < \ldots < t_{m_d}$ where $m_d = m - 1$ which will minimize

$$\sum(\beta_{0t} - \hat{\beta}_{0t})^2 + \theta \int \hat{\beta}''_{0t} dt,$$

where θ is a smoothing parameter and $\hat{\beta}_{0t}$ is the estimated spline.

2.3 State Transition Models for Recurrent Events (Discrete Frailty Survival Model)

The model discussed in section 2.2 considered absorbing destination states. In real life, however, some processes have recurrent events, for example in family formation and dissolution studies (Steele et al., 2005) or in employment studies of transitions between employment and no employment (McCall, 1996; Cameron & Trivedi, 2005; Steele, 2011). It is common to assume that these recurrent events are correlated as they occur on the same subject (Raudenbush & Bryk, 2002; Han & Boves, 2007; Goldstein, 2011). There is an important biological or psychological variation between subjects, hence the need of an additional parameter, the random effect, which may capture and measure heterogeneity among the subjects to avoid bias (Scheike & Jensen, 1997; Heckman & Singer, 1984; Hougaard, 1986; Hougaard et al., 1994; Hougaard, 1995; Julian, 2001).

There is extensive literature available that has attempted to consider a discrete-time survival model with recurrent events, while considering subject-specific random effects: also referred to as multilevel models. Here, the data are hierarchical (Raudenbush & Bryk, 1986; Bryk & Raudenbush, 1988; Raudenbush & Bryk, 1988, 2002; Molenberghs & Verbeke, 2000, 2005; Fitzmaurice et al., 2008) and have more than one level, be these natural clusters, repeated measures or longitudinal. Hence, a level may be defined as a unit of analysis which can be a subject or a cluster. The probability distribution for the observation errors at the first level is specified and another probability distribution is also specified for the parameters (random effects in the model at subsequent higher levels) (Laird & Ware, 1982; De Leeuw & Kreft, 1986; Rice & Leyland, 1996; Langford & Lewis, 1998; Zhang & Steele, 2004; Steele, 2008; Molenberghs & Verbeke, 2005). The random parameters belong to higher levels and are assumed to vary across clusters, since the observed clusters are a random sample from all clusters in the population (Kaombe & Manda, 2021). The subject's group effect (random covariates) are to be estimated by the model under the assumption that there are interactions between the fixed covariates and the subject's group effect. Though the random effects may be estimated for each group of subjects, focus may usually be on simply measuring the variation in the outcome variable that is contributed by clustering in the data. So, a covariate may enter the model as a fixed or random effects variable.

According to Molenberghs & Verbeke (2000), there are basically three approaches to handling subject-specific parameters. Firstly, they may be treated as unknown

fixed parameters. Secondly, they may be considered as nuisance parameters and estimation of the natural parameters is done by maximizing the likelihood of the data conditional on the sufficient statistic for the subject-specific parameters. Lastly, the random effect approach is used where interest is in drawing inferences with respect to the subject specific parameters, including making some subject-specific predictions. Here, the subject-specific parameters are drawn from a population of subject-specific parameters simultaneously with randomly sampling subjects from a general population of subjects, parameters. This will mean that the sample of random-specific parameters may be considered as random vectors, drawn independently from a distribution function, say $f(u_k)$, which is known as the mixing distribution. By integrating them out over their assumed distribution, elimination of the subject-specific parameters is then obtained. In this study, focus is only made on the latter approach.

Based on the random effects approach, equations 2.1 and 2.2 may be modified by introducing a random effect, u_k which is conveniently assumed to follow a normal distribution, i.e. $u_k \sim N(0, \sigma_u^2)$. Let $T_{k1}, T_{k2}, ..., T_{kl_k}$ be the survival times for subject k, where $l = 1, 2, .., l_k$ denote the episode count. An episode is defined as a continuous period during which a subject is at risk of experiencing an event (Steele, 2011). Tutz et al. (2016) describes the idea of multiple spells (episodes) from a 2-state model which we will adopt. There may exist some unobserved factors that cause subject k to transition between states. Therefore, due to this unobserved heterogeneity, we define the hazard function for the l^{th} episode for subject k by extending model 2.1 as

$$\lambda^{(l)}\left(t|\boldsymbol{Z}_{kl}, u_k\right) = P\left(T_{kl} = t|T_{kl} \geq t; \boldsymbol{Z}_{kl}, u_k\right), \tag{2.15}$$

where u_k is the subject-specific random effect. This hazard is also linked to the covariates by a link function in a similar manner as in equation 2.3 by

$$\lambda^{(l)}\left(t|\boldsymbol{Z}_{kl}, u_k\right) = h(\boldsymbol{\beta}_{0t} + \boldsymbol{Z}'_{kl}\boldsymbol{\beta} + u_k). \tag{2.16}$$

2.3.1 The Likelihood

Traditionally, the maximum likelihood estimation of parameters has been commonly done (Czepiel, 2002), which requires that the estimates are derived as modes of the log-likelihood function corresponding to the distribution of the observed outcomes.

Define the censoring variable, δ_{kl}, such that

$$\delta_{kl} = \begin{cases} \delta_{kl} = 1, & \text{for observed spell } 1, 2, ...l_k - 1 \\ \delta_{kl_k} = 0, & \text{for the last spell, } l_k \text{ .} \end{cases}$$

We therefore, note that after the last episode, censoring of the episodes is observed, meaning that $\delta_{ijk} = 0$ in the last episode. The l^{th} spell may be represented as a sequence of binary variables $\varepsilon_{kl1}, \varepsilon_{kl2}, \ldots, \varepsilon_{kl\tau_{kl}}$ where

$$\varepsilon_{klt} = \begin{cases} 1, & \text{if subject } k \text{ experiences the event in episode } l \text{ given it reaches interval } t \\ 0, & \text{otherwise} \end{cases}$$

The contribution to the joint likelihood for subject k is then given by

$$L_k = \prod_{l=1}^{l_k} \prod_{b=1}^{t_k-(1-\delta_{kl})} \left\{ \lambda(b|\boldsymbol{Z}_k)^{\varepsilon_{klb}} \left(1 - \lambda(b|\boldsymbol{Z}_k)\right)^{1-\varepsilon_{klb}} \right\} \times f(u_k|\sigma_u^2), \tag{2.17}$$

The overall joint likelihood of the data and random effects over all subjects is then the product of the likelihood in equation 2.17 over all subjects and the density function of the random effect, $f(u_k)$. This is given by,

$$L = \prod_{k=1}^{K} \prod_{l=1}^{l_k} \prod_{b=1}^{t_k-(1-\delta_{kl})} \left\{ \lambda(b|\boldsymbol{Z}_k)^{\varepsilon_{klb}} \left(1 - \lambda(b|\boldsymbol{Z}_k)\right)^{1-\varepsilon_{klb}} \right\} \times \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{u_k^2}{2\sigma_u^2}\right). \tag{2.18}$$

The corresponding marginal log-likelihood which is integrated over the random effect is then given by

$$\ell(\boldsymbol{\beta}, \sigma^2) = \sum_{k=1}^{K} \log\left(\int_{-\infty}^{\infty} \prod_{l=1}^{l_k} \prod_{b=1}^{t_k-(1-\delta_{kl})} \left\{ \lambda(b|\boldsymbol{Z}_k)^{\varepsilon_{klb}} \left(1 - \lambda(b|\boldsymbol{Z}_k)\right)^{1-\varepsilon_{klb}} \right\} du_k \right), \tag{2.19}$$

where

$$\lambda(t|\boldsymbol{Z}_k, u_k) = \frac{\exp\left(\boldsymbol{\beta}_0(t) + \boldsymbol{Z}_k'(t)\boldsymbol{\beta} + u_k\right)}{1 + \exp\left(\boldsymbol{\beta}_0(t) + \boldsymbol{Z}_k'(t)\boldsymbol{\beta} + u_k\right)}.$$

The score functions for σ_u^2 and $\boldsymbol{\beta}$ follow from the first-order derivatives of the log-likelihood in equation 2.19 and are given by

$$U_{\sigma_u^2} = \frac{\partial \ell(\boldsymbol{\beta}, \sigma_u^2)}{\partial \sigma_u^2}$$

and

$$U_{\boldsymbol{\beta}} = \frac{\partial \ell(\boldsymbol{\beta}, \sigma_u^2)}{\partial \boldsymbol{\beta}},$$

respectively. Setting these score functions equal to 0 and solving them simultaneously will maximize the likelihood to obtain the estimates of σ_u^2 and $\boldsymbol{\beta}$.

2.3.2 Numerical Approximation Techniques

Solving a system of nonlinear equations algebraically may not be as easy and straight forward as solving linear equations (Czepiel, 2002). The estimation of parameters can be evaluated analytically by maximizing marginal likelihood of the data which is performed by integrating over the random effects (Verbeke, 1997; Maqutu, 2010). Also, the integrals of the log-likelihood functions in equation 2.19 are not in closed form. For that reason, iterative processes are viable alternatives that may be used to numerically estimate the solution to the system of nonlinear equations which results from maximization of the log-likelihood functions (numeric or stochastic integration). These include the Newton-Raphson method (Ripatti & Palmgren, 2000) of the data, the Laplace method which approximates the integrand (Breslow & Clayton, 1993), the Gausian quadratures approximation for the integral (Liu & Yu, 2008; Liu & Pierce, 1994) as well as the penalized and marginal quasi-likelihoods which approximate the data. In terms of computational burden as well as precision, each approach has its own advantages and disadvantages, although they might produce slightly different estimates, especially for the variance-covariance parameters of the random effects. In this chapter, we will review the Laplace, Quadrature and Quasi-Likelihood methods.

Laplace Approximation of the Integrand In the Laplace method (Laplace, 1986), the integral is approximated where the exact likelihood is difficult to evaluate. The integrand is manipulated to approximate the normal integrand, so that is has a normal-like quantity which is easy to integrate, then exploit the normal solution type iteratively until convergence is reached. The principle is to fit a normal distribution to some function by fitting some constant, say $C\times$ the Gaussian. It is also used to find the integral of some function (the normalizing function). This is because we know how to calculate the integral of the normal distribution, we can now find the normalizing constant of the normal distribution. Suppose $f(u)$ is a unimodal function which can be transformed as $Q(u) = \log f(u)$, so that $f(u) = e^{Q(u)}$. The second order Taylor series is then expanded around the mode of $f(u)$, say $\hat{u}$:

$$Q(u) \approx Q(\hat{u}) + \frac{(u-\hat{u})^1}{1!}\cancelto{0}{Q'(\hat{u})} + \frac{(u-\hat{u})^2}{2!}Q''(\hat{u}). \tag{2.20}$$

It is known that $Q'(\hat{u}) = 0$ at the unique global maxima, so the middle term on the right side of equation 2.20 will be 0 and equation 2.20 will reduce to

$$Q(u) \approx Q(\hat{u}) + \frac{Q''(\hat{u})}{2}(u - \hat{u})^2. \tag{2.21}$$

So, according to the example by Molenberghs & Verbeke (2005), the integral of the form

$$\begin{aligned} I = \int f(u)du &= \int e^{Q(u)}du \\ &\approx \int e^{Q(\hat{u}) + \frac{Q''(\hat{u})}{2}(u-\hat{u})^2}du \\ &= \underbrace{e^{Q(\hat{u})}}_{f(\hat{u})} \underbrace{\int e^{\frac{Q''(\hat{u})}{2}(u-\hat{u})^2}du}_{\sqrt{\frac{2\pi}{|Q''(\hat{u})|}}} \\ &= f(\hat{u})\sqrt{\tfrac{2\pi}{|Q''(\hat{u})|}}, \end{aligned} \tag{2.22}$$

is derived, which is the normalizing constant that is used to approximate the integral I, where $\hat{u}$ is the value of u for which $Q(u)$ is maximized. Here, $Q(u)$ is a twice-differentiable function and $Q''(\hat{u})$ is a vector of the second order derivatives of Q which is evaluated at $\hat{u}$ (equivalent to the Hessian of Q). It can be noted that the integral in the third step of equation 2.22 is obtained by noting that the integral of the probability density function of a random variable u from a normal distribution with $mean = mode = \hat{u}$ and variance $\frac{1}{|Q''(\hat{u})|}$ over all the area is 1, i.e.

$$\frac{1}{\sqrt{2\pi \times \frac{1}{Q''(\hat{u})}}} \int e^{\frac{1}{2}\left(\frac{(u-\hat{u})^2}{\left(\frac{1}{Q''(\hat{u})}\right)}\right)}du = 1$$

$$\Rightarrow \int e^{\frac{1}{2}\left(\frac{u-\hat{u}}{Q''(\hat{u})}\right)^2}du = \sqrt{2\pi \times \frac{1}{Q''(\hat{u})}} = \sqrt{\frac{2\pi}{|Q''(\hat{u})|}}\ , \tag{2.23}$$

which may now be substituted in the third step of equation 2.22 to obtain the normalizing constant in the fourth step of equation 2.22. This may now be used to approximate the integrand in the likelihood in equation 2.19.

In cases where $Q(.)$ is bimodal, the improved Laplace approximation may be used. Here, as many different estimates of u may be used as the different modes of the $Q(.)$

function. Where a multivariate case is considered, the mode will be represented by a vector, $\boldsymbol{u}$, and the second derivative will be a negative definite matrix (which will produce the Hessian matrix). Note that the Laplace approximation for the integrals in the marginal likelihood may also be specified by choosing the adaptive Gaussian quadrature with only $q = 1$ quadrature point (Molenberghs & Verbeke, 2000).

Gaussian Quadrature Gaussian quadratures are common in approximating integrals of the form

$$\int_a^b W(u)f(u)du \approx \sum_{k=1}^{\infty} \omega_k f(u_k)$$

where $W(u)$ is a known weight function, u_k are the nodes, and ω_k are the weights at the nodes. There are broadly four types of Gaussian quadratures which are used in different special situations, namely:

- Hermite-Gaussian quadrature where the weight function is of the form $W(u) = e^{-u^2}$, so that the integral becomes $\int_{-\infty}^{\infty} e^{-u^2} f(u)du$,
- Legendre-Gaussian quadrature where the weight function is of 1, so that the the integral becomes $\int_{-1}^{1} f(u)du$,
- Laguerre-Gaussian quadrature where the weight function is of the form $\int_0^{\infty} e^{-u} f(u)du$ and
- Chebyschev-Gaussian quadrature where the weight function is of the form $W(u) = \frac{1}{\sqrt{1-u^2}}$, so that the the integral becomes $\int_{-1}^{1} \frac{1}{\sqrt{1-u^2}} f(u)du$.

However, the Legendre-Gaussian quadrature may be used as a general purpose integration routine.

In a univariate case (Stata.com, 2022), the integral of a function multiplied by the kernel of the standard normal distribution may be approximated using Gauss–Hermite quadrature. It is often used in performing maximum likelihood estimation, particularly in random effects models, because of its relation to Gaussian densities (Pinheiro & Bates, 1995; Liu & Pierce, 1994). For a Gauss–Hermite quadrature with q-points, we let the abscissa and weight be (u_k^*, ω_k^*), where $k = 1, ..., q$. Then the Gauss–Hermite quadrature is given by

$$\int_{-\infty}^{\infty} e^{-u^2} f(u)du \approx \sum_{k=1}^{q} \omega_k^* f(a_k^*). \tag{2.24}$$

In the Gauss-Hermite, the nodes, a_k's, are the roots of the q^{th} order Hermite polynomial

$$H_{n+1}(u) = (-1)^{n+1} e^{u^2} \frac{\partial^{n+1}}{\partial u^{n+1}} e^{-u^2}. \tag{2.25}$$

These Hermite polynomials are orthogonal with respect to the weight function, e^{u^2}, in $(-\infty; \infty)$ i.e.

$$\int_{-\infty}^{\infty} e^{2} H_m(u) H_n(u) du = 0 \qquad \text{for } n \neq m. \tag{2.26}$$

The corresponding weights are then determined from the relation

$$a_k = \int_{\infty}^{\infty} \frac{e^{u^2} H_{n+1}(u)}{(u - u_k) H'_{n+1}(u_k)} du. \tag{2.27}$$

The exact results for polynomials of degrees $\leq 2n + 1$ are produced by equation 2.24. In the multivariate case, the change of variables technique may then be used to transform the multivariate integral into a set of nested univariate integrals. Each univariate integral can then be evaluated using Gauss–Hermite quadrature.

Penalized and Marginal Quasi-Likelihood According to McCulloch & Searle (2004) and Molenberghs & Verbeke (2005), the quasi-likelihood approach is based on the approximation of the data, rather than the integrand from a parametric-likelihood. It is preferred due to its ability to generate highly efficient estimators without making precise distributional assumptions. Here, the data is decomposed into the mean and an appropriate error term with a Taylor series expansion of the mean (Agresti et al., 2000; Agresti, 2003; Molenberghs & Verbeke, 2005). Rather than specifying a distribution, the quasi-likelihood only specifies the mean and variance of the data. Various approximations using quasi-likelihood differ in the order of the Taylor approximation and/or the point around which the approximation is expanded (Molenberghs & Verbeke, 2005).

In penalized quasi-likelihood (PQL) for generalized linear models, optimization of the quasi-likelihood function is augmented with a penalty term on the random effects. Following Agresti et al. (2000), the PQL approach is an iterative approach which requires that each iteration contains two steps, namely: updating the parameter ($\boldsymbol{\beta}$) and then updating the variance parameter (σ). Let $(\boldsymbol{\beta}^{(r)}, \sigma^{(r)})$ be the values after r iterations. The Henderson's mixed-model equations (Henderson et al. 1959) may now be used to update $\boldsymbol{\beta}$. It is based on the normal theory mixed model where both $f(t|u_k; \boldsymbol{\beta})$ and $f(u_k; \sigma)$ are densities for random variables that are normally distributed. The values of u and $\boldsymbol{\beta}$ that jointly maximize the function $f(t|u_k, \boldsymbol{\beta}) f(u_k; \hat{\sigma})$

are the empirical estimated best linear unbiased predictor (EBLUP) of u and the maximum likelihood estimates (MLE) of $\boldsymbol{\beta}$, where $\hat{\sigma}$ is the MLE of σ (Agresti et al., 2000; Searle et al., 2009). This now makes it possible to update for $\boldsymbol{\beta}$ in the generalized linear mixed models context. Given $\sigma^{(r)}$, the function $f(t|u_k, \boldsymbol{\beta})f(u_k; \sigma^{(r)})$ is then maximized with respect to u_k and $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^{(r+1)}$ is then assigned the maximizing value of $\boldsymbol{\beta}$. Since this maximization is not trivial, it will require an iterative method such as the Newton-Raphson method.

A further normal approximation is used to update σ through a working dependent variable, say z, which is constructed similar to the usual iteratively reweighted least squares algorithm used in generalized linear models. Under the assumption that z follows a normal linear mixed model, the variance component is now estimated using MLE (Searle et al., 2009; Agresti et al., 2000) and the corresponding values of these MLEs are then the components of $\sigma^{(r)}$.

Marginal quasi-likelihood approximation is another alternative, which is similar to the PQL approach, though based on a linear Taylor expansion of the mean around current estimates of $\hat{\beta}$ for the fixed effects, and around $u_k = 0$ for the random effects. It produces similar expressions to those of PQL except that instead of the conditional mean taking the form $g^{-1}\left(\boldsymbol{Z}'\hat{\beta} + \hat{u_k}\right)$, the mean takes the form $g^{-1}\left(\boldsymbol{Z}'\hat{\beta}\right)$. The MQL estimates are similar to PQL estimates in that they both may be obtained by optimizing a quasi-likelihood function which involves only first and second order moments, but in MQL, they are are now evaluated using the marginal linear predictor rather than the conditional predictor (Maqutu, 2010).

The quasi-likelihood approach has the drawback of preventing likelihood-based inference since the likelihood function is not evaluated (Grilli & Rampichini, 2007). In addition, as a result of the wrong (quasi) likelihood being used, quasi-likelihood estimators lose asymptotic efficiency when compared to maximum likelihood estimators. For these reasons, they will not be used in this study.

2.4 Higher Order Level Random Effects Model

In a case where a 3^{rd} level of hierarchy is considered, level 1 units are nested within level 2 units, which in turn are nested within level 3 units (Julian, 2001; Raudenbush & Bryk, 1986, 2002; Bryk & Raudenbush, 1988; De Leeuw & Kreft, 1986; Steele et al., 2005). Here, an extra random effect is included in the model so that the hazard function for spell $l(l = 1\ldots., L_{hk})$ from subject $k(k = 1, \ldots, K_h)$ in household $h(h =$

$1, \ldots, H)$ is given by

$$L_{hkl}\lambda\left(t|\boldsymbol{Z}_{hkl}, u_k, u_{hk}\right) = P\left(T_{hkl} = t|T_{hkl} \geq t; \boldsymbol{Z}_{hkl}, u_{hk}, u_h\right), \tag{2.28}$$

where u_{hk} and u_h denote the random effects for subject hk and household h , respectively, which may be assumed independent (Biggeri et al., 2001). A dummy variable, $\varepsilon_{hkt} = 1$, is defined if subject k from household h has experienced an event in the l^{th} spell at time interval t and $\varepsilon_{hkt} = 0$ otherwise. Their contribution to the likelihood is

$$\prod_{b=1}^{t_{hkl}-(1-\delta_{hkl})} \left\{\lambda(b|\boldsymbol{Z}_k)^{\varepsilon_{hklb}} \left(1 - \lambda(b|\boldsymbol{Z}_{hkl})\right)^{1-\varepsilon_{hklb}}\right\} \times f(\boldsymbol{u}), \tag{2.29}$$

where $\boldsymbol{u} = \begin{pmatrix} u_h \\ u_{hk} \end{pmatrix} \sim MVN(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$ and $\boldsymbol{\Sigma}_u = \begin{pmatrix} \sigma_{u_h} & 0 \\ 0 & \sigma_{u_{hk}} \end{pmatrix}$. Also, $\delta_{hkl} = 1$ in spells $l = 1...l_{hk} - 1$ and $\delta_{hkl} = 0$ in the last spell for subject k_h.
Estimation procedure is similar to what has been discussed in section 2.3 except that this model results in one extra parameter to be estimated, u_h.

2.5 Summary

In this chapter, standard methods of analyzing univariate multilevel discrete time-to-event models were reviewed. Standard univariate survival analysis provides an important conceptual and analytic framework from which to evaluate if and when transitions (events) occur. Methods of modeling the baseline hazard and integration of the likelihood were also discussed. Application using a real-life data set is made in the next chapter.

CHAPTER 3

APPLICATION TO AGE AT FIRST MARRIAGE DATA

In this chapter, application of the standard methods discussed in Chapter. 2 will be made to a routine data set based on age at first marriage. In section 3.1, a motivation of the data is presented. and a brief description of the data is made in section 3.2. Techniques of handling missing data are also presented in section 3.3 and finally, application of these methods to real-life data is then done in section 3.4. Procedures on data preparation for discrete-time analysis are given in Appendix C.

3.1 Motivating example: Context of Marriage Formation and dissolution

Multi-state models have commonly been used in clinical settings, for example, in studying the transitions between disease stages such as HIV progression (Reddy et al., 2011). In cancer studies recurrent events arise when patients' tumors progress and metastasize or recover as the patient moves between the cancer stages. These models can also be used in modeling movement between marital states where, for example a married person may divorce and remarry, thus recurrent events.

The timing of a first marriage has a huge impact on the health, social, psychological, educational and economic profiles of families and their children (Gähler & Palmtag, 2015; Uecker, 2012; Ikamari, 2005). A common example in the African context is observable when a (probably forced) marriage is a result of teenage pregnancy followed by parents sending their girl child away for an early marriage where she will be forced to stay with a totally new family. Early marriages potentially have consequences such as school dropouts, teenage pregnancies, abortions as well as emotional disorders such as hypertension, stress, bipolar and other physical health related diseases like high blood pressure. In worst-case scenarios, suicidal thoughts which may lead to committing suicide may result (Gage, 2013). Poor relationship qualities (sometimes leading to marital separation) potentially compromise both physical and mental well-being (Waltz et al., 1991), whereas good marriages pro-

mote a good well-being. These changes in status, into marriage, divorce, separation or being widowed, may also affect children and relatives involved (Hasselmo et al., 2015; Sbarra et al., 2009; Wang et al., 2015). It is therefore, of special interest to study the interrelated factors leading to these changes in naptiality. using multi-state models.

Multi-state models have commonly been used in clinical settings, for example, in studying the transitions between disease stages such as HIV progression (Reddy et al., 2011). In cancer studies recurrent events arise when patients' tumors progress and metastasize or recover as the patient moves between the cancer stages.

3.2 The Data

3.2.1 Data and Study Area

The Africa Health Research Institute (ARHI, previously Africa Center for Population Health studies), is situated in the rural area of northern KwaZulu-Natal in South Africa where it collects the data (Tanser et al., 2007). This area is largely dominated by a black community of the Zulu cultured South Africans. As represented in the figure displayed in Appendix B, South Africa is found at the bottom of the African continent and has nine provinces including KwaZulu-Natal which also has areas including the Zululand and the Elephant Coast where the data set was collected. The Population Research Department provides world class coordinated research operational support for population-level research studies, including development and implementation of innovative, efficient and cost-effective field-based and telephonic data collection systems, anchored on robust quality assurance and quality control practices (AHRI, 2021, 2018).

The AHRI is one of the many Health and Demographic Surveillance Systems (HDSS) sites under the INDEPTH network. It collects longitudinal data including, in addition to HIV and TB, data on the life course history of marriage formation and dissolution events and duration, and has been in operation since 1996. The surveillance area is near Mtubatuba, in the Umkanyakude rural district. It covers an area of 438 square kilometres with a highly mobile population of approximately $90\ 000$ people who are members of approximately 11 000 households. These include all individuals reported by household informants as household members regardless of them being resident or non-resident (Tanser et al., 2007; Hosegood et al., 2009). The Africa Centre Demographic Information System (ACDIS) started data collection on 1 Jan-

uary 2000 and is conducting an ongoing study. However, for the purposes of this research, we only considered subjects who were followed up from January 2004 to December 2016. For all registered households and individuals, demographic and health information was collected every 4 months. Because the surveillance cohorts are dynamic, subjects may enter or leave the cohort through migration or births and deaths at any time (Tanser et al., 2007), thus ragged study entries. The migration might be within the surveillance area itself (where subjects change households) or as a result of in-migration and out-migration, as defined by Dobra et al. (2017). As such, the participation rates at each wave is about 95% for household data collection. Therefore to minimize non-response, where respondents are either non-resident or unavailable, suitable household members are selected as alternative informants.

This study considers 50 698 eligible subjects who were enrolled between January 2004 and December 2016, some of whom got lost to follow-up. To the best of our knowledge, this data set has been used to study family formation and dissolution, but using different statistical methods such as Hosegood et al. (2009) who compared the 2000 and the 2006 cohorts. In the case where discrete time event history analysis methods were used, sometimes it would be on HIV or some other areas of study (Tomita et al., 2017) or different locations (Houle et al., 2014; Clark & Brauner-Otto, 2015; Clark et al., 2013). From the analyzed data, 4 marital states were considered (*Never Married, Married, Separated* and *Widowed*) and each subject would be in one of these states at each time of visit. The possible paths of transitions are graphically displayed in Figure 3.1 below:

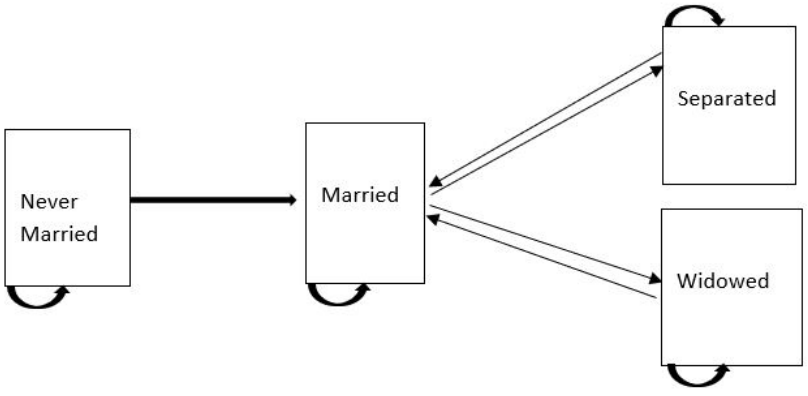


Figure (3.1) MultiState model for family formation and dissolution.

3.2.2 Data Availability

The data is available upon request from Africa Health Research Institute from ARHI (Tanser et al., 2007) whose website is AHRI (2021). Data sets with restricted access require a data access agreement to be completed. A request is then submitted to the applicable data custodian for specific data sets on the repository. Requests for ad hoc data sets, beyond the data sets archived on the Africa Centre data repository, have to be directed to Africa Centre's Helpdesk (help@africacentre.ac.za). (Tomita et al., 2017)

3.2.3 Ethical Clearance

Ethical approval for all the data collected by AHRI was obtained from the University of KwaZulu-Natal's Ethics Committee (BE 169/15) (Tanser et al., 2007; Tomita et al., 2017).

3.2.4 Statistical Software

Discrete-time analysis for this chapter was done in STATA17 using the ***xtlogit*** for the random effects model where the clustering variable (*ID*) was specified.

3.3 Handling Missing Data

The data from AHRI was collected and recorded as different data sets for different purposes. As a result, not all variables needed for this study were found in the same data set. Different data sets were therefore merged and/or appended. These are Women's General Health-All (from which the variables Marital status, Age at visit, Age at first marriage, Age at first sex, Gender and BP were extracted for females); Men's General Health-All (from which the variables Marital status, Age at visit, Age at first marriage, Age at first sex, Gender and BP were extracted for males); RD07-99 ACDIS HSE-I All (from which the variables Education level, Is income earned, Is employed, were extracted) and the DayDataSetEpisodesAbridgedV0 (which contains the variables household and bounded structure to which each subject belongs).

All duplicates (of subjects on visit dates) as a result of merging were dropped since they gave the same information. Merging these data sets resulted in a lot of missing data, as other data sets did not record all subjects for all the required variables. On the other hand, since this is an ongoing study like most longitudinal studies, some subjects entered the study late and others left earlier than December 2016 (due to

death, migration and so forth). Others exhibit intermittent missingness which might have resulted from non-response, refusal from subjects in a specific variable and time point. All this missingness can cause serious bias to results, loss of power and underestimation of standard errors and loss of valuable statistical information (Satty et al., 2015; Chinomona & Mwambi, 2015; Molenberghs & Kenward, 2007; Satty et al., 2013; Satty & Mwambi, 2014) .

Missing data is a common feature in this study as it is in most areas of research that use longitudinal data. There are a different mechanisms of missingness such as Missing Not at Random, Missing at Random Analysis and Missing Completely at Random. These and their methods of handling their analysis are well documented in Allison (2001); Molnar et al. (2008); Molenberghs & Kenward (2007); Satty & Mwambi (2012); Satty et al. (2013); Little & Rubin (2014); Satty et al. (2015).

Various methods of imputation have been suggested in literature, such as multiple imputation (Rubin, 1987, 2004). Most survey data analyses are done using a complete-case method which is also a default by most statistical software. In this approach, a list-wise deletion of all cases with missing values is performed under the assumption that the values are missing completely at random. As a result of reduced sample sizes, this approach has a disadvantage of potential loss of power to detect an association between exposures and an outcome (Cleves, 2008). The remaining observations might not be a representative of the population.

The Last Observation Carried Forward (LOCF) approach (Molenberghs & Kenward, 2007; Molnar et al., 2008; Siddiqui & Ali, 1998; Mallinckrodt et al., 2003) is also commonly applied to longitudinal data where incompleteness usually results from attrition or loss to follow up. This method can be applied to both monotone and non-monotone missing data and it mainly assumes that missing values do so completely at random. This naive approach is the method that was utilized to impute missing values on the categorical variables in this study. To do so, it is also assumed that the observations would not significantly change in a year's period. For instance, if a subject's marital status in 2012 is not known, but it was known to be *Married* in 2011, then the value for 2012 for that subject will be imputed to be *Married*. One danger of using this technique lies in the possible underestimation of the variance for parameter estimates (Satty & Mwambi, 2014).

3.4 Results

The analyzed data had $268\ 557$ person-years for subject who were considered in the *Never married* and *Married* states and they were followed up over a mean period of 6.8 person-years. Of these, $1,274$ *Never Married* $\rightarrow$ *Married* transitions were made. Total time at risk was $147\ 651.2$ person-years with $24\ 332$ subjects who were right censored. Table 3.1 below displays the distribution of some of the covariates among the subjects at entry where applicable. It is clear that the majority of the population occupied a *Never Married* state compared to a *Married* state .

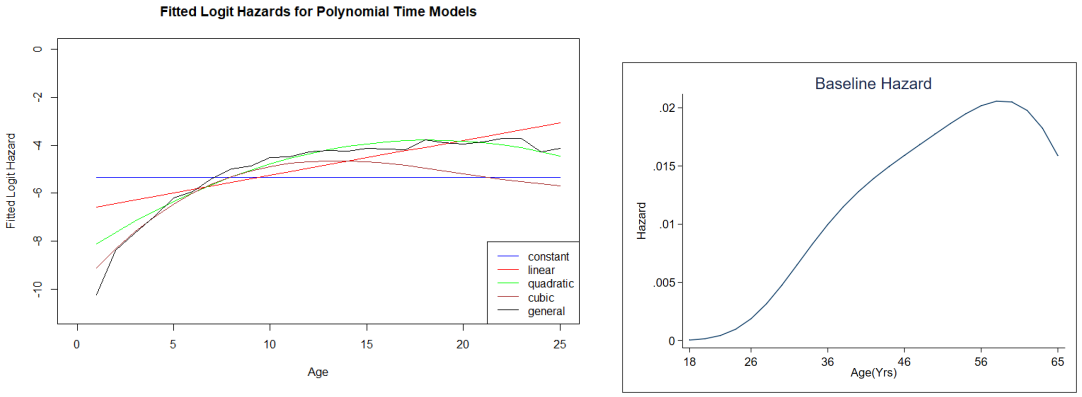
Table (3.1) Distribution of sampled individuals by explanatory variables at study entry

Variable	**Never Married** N (%)	**Married** N (%)
Gender		
Female	17 130 (95.33)	839 (4.67)
Male	13 145 (97.15)	386 (2.85)
Income		
Yes	2 791 (94.77)	154 (5.23)
No	24 742 (95.87)	1 065 (4.13)
Is Employed		
Yes	10 526 (95.75)	467 (4.25)
No	18 314 (96.06)	751 (3.94)
Highest Education		
Never went to school	198 (76.45)	61 (23.55)
Primary	1 884 (91.59))	173 (8.41)
High School	2 634 (94.17)	163 (5.83)
Tertiary	86 (86.87)	13 (13.13)

Regarding the parametric functional form of the baseline hazards, choices were made between a constant baseline, polynomial functions of age up to the 5^{th} order, a logarithmic function of age, as well as a piece-wise constant. Using the smallest AIC and BIC (Chakrabarti & Ghosh, 2011) coupled with fewer degrees of freedom, a

Table (3.2) Information criterion tables for the different functional forms of the baseline hazard

Model	No. of Parameters	Deviance	Dev.diff. from Constant	AIC	BIC
Constant	1	16176.04	0.00	16178.04	16188.54
Linear	2	14746.39	1429.66	14750.39	14771.39
Quadratic	3	14254.29	492.09	14260.29	14291.79
Cubic	4	14195.84	58.45	14203.84	14245.84
Quartic	5	14180.38	15.46	14190.38	14242.89
Quintic	6	14179.77	0.69	14191.77	14254.78
Logarithmic	2	14302.13	0.00	14306.13	14327.13
Piece-wise constant	25	14167.87	0.00	14217.87	14480.39



(a) Polynomial **(b)** Smoothed

Figure (3.2) Baseline hazards for the modeling age at first marriage

quartic baseline hazard model was preferred. Results are displayed in Table 3.2 and Figure 3.2a. For non-parametric model, splines were used as displayed in Figure 3.2b. All confidence intervals in the thesis were reported at 95% level and p-values < 0.05 were considered statistically significant.

Table (3.3) Results for *Never Married* to *Married* Transition.

	Fixed Effects		Random Effects (Gauss-Hermite)		3 Level Random Effects (Gauss-Hermite)	
	Odds Ratio (SE)	*95% CI*	*Odds Ratio(SE)*	*95% CI*	*Odds Ratio(SE)*	*95% CI*
Age^1	6.94 (3.31)	2.73; 17.66	6.94 (3.31)	2.73;17.66	6.94 (3.38)	2.70; 18.05
Age^2	0.84 (0.05)	0.75; 0.93	0.84 (0.05)	0.75; 0.93	0.84 (0.05)	0.75; 0.93
Age^3	1.01 (0.00)	1.00, 1.01	1.01 (0.00)	1.00, 1.01	1.01 (0.00)	1.00;1.01
Age^4	1.00 (0.00)	1.00; 1.00	1.00 (0.00)	1.00; 1.00	1.00 (0.00)	1.00; 1.00
Gender(Ref:Female)						
No	1.16 (0.12)	0.94 ; 1.42	1.16 (0.12)	0.94 ; 1.42	1.14 (0.13)	0.92; 1.41
Income(Ref:Yes)						
No	0.91 (0.12)	0.71; 1.18	0.91 (0.12)	0.70; 1.20	0.92 (0.12)	0.71; 1.20
Is Employed(Ref:Yes)						
No	1.11 (0.12)	0.90; 1.37	1.11 (0.12)	0.90; 1.37	1.11 (0.12)	0.89; 1.38
Highest Education(Ref:Nev went)						
Primary School	0.88 (0.13)	0.67; 1.17	0.88 (0.13)	0.67; 1.17	0.89 (0.14)	0.66; 1.22
High School	1.00 (0.15)	0.73; 1.35	1.00 (0.15)	0.73; 1.35	1.03 (0.18)	0.73; 1.45
Tertiary	1.07 (0.35)	0.57; 2.01	1.07 (0.35)	0.57; 2.01	1.10 (0.38)	0.55; 2.18
Age at 1^{st} ***Sex***	1.04 (0.01)	1.02; 1.05	1.04 (0.01)	1.02; 1.05	1.04 (0.01)	1.02; 1.05
Var_{ID}	.	.	0.00 (0.00)	0.00; 0.00		
Var_{HID}	.	.	0.00 ()	0.00; 0.00	1.03 (0.35)	0.52; 3.04
AIC	4 264.52		4 264.52		4 254.43	
Likelihood Ratio Test			$\chi^2 = 0.00$		$\chi^2 = 0.00$, p-value$= 0.005$	

$Age^1 - Age^4$ are coefficients of the quartic model

Table 3.3 displays the results from the discrete-time fixed effects survival model, the random effects model with 2-levels and a random effects model with 3-levels, using Gauss-Hermite quadratures with 7 quadrature points where the baseline was modeled using a polynomial of order 4. The multilevel model which considered 2-levels produced a subject-variance of 0 and similar estimates to that of the standard survival model, implying no improvement as a result of the inclusion of a subject-specific random effect. Hence, the standard survival model was a better option, due to parsimony. The model with 3-levels of hierarchy produced estimates of covariates effects that are almost comparable to that with 2-levels and the fixed effects model. In addition to the subject-to-subject variability, the model produced a household random effect ($Var_{HHID} = 1.03; (CI = 0.52 : 3.04)$). The likelihood ratio test showed that including a household random effect improved the model.

Age at first sex was the only statistically significant variable. Subjects who had an early sexual debut were significantly more likely to enter a first marriage at an older age ($OR = 1.04$, $CI = 1.02; 1.05$). Effects of gender, employment and education were not statistically significant on the transition into a first marriage.

3.5 Summary

This chapter presented results from the standard existing methods for analyzing univariate discrete-time survival models with application to data on first marriages. From the three models considered, results were almost comparable, showing that age of sexual debut had a positive effect on the transition into a first marriage in this cohort. In addition, the inclusion of the subject random effect did not show if there is unobserved heterogeneity due to the subjects, but the presence of the household implyied a slight variability among the households. In this chapter, only one type of transition was considered, there could exist different transitions in a stochastic process which may be of interest. The next chapter reviews statistical methods used where not only one transition is possible, but transitions between various states of a multi-state process.

CHAPTER 4

MODELING STATE OCCUPANCY AND TRANSITIONS

4.1 Introduction

This chapter explores the various discrete time-to-event parameterizations that exist in modeling a multi-state process. Three models are reviewed, namely: multinomial model for state occupation in section 4.2, a series of polytomous logistic models for two-way transitions in section 4.3 as well as a competing risks model for exiting a particular state in section 4.4. In section 4.5, a competing risks model with a subject-specific random effect is discussed. Application to the data on marriage formation and dissolution is then done for each model and results are presented in section 4.6.

4.2 State Occupation Model: *Multinomial Logit Model*

In cases where the outcome variable is polytomous (multi-state), a multinomial logit model (Bock, 1970; Chamberlain, 1979) may be used to simultaneously find determinants of the outcome (Agresti, 2003). It basically estimates the log-odds of the outcome of interest versus a base outcome (reference category) and separate odds ratios are then obtained for each of the remaining outcomes.

Denote P_{kt_s} as the probability that subject $k, (k = 1, 2, ..., K)$ occupies state $s(s = 1, \ldots, m)$ at any time interval $t(t = 1, \ldots, \tau)$. The log-odds for state s occupation versus occupation of the reference state (s^*) is then given by

$$log\left(\frac{P_{kt_s}}{P_{kt_{s^*}}}\right) = \beta_{0ks} + \boldsymbol{\beta'}_{ks}\boldsymbol{Z}_k + u_k, \tag{4.1}$$

where β_{0ks} is the intercept which will be assumed constant, u_k is the subject-specific random effect and $\boldsymbol{\beta'}_{ks}$ is the covariates effect for the p covariates which are associated with state s occupation. Let Y_{kt} be a random variable representing the value of

the nominal outcome variable such that

$$Y_{kt} = \begin{cases} s, & \text{if subject } k \text{ is occupying state } s \text{ at time interval } t. \\ s^*, & \text{if subject } k \text{ is occupying state } s^* \text{ at time interval } t \end{cases}.$$

where the state space $s = 1, 2, ..., m$ (usually, $s \in \mathbb{N}$). Solving for $P_{kt_s}(s \neq 1)$ in equation 4.1, the probability may now be written as

$$P_{kt_s} = \frac{\exp\left(\beta_{0ks} + \boldsymbol{\beta'}_{ks}\boldsymbol{Z}_k + u_k\right)}{1 + \sum_{s=2}^{m}\left(\beta_{0ks} + \boldsymbol{\beta'}_{ks}\boldsymbol{Z}_k + u_k\right)}, \tag{4.2}$$

where $s = 1$ is the reference category denoted as s^*.

One potential drawback of the multinomial logit model lies in its requirement of the independence of irrelevant alternatives assumption (IIA). The odds ratio for any pair of outcomes is assumed independent of any third alternative. In other words, elimination of any one of the outcomes should not change the ratios of probabilities for the remaining outcomes. However, Grilli & Rampichini (2007) do not view this as a restrictive feature of the multinomial logit models as this assumption holds conditionally on all the covariates and errors, hence can partially be relaxed by introducing random terms in the linear predictors. Another major shortcoming of the multinomial logit model, which motivates this study, lies in its inability to detect and capture the dynamic nature of a multi-state process. Rather, the investigator has to handle transitions the way it is done in the cross-sectional case. Section 4.3 considers modeling of transitions in a multi-state model.

4.3 State Transition Models:. *Binary Transition Models*

Discrete-time-to event models have been commonly used to model event histories (Fahrmeir & Knorr-Held, 1996, 1997; Fahrmeir, 1997; Steele et al., 1996; Goldstein, 1986, 2011) where the two states are recurrent as previously discussed in Chapter 2. One basic way of modeling transitions in a multi-state process is through fitting separate survival models (Allison, 1982), which were discussed in Chapter 2, for each possible transition in the process as used by Kryscio et al. (2006) through polytomous logistic models. In this section, we adopt this approach in separately modeling age at marital dissolution (through separation or death of a partner) as well as age at remarriage state (after a marital separation or being widowed). One disadvantage of this approach lies in its inability to account for correlation between the separate

transitions, if it exists.

4.4 State Transition Models: *Competing Risks Model*

In a competing risks model, two or more hazards exist that may cause failure (Prentice et al., 1978; Tuma et al., 1979; Han, 1987; Han & Hausman, 1988, 1990). A subject who is in the origin state, i, may fail by making a transition into one of the states $j = 1, 2, ..., m$. Figure 4.1 below, which was extracted from Allignol et al. (2011), shows an example sketch of states considered in a typical competing risk process where $\alpha_{01}(t)$ and $\alpha_{02}(t)$ are representing hazards of transition from state 0 to state 1 and transition from state 0 to state 2, respectively. If there are more than two competing events, then the diagram will have as many arrows as there are number of endpoints. Unlike in survival analysis methods where time to a single endpoint is modeled, a competing risks model has multiple possible end points. Hence, hazards for each transition type will need to be modeled.

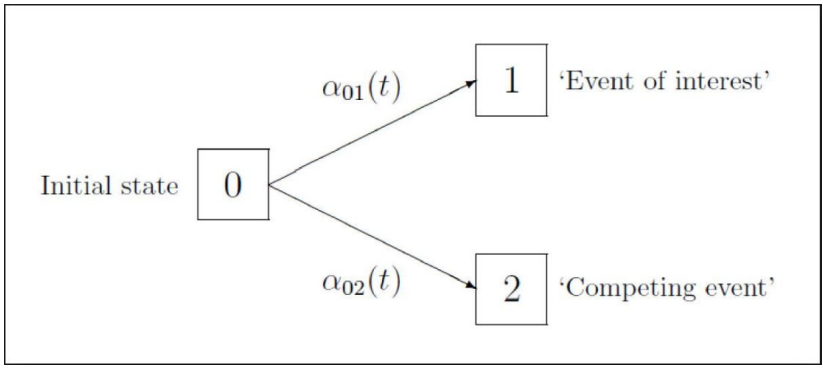


Figure (4.1) A multistate competing risk process

Competing risks models have been developed (Tsiatis, 1998; Sampford, 1952; Prentice et al., 1978; Andersen et al., 2002; Allison, 1982) and are widely used in continuous-time cases where the cumulative incidence technique (Lin, 1997; Gutierrez, 2010) and the Fine and Gray technique (Fine & Gray, 1999) are the most common approaches. In the discrete-time case, Tutz et al. (2016) modeled a cause-specific hazard for competing risks while recently Berger et al. (2018, 2020) modeled a subdistribution hazard for a competing risks model which is an analogue of the Fine and Gray subdistribution hazard of the continuous-time analysis by Fine & Gray (1999). Based on the cause-specific hazard approach, the theory of GLMs is also utilized through the use of multinomial logistic regression (Allison, 1982; Jenkins, 1995; Steele, 2011; Tutz

et al., 2016). Here, the transitions are treated as the outcome.

4.4.1 The Model

Suppose a multi-state process has $m+1$ states, $J \in 0, 1, \ldots, m$, where $j = 1, 2, \ldots, m$ are the distinct target states from the state of origin, i (which is labeled as state 0). The cause specific hazard function which results from failure due to any one of the causes, j, (making a type ij transition) is given by $\lambda_{ij}(t|\boldsymbol{Z})$. For simplicity, the subscript i in the notation is dropped since all transitions are considered to start from state i. Therefore, the hazard is given by

$$\lambda_j(t|\boldsymbol{Z}) = P\left(T = t, J = j | T \geq t, \boldsymbol{Z}\right), \tag{4.3}$$

where $\lambda_1(t|\boldsymbol{Z}), \ldots, \lambda_m(t|\boldsymbol{Z})$ are the m possible hazard functions which may be combined into one overall hazard function which defines the hazard of leaving state i in general, regardless of the destination state, as

$$\lambda(t|\boldsymbol{Z}) = \sum_{j=1}^{m} \lambda_j(t|\boldsymbol{Z}) = P\left(T = t | T \geq t, \boldsymbol{Z}\right). \tag{4.4}$$

Subsequently, the survival function, which is defined as the probability of survival beyond t (not experiencing any event by time interval t), is given by

$$S(t|\boldsymbol{Z}) = P\left(T > t|\boldsymbol{Z}\right) = \prod_{b=1}^{t} \left(1 - \lambda(b|\boldsymbol{Z})\right). \tag{4.5}$$

A subject who reaches interval t will either fail due to one of the m causes or survives beyond t with respective probabilities $\lambda_1(t|\boldsymbol{Z}), \ldots, \lambda_m(t|\boldsymbol{Z}), 1 - \lambda(t|\boldsymbol{Z})$.

4.4.2 Parametric Modeling of Competing Risks

The $m+1$ events may now be modeled using the common multinomial logit regression model (Hartzel et al., 2001; Tutz et al., 2016; Steele, 2005) or any other model for categorical responses (see Agresti (2003)). Here, the hazard function will be modeled as

$$\lambda_j(t|\boldsymbol{Z}) = h_j(\boldsymbol{Z}_t\boldsymbol{\beta}) = \frac{\exp\left(\beta_{0tj} + \boldsymbol{Z}'\boldsymbol{\beta}_j\right)}{1 + \sum_{b=1}^{m} \exp\left(\beta_{0tb} + \boldsymbol{Z}'\boldsymbol{\beta}_b\right)}, \tag{4.6}$$

where β_{0tj} is the baseline hazard function $\forall t$ and b in the denominator represents ib, which are all the other possible transition types except ij. We consider the log-odds

of experiencing an ij type of transition versus no transition (staying in state i) using the multinomial logit for the $m+1$ categories as

$$\log\left(\frac{\lambda_j(t|\boldsymbol{Z})}{\lambda_0(t|\boldsymbol{Z})}\right) = \beta_{0tj} + \boldsymbol{Z}'\boldsymbol{\beta}_j. \tag{4.7}$$

This modeling framework allows us to interpret the coefficients, β_j, in terms of odds ratios. It will give us the odds of failing due to cause j versus survival (staying in the state origin). So J equations will be estimated using the hazard function for the J transitions. Each transition type is allowed to have its own form of the baseline hazard. The multinomial logit provides a good approach to estimating a competing risk model. It treats the response variable as a polytomous qualitative choice variable.

4.4.3 The Likelihood Function

The contribution to the likelihood for subject k towards the j^{th} transition under the assumption of random censoring is

$$L_{jk} = \lambda_{j_k}(t_k|\boldsymbol{Z}_k)^{\delta_k}\left(1-\lambda_{j_k}(t_k|\boldsymbol{Z}_k)\right)^{1-\delta_k}\prod_{t=1}^{t_k-1}\left(1-\lambda(t|\boldsymbol{Z}_k)\right), \tag{4.8}$$

where

$$\delta_k = \begin{cases} 0, & \text{if subject } k \text{ is still in state } i \text{ at interval } t\ . \\ 1, & \text{if subject } k \text{ has made a transition from state } i \text{ to } j \text{ at interval } t. \end{cases} \tag{4.9}$$

Another binary indicator, ε_{ktj}, may be defined such that

$$\varepsilon_{ktj} = \begin{cases} 0, & \text{if subject } k \text{ is not occupying state } j \text{ at interval } t \\ 1, & \text{if subject } k \text{ is occupying state } j \text{ at interval } t \end{cases},$$

we define for each observation in each interval $t < t_k$

$$\boldsymbol{\varepsilon}'_{kt} = \varepsilon_{kt0}, \varepsilon_{kt1}, \ldots, \varepsilon_{ktm} = (1, 0, \ldots, 0)$$

to indicate that the subject is still occupying state i i.e. survival in all intervals before t_k. However, for interval $t = t_k$, where $\delta_k = 0$, we define

$$\boldsymbol{\varepsilon}'_{kt_k} = \varepsilon_{kt_k0}, \varepsilon_{kt_k1}\ldots_{kt_km} = (1, 0, \ldots, 0),$$

if subject k is censored. Otherwise, for interval $t = t_k$, where $\delta_k = 1$, we define

$$\boldsymbol{\varepsilon}'_{kt_k} = \varepsilon_{kt_k0}, \varepsilon_{kt_k1} \ldots, \varepsilon_{kt_km} = (0, \ldots, 1, \ldots, 0),$$

so that $\varepsilon_{kt_kj_k} = 1$ corresponds to the state that subject k is now occupying. These binary indicators enable the likelihood function for subject k to be rewritten in the GLM framework as

$$\begin{aligned} L_k &= \prod_{t=1}^{t_k} \left(\prod_{j=1}^{m} \lambda_j(t|\boldsymbol{Z}_k)^{\varepsilon_{ktj}} \right) (1 - \lambda(t|\boldsymbol{Z}_k))^{\varepsilon_{kt0}} \\ &= \prod_{t=1}^{t_k} \left[\left(\prod_{j=1}^{m} \lambda_j(t|\boldsymbol{Z}_k)^{\varepsilon_{ktj}} \right) \left(1 - \sum_{j=1}^{m} \lambda_j(t|\boldsymbol{Z}_k) \right)^{\varepsilon_{kt0}} \right], \end{aligned} \tag{4.10}$$

Note that the likelihood for subject k in equation 4.10 is similar to that one for the t_k observations $\boldsymbol{\varepsilon}_{k1}, \ldots, \boldsymbol{\varepsilon}_{kt_k}$ of a multinomial response model where the indicator variables represent the distributions, given that a particular interval has been reached. Suppose subject k reaches interval t, the response has a multinomial distribution (Czepiel, 2002; Tutz et al., 2016) with

$$\boldsymbol{\varepsilon}'_{kt} = \varepsilon_{kt0}, \varepsilon_{kt1}....., \varepsilon_{ktm} =\sim M\left(1, 1 - \lambda(t|\boldsymbol{Z}_k), \lambda_1(t|\boldsymbol{Z}_k), ..., \lambda_m(t|\boldsymbol{Z}_k)\right).$$

The dummy variable $\varepsilon_{kt0} = 1 - \varepsilon_{kt1} - \ldots - \varepsilon_{ktm}$ assumes a value of 1 if subject k does not experience the ij transition in interval t and 0 if the ij transition has been made in interval t. This means that the likelihood is similar to that of a multi-categorical model for the probability of an ij transition occurring, $P(\boldsymbol{\varepsilon}_{kt} = j) = h_j(\boldsymbol{Z}_t\boldsymbol{\beta})$, where $\boldsymbol{\varepsilon}_{kt} = j$ if $\boldsymbol{\varepsilon}_{ktj} = 1$. The calculation of MLEs may be made augmenting the design matrices within the framework of multivariate generalized linear models (GLM),

$$\begin{pmatrix} \boldsymbol{\varepsilon}_{k1} \\ \vdots \\ \boldsymbol{\varepsilon}_{kt_k} \end{pmatrix}, \begin{pmatrix} \boldsymbol{Z}_1 \\ \vdots \\ \boldsymbol{Z}_{t_k} \end{pmatrix},$$

just as when modeling a single cause discrete survival. With this, the complete log-

likelihood function is then given by

$$\ell = \sum_{k=1}^{K} \sum_{t=1}^{t_k} \left(\sum_{j=1}^{m} \varepsilon_{ktj} \log \lambda_j(t|\boldsymbol{Z}_k) + \varepsilon_{kt0} \log \left(1 - \sum_{j=1}^{m} \lambda_j(t|\boldsymbol{Z}_k) \right) \right) \tag{4.11}$$

and may be maximized to obtain the parameter estimates. These methods are easily implemented using standard software modeling GLM's (Tutz et al., 2016; Molenberghs & Verbeke, 2005, 2000). The likelihood may be maximized to obtain the MLEs for the parameters.

4.4.4 Alternative Approaches to Modeling Competing Risks

This approach for discrete-time competing risk is referred to as the discrete cause-specific competing risks (Mills, 2011b; Steele, 2005). Recently, Berger et al. (2020) proposed a technique for modeling the discrete-time subdistribution hazard in competing risks using weighted maximum likelihood estimation as

$$\lambda_r(t|\boldsymbol{Z}_k) = P\left(T_k = t, \xi = r | (T_k \geq t) \cup (T_k \leq t-1, \xi \neq r), \boldsymbol{Z}_k\right), \tag{4.12}$$

which is an analogue to Fine & Gray (1999). It is defined as the probability of event of type r occurring in interval t given that either no event has occurred or if it occurred, it was an event other than type r. The hazard for only one type of transition, out of the many possible transitions, was modeled, hence the name subdistribution hazard. Therefore, only one model was considered for interpretation, with focus on one specific event, which is easier to implement and interpret. This approach, however, has a drawback in that it does not provide insight into the characteristics of the cause-specific hazard functions as it only considers one type of transition in the presence of other possible kinds of transitions.

Other discrete-time approaches to the competing risks model have been reviewed by Schmid & Berger (2020). These include non-parametric models using trees (random forests) in cases where the assumption that predictors are linear functions of covariates is violated. A popular tree method is the Classification and Regression Trees (CART) where the covariate space is sequentially subdivided into a set of disjoint rectangles (Breiman et al., 2017; Ma, 2018). Simple models are then fitted in each rectangle and the models are then combined to produce non-parametric estimates of the sub-distribution hazard or the cause-specific hazards. A CART algorithm starting with a *root node* which represents the whole covariate space is used to determine

the best partition of the covariate space. A tree is then grown from the root node in a hierarchical manner thereby generating *child nodes*. Essentially, this is achieved by recursively splitting the covariate space into two smaller subsets. Now, in each split, a single covariate is chosen to define the split rule. The tree algorithms for discrete competing risks data will conveniently be based on CART approaches for multi-categorical or binary outcomes. Machine learning techniques are advancing and maybe be used (see more details in Schmid & Berger (2020)).

4.5 Competing Risks Model for Recurrent Events.

Most approaches to modeling competing risks events are based on transitions into absorbing competing states. When the states are transient, modification to the model needs to be made to account for variability within subjects. Masyn (2009) modeled a survival factor mixture model for low-frequency recurrent event histories. Muenz & Rubinstein (1985) also modeled covariance dependence on binary variables using MLE of the parameters while allowing for non-stationary or second order Markov chains. A more favorable approach was used by Regier (1968) where the 2-state transition matrix was re-parameterized to include the odds ratio of staying in one state versus that of staying in the other, as a parameter to allow for detection of tendencies to migrate to one state.

Here, the approach by Tutz et al. (2016) will be used to incorporate the subject-specific random effects which will quantify the variability due to subjects on each transition (Hedeker, 2003; Hartzel et al., 2001). The hazard function in 4.3 may now be modified to be defined as the probability of subject $k(k = 1, \ldots, K)$ failing due to cause $j(j = 1, \ldots, mjz)$ the l^{th} $(l = 1, \ldots, l_{kj})$ time, as follows:

$$\lambda_j^{(l)}(t|\boldsymbol{Z}_{kl}, u_{kj}) = P\left(T_{kl} = t, J = j | T_{kl} \geq t, \boldsymbol{Z}_{kl}, u_{kj}\right) = h(\beta_{0jt} + \boldsymbol{Z}'_{kl}\boldsymbol{\beta} + u_{kj}), \quad (4.13)$$

where $u_{kj} \sim N(0, \sigma_{u_{kj}})$ are the random effects associated with the occurrence of the j^{th} event such that

$$\boldsymbol{u} = \begin{pmatrix} u_{k1} \\ \vdots \\ u_{kJ} \end{pmatrix} \sim MVN(\mathbf{0}, \boldsymbol{\Omega}_u), \quad (4.14)$$

and are assumed independent, thus allowing for shared unobserved risk factors.

4.5.1 Likelihood Construction

To model the hazard, a multinomial logit provides a good approach to estimating a competing risk model as previously described in section 4.4. In the GLM framework, the response variable for subject i in episode l of event j during time interval t of the episode is

$$\varepsilon_{kltj} = \begin{cases} 1, & \text{if subject } k \text{ fails in episode } l, \text{ due to cause } j \text{ at interval } t \ . \\ 0, & \text{otherwise.} \end{cases}$$

The contribution of subject k to the likelihood is

$$L_k = \prod_{l=1}^{l_{kj}} \prod_{t=1}^{t_{kj}-(1-\sigma_{kjl})} \left(\prod_{j=1}^{J} \lambda_j(t|\boldsymbol{Z}_k)^{\varepsilon_{kltj}} \right) [1 - \lambda(t|\boldsymbol{Z}_k)]^{\varepsilon_{klt0}} \times f(\boldsymbol{u}), \tag{4.15}$$

where $\delta_{kjl} = 1$ for $l = 1, \ldots, l_{kj} - 1$ and $\delta_{kjl} = 0$ for the last episode.

4.5.2 Estimation with Statistical Software for Regression

With this representation of multinomial variables, the model may now be estimated in the same way as a multinomial mixed effects model for repeated measurements where, in this case, repeated measurements are the separate episodes and time points. Standard software for multinomial distributions with random effects may be used. These include, but are not limited to **proc LOGISTIC** in SAS, with a **random intercept** statement, **Xtmlogit** in STATA or the **mclogit** in R.

4.6 Results

We consider a 4-state process with two communicating classes, some of which have recurrent states. The process to be considered is represented in Figure 3.1 which was discussed earlier. For each year, subjects are followed up, they may be in one of the marital states: *Never married*, *Married, Separated* or *Widowed*. As shown in Figure 3.1, subjects would move between the four marital states and the possible transition types to be considered are entry into a first-marriage (*Never Married* $\rightarrow$ *Married*), exiting a marriage which will be used for the competing risks model (*Married*$\rightarrow$*Separated* and *Married*$\rightarrow$*Widowed*), and remarriage (*Separated*$\rightarrow$*Married* and *Widowed*$\rightarrow$*Married*).

4.6.1 Descriptive Statistics

Table 4.1 below shows the baseline distribution of some variables of interest where applicable. In all categories of the variables, never married subjects had the highest proportion. At the beginning of episodes of the study, 21.05% of the female participants were married and 15.07% of the males were married. Most participants did not earn an income and of those who did, 27.28% were married and 64.58% were never married. The number of transitions (in person-years) into the different marital states are represented in Table 4.2 below, where most subjects remained in the *Never Married* state and the most common transition type was the *Never Married* to *Married* transition with $1\ 225$ transitions followed by the *Married* to *Widowed* with $1\ 149$ transitions.

Figure 4.2 displays the proportion at each age of subjects who were occupying each state. It is clear that below 46 years of age, the biggest part of the population was constituted by subjects who were never married. Figure 4.3 also displays the trends over time for marital state occupation. Over the whole study period, the proportion of never married subjects was the highest. It decreased until 2008 and then stabilized afterwards. Proportion of married subjects was considerably low but slightly increased with time, then declined after 2009. The rates of marital separation were almost constant over time while those of widowhood started to increase slightly after 2008.

Table (4.1) Distribution of explanatory variables at study entry by age

Variable	Total	Marital Status			
		Never Married *N* (%)	Married *N* (%)	Separated *N* (%)	Widowed *N* (%)
Gender					
Female	24 637	17 130 (69.53)	5 185 (21.05)	117 (0.47)	2 205 (8.95)
Male	15 717	13 145 (83.64)	2 368 (15.07)	43 (0.27)	161 (1.02)
Income Is Earned					
Yes	4 322	2 791 (64.58)	1 179 (27.28)	26 (0.60)	326 (7.54)
No	33 231	24 742 (74.45)	6 324 (19.03)	134 (0.40)	2 031 (6.11)
Is Employed					
Yes	14 220	10 526 (74.02)	2 948 (20.73)	79 (0.56)	667 (4.69)
No	24 459	18 314 (74.88)	4 398 (17.98)	76 (0.31)	1 671 (6.83)
Education					
Nev Went to Sch	567	198 (34.92)	207 (36.51)	4 (0.71)	158 (27.87)
Primary	3 225	1 884 (58.42)	835(25.89)	16 (0.50)	490 (15.19)
High School	3 614	2 634 (72.88)	727 (20.12)	24 (0.66)	229 (6.34)
Tertiary	245	86 (35.10)	123 (50.20)	2 (0.82))	34 (13.88)

Table (4.2) Number of transitions in the entire study period

Previous Marital Status	Current Marital Status				
	Never Married	Married	Separated	Widowed	Total
Never Married	242 486	1 225	0	0	243 711
Married	0	47 471	61	1 149	48 681
Separated	0	47	600	0	647
Widowed	0	372	0	9 029	9 401
Total	242 486	49 115	611	10 178	302 440

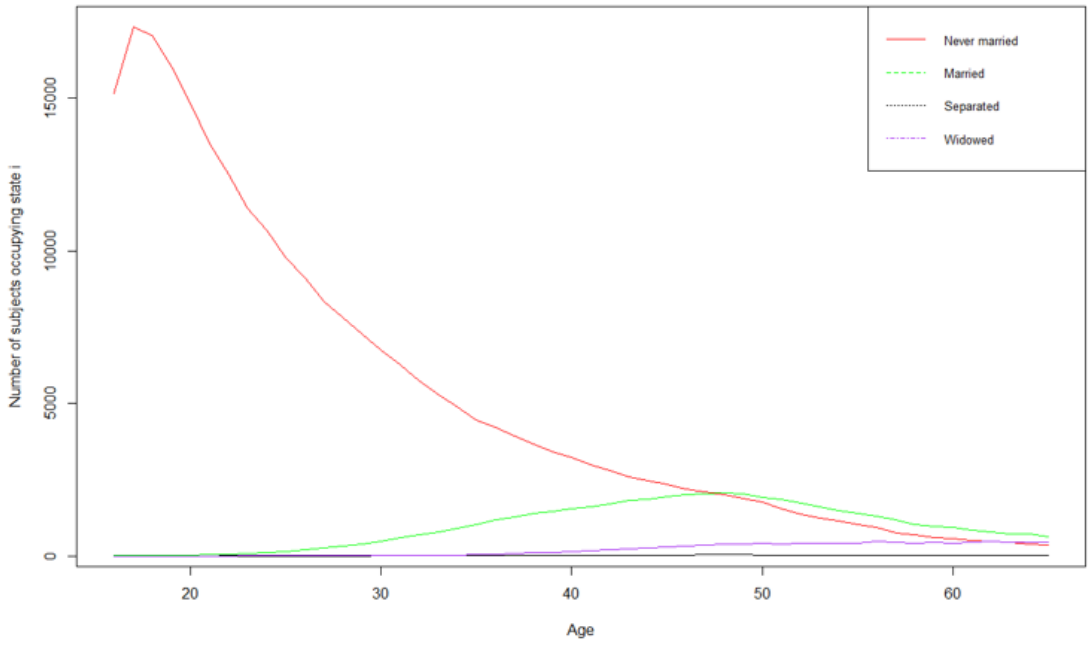


Figure (4.2) Distribution of subjects in each state

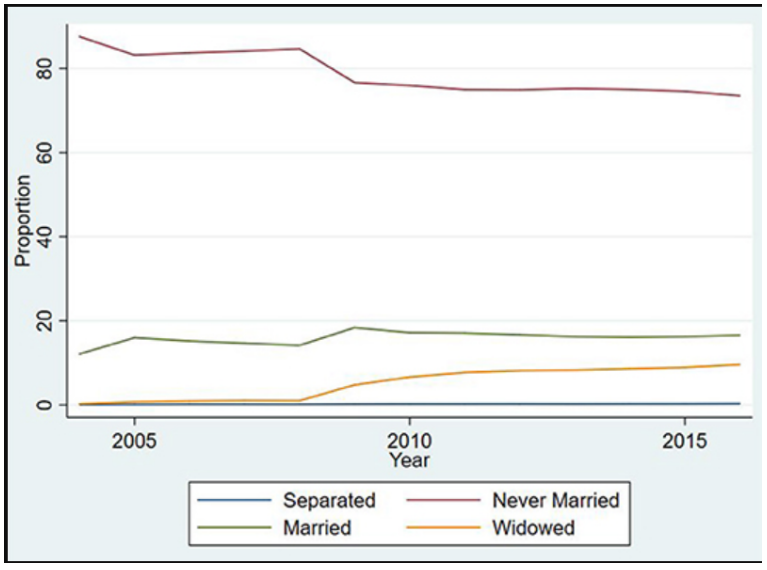


Figure (4.3) Trends for various marital state occupations over time

Table (4.3) Results for the marital state occupations for subjects aged between 17 and 65 years from $2004 - 2016$ for the 56 308 subjects

.

Covariates	**Married** vs **N.Married** OR (95% *CI*)	**Separated** vs **N.Married** OR (95% *CI*)	**Widowed** vs **N.Married** OR (95% *CI*).
Constant	0.01(0.01; 0.01)	0.00 (0.00; 0.00)	0.00 (0.00; 0.00)
Age Interval	1.28 (1.27; 1.29)	1.35 (1.27; 1.44)	1.50 (1.46; 1.53)
Gender(Ref:Female)			
Male	0.84 (0.76; 0.93)	0.42 (0.24; 0.73)	0.09 (0.07; 0.11)
Income(Ref:Yes)			
No	0.93 (0.87; 1.01)	0.96 (0.70; 1.32)	1.03 (0.92; 1.16)
Is Employed(Ref:Yes)			
No	0.93 (0.86; 0.99)	0.52 (0.35; 0.78)	1.08 (0.97; 1.21)
Highest Education(Ref:Never went to Sch)			
Primary	1.68 (1.43; 1.98)	1.62 (0.59; 4.46)	1.34 (1.11; 1.62)
High School	2.42 (2.04; 2.86)	4.50 (1.52; 13.34)	1.62 (1.29; 2.02)
Tertiary	6.43 (4.89; 8.47)	14.86 (4.14; 53.31)	2.96 (1.96; 4.46)
Age at First Sex	1.00 (0.99; 1.01)	1.01 (0.98; 1.03)	1.00 (0.99; 1.01)
σ_i	0.22 (0.21; 0.23)	0.09 (0.08; 0.10)	0.43 (0.23; 0. 63)

4.6.2 Results from the Models

Results for Marital State Occupation

From Table 4.3, results of the multinomial models show that the likelihood of being married, compared to never married, significantly increased with age ($OR = 1.28; CI = (1.27 : 1.29)$) and with higher levels of education ($OR = 1.68, CI = (1.43 : 1.98), OR = 2.42; CI = (2.04 : 2.86)$ and $OR = 6.43; CI = (4.89 : 8.47)$,respectively), but was lower for males ($OR = 0.84; CI = (0.76 : 0.93)$), unemployed subjects ($OR = 0.93; CI = (0.86 : 0.99)$) and those with no income ($OR = 0.93; CI = (0.87 : 1.01)$). The same direction of odds ratios was found when comparing the likelihood of separation and widowhood with never married. There existed some subject-to-subject variation in the likelihood of all state occupations when compared to staying in the $Never\ Married$ state.

Results for Binary Marital State Transition Models

Results of the separate binary transitions with subject-specific random effects are displayed in Table 4.5, except those of entering a first time marriage which were already covered in Chapter 3. Using the smallest AIC, they all had a polynomial baseline effect of age as indicated in Table 4.4.

Table (4.4) Information criterion tables for the different functional forms of hazard for the binary transition models

Baseline	*Marr → Sep*	*Marr → Wid*	*Sep → Marr*	*Wid → Marr*
Constant	**940.4**	10 956.3	347.8	3 221.4
Linear	941.5	10 770.2	**347.0**	**3195.6**
Quadratic	943.4	10 761.9	348.7	3 197.5
Cubic	945.1	**10756.7**	348.5	3 198.2
Quartic	945.2	10 757.2	348.2	3 198.7
Quintic	947.1	10 759.2	347.9	3 200.7
Logarithmic	941.5	10 759.0	347.2	3 197.1
Piece-wise	947.5	14 728.1	354.0	5 863.9

Inclusion of covariates improved the models and a random effect for a $Widowed \rightarrow Married$ transition had a variance of 0. The random effects for the $Married \rightarrow Separated$ and $Married \rightarrow Widowed$ transitions are statistically significant as their confidence intervals do not include 0's ($Var_{MS} = 7.20; CI = (2.18 : 23.74)$ and ($Var_{MW} = 1.80; CI = (1.20 : 2.70)$), respectively). This implies that there could be unobserved heterogeneity due to other factors associated with these transitions which were not captured by the models. Older aged individuals were significantly

more associated with marital separation ($OR = 1.05; CI = (1.01 : 1.09)$) and remarrying after death of a partner ($OR = 1.06; CI = (1.01 : 1.10)$) but had a significant negative effect on the $Married \rightarrow Widowed$ transition ($OR = 0.92; CI = (0.88 : 0.97)$). Unemployment had a negative effect on marital separation ($OR = 0.34; CI = (0.14; 0.83)$) Age at first sex also played a significant role in the transitions where it has a slight positive effect on the odds of the $Married \rightarrow Widowed$ transition ($OR = 1.01; CI = (1.00 : 1.03)$) and the $Widowed \rightarrow Married$ transition ($OR = 1.07; CI = (1.09 : 1.09)$).

Table (4.5) Results for pairwise transitions for the different family dynamics.

	Married $\longrightarrow$ **Separated**		**Married** $\longrightarrow$ **Widowed**		**Separated** $\longrightarrow$ **Married**		Widowed $\longrightarrow$ **Married**	
	OR (SE)	*95% CI*	*OR (SE)*	*95% CI*	*OR (SE)*	*95% CI*	*OR (SE)*	*95% CI*
Constant	0.00 (0.00)	0.00; 0.00	0.00 (0.00)	0.00; 0.00	0.03 (0.05)	0.00; 0.83	0.07 (0.04)	0.21; 0.19
Age Interval			0.92	(0.88; 0.97)	1.05	(1.01; 1.09)	1.06	(1.01; 1.10)
Age interval^2			0.051	(0.0428; 0.0607)				
textit**Age interval**^{3}			1.729	(0.134; 22.240)				
Gender(Ref:Female)								
Male	1.15 (0.55)	0.44; 2.96	0.12 (0.02)	0.08; 0.17	0.82 (0.55)	0.22; 3.07	1.27 (0.41)	0.68; 2.39
Income(Ref:Yes)								
No	0.90 (0.46)	0.33; 2.48	1.09 (0.15)	0.84; 1.42	0.45 (0.30)	0.12; 1.65	0.84(0.17)	0.57; 1.24
Is Employed(Ref:Yes)								
No	0.34 (0.16)	0.14; 0.83	0.98 (0.11)	0.79; 1.21	1.33 (0.86)	0.37; 4.71	1.40 (0.25)	0.99;1.98
Highest Educ(Ref: Never Went Sch)								
Primary	3.13 (3.65)	0.32; 0.64	0.93 (0.14)	0.70; 1.25	1.66 (1.02)	0.38; 7.27	1.18 (0.25)	0.79; 1.78
High School	3.47 (1.07)	0.35; 34.43	0.65 (0.11)	0.47; 0.91	0.48 (0.43)	0.08; 2.76	0.92 (0.23)	0.57; 1.50
Tertiary	0.87 (1.39)	0.04; 19.85	0.34 (0.11)	0.18; 0.63	0.00 (0.00)	0.18; 0.63	0.99 (0.47)	0.39; 2.50
Age at First Sex	1.04 (0.03)	0.98; 1.10	1.01 (0.01)	1.00; 1.03	1.02 (0.05)	0.93; 1.13	1.07 (0.01)	1.05; 1.09
Var_k	7.20 (4.38)	2.18; 23.74	1.80 (0.37)	1.20; 2.70	0.22 (1.08)	0.00; 3474.70	0.00 (0.00)	0.00; 0.00
AIC	420.20		5204.42		138.67		1680.57	

Results for Termination of a Marriage

Results of the fixed effects competing risks model are displayed in Table 4.6 where the random effects are assumed to be independent of each other (AIC $= 5682.75$). The direction of the estimate are similar to those produced by binary transition models, although the magnitude of the estimated differ slightly. Unemployed subjects had significantly lower odds of exiting a marriage through separation than the employed subjects ($OR = 037; CI = (0.17 : 0.84)$). For the $Married \rightarrow Widowed$ transitions, males had significantly lower odds ($OR = 016; CI = (0.12 : 0.22)$) and age at first sexual debut had a slight positive effect ($OR = 1.02; CI = (1.00 : 1.03)$) on the $Married \rightarrow Widowed$ transitions. Education had a negative effect on the $Married \rightarrow Widowed$ transitions (($OR = 0.90; CI = (0.72 : 1.13)$), ($OR = 0.63; CI = (0.49 : 0.82)$) and ($OR = 0.37; CI = (0.22 : 0.63)$)) for primary, high school and tertiary education, respectively.

The random effects competing risks model produced results that are displayed in Table 4.7. Here, substantial unobserved heterogeneity among subjects was noticed for both events as both variances were statistically significant ($Var_{MS} = 7.46; (CI = 2.28 : 24.38)$ and $Var_{MW} = 1.80; (CI = 1.21 : 2.70)$) and the values of the variances of the random effects did not deviate much from the ones obtained in binary transitions models. Hence, the inclusion of the subject-specific random effect improved the fit of the model, which is noticed in the lower AIC ($= 5\ 628.97$) compared to that of the fixed effects competing risks model (AIC$= 5\ 682.75$), which makes it a more preferable model. The model resulted in standard errors that are comparable to those from a fixed effects model. Therefore, reporting will not be repeated in this section.

Table (4.6) Fixed effects competing risks: Results for exiting a marriage (AIC $= 5682.75$)

	Married $\longrightarrow$ **Separated**		**Married** $\longrightarrow$ **Widowed**	
	OR (SE)	*95% CI*	*OR (SE)*	*95% CI*
Gender(Ref:Female)				
Male	1.05 (0.44)	0.46; 2.36	0.16 (0.03)	0.12; 0.22
Income(Ref:Yes)				
No	0.92 (0.44)	0.37; 2.33	1.04 (0.13)	0.82; 1.32
Is Employed(Ref:Yes)				
No	0.37 (0.15)	0.17; 0.84	0.98 (0.09)	0.81; 1.18
Highest Education(Ref:Never went to sch)				
Primary	2.99 (3.11)	0.39; 23.03	0.90 (0.10)	0.72; 1.13
High School	3.03 (3.20)	0.38; 24.11	063 (0.86)	0.49; 0.82
Tertiary	0.86(1.2)	0.05; 14.43	0.37 (0.10)	0.22; 0.63
Age at First Sex	1.04 (0.03)	0.99; 1.10	1.02 (0.01)	1.00; 1.03

Table (4.7) Random effects competing risks: Results for exiting a marriage (AIC $= 5628.97$)

	Married $\longrightarrow$ **Separated**		**Married** $\longrightarrow$ **Widowed**	
	OR (SE)	*95% CI*	*OR (SE)*	*95% CI*
Gender(Ref:Female)				
Male	1.14 (0.56)	0.43; 3.00	0.012 (0.02)	0.08; 0.17
Income(Ref:Yes)				
No	0.91 (0.46)	0.33; 2.47	1.08 (0.15)	0.83; 1.41
Is Employed(Ref:Yes)				
No	0.33 (0.16)	0.14; 0.83	0.97 (0.10)	0.78; 1.20
Highest Education(Ref:Never went to sch)				
Primary	3.16 (3.70)	0.31; 31.37	0.95 (0.14)	0.71; 1.27
High School	3.54 (4.23)	0.34; 36.91	0.65 (0.11)	0.46; 0.90
Tertiary	0.88 (1.42)	0.04; 20.81	0.34 (0.11)	0.18; 0.64
Age at First Sex	1.04 (0.03)	0.99; 1.11	1.01 (0.01)	1.00;1.02
$\boldsymbol{Var_k}$	7.46 (4.51)	2.28; 24.38	1.80 (0.37)	1.21; 2.70

4.7 Summary

This chapter reviewed a multinomial model for state occupation, state transition model using a series of discrete-time survival models and a discrete-time competing risks model for transitions out of an origin state. These were executed while incorporating subject-specific random effects to capture the variability due to subjects. The transition models that were considered were done by separating the different transitions in the process. Additionally, the random effects were assumed independent. As a result, the dependencies between the different transitions in the process were not measured. The next chapter proposes a single discrete-time multi-state model which, in addition to modeling covariate effects, will model the correlation between the transitions in a multi-state process.

CHAPTER 5

MULTIVARIATE DISCRETE TIME-TO-EVENT TRANSITIONS

5.1 Introduction

This chapter seeks to construct a multivariate discrete-time multi-state model where the transitions may occur in any direction, with possible recurrent events, while incorporating covariates. Specifically, a novel joint modeling of discrete-time survival models is constructed, which is derived from the existing literature. Section 5.2 provides a review of multi-state models and section 5.3 describes the existing methodology for multi-state models. In section 5.4, a multivariate model is proposed whose baseline hazard form is described in section 5.5 and likelihood function is constructed in section 5.6. Methods of approximating the likelihood are discussed in section 5.7. A simulation study was then done in section 5.10 to evaluate performance of the model and demonstration of the usefulness of the model through an empirical analysis and the results are presented in section 5.11. The proposed method is a novel approach to examine a common research question in the literature for family formation and dissolution and model diagnosis is presented in section 5.12.

5.2 Discrete-Time Multi-State Models

Life histories analysis using multi-state models has been extensively used in biostatistics, demography and economics (Helbert, 2015; Steele et al., 2004; Willekens, 2014; Aalen et al., 2008, and the references therein) and may be used to inform how subjects evolve through various states in a stochastic process. Multi-state models are important in providing the general trajectories through intermediate states (Matsena Zingoni et al., 2019). A common case scenario is observed where time is treated as continuous, such that the exact time of event occurrence is known, since it is modeled at every instant in time (Jackson, 2007; Willekens et al., 1982; Mills, 2011b; Meira-Machado et al., 2009; Putter et al., 2007; Andersen & Keiding, 2002; Reddy et al., 2011; Blossfeld, 2012; Blossfeld et al., 2007; Matsena Zingoni et al., 2019; Sutradhar et al., 2010, among others). Willekens (2014) gives a summary of commonly used statis-

tical methods (with their modifications) and packages that have been developed to analyze such continuous data viz; **mstate** (De Wreede et al., 2010; de Wreede et al., 2011; Putter et al., 2007) for multistate models using Coxph models, the **msm** (Jackson et al., 2011; Therneau et al., 2018) for Markov models for panel data and the **mSurv** for multistate models with non-parametric estimation (Ferguson et al., 2012). Other methods including marginal regression models by Anderson-Gill model, such as the Prentice, Williams, Peterson Total Time (PWP-TT) recurrent event extensions of the Cox model and PWP-GT) in clinical setting are often used(Andersen & Gill, 1982) to account for the effect of order of events by adjusting the risk of subsequent events on the basis of the number of previous events.

Limited research has been conducted on how to use panel data to estimate parameters for a discrete-time multistate model (for example Barbu et al. (2018) who simulated data for a discrete-time semi-markov model). Spedicato et al. (2016) and Nicholson (2013) developed methods which handle a discrete-time Markov model with maximum likelihood estimation, although they did not incorporate covariate effects in their models. It is often important to include and analyze the covariate effects associated with each transition type in order to distinguish the effect of various demographic, clinical or socio-economic factors on transitions (Jackson, 2007).

In Chapter 4, analysis of a multistate process was done by considering separate binary discrete-time survival models for each transition type, which have a disadvantage of not allowing for the simultaneous analysis of transitions between states and adjustments for intermediate events (Eulenburg et al., 2016), hence cannot be used to directly elucidate pathways of association across multiple states. Competing risks models were also considered according to Tutz et al. (2016), which can only determine odds of transition (hence transition probabilities) as well as covariates effects. These competing risks models were further extended to accommodate the recurring nature of events by including a subject-specific random effect for each transition type, since repeated events occur within the same subjects. All these models, however, were not able to capture and quantify the possible correlation between the transitions themselves. This is necessary as the occurrence of one transition might have an effect on the occurrence of another. For example, early entrance into a first marriage might be associated with an early marital separation. A more advanced model is required to account for such dependencies and will be derived in this chapter.

5.3 Theoretical Framework for a Multistate Model

Let T_k be the event time for subject k ($k = 1, 2, ..., K$). Then T_k is a random variable which takes on τ discrete values such that $T_k = \{0, 1, ..., t, ...\tau_k\}$, where $\tau_k \in \mathbb{N}$. In cases where data are originally continuous, the timescale, $0 = a_0 < a_1 < ...., < a_{\tau-1} < \infty$, is partitioned into τ_k intervals such that $I_t = [a_t; a_{t+1})$ represents the discrete time intervals which can be monthly, yearly, 2-yearly or so on. The intervals do not necessarily need to be of equal length. $T_k = t$ implies that the event for subject k has occurred in time interval $I_t = [a_t, a_{t+1})$ with $a_\tau = \infty$. A discrete-time multistate process is then defined as a stochastic process, $(X_t, t \in T_k)$, with a finite state space, $S = \{1, 2, ..., m\}$ (usually, $S \in \mathbb{N}$). It differs from a continuous-time process whose time parameter is allowed to assume any non-negative real number, $t \geq 0$ (Doytchinov & Irby, 2010). As the discrete-time multistate process evolves over time, a history, $\mathbf{H_{s^-}}$, of observations is generated and may provide information regarding all the states that were previously visited by subjects. For a subject who is in state i at discrete time t to move to state j in the next discrete time $t + 1$, we define a one-step transition probability as

$$p_{ij} = P(X_{t+1} = j | X_t = i, \mathbf{H_{s^-}}), \tag{5.1}$$

where $i, j \in S$.
Various assumptions may be made regarding the dependence of the probability on history or time. For example, a Markov model assumes that the probability of transition into the future only depends on the current state, hence history is not important (Markov, 1954). As a result, equation 5.1 may equivalently be rewritten as:

$$p_{ij} = P(X_{t+1} = j | X_t = i, \mathbf{H_{s^-}}) = P(X_{t+1} = j | X_t = i) \tag{5.2}$$

or even further as

$$p_{ij} = P(X_1 = j | X_0 = i).$$

The probability of moving from state i to state j in n steps is then denoted by

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i).$$

The process, X_t, is said to be time-homogeneous if

$$P(X_{t+1} = j | X_t = i) = P(X_1 = j | X_0 = i) \quad \forall i, j \in S.$$

This means that the transition probabilities do not change with time. As a consequence of the Markov property, these transition probabilities satisfy the Chapman-Kolmogorov equations as follows:

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} * p_{kj}^{(n)}$$

$\forall m, n$ with $i, k, j \in S$.
For all the transition types in a multi-state process, the transition probability in equation 5.1 will form the ij^{th} element of the one-step square transition matrix,

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0n} \\ p_{10} & p_{11} & \cdots & p_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n0} & p_{n1} & \cdots & p_{nn} \end{pmatrix}.$$

Since X_t is a stochastic process, we note that when the process leaves state i, it must move to one of the states, $j \in S$, which implies that each row of the transition matrix sums to one i.e. $\sum_{j \in S} p_{ij} = 1, \quad \forall i \in S$. Any model summary (such as sojourn times, expected duration times in a state and lifetime risk (i.e.*P(ever visiting a particular state*)) and steady states) will be a function of the resulting one-step transition matrix since it describes progression (Craig & Sendi, 2002).

Relationship Between the Discrete- and Continuous-time Multi-state Models

Unlike in discrete-time multi-state modeling (DTMSM), continuous-time multi-state modeling (CTMSM) considers the infinitesimal time so that the transition intensity, $\lambda_{ij}(t)$, will represent the instantaneous probability of transitioning from state i at time t to state j in the next infinitesimal possible time $s = t + \delta t$. These transition intensities fully characterize the transition probabilities between times (say t and s), $p_{ij}(s, t)$, which are elements of the transition matrix $\mathbf{P}(s, t)$ (Meira-Machado et al., 2009; Jackson, 2007). In a time-homogeneous model, $\mathbf{P}(s, t) = \mathbf{P}(t - s)$. The DTMSM, on the other hand, is characterized by transition probabilities themselves and the initial probability, $\pi_i^{(0)} = P(X_0 = i)$. The relationship between the DTMSM and the CTMSM is linked through their one-step transition probabilities. For the DTMSM, $\mathbf{P}$ is naturally a one-step transition probability, but for the CTMSM, $\mathbf{P}$ is a one-step transition probability iff the time interval between two observation periods under consideration equals one step (unit), which can be one year interval, two year interval, 6-month interval, etc. Thus, we will have $\mathbf{P} = \mathbf{P}(1)$. For that reason,

in DTMSM models the hazard functions modeled are equivalent to the transition probabilities. This is in contrast to CTMSM where transition intensities are modeled first and these will then be used to compute transition probabilities through Kolmogorov's differential equations.

The theoretical framework reviewed in this section has been well documented and applied in literature, where the covariate effect for a DTMSM is not evaluated (Spedicato et al., 2016). Asanjarani et al. (2021) notes that the complexity of any model always greatly depends on the number of states as well as the possible transitions.

5.3.1 Modeling Transition Probabilities for the DTMSM with Covariates

Kryscio et al. (2006) modeled the one-step transition probabilities for a DTMSM while including covariates using a series of polytomous logistic models which is similar to Chapter 4 of this thesis. These were in contrast to the approach by Korn & Whittemore (1979) where the probability of occupying a current state was modeled using the state in the previous time point as one of the covariates so that intercepts vary according to the previous state. For a basic discrete-time Markov model, Spedicato et al. (2016) then used maximum likelihood (ML) and ML with Laplace smoothing and bootstrap for the estimation of transition probabilities. However, this basic model did not consider the influence of covariates. Other works which model the covariate effects have been developed such as the discrete cause-specific competing risks (Tutz et al., 2016; Steele, 2005) whose destination states were absorbing. Dean et al. (2014) also modeled a multiple event process survival mixture model for analyzing non-repeatable events measured in discrete-time that may occur at the same point in time. Recently, Berger et al. (2020) and Schmid & Berger (2020) have modeled a discrete-time sub-distribution hazard in competing risks using weighted maximum likelihood estimation. Again, the models had absorbing destination states, and considered only one cause of failure in the presence of other possible causes of failure.

Recurrent events Cook et al. (2007) in the discrete-time scenario have been modeled by Tutz et al. (2016) , with only two recurrent states in the process (employment and non-employment) and this was adopted in section 2.3 of this thesis. They suggested a model which considered a subject-specific random effect since observations are made on the same subject. For a process with $S > 2$, this chapter will perform a multivariate model of the discrete survival models from a multi-state process, each with its own subject-specific random effect. These various random effects are also given a distribution so that they can measure the dependencies between the vari-

ous transitions. This will be executed by modeling the ij-transition probability for a subject, say k, in time interval t with covariate vector $\boldsymbol{Z}_k$ and random effects u_{kij} at the same time modeling the correlation between the various transition types in the process. The idea is to extend the transition intensity approach (TIA) by Esquível et al. (2021) to a discrete time case. Here, transition intensities, λ_{kij} where $i \neq j$, in one unit interval are similar to the one-step transition probabilities in that interval. In the case of a discrete-time Markov model (DTMC), to allow incorporation of covariates, $\boldsymbol{Z}_k$, into modeling of the transition intensities, the partial Markov property (Aralis, 2016; Duffy et al., 2014) is used, where the one-step transition probability in equation 5.1 may further be written as:

$$p_{ij} = P\left(X_{t+1} = j | X_t = i, \boldsymbol{Z}_k\right). \tag{5.3}$$

Without a Markov assumption, history would be included and model 5.3 may be written as

$$p_{ij} = P\left(X_{t+1} = j | X_t = i, \boldsymbol{Z}_k, \boldsymbol{H}_{s^*}\right), \tag{5.4}$$

where the history, $\boldsymbol{H}_{s^*}$ may be captured by a random effect, u_{kij} as discussed in Chapter 2. Subsequently in a general multi-state model, the ij^{th} discrete hazard function would be $\lambda_{kij}(t|\boldsymbol{Z}_k)$ where $\boldsymbol{Z}_k$ is a multivariate set of covariates associated with the ij transition at time interval, t for subject k. The discrete hazard function in equation 2.15 for a discrete random effect survival model may now conveniently be re-written to specify the type of transition (say ij) for subject k in the process as

$$\lambda_{kij}(t|\boldsymbol{Z}_k, u_k) = P(T_{kijl} = t | T_{kijl} \geq t, \boldsymbol{Z}_k, u_{kij}). \tag{5.5}$$

This will now be the sub-model used in the multivariate model.

5.4 Multivariate Modeling of the Different Transitions

With all the extensive literature available, none of them , according to our knowledge, has attempted to consider a discrete-time multistate model with recurrent events, while incorporating covariates by considering subject-specific random effects. For this reason, we propose a multivariate discrete-time survival model where the discrete-time one-step transition probability for the l^{th} episode of the ij^{th} transition for subject k is modeled, and the possible correlation between transitions is accounted for.

Multivariate models through a random effect (Fang et al., 2018; Fahrmeir et al., 1994; Pinheiro & Bates, 2006; Diggle, 2002; Cook et al., 2007) have become common in longitudinal studies (Molenberghs & Verbeke, 2005; Fahrmeir et al., 1994) and have several advantages over univariate models. These include an improved control over type-I error rates during multiple testing and efficiency in estimating the parameters. In addition, the correlation between outcomes can be quantified and controlled for (Ayele et al., 2014; Mchunu et al., 2020). Joint models involve simultaneously modeling multiple outcomes through a random effect. Various approaches to joint modeling have been proposed in literature; the use of a multivariate model being the most common. Here the univariate models for each response are combined through the specification of a joint multivariate distribution for the random effect (Fang et al., 2018; Habyarimana et al., 2016; Ayele et al., 2014). However, all these have been formulated for cross sectional data and continuous time-to-event data (Pinheiro & Bates, 2006; Diggle, 2002; Maqutu, 2010; Mchunu et al., 2020). In the case of joint modeling of binary transition outcomes for longitudinal data, a more complex model is required to account for the correlation both between and within transitions (Bel & Paap, 2014; Hedeker, 2003; Fahrmeir et al., 1994).

5.4.1 Notation

Consider a population with k $(k = 1, ..., K)$ subjects who are followed over an interval $[0, \tau)$. The subjects are at risk of experiencing episode l $(l = 1, \ldots, l_k)$ of a transition from state i to state j $(i, j = 1, ..., m)$ at any given time interval t $(t = 1, ..., \tau_{kijl})$. Let $X_k = \min(C_k, \tau)$ be the observation time for the transitions, C_k is the right censoring time which is assumed to be non-informative and independent of the event times. The assumption of non-informative censoring is important as it allows us to assume that all non-censored subjects at each time interval are representative of all subjects who would have remained in the study had censoring not occurred (Dean et al., 2014). If censoring occurs, it is assumed to do so at the end of the interval. An indicator variable, $\delta_{kijl} = I(t_{kijl} \leq C_k)$, is also defined which equals 1 if transition of type ij occurs and 0 otherwise.

5.4.2 Construction of the Joint Discrete-Time Transition Models

For subject k, the hazard of experiencing a recurring transition of type ij at time interval t may be modeled through

$$\text{logit}(\lambda_{kij}(t)) = \beta_{ij0}(t) + \boldsymbol{Z}_{kij}^T \boldsymbol{\beta}_{ij} + u_{kij}, \quad \text{where} \quad i \neq j \tag{5.6}$$

where $\beta_{ij0}(t)$ is the baseline hazard function for the ij^{th} transition to be modeled, and $\boldsymbol{Z}_{ijk} = (Z_{ijk1}, \ldots, Z_{ijkp})$ is a $p \times 1$ vector of covariates associated with the $p \times 1$ vector of regression coefficients $\boldsymbol{\beta}_{ij} = \beta_{ij1} \ldots, \beta_{ijp}$. The subject-specific variability on the transition of type ij is explained by u_{kij}.

Considering all transitions in the multi-state process, the multivariate hazard for subject k comprises of all possible hazards that k may experience and can be rewritten compactly as

$$\text{logit}(\boldsymbol{\lambda}_k) = \begin{pmatrix} \text{logit}\lambda_{k12}(t) \\ \vdots \\ \text{logit}\lambda_{k1m}(t) \\ \vdots \\ \text{logit}\lambda_{m1}(t) \\ \vdots \\ \text{logit}\lambda_{m(m-1)}(t) \end{pmatrix}$$

$$= \begin{pmatrix} \beta_{012}(t) \\ \vdots \\ \beta_{01m}(t) \\ \vdots \\ \beta_{0m1}(t) \\ \vdots \\ \beta_{0m(m-1)}(t) \end{pmatrix} + \begin{pmatrix} Z_{k121} & Z_{k122} & \cdots & Z_{k12p} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{k1m1} & Z_{k1m2} & \cdots & Z_{k1mp} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{km11} & Z_{km12} & \cdots & Z_{km1p} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{km(m-1)1} & Z_{km(m-1)2} & \cdots & Z_{km(m-1)p} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{121} & \cdots & \boldsymbol{\beta}_{m(m-1)1} \\ \boldsymbol{\beta}_{122} & \cdots & \boldsymbol{\beta}_{m(m-1)2} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\beta}_{12p} & \cdots & \boldsymbol{\beta}_{m(m-1)p} \end{pmatrix} + \begin{pmatrix} u_{k12} \\ \vdots \\ u_{k1m} \\ \vdots \\ u_{km1} \\ \vdots \\ u_{km(m-1)} \end{pmatrix}$$

$$\text{logit}\boldsymbol{\lambda}_k = \boldsymbol{\beta}_0(t) + \boldsymbol{Z}_k'\boldsymbol{\beta} + \boldsymbol{u}_k, \tag{5.7}$$

where $\boldsymbol{\beta}_0(t)$ is an $m(m-1) \times 1$ vector of the baseline hazard functions for the $m(m-1)$ transitions, $\boldsymbol{Z}_k$ is an $p \times m(m-1)$ matrix of the $m(m-1)p$ fixed effects augmenting all covariates vector for subject k for each ij transition and $\boldsymbol{\beta}$ is an $p \times m(m-1)$ matrix of regression coefficients (p regression coefficients for each transition) to be estimated. The multivariate subject-specific random effects, $\boldsymbol{u}_k = \left(u_{k12}, \ldots, u_{k1m}, \ldots, u_{km1}, \ldots, u_{km(m-1)}\right)'$ account for the unobserved heterogeneity, the inter-recurrence dependencies within a subject as well as the dependency between different transitions (Mazroui et al., 2013). As highlighted in Chapter 2, a logit link function will be used to model the multivariate hazard function. The joint model is performed with separate random effects while imposing an associational structure across transitions. Without loss of generality, the joint distribution of the random effects is assumed to be a $\mathbf{0}$-centered multivariate normal; $\boldsymbol{u}_k \sim MVN(\mathbf{0}, \boldsymbol{\Omega}_u)$ where $\boldsymbol{\Omega}_u$ is an $m(m-1) \times m(m-1)$ positive-definite variance-

covariance matrix with $m(m-1)$ variances and $m(m-1)((m^2-m-1))/2$ covariances.

For a process in Figure 3.1 with $m = 5$ possible transitions (between states $1 \rightarrow 2, 2 \rightarrow 3, 2 \rightarrow 4, 3 \rightarrow 2$ and $4 \rightarrow 2$), the joint discrete-time survival model with a logit link as discussed in chapter 2 is constructed as:

$$\lambda_{k12}(t) = \frac{\exp(\beta_{120}(t) + \boldsymbol{Z}'_{k12}\boldsymbol{\beta}_{12} + u_{k12})}{1 + \exp(\beta_{120}(t) + \boldsymbol{Z}'_{k12}\boldsymbol{\beta}_{1} + u_{k12})}$$

$$\lambda_{k23}(t) = \frac{\exp(\beta_{230}(t) + \boldsymbol{Z}'_{k23}\boldsymbol{\beta}_{23} + u_{k23})}{1 + \exp(\beta_{230}(t) + \boldsymbol{Z}'_{k23}\boldsymbol{\beta}_{1} + u_{k23})}$$

$$\lambda_{k24}(t) = \frac{\exp(\beta_{240}(t) + \boldsymbol{Z}'_{k24}\boldsymbol{\beta}_{24} + u_{k24})}{1 + \exp(\beta_{240}(t) + \boldsymbol{Z}'_{k24}\boldsymbol{\beta}_{1} + u_{k24})}$$

$$\lambda_{k32}(t) = \frac{\exp(\beta_{320}(t) + \boldsymbol{Z}'_{k32}\boldsymbol{\beta}_{32} + u_{k32})}{1 + \exp(\beta_{320}(t) + \boldsymbol{Z}'_{k32}\boldsymbol{\beta}_{1} + u_{k32})}$$

$$\lambda_{k42}(t) = \frac{\exp(\beta_{420}(t) + \boldsymbol{Z}'_{k42}\boldsymbol{\beta}_{42} + u_{k42})}{1 + \exp(\beta_{420}(t) + \boldsymbol{Z}'_{k42}\boldsymbol{\beta}_{42} + u_{k42})}, \tag{5.8}$$

where

$$\boldsymbol{u}_k = \begin{pmatrix} u_{k12} \\ u_{k23} \\ u_{k24} \\ u_{k32} \\ u_{k42} \end{pmatrix} \sim MVN \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{u_{k12}} & \sigma_{u_{k12},u_{k23}} & \sigma_{u_{k12},u_{k24}} & \sigma_{u_{k12},u_{k32}} & \sigma_{u_{k12},u_{k42}} \\ & \sigma^2_{u_{k23}} & \sigma_{u_{k23},u_{k24}} & \sigma_{u_{k23},u_{k32}} & \sigma_{u_{k23},u_{k42}} \\ & & \sigma^2_{u_{k24}} & \sigma_{u_{k24},u_{k32}} & \sigma_{u_{k24},u_{k42}} \\ & & & \sigma^2_{u_{k32}} & \sigma_{u_{k32},u_{k42}} \\ & & & & \sigma^2_{u_{k42}} \end{pmatrix} \right). \tag{5.9}$$

The diagonal entries of the symmetric variance-covariance matrix, $\boldsymbol{\Omega}_u$, measure the variance of each random effect. The off-diagonal entries, which may or may not be equal depending on the specification of the correlation structure imposed, will determine the covariance between pairs of multiple random effects, hence correlations between events.

Correlation Structures

The distinct nature of recurring survival model analysis is the correlation structure of the observed data. Since observations are made on the same subject, they are likely to be more similar than those made on different subjects (Tutz et al., 2016). Although it is not of primary interest in the analysis, the correlation structure must be correctly modeled for the analysis to be valid since it plays a huge role in the validity

of inferences. Various choices of the form of the working correlation structures have been proposed in literature Fahrmeir et al. (1994); Littell et al. (2000). The most commonly used ones are: independent, unstructured (UN), compound symmetry (CS), variance component and autoregressive (AR(1)).

Independent structure Here, the assumption is that the random effects are independent. It is the most simple correlation structure, which is similar to a marginal model.

Unstructured This is the most flexible but complex correlation structure where the pattern of correlations is unstructured. It assumes unconstrained pairwise correlations to be estimated from the data so that all variances and correlations are different. It is more versatile because it assumes no pattern at all for the association between pairs of multiple random effects, hence unconstrained pair-wise correlations. Each correlation is estimated directly from the data. This lets the data dictate what they should be and requires the estimation of many parameters (Molenberghs & Verbeke, 2000; Littell et al., 2000; Mchunu, 2018; Diggle, 2002). A major disadvantage for this assumption is that it increases the number of parameters to be estimated in the overall model. This in turn, causes possible non-convergence problems, particularly those associated with boundary values. One way to try and reduce the number of parameters is to further assume that all the variances along the diagonal have a constant variance (Littell et al., 2000; Fahrmeir et al., 1994; Mchunu, 2018; Diggle, 2002). Following the principle of parsimony, an analysis that uses this correlation structure usually has lower statistical power than that which uses a less parametric but more realistic structure. In addition, if the transitions are sparse, the model might not work. For this reason, analysis using this kind of structure will not be considered.

Compound symmetry structure This structure, also known as the exchangeable, specifies that observations on the same subject have homogeneous variance and homogeneous covariance. It assumes that the association between the random effects is the same across all pairs. Hence the variance-covariance matrix will be of the form

$$\boldsymbol{\Omega}_u = \begin{pmatrix} \sigma^2_{u_k} + \sigma & \sigma & \sigma & \sigma & \sigma \\ & \sigma^2_{u_k} + \sigma & \sigma & \sigma & \sigma \\ & & \sigma^2_{u_k} + \sigma & \sigma & \sigma \\ & & & \sigma^2_{u_k} + \sigma & \sigma \\ & & & & \sigma^2_{u_k} + \sigma \end{pmatrix}.$$

Variance component structure Here, the random effects are assumed to have their own variances while there is no association between them. Hence the variance-covariance matrix will be of the form

$$\boldsymbol{\Omega}_u = \begin{pmatrix} \sigma^2_{u_{k1}} & 0 & 0 & 0 & 0 \\ & \sigma^2_{u_{k2}} & 0 & 0 & 0 \\ & & \sigma^2_{u_{k3}} & 0 & 0 \\ & & & \sigma^2_{u_{k4}} & 0 \\ & & & & \sigma^2_{u_{k5}} \end{pmatrix}.$$

First-order autoregressive (AR(1)) structure The AR-1 correlation structure depends on the distance between the measures (Maqutu, 2010). The specification assumes that the random effects have a homogeneous variance, but covariances between observations on the same subject decrease towards zero with increasing lag. Values of the correlations decline over time as the separation between pairs of repeated measures increases. Correlation between any two responses that are h measurements apart is ρ^h. This structure assumes that the measurement occasions are equally spaced (Molenberghs & Verbeke, 2005; Fitzmaurice et al., 2008; Maqutu, 2010) and is appropriate for repeated measurements (longitudinal studies), hence will not be considered in this study. The variance-covariance matrix will be of the form

$$\boldsymbol{\Omega}_u = \begin{pmatrix} 1 & \rho_{u_k} & \rho^2_{u_k} & \rho^3_{u_k} & \rho^4_{u_k} \\ & 1 & \rho_{u_k} & \rho^2_{u_k} & \rho^3_{u_k} \\ & & 1 & \rho_{u_k} & \rho^2_{u_k} \\ & & & 1 & \rho_{u_k} \\ & & & & 1 \end{pmatrix}.$$

Selecting the best working correlation structure Choosing a correlation structure which is too simple may lead to increased Type I error rates and selecting a too complex correlation structure may compromise statistical power and efficiency (Littell et al., 2000). This leaves the researcher with a dilemma regarding the choice of the correlation structure. As explained in Chapter 2, the use of information criterion such as the AIC or BIC is recommended. These are based on assessing model fit while penalizing the number of estimated parameters. The model with a correlation structure that produces the smallest value of the AIC is preferred.

5.5 Functional Form of the Baseline Hazard

As discussed in Chapter 2, we will consider the modeling of $\boldsymbol{\beta}_0(t)$ - a vector of the real-valued time-dependent intercepts $\beta_{ij0}(t))$ for each of the transitions (sometimes referred to as the transition-specific baseline coefficients) with $t = 1, 2...\tau$. In cases where the number of event times is large relative to the sample size, the baseline function may sometimes be represented by a smooth function of an unspecified form, such as smoothing splines or P-splines (Berger et al., 2020; Tutz et al., 2016). The linear relationship of the baseline with log time is relaxed through the use of splines (Crowther et al., 2014).

We will consider two forms of baseline, which are assumed similar across transitions, namely: discrete piece-wise constant and a smooth baseline function represented by cubic P-splines with a second-order difference penalty. For a piece-wise constant, the baseline hazard for each ij transition is represented by the parameters $\beta_{ij0} = \beta_{ij0}(1), \beta_{ij0}(2), \ldots, \beta_{ij0}(\tau)$, which comprises constant functions in each time interval, $t = 1, \ldots, \tau$, that result from partitioning the time scale. This approach has a risk of producing a large number of parameters as the number of parameters in the model is determined by the number of intervals, because for each interval there is a separate intercept hazard.

Following Berger et al. (2020), a smoothing function can be specified by a simpler parameterization which contains fewer parameters than the piece-wise specification :

$$\beta_{0ij}(t) = \sum_{g=1}^{G} \beta_{0ijg}(t)\phi_{ijg}(t), \tag{5.10}$$

where $\phi_{ijg}(.), g < \tau$ are fixed basis functions for the ij^{th} transition where polynomial splines in the form of the truncated power series basis and B-splines are the most common choices (Tutz et al., 2016).

The algorithm for obtaining polynomial regression splines is as follows Tutz et al. (2016):

1. Partition the time domain into continuous intervals $[\tau_k; \tau_{k+1}]$
2. Represent the unknown function by a separate polynomial of degree d in each interval, which are supposed to join smoothly at the knots $\tau_1, \ldots, \tau_{G-1}$ (determining the boundaries of the intervals). Note that these knots are selected from the time domain $[0, T]$

3. Represent the polynomial splines of degree d with the truncated power series basis to obtain

$$\beta_{0ij}(t) = \beta_{00ij} + \beta_{01ij}t + \ldots + \beta_{0dij}t^d + \sum_{k=1}^{G-1} \beta_{ij(\ell+k)}(t-\tau_k)_+^d \tag{5.11}$$

where the truncated power functions, $(t-\tau_k)_+^d$, are defined by

$$(t-\tau)_+^d := \begin{cases} (t-\tau)^d & \text{if } t \geq \tau, \\ 0 & \text{if } t < \tau\ . \end{cases}$$

This representation from equation 5.11 uses $G = (G-1) + d + 1$ basis functions

$$\phi_{ij1} = 1, \quad \phi_{ij2}(t) = t, \ldots, \phi_{ij,d+1}(t) = t^q, \quad \phi_{ij,d+2}(t) = (t-\tau_1)_+^d, \ldots \phi_{ij,d+(G-1)+1}(t) = (t-\tau_{G-1})_+^d, \tag{5.12}$$

which form the truncated power series basis of degree d.
Again, the AIC will determine which model best suites the data.

5.6 Likelihood Construction

Literature has documented various estimation techniques for parameters of joint time-to-event models which may either be Bayesian (Kneib & Hennerfeind, 2008; Matsena Zingoni et al., 2021; Sweeting et al., 2005, in continuous case analysis) or frequentist. There is vast literature within the frequentist framework for likelihood-based estimation. It has often been used in discrete-time multi-state models, specifically maximum likelihood estimation (MLE) which has a variety of optimality properties (Spedicato et al., 2016). Other estimation techniques also exist such as the methods of moment estimation, which may be explored. In this section, the complete and marginal likelihood functions for the joint model 5.7 are constructed, where all the events are modeled with separate random effects and an associational structure across transitions is imposed by assuming that the random effects have a multivariate normal distribution with the off-diagonal elements of the variance-covariance matrix measuring dependence. The complete joint likelihood for subject k experiencing a transition of type ij is given by

$$L_{kij}(t, \boldsymbol{\beta}_{ij}, u_{kij}, \sigma_u^2) = \prod_{h=1}^{\tau_k} \left\{ \lambda_{kij}(h|\boldsymbol{Z}_k)^{\delta_{kijh}} \left(1 - \lambda_{kij}(h|\boldsymbol{Z}_k)\right)^{1-\delta_{kijh}} \right\} \times \left\{ \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(\frac{-u_{kij}^2}{2\sigma_u^2} \right) \right\}, \tag{5.13}$$

where δ_{kijh} is a recurrent event indicator as defined earlier in section 5.4.1 and a logit link is used from the hazard as mentioned in 5.4.2. An estimate of the baseline hazard function is used as defined in section 5.5. This yields a complete joint likelihood over all the subjects for all the transitions as

$$L_{kij}(t,\boldsymbol{\beta},\boldsymbol{u}) = \prod_{k=1}^{K}\prod_{ij=1}^{m}\prod_{h=1}^{\tau_k-(1-\delta_{kijl})}\left\{\lambda_{kij}(h|\boldsymbol{Z}_k)^{\delta_{kijh}}\left(1-\lambda_{kij}(h|\boldsymbol{Z}_k)\right)^{1-\delta_{kijh}}\right\}$$

$$\times f(u_{k12},\ldots,u_{k1m},\ldots,u_{km1},\ldots,u_{km(m-1)}), \tag{5.14}$$

where $f(u_{k12},\ldots,u_{k1m},\ldots,u_{km1},\ldots,u_{km(m-1)})$ is the joint density function of the random effects $u_{k12},\ldots,u_{k1m},\ldots,u_{km1},\ldots,u_{km(m-1)}$ which is assumed to follow a multivariate normal distribution with mean $\mathbf{0}$ and a variance covariance matrix $\boldsymbol{\Omega}_u$ so that

$$L(\beta,\boldsymbol{\Omega}) \propto \prod_{k=1}^{K}\prod_{ij=1}^{m}\prod_{l=1}^{l_k}\prod_{h=1}^{\tau_k-(1-\delta_{kijl})}\left\{\lambda_{kij}(h|\boldsymbol{Z}_k)^{\delta_{kijh}}\left(1-\lambda_{kij}(h|\boldsymbol{Z}_k)\right)^{1-\delta_{kijh}}\right\}$$

$$\times (2\pi)^{-\frac{m}{2}}|\boldsymbol{\Omega}|^{-\frac{m}{2}}\exp\left(-\frac{1}{2}\mathbf{u}'\boldsymbol{\Omega}^{-1}\mathbf{u}\right) \tag{5.15}$$

The corresponding joint marginal log-likelihood (integrated over all random effects) is given by

$$\ell(\boldsymbol{\beta},\boldsymbol{\Omega}) \propto \sum_{k=1}^{K}\sum_{ij=1}^{m}\log\int....\int\left(\prod_{h=1}^{\tau_k-(1-\delta_{kijl})}\left\{\lambda_{kij}(h|\boldsymbol{Z}_k)^{\delta_{kijh}}\left(1-\lambda_{kij}(h|\boldsymbol{Z}_k)\right)^{1-\delta_{kijh}}\right\}\right)$$

$$\times\left((2\pi)^{-\frac{m}{2}}|\boldsymbol{\Omega}|^{-\frac{m}{2}}exp\left(-\frac{1}{2}\mathbf{u}'\boldsymbol{\Omega}^{-1}\mathbf{u}\right)\right)du_{k12},\ldots,du_{k1m},\ldots,du_{km1},\ldots,du_{km(m-1)} \tag{5.16}$$

where

$$\lambda_{kij}(t|\boldsymbol{Z}_k;u_{kij}) = \frac{\exp\left(\hat{\beta}_{0ij}(t)+\boldsymbol{\beta}_{ij}^{'}\boldsymbol{Z}_{kij}+u_{kij}\right)}{1+\exp\left(\hat{\beta}_{0ij}(t)+\boldsymbol{\beta}_{ij}^{'}\boldsymbol{Z}_{kij}+u_{kij}\right)}.$$

The score functions for $\boldsymbol{u}$ and $\boldsymbol{\beta}$ follow from the log-likelihood in equation 5.16 and are given by

$$U_{\boldsymbol{u}} = \frac{\partial\ell(\boldsymbol{\beta},\boldsymbol{\Omega})}{\partial\boldsymbol{u}}$$

and

$$U_{\boldsymbol{\beta}} = \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\Omega})}{\partial \boldsymbol{\beta}},$$

respectively. Solving these score functions simultaneously and setting them equal to 0 will maximize the likelihood to obtain the estimates of $\boldsymbol{u}$ and $\boldsymbol{\beta}$.

5.7 Numerical Approximation Techniques

Maximum likelihood estimation is used in this study because of its optimality properties. However, one of its major drawbacks lies in the fact that most of the integrals do not have closed form solutions, hence numerical integration techniques may be a plausible approach for approximation. Solving a system of nonlinear equations algebraically may not be as easy and straight forward as solving linear equations (Czepiel, 2002). For that reason, iterative processes are viable alternatives that may be used to numerically estimate the solution to the system of nonlinear equations which results from maximization of the log-likelihood functions (numeric or stochastic integration). As discussed in Chapter 2, this study will consider the Gauss-Hermite quadrature method for the random effects model. Since model 5.7 has a relatively extensive random-effects structure, a non-adaptive Gaussian quadrature with 3 quadrature points will be used for the ease of convergence. Higher order quadrature points may, however, be used until numerical convergence is reached. In addition, selection of decent seed values of the parameters is of utmost importance in achieving convergence. For this reason, parameter estimates from the univariate random effects models obtained in Chapter 3 will be used as starting values for the $\boldsymbol{\beta}$ estimates. Lastly, the maximum number of iterations was set to 50 since time to reach convergence is too long for this model.

5.8 Estimation using Software for Regression

With the representation of binary variables, it implies that the model may be estimated in the same way as binary mixed effects models for repeated measurements, where repeated measurements in this case are the separate episodes and different time points. Standard computer software for binary distributions with random effects may be used as a basis such as **WinBUGS** (Spiegelhalter et al., 2003) or **PROC NLMIXED** (SAS Institute, 2015), in SAS, whose advantage is that it allows a very high degree of flexibility in the way the model is specified and parameterized. However, the downside might be the fact that not only does one need to specify the model, but also has to specify the names for all the parameters in the model (Molen-

berghs & Verbeke, 2005). **PROC GLIMMIX** in SAS may also be used where approximation of the data is required (such as penalized quasi-likelihood or marginal quasi-likelihood) rather than approximation of the integral. These may be computationally intensive and their runtime is longer. **MLwin** (Rasbach et al., 2003) in STATA is computationally faster and may also be used for MCMC estimation, but might be costly to get the license.

For demonstration purposes, the joint model will consider only the covariates that are common across events. For this cohort, there were very few transitions relative to the sample size. For this reason, convergence of the unstructured covariance structure was not being reached as the transitions were sparse for the subjects. To overcome this, a selection of 27 subjects who had experienced more than 5 transitions was made and the results reported were based on them. It is crucial for such an analysis to have as many transitions as possible, for the estimation to run fully.

5.9 Model Comparisons.

For each method, a parsimonious model of best fit with covariates was selected using a likelihood ratio test defined as

$$-2\frac{ln(L_R(\hat{\theta})}{lnL_F(\hat{\theta})} \sim \chi^2_n,$$

where $L_R(\hat{\theta})$ is the likelihood of the reduced model with no (or a subset of) covariates and $L_F(\hat{\theta})$ is the likelihood of the full model with all covariates. Thus the LR test is most powerful for comparison of nested models. In all tests, we used 5% level of significance. Model comparison based on MLEs can be done using Aikake information criteria (AIC) and the Bayesian information criteria (BIC) for non-nested models, where AIC is defined by

$$2\left(-\log L\left(\theta\right)+K\right)$$

and BIC is defined by

$$-2\log L\left(\theta\right)+K\log\left(N\right)$$

where $L(\theta)$ is the likelihood function, K is the number of parameters estimated from the model and N is the number of observations in the study. When models are complex, the BIC penalizes them more than the AIC does. This is obvious since we observe that $log(N)$ in the BIC replaces the factor 2 in the AIC in the penalty term.

For this reason, we will use the AIC for model comparison in this study.

5.10 Simulation study

Simulation studies are essential for understanding and evaluating both current and new statistical models, hence are commonly used to evaluate their performance. (Burton et al., 2006; Crowther & Lambert, 2012). In this section, a simulation of the model proposed in section 5.4 is conducted. The generated data is typical of real data which has multiple discrete-time survival outcomes to mimic the data that we analyzed using the proposed method.

The simulated model is of the form:

$$
\begin{aligned}
logit(\lambda_{k1}) &= \beta_{0,11}(t) + \beta_{0,12}(t) + \beta_{0,13}(t) + \beta_{0,14}(t) + \beta_{0,15}(t) + Z'_{k1}\beta_{11} + Z'_{k2}\beta_{12} + Z'_{k3}\beta_{13} + u_{k1} \\
logit(\lambda_{k2}) &= \beta_{0,21}(t) + \beta_{0,22}(t) + \beta_{0,23}(t) + \beta_{0,24}(t) + \beta_{0,25}(t) + Z'_{k1}\beta_{21} + Z'_{k2}\beta_{22} + Z'_{k3}\beta_{23} + u_{k2} \\
logit(\lambda_{k3}) &= \beta_{0,31}(t) + \beta_{0,32}(t) + \beta_{0,33}(t) + \beta_{0,34}(t) + \beta_{0,35}(t) + Z'_{k1}\beta_{31} + Z'_{k2}\beta_{32} + Z'_{k3}\beta_{33} + u_{k3} \\
logit(\lambda_{k4}) &= \beta_{0,41}(t) + \beta_{0,42}(t) + \beta_{0,43}(t) + \beta_{0,44}(t) + \beta_{0,45}(t) + Z'_{k1}\beta_{41} + Z'_{k2}\beta_{42} + Z'_{k3}\beta_{43} + u_{k4} \\
logit(\lambda_{k5}) &= \beta_{0,51}(t) + \beta_{0,52}(t) + \beta_{0,53}(t) + \beta_{0,54}(t) + \beta_{0,55}(t) + Z'_{k1}\beta_{51} + Z'_{k2}\beta_{52} + Z'_{k3}\beta_{53} + u_{k5},
\end{aligned}
$$

for the 5 transition types, namely $1, 2, ...$ and 5, respectively.

The simulated multivariate discrete-time survival model was evaluated using the mean of the point estimates, mean bias, mean square error (MSE) as well as the coverage probability (CP) of the 95% confidence interval. Mean bias was calculated as the average difference between the true mean and its estimate across the total simulated replicates. MSE was calculated as the average squared difference between the true mean and its estimate across the total simulated replicates. The $95\%CP$ gives the proportion of time which the confidence interval contains the true value.

5.10.1 Simulation Protocol

For demonstration, purposes, a simulation of a multivariate survival model was done for the 4 states (*Never Married, Married, Separated* and *Widowed*) with five possible transitions. A few assumptions were made. These are:

- the baseline hazard is a fully discrete function of time i.e. each time period has it's own conditionally independent intercept term
- the hazard function is conditioned linearly on any predictors.

A piece-wise constant baseline hazard was set with 4 intervals whose parameters were set at
$\beta_{0,11}(t) = 0.1, \beta_{0,12}(t) = 0.1, \beta_{0,13}(t) = 0.1, \beta_{0,14}(t) = 0$ and $\beta_{0,15}(t) = 0.2$ for transition sub-model 1,
$\beta_{0,21}(t) = 0.5, \beta_{0,22}(t) = 0, \beta_{0,23}(t) = 0.5, \beta_{0,24}(t) = 0.25$ and$\beta_{0,25}(t) = 0, 1$ for transition sub-model 2,
$\beta_{0,31}(t) = 0.2\beta_{0,32}(t) = 0.6\beta_{0,33}(t) = 0.8\beta_{0,34}(t) = 0.25$ and $\beta_{0,35}(t) = 0.7$ for transition sub-model 3,
$\beta_{0,41}(t) = 0, \beta_{0,42}(t) = 0.1, \beta_{0,43}(t) = 0.2, \beta_{0,44}(t) = 0.3$ and $\beta_{0,45}(t) = 0.08$ for transition sub-model 4,
and $\beta_{0,51}(t) = 0.1\beta_{0,52}(t) = 0.3, \beta_{0,53}(t) = 0.1, \beta_{0,54}(t) = 0$ and $\beta_{0,55}(t) = 0.3$ for sub-model 5. The fixed effects parameters were set at $\beta_{11} = 0, \beta_{12} = -1.5$ and $\beta_{13} = 2$ for transition 1, $\beta_{21} = 2, \beta_{22} = -0.3$ and $\beta_{23} = -0.1$ for transition 2, $\beta_{31} = -03, \beta_{32} = 1$ and $\beta_{33} = 0.4$ for transition 3, $\beta_{41} = 0.2, \beta_{42} = 1$ and $\beta_{43} = -0.1$ for transition 4 and $\beta_{51} = -0.1, \beta_{52} = -0.4$ and $\beta_{53} = -2$ for transition 5
and were associated with the 3 time-independent covariates, similar across transitions, whose distributions are $Z_{k1} \sim Bernoulli(0.5), Z_{k2} \sim Normal(0, 1)$ and $Z_{k3} \sim Poison(0.5)$, respectively. For the random effects, a **0**-centered multivariate normal distribution was assumed and the variance-covariances were fixed at

$$\begin{pmatrix} u_{k12} \\ u_{k23} \\ u_{k24} \\ u_{k32} \\ u_{k42} \end{pmatrix} \sim MVN \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & -0.5 & 0.25 & -0.25 \\ & 1 & -0.5 & 0.25 & 0.25 \\ & & 1 & 0.5 & -0.5 \\ & & & 1 & 0.5 \\ & & & & 1 \end{pmatrix} \right).$$

A sample of $n = 1000$ was simulated with 100 replicates and a maximum of 5 transitions per subject was allowed. Data simulation was done using the package **discSurv** (Welchowski & Schmid, 2015) in R and the analysis was done in SAS using the **NLMIXED** procedure which would also be used for analyzing the real-life data set in the next section (SAS Institute, 2015). This procedure is used to fit non-linear mixed models by numerically maximizing an approximation to the marginal likelihood (Molenberghs & Verbeke, 2005; Littell & Milliken, 2006). This is done by integrating the likelihood over the random effects.

5.10.2 Results from the Simulation

Fixed effects results from the simulations of a multivariate discrete-time survival model described in section 5.10.1 are displayed in Table 5.1, while random effects

results are shown in Table 5.2. For each parameter, we show its true value (θ), its average estimate ($\hat{\theta}$), its bias (**B**), its mean square error (MSE) and its 95% coverage probability (95%CP). Results show a fair performance of the model as noticed in small (near zero) biases and mean square errors of the estimates and coverage probabilities close to the nominal value of 0.95. These high coverage probabilities of the 95% confidence intervals constructed with the estimates imply a good covering, hence satisfactory outcomes. A few estimates had a poor fit such as the outcome for random effects covariances between transition 4 and 5 whose $95\%CP$ was 0.51, although the mean square error was 0.06 which is close to 0. The outcome variance of transition 3 with also has a smaller $95\%CP$ of 0.68. For transition 1 and 2, all the average estimates have a small bias and MSE with a $95\%CP$ between 0.90 and 0.99. In general, the model performs well since it produces satisfactory results.

Table (5.1) Results of fixed effects from the Simulations for a multivariate discrete-time survival model

Parameter	θ	$\hat{\theta}$	**B**	MSE	$95\%CP$
Transition 1					
$\beta_{0,11}(t)$	0.10	1.05	0,05	0.02	0.92
$\beta_{0,12}(t)$	0.10	0.98	-0.02	0.01	0.98
$\beta_{0,13}(t)$	0.10	0.14	0.04	0.01	0.95
$\beta_{0,14}(t)$	0.00	0.03	0.03	0.03	0.90
$\beta_{0,15}(t)$	0.20	0.19	0.01	0.02	0.92
β_{11}	0.00	0.01	0.01	0.02	0.91
β_{12}	-1.50	-1.88	-0.38	0.02	0.94
β_{13}	2.00	2.01	0.01	0.01	0.99
Transition 2					
$\beta_{0,21}(t)$	0.50	0.52	0.02	0.01	0.91
$\beta_{0,22}(t)$	0.00	0.01	0.01	0.03	0.90
$\beta_{0,23}(t)$	0.50	0.49	-0.01	0.01	0.99
$\beta_{0,24}(t)$	0.25	0.22	-0.03	0.03	0.97
$\beta_{0,25}(t)$	0.10	0.98	-0.02	0.01	0.96
β_{21}	2.00	2.20	0.20	0.01	0.95
β_{22}	-0.30	-0,31	-0.01	0.02	0.95
β_{23}	-0.10	-0.14	-0.04	0.02	0.92
Transition 3					
$\beta_{0,31}(t)$	0.20	0.17	-0.03	0.04	0.87
$\beta_{0,32}(t)$	0.60	0.50	-0.10	0.04	0.72
$\beta_{0,33}(t)$	0.80	0.77	-0.03	0.03	0.90
$\beta_{0,34}(t)$	0.25	0.27	0.02	0.01	0.93
$\beta_{0,35}(t)$	0.70	0.66	-0.04	0.01	0.94
β_{31}	-0.30	-0.25	0.05	0.03	0.92
β_{32}	1.00	1.01	0.01	0.01	0.95
β_{33}	0.40	0.39	-0.01	0.02	0.94
Transition 4					
$\beta_{0,41}(t)$	0.00	0.02	0.02	0.01	0.96
$\beta_{0,42}(t)$	0.10	0.09	-0.01	0.01	0.96
$\beta_{0,43}(t)$	0.20	0.21	0.01	0.03	0.92
$\beta_{0,44}(t)$	0.30	0.35	0.05	0.02	0.90
$\beta_{0,45}(t)$	0.80	0.78	-0.02	0.04	0.88
β_{41}	0.20	0.18	-0.02	0.03	0.91
β_{42}	1.00	1.03	0.03	0.01	0.96
β_{43}	-0.10	-0.09	0.01	0.01	0.99
Transition 5					
$\beta_{0,51}(t)$	0.10	0.14	0.04	0.02	0.97
$\beta_{0,52}(t)$	0.30	0.32	-0.02	0.02	0.95
$\beta_{0,53}(t)$	0.10	0,11	0.01	0.03	0.93
$\beta_{0,54}(t)$	0.00	0,05	-0.05	0.04	0.88
$\beta_{0,55}(t)$	0.30	0.29	-0.01	0.02	0.95
β_{51}	-0.10	-0,10	-0.00	0.01	0.96
β_{52}	-0.40	-0.44	0.04	0.03	0.92
β_{53}	-2.00	-2.05	-0.05	00.01	0.95

Table (5.2) Random Effects Results from the Simulations for a multivariate discrete-time survival model

Parameter	θ	$\hat{\theta}$	**B**	MSE	$95\%CP$
u_{k1}	1.00	1.15	0.15	0.01	0.98
u_{k2}	1.00	0.99	-0.01	0.01	0.99
u_{k3}	1.00	0.88	-0.12	0.04	0.68
u_{k4}	1.00	0.96	-0.04	0.02	0.95
u_{k5}	1.00	1.02	0.02	0.02	0.93
$u_{k1,k2}$	0.50	0.55	0.05	0.03	0.90
$u_{k1,k3}$	-0.50	-0.44	0.06	0.01	0.91
$u_{k1,k4}$	0.25	0.26	0.01	0.02	0.92
$u_{k1,k5}$	-0.25	-0.31	-0.06	0.01	0.92
$u_{k2,k3}$	-0.50	-0.38	0.12	0.01	0.95
$u_{k2,k4}$	0.25	0.11	-0.14	0.05	0.62
$u_{k2,k5}$	0.25	0.27	0.02	0.01	0.90
$u_{k3,k4}$	0.5	0.49	-0.01	0.02	0.92
$u_{k3,k5}$	-0.5	-0.36	0.14	0.04	0.74
$u_{k4,k5}$	0.5	0.22	0.28	0.06	0.51

5.11 Empirical Results

To illustrate the application of the multivariate survival model in discrete-time, a population-based data set from Africa Health Research Institute (ARHI), whose description was presented in Chapter 2, was analyzed. Most of the demographical features for this cohort are well documented in Batidzirai et al. (2020). Covariates considered included gender, income earned, employment, highest education and age at first sex. Various baseline hazard functions and covariance structures were considered. Based on the compound symmetry covariance structure, Table 5.3 displays the results for the variance and correlation estimates under a piece-wise baseline hazard. Variances of the random effects are assumed equal with a value of 0.36. The correlation between the various transitions is also assumed equal with a value of 0.14, implying a weak positive dependence of transitions.

Table (5.3) Compound symmetry variance components with a piece-wise baseline hazard

	Estimate	Standard Error	P-value
Variances	0.36	0.008	<0.0001
Correlations between transitions	0.14		

Regarding the unstructured covariances, a few subjects were used for analysis and Table 5.4 displays the estimates of covariance matrices produced by a model with a piece-wise baseline hazard function.

Table (5.4) Correlation matrix: Unstructured with a piece-wise baseline hazard using only 27 subjects

	Estimate (Standard Error)				
	NM	**MS**	**MW**	**SM**	**WM**
NM	3.28 (1.81)	0.66 (0.41)	0.72 (0.16)	-0.54 (0.07)	-0.60 (0.03)
MS		8.63 (3.22)	0.47 (0.01)	-0.08 (0.01)	0.28 (0.05)
MW			5.41 (1.28)	-0.18 (0.01)	-0.649 (0.09)
SM				9.8 (1.12)	0.85 (0.10)
WM					0.72 (0.06)

Based on its small AIC, results from a model with a cubic spline baseline will not be presented. Table 5.5 displays results from a multivariate discrete-time multi-state model using joint survival models under the assumption of the compound symmetry covariance structure, while the baseline hazard was modeled using piece-wise constant form. Results show that the direction and magnitude of the estimates did not deviate significantly from those of the univariate models from chapter 2. How-

ever, a few of the estimates which were not statistically significant in the univariate models are now statistically significant in the multivariate model and vice-versa. For example, the positive effect of employment on the *Widowed* $\rightarrow$ *Married* (OR=1.43, CI=$1.22; 1.64$) and *Never Married* $\rightarrow$ *Married* (OR=1.20, CI=$1.01; 1.40$) transitions were now statistically significant. For all transitions except the *Separated* $\rightarrow$ *Married* transition, age at first sex had a positive and significant effect. Males were significantly less likely to get widowed (OR=0.18; CI=$0.15; 0.33$) than females. Compared to subjects who never went to school, the odds of experiencing a marital separation were significantly high for those with primary education and with high school education (OR=4.21, CI=$2.19; 6.23$ and OR=2.98, CI=$1.97; 3.99$ for primary and secondary education, respectively) .

Table (5.5) Results for the multivariate discrete-survival model under piece-wise constant baseline hazard

	N.Married ⟶ Married		**Married ⟶ Separated**		**Married ⟶ Widowed**		**Separated ⟶ Married**		**Widowed⟶Married**	
	OR (SE)	*95% CI*	*OR (SE)*	*95% CI*	*OR (SE)*	*95% CI*	*OR (SE)*	*95% CI*	*OR (SE)*	*95% CI*
Gender(Ref::Female)										
Male	1.19 (0.10)	0.90; 3.36	1.55 (0.31)	0.93; 4.03	0.18 (0.01)	0.15;0.33*	0.88 (0.44)	00.46; 1.34	1.31 (0.39)	0.92; 2.23
Income(Ref:Yes)										
No	0.77 (0.10)	0.574; 0.966	0.77 (0.48)	0.29; 1.25	1.08 (0.09)	0.9 ;1.26	0.47 (0.3)	0.02; 0.92	0.88(0.15)	0.73; 1.03
Is Employed(Ref:Yes)										
No	1.20 (0.1)	1.01; 1.40	0.32 (0.13)	0.19; 0.45	0.92 (0.09)	0.9; 1.1	1.46 (0.66)	0.47; 2.45	1.43(0.21)	1.22; 1.64
Highest Educ(Ref: Never Went Sch)										
Primary	0.65 0.22	0.22 1.08	4.21 (2.02)	2.19; 6.23	0.87 (0.09)	0.90; 1.05	1.82 (0.96)	0.38; 3.26	1.07 (0.12)	0.95 ; 1.19
High School	0.91 (0.12)	0.67; 1.15	2.98 (1.01)	1.97; 3.99	0.57 (0.08)	0.49; 0.73	0.61 (0.38)	0.04; 1.18	0.88 (0.19)	0.69; 1.07
Tertiary	1.04 (0.33)	0.60; 1.98	0.87 (1.39)	0.04; 19.85	0.34 (0.11)	0.18; 0.63	0.22 (0.00)	0.18; 0.63*	0.99 (0.47)	0.39; 2.50
Age at First Sex	1.04 (0.01)	1.02; 1.06	1.03 (0.02)	1.01; 1.05	1.01(0.01)	1.00; 1.03	1.02 (0.04)	0.96; 1.08	1.08 (0.05)	1.03; 1.13
Var_{event}	0.36 (0.08)	0.25; 0.47	0.36 (0.08)	0.25; 0.47	0.36 (0.08)	0.25;0.47	0.36 (0.08)	0.25; 0.47	0.36 (0.08)	0.25; 0.47
Cov_{events}	0.05 (0.02)	0.01; 0.09	0.05 (0.02)	0.01; 0.09	0.05 (0.02)	0.01; 0.09	0.05 (0.02)	0.01; 0.09	0.05 (0.02)	0.01; 0.09
AIC	1165									

5.12 Model Diagnostics

The AIC the multivariate discrete-time suvrival model with a piece-wise baseline hazard and a compound symmetry covariance structure was 1827.01 which is bigger than the AIC's from any of the univariate models.

5.13 Conclusions

The chapter aimed to construct a multivariate discrete-time multi-state model using binary survival models. This is attractive because in addition to estimates of parameters associated with the covariates, it produced estimates of correlations between and within events. Although it produced smaller standard errors than univariate models, this approach resulted in high dimensional joint models which are computationally intensive. Manipulations such as reducing number of quadrature points, number of iterations and choosing a simple working covariance structure were done in an attempt to make the model run. The next chapter attempts to reduce the dimensionality of the joint models, although at the expense of simplicity of the sub-models. Instead of binary survival sub-models, competing risks sub-models will be used.

CHAPTER 6

MULTIVARIATE COMPETING RISKS MODEL IN DISCRETE TIME

6.1 Introduction

Instead of considering joint models of binary transitions in a multi-state process, one may model the hazard of all possible transitions on the condition of being in one particular state, for all states (Jackson et al., 2011). As a result, multivariate competing risks models are considered. Section 6.2 gives a joint model for competing risks. Results are then presented in section 6.3, where compound symmetry covariance structure was utilized and it's model diagnostic is done in section 6.4.

6.2 Multivariate Model for Multinomial Responses

As discussed in Chapter 2, discrete-time competing risks models may utilize the multinomial logit framework (Hartzel et al., 2001; Tutz & Hennevogl, 1996; Tutz et al., 2016). For subject k who is in state i at time interval t, the hazard of failure due to cause j ($t = 1, \ldots, \tau_{kij}; \quad j = 1, \ldots, J_i; \quad i = 1, \ldots, m; \quad k = 1, \ldots K$) is given by

$$\lambda_{kij}(t|\boldsymbol{Z}, \boldsymbol{u}_{kij}) = \frac{\exp\left(\beta_{0ijt} + \boldsymbol{Z}'\boldsymbol{\beta}_{ij} + u_{kij}\right)}{1 + \sum_{b=1}^{J_i} \exp\left(\beta_{0ib} + \boldsymbol{Z}'\boldsymbol{\beta}_{ib} + u_{kij}\right)}, \tag{6.1}$$

where β_{0ijt} is the baseline hazard for transition ij occurring, $\boldsymbol{\beta}_{ij}$ is a $p \times 1$ vector of parameters associated with the covariates and $\boldsymbol{u}_{ki} = (u_{ki1}, \ldots, u_{kiJ_i})$ is a $J_i \times 1$ vector of the random effects for subject k, such that $\boldsymbol{u}_{ki} \sim \text{MVN}\left(\boldsymbol{0}, \boldsymbol{\Omega}_{ui}\right)$

$$\boldsymbol{u}_{ki} \sim \text{MVN}\left(\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\sigma}_{u_{i1}} & \boldsymbol{\sigma}_{u_{i1}u_{i2}} & \cdots & \boldsymbol{\sigma}_{u_{i1}u_{iJ_i}} \\ & \boldsymbol{\sigma}_{u_{i2}} & \cdots & \boldsymbol{\sigma}_{u_{i2}u_{iJ_i}} \\ & & \ddots & \vdots \\ & & & \boldsymbol{\sigma}_{u_{iJ_i}} \end{pmatrix}\right).$$

The likelihood for subject k leaving the origin state i, over all states in the process, which is modeled by a multinomial distribution in 6.1, will then have the form

$$L_k(\boldsymbol{\beta}, \boldsymbol{\Omega}_{ui}) = \prod_{i=1}^{m} \prod_{t=1}^{\tau_{kij}} \left(\prod_{j=1}^{J_i} \lambda_{kij} \left(t|\boldsymbol{Z}, \boldsymbol{u}_{kij}\right)^{\epsilon_{ktij}} \right) \left(1 - \sum_{j=1}^{J_i} \lambda_{kij} \left(t|\boldsymbol{Z}, \boldsymbol{u}_{kij}\right)^{\epsilon_{ktii}} \right) \times f(\boldsymbol{u}_{kij}) \tag{6.2}$$

where ϵ_{ktij} has been defined earlier in section 4.4.3. The complete joint likelihood over all subjects is then

$$L(\boldsymbol{\beta}, \boldsymbol{\Omega}_{ui}) \propto \prod_{k=1}^{K} \prod_{i=1}^{m} \prod_{t=1}^{\tau_{kij}} \left(\prod_{j=1}^{J_i} \lambda_{kij} \left(t|\boldsymbol{Z}, \boldsymbol{u}_{kij}\right)^{\epsilon_{ktij}} \right) \left(1 - \sum_{j=1}^{J_i} \lambda_{kij} \left(t|\boldsymbol{Z}, \boldsymbol{u}_{kij}\right)^{\epsilon_{ktii}} \right), \times -\frac{1}{2} \boldsymbol{u}'_{ki} \boldsymbol{\Omega}_{ui}^{-1} \boldsymbol{u}_{ki}. \tag{6.3}$$

and the corresponding marginal likelihood integrated over all state origins and all random effects is

$$L(\boldsymbol{\beta}, \boldsymbol{\Omega}_{ui}) \propto \sum_{k=1}^{K} \sum_{i=1}^{m} \int \dots \int \prod_{t=1}^{\tau_{kij}} \left(\prod_{j=1}^{J_i} \lambda_{kij} \left(t|\boldsymbol{Z}, \boldsymbol{u}_{kij}\right)^{\epsilon_{ktij}} \right) \left(1 - \sum_{j=1}^{J_i} \lambda_{kij} \left(t|\boldsymbol{Z}, \boldsymbol{u}_{kij}\right)^{\epsilon_{ktii}} \right) \times -\frac{1}{2} \boldsymbol{u}'_{ki} \boldsymbol{\Omega}_{ui}^{-1} \boldsymbol{u}_{ki} \quad d\boldsymbol{u}_{ki1} \dots d\boldsymbol{u}_{kiJ_i} \tag{6.4}$$

To approximate these integrals, various methods may be used. These have been discussed in Chapter 2 and 3. In this section, focus is on the Gauss-Hermite quadrature method which uses weights and nodes as previously described. To increase efficiency, this approach centers the nodes with respect to the mode of the function that is being integrated and scales them according to the estimated curvature at the mode. By so doing, the number of quadrature points needed to approximate the integrals effectively is dramatically reduced (Hartzel et al., 2001; Liu & Pierce, 1994;

Pinheiro & Bates, 1995). Let

$$\left(\prod_{j=1}^{J_i}\lambda_{kij}\left(t|\boldsymbol{Z},\boldsymbol{u}_{kij}\right)^{\epsilon_{ktij}}\right)\left(1-\sum_{j=1}^{J_i}\lambda_{kij}\left(t|\boldsymbol{Z},\boldsymbol{u}_{kij}\right)^{\epsilon_{ktii}}\right)$$

from likelihood 6.2 be re-written as $h(\bar{\boldsymbol{y}}_{kij}|\boldsymbol{u}_{ki};\Omega_{ui})$, according to properties of the exponential family of distributions, formerly Koopman-Darmois family. Then the integrand of likelihood 6.2 may now be written as

$$I=\prod_{t=1}^{\tau_{kij}}\left(h(\bar{\boldsymbol{y}}_{kij}|\boldsymbol{u}_{ki};\Omega_{ui})\right)\times f(\boldsymbol{u}_{ki};\Omega_{ui}).$$

The mode $\boldsymbol{\omega}_{ij}$ of the integrand I is calculated for category j and the original Gauss-Hermite nodes are then centred about that point. These centred nodes are scaled according to the curvature of the integrand around the mode to further localize the quadrature points closer to the bulk of the integrand. The inverse of the negative of the second derivative matrix of the integrand which was evaluated at the estimated mode, will then produce an estimate of the curvature , $\hat{\boldsymbol{Q}}_{ij}$, at the mode of the integrand. Denoting the original Gauss-Hermite nodes by $\mathbf{z_l}=(z_{l_1},\ldots,z_{l_{J_i}})$, where J_i is the dimension of the vector $\boldsymbol{u}_{ij}$, the adaptive Gauss-Hermites nodes for the ij^{th} transition are

$$\mathbf{z}^*_{ij\mathbf{l}}=\hat{\boldsymbol{\omega}}_{ij}+\sqrt{2}\hat{\boldsymbol{Q}}_{ij}^{\frac{1}{2}}\mathbf{z_l},$$

where $\hat{\boldsymbol{Q}}_{ij}^{\frac{1}{2}}$ comes from the Cholesky decomposition of $\hat{\boldsymbol{Q}}_{\boldsymbol{ij}}$, curvature that had previously been estimated. Now, the contribution to the likelihood of the ij^{th} transition, using the transformed Gauss–Hermite nodes for each category, will be approximated by

$$L_{ij}(\boldsymbol{\beta},\boldsymbol{\Sigma}_u)\approx|\hat{\boldsymbol{Q}}|^{\frac{1}{2}}2^{\frac{J_i}{2}}\sum_{\forall\mathbf{l}}\mathrm{w_l}\left(\prod_{t=1}^{\tau_{kij}}\left(h(\bar{\boldsymbol{y}}_{kij}|\mathbf{z}^*_{ij\mathbf{l}};\boldsymbol{\phi})\right)\times f(\mathbf{z}^*_{ij\mathbf{l}};\boldsymbol{\Omega}_{ui})\exp(\mathbf{z}'_{\mathbf{l}}\mathbf{z_l}),\right.$$

where $\mathrm{w_l}=\prod_{t=1}^{J_i}(\mathrm{w}_{1_t})$ are the weights for the original Gauss-Hermite. Therefore, the intractable integrals are approximated by a finite summation, with each of the J_i summations taken over K quadrature points. The number of quadrature points, K, is sequentially increased until both the estimates and standard errors produce negligible changes.

6.2.1 Inference and Estimation Using Statistical Software

This approach produces a smaller dimension of the models to be jointly modeled than the multivariate binary survival models in Chapter 5. Though more complicated, it possibly becomes less computationally intensive. For analysis, ***proc NLMIXED*** will be used and this, by default, performs MLE via adaptive Gauss-Hermite quadrature. Piece-wise constant (general) baseline hazards were conveniently considered. In addition, a compound symmetry covariance structure for the random effects effects was considered. Heavy duty machines would be required for more complicated covariance structures. Random effects for transitions of leaving each state are assumed not correlated between origin states. From Figure 3.1, it can be noticed that there are $i = 1, .., 4$ states, where $i = 1$ is the *Never Married* state which has only one possible transition ($j = 1$, into a *Married* state). On condition of occupying the *Married* state, $i = 2$ and $j = 1, 2$ which denote transitions into the *Separated* and *Widowed* states, respectively. For $i = 3$ and $i = 4$, they both have only one possible transition into the $Married$ each.

6.3 Results of a Multivariate Competing Risks Model

Results from the multivariate model with constant baseline under the assumption of compound symmetry variance structure are displayed in Table 6.1. Estimates are not significantly deviating from results of the univariate competing risks models presented in Chapter 4. They have negligibly smaller standard errors and narrower confidence intervals compared to estimates from the univariate models. The model gives rise to additional parameters measuring dependencies of events, namely, variance and covariance of the transitions. The variance for each transition, under the compound symmetry assumption of the covariance structure, was 9.56 (CI=$3.27; 15.85$), which is slightly higher than the one produced by the multivariate survival model. Covariance was 4.54(CI=$-0.25; 9.32$) which gives a correlation coefficient of 0.47. This implies a weak and positive correlation between transitions. The value of the AIC was reduced, from those produced by the univariate models.

Results show that the odds of experiencing a *Married* $\rightarrow$ *Widowed* transition were significantly lower for males than females (OR=0.07, CI=$0.07; 0.09$). Higher levels of education were associated with lower odds of marital dissolution through widowhood (OR=0.99 CI=$0.76; 1.22$, OR=0.82 CI=$0.68; 0.96$ and OR=0.51 CI=$0.37; 0.65$ for primary, secondary and tertiary education, respectively), although the effect was not statistically significant. Subjects with no employment had significantly higher odds of entering a first marriage (OR=1.26, CI=$1.02; 1.50$) and exiting a marriage through

separation (OR=1.25, CI=1.17; 1.33).

Table (6.1) Results for the multivariate competing risks model

	N.Married ⟶ Married		**Married ⟶ Separated**		**Married ⟶ Widowed**		**Separated ⟶ Married**		Widowed⟶**Married**	
	OR (SE)	*95% CI*	*OR (SE)*	*95% CI*	*OR (SE)*	*95% CI*	*OR (SE)*	*95% CI*	*OR (SE)*	*95% CI*
Gender(Ref:Female)										
Male	1.18 (0.12)	0.94; 1.42	1.17 (0.47)	0.72; 1.62	0.07 (0.01)	0.05; 0.09*	0.97 (0.55)	0.44; 1.50	1.69 (0.38)	0.95; 3.65
Income(Ref:Yes)										
No	0.88 (0.12)	0.64; 1.12	0.88 (0.32)	0.58; 1.18	1.05 (0.10)	0.86; 1.25	0.72 (0.30)	0.43; 1.01	0.76 (0.11)	0.54; 2.72
Is Employed(Ref:Yes)										
No	1.26 (0.12)	1.02; 1.50	1.25 (0.08)	1.17; 1.33	0.99 (0.08)	0.83; 1.15	1.49 (0.77)	0.75; 2.23	1.45 (0.19)	1.08; 3.41
Highest Educ(Ref: Never Went Sch)										
Primary	0.86 (0.11)	0.64; 1.08	4.28 (4.14)	0.35; 8.21	0.99 (0.12)	0.76; 1.22	1.82 (0.99)	0.87; 2.77	1.25 (0.19)	0.87; 1.63
High School	0.99 (0.1)0	0.79; 1.19	4.77 (5.00)	0.02; 9.52	0.82 (0.07)	0.68; 0.96	0.66 (0.37)	0.30; 1.02	0.98 (0.2)	0.59; 1.37
Tertiary	1.05 (0.33)	0.40; 1.70	0.98 (1.02)	0.01; 1.95	0.51 (0.07)	0.37; 0.65	0.00 (0.00)	0.00;0. 0 0	0.95 (0.38)	0.21; 1.70
Age at First Sex	1.02 (0.01)	1.00; 1.04	1.03 (0.02	) 1.01; 1.05	1.05 (0.01)	1.03; 1.07	1.04 (0.04)	1.00; 1.08	1.15 (0.01)	1.13; 1.17
Var_{event}	9.56 (3.21)	3.27;15.85	9.56 (3.21)	3.27;15.85	9.56 (3.21)	3.27;15.85	9.56 (3.21)	3.27;15.85	9.56 (3.21)	3.27;15.85
Cov_{events}	4.54 (2.31)	-0.25; 9.32	4.54 (2.31)	-0.25; 9.32	4.54 (2.31)	-0.25; 9.32	4.54 (2.31)	-0.25; 9.32	4.54 (2.31)	-0.25; 9.32
AIC	1 082.36									

6.4 Model Diagnostics

The AIC the multivariate discrete-time competing risks model under a compound symmetry covariance structure was 5931.14 which is bigger than the AIC's from the univariate model.

6.5 Conclusion

This chapter attempted to minimize the run time of a multi-state model using the joint synthesis of competing risks models. On the condition that one is in a particular state, the hazard of transition out of it was modeled, and this was done for each state in the multi-state process. Of course, out of some states, it is possible to have only one possible transition and from other states (which are absorbing), it is possible to have no possible transitions. The associational structure across transitions was imposed by assuming a multivariate normal distribution of the random effects, whose covariance structure was a compound symmetry. This model, although it had fewer sub-models, produced comparable estimates to that of competing risks model from Chapter 2 and univariate survival models. Age at first sexual debut was still a significant factor with a positive effect on most transitions.

CHAPTER 7

DISCUSSION AND CONCLUSION

7.1 Final Discussions

Univariate discrete time-to-event models are now commonly used to model transitions between recurrent states, thus measuring within-subject heterogeneity. These have been preferred over the continuous-time methods as they have a special advantage of automatic handling of time-varying covariates, thus avoiding tie-handling issues (Allison, 1982). However, not much has been done in modeling a discrete-time multi-state process to investigate covariate effect on the various transitions while accounting for possible dependencies between the transitions themselves.

This thesis was set out to construct a multivariate discrete time-to-event model that jointly estimates transitions between a multi-state process through random effects which followed a multivariate normal distribution, to account for the correlation between transitions. Unlike univariate discrete-time survival models (Steele, 2008; Manda & Meyer, 2005; Jenkins, 2005), the proposed models offer a better understanding of the evolution of all the transitions in a multi-state process simultaneously. The first multivariate model was executed via joint modeling of all binary transitions in the multi-state process. Since this resulted in high dimensional joint models which are computationally intensive, the second multivariate model comprised of joint models in competing risks, which model transitions out of the respective states. In both models, maximum likelihood based estimation of parameters was developed using Gauss-Hermite quadratures numerical integration of the likelihood. These models provide substantive extensions to the basic time-to-event models which may now be routinely implemented using standard statistical software packages.

The proposed methods were demonstrated using a prospective data set to model transitions between marital states among rural South Africans. Consistent with previous research (Hosegood et al., 2009; Claassens et al., 2013), effects of risk factors of the transitions between various marital states were produced. Standard errors from the multivariate survival and multivariate competing risks models were negligibly smaller than those from univariate models, thus exhibiting more precision. Results from these models did not show much of a difference in the estimates of the param-

eters, though there were extra parameters produced to measure dependence of the transitions. These showed the presence of very small unobserved heterogeneity due to subjects for each transition and a weak positive correlation between events. A key finding which was consistent across the models and with previous studies, was that the age at sexual debut was one of the prominent factors that significantly contributed to marital dissolution. Subjects who had their first sex at an early age were more likely to experience a marital dissolution than those who had a late sexual debut. As expected, the odds of experiencing a *Married* $\rightarrow$ *Widowed* transition increases as one gets older.

The proposed methods offer a useful extension of the methodology developed by Spedicato et al. (2016) to incorporate covariate effects into a discrete-time multi-state model beyond Markov assumptions. An option for modeling dependence structures through random effects with a multivariate normal distribution was provided. The models and estimation give significant and substantive extensions to the standard univariate time-to-event models and the associated estimation techniques. They may now be implemented using statistical software packages.

7.2 Limitations, Future Research and Recommendations

Spatial analysis was not done. The study was conducted in the rural areas of KwaZulu-Natal which is a small part of a province in South Africa where there might be heavy correlation due to culture, race or traditional beliefs. A model with a spatial component would also be appropriate to determine which areas need more interventions than others. It would be of interest to extend the study to different parts of the country including subjects in urban areas and with various cultures and economic or health systems. This would make it possible to incorporate spatial analysis which can be done to allow for any possible correlations due to geographical locations. This would be necessary if results need to be comparable and be able to identify exactly where resources need to be allocated. In addition, the results may not be generalized to the whole of South Africa as it wa only collected in one part of the province. Hence there might be a need to perform a study that covers the whole country for the results to be generalized.

The data used was population-based. In cases where sampling is done to reduce bias, such as in Demographic Health Survey (DHS) data, it would also be recommended to apply statistically correct survey weights on the methods.

Additionally, the random effects in this study were conveniently modeled using a normal distribution but a different distribution (such as gamma mixture, by Jenkins (1997)) may also be considered, where different assumptions are made about the subjects. A non-parametric approach (in the sense of lack of assumption about the random effects distribution), such as Manda (2011), may be considered as there might exist some unknown effects of misspecification of the random effects distribution (Hartzel et al., 2001).

When considering transitions into a first marriage, a very large proportion of subjects remained in the *Never Married* state and never made a transition. Cure models (Schneider, 2019) may also be considered where some individuals never experience the events of interest (non-susceptible group) and the survival function never turns to zero.

The data used possibly has its own shortcomings, in terms of low rates of transition relative to the large sample size. The study did not employ simulation to assess the behavior of the proposed method under varying number of events relative to sample sizes. This may be considered in future research.

7.3 Conclusion

In conclusion, this thesis has extended the multilevel discrete time-to-event to model highly complex multi-state structures. In the studied context of family formation and dissolution transitions, the complex structure arose because several types of transition occurred repeatedly over time, with interdependences between the different event transitions. Our multivariate approach has offered a better understanding of the processes and dynamics of family formations and dissolutions in rural South Africa. The use of multivariate analyses of the different transitions of a multi-state process resulted in slightly smaller standard errors of the parameter estimates and negligibly narrower confidence intervals of the fixed effects parameter estimates as well as smaller AIC values compared to the univariate analyses, which could point to better performance. We conclude that the proposed multivariate discrete-time-to event model which takes into account the between- and within-transition correlation is better than separate univariate analyses. The proposed multivariate discrete-time multi-state models provides an alternative method for analyzing such data and may be recommended to avoid biased parameter estimates.

REFERENCES

Aalen, O., Borgan, O., & Gjessing, H. (2008). *Survival and event history analysis: a process point of view*. Springer Science & Business Media.

Agresti, A. (2003). *Categorical data analysis*, vol. 482. John Wiley & Sons.

Agresti, A., Booth*, J. G., Hobert*, J. P., & Caffo*, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*, *30*(1), 27–80.

AHRI (2018). Africa health research institute. `https://www.ahri.org/`. [Online; accessed 28 September - 2018].

AHRI (2021). . `https://www.ahri.org/`. [Online; accessed 12-October-2021].

Ali, R., Gabr, A., Abouchaleh, N., Al Asadi, A., Mora, R. A., Kulik, L., Abecassis, M., Riaz, A., Salem, R., & Lewandowski, R. J. (2018). Survival analysis of advanced hcc treated with radioembolization: Comparing impact of clinical performance status versus vascular invasion/metastases. *Cardiovascular and interventional radiology*, *41*(2), 260–269.

Allignol, A., Schumacher, M., Wanner, C., Drechsler, C., & Beyersmann, J. (2011). Understanding competing risks: a simulation point of view. *BMC medical research methodology*, *11*(1), 86.

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological methodology*, *13*, 61–98.

Allison, P. D. (2001). *Missing data*, vol. 136. Sage publications.

Allison, P. D. (2014). *Event history and survival analysis: Regression for longitudinal event data*, vol. 46. SAGE publications.

Andersen, P. K., Abildstrom, S. Z., & Rosthøj, S. (2002). Competing risks as a multi-state model. *Statistical methods in medical research*, *11*(2), 203–215.

Andersen, P. K., Borgan, O., Gill, R. D., & Keiding, N. (2012). *Statistical models based on counting processes*. Springer Science & Business Media.

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

Andersen, P. K., & Keiding, N. (2002). Multi-state models for event history analysis. *Statistical methods in medical research*, *11*(2), 91–115.

Aralis, H. J. (2016). *Modeling Multistate Models with Back Transitions: Statistical Challenges and Applications*. Ph.D. thesis, UCLA.

Asanjarani, A., Liquet, B., & Nazarathy, Y. (2021). Estimation of semi-markov multistate models: a comparison of the sojourn times and transition intensities approaches. *The International Journal of Biostatistics*.

Austin, P. C. (2017). A tutorial on multilevel survival analysis: methods, models and applications. *International Statistical Review*, *85*(2), 185–203.

Ayele, D., Zewotir, T., & Mwambi, H. (2014). Modelling the joint determinants of a positive malaria rapid diagnosis test result, use of mosquito nets and indoor residual spraying with insecticide. *Occupational Health Southern Africa*, *20*(4), 20–27.

Barbu, V., Bérard, C., Cellier, D., Sautreuil, M., & Vergne, N. (2018). Smm: An r package for estimation and simulation of discrete-time semi-markov models.

Batidzirai, J. M., Manda, S. O., Mwambi, H. G., & Tanser, F. (2020). Discrete survival time constructions for studying marital formation and dissolution in rural south africa. *Frontiers in Psychology*, *11*.

Bel, K., & Paap, R. (2014). A multivariate model for multinomial choices. Tech. rep.

Berger, M., Schmid, M., Welchowski, T., Schmitz-Valckenberg, S., & Beyersmann, J. (2018). Subdistribution hazard models for competing risks in discrete time. *Biostatistics*.

Berger, M., Schmid, M., Welchowski, T., Schmitz-Valckenberg, S., & Beyersmann, J. (2020). Subdistribution hazard models for competing risks in discrete time. *Biostatistics*, *21*(3), 449–466.

Beyersmann, J., Allignol, A., & Schumacher, M. (2011). *Competing risks and multistate models with R*. Springer Science & Business Media.

Biggeri, L., Bini, M., & Grilli, L. (2001). The transition from university to work: a multilevel approach to the analysis of the time to obtain the first job. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *164*(2), 293–305.

Bijwaard, G. E. (2014). Multistate event history analysis with frailty. *Demographic Research*, *30*, 1591.

Blossfeld, H.-P. (2012). *Event history analysis with Stata*. Psychology Press.

Blossfeld, H.-P., Golsch, K., & Rohwer, G. (2007). Techniques of event history modeling using stata: New approaches to causal analysis.

Bock, R. D. (1970). Estimating multinomial response relations. *Contributions to statistics and probability*, (p. 453479).

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, *88*(421), 9–25.

Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education*, *97*(1), 65–108.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in medicine*, *25*(24), 4279–4292.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Chakrabarti, A., & Ghosh, J. K. (2011). Aic, bic, and recent advances in model selection. *Handbook of the philosophy of science*, *7*, 583–605.

Chamberlain, G. (1979). Analysis of covariance with qualitative data. Tech. rep., National Bureau of Economic Research.

Chinomona, A., & Mwambi, H. (2015). Multiple imputation for non-response when estimating hiv prevalence using survey data. *BMC public health*, *15*(1), 1059.

Claassens, A., Smythe, D., & Bradfield, G. (2013). *Marriage, Land and Custom: Essays on Law and Social Change in South Africa*. Juta and Company Ltd.

Clark, S., & Brauner-Otto, S. (2015). Divorce in sub-saharan africa: Are unions becoming less stable? *Population and Development Review*, *41*(4), 583–605.

Clark, S. J., Kahn, K., Houle, B., Arteche, A., Collinson, M. A., Tollman, S. M., & Stein, A. (2013). Young children's probability of dying before and after their mother's death: a rural south african population-based surveillance study. *PLoS medicine*, *10*(3), e1001409.

Cleves, M. (2008). *An introduction to survival analysis using Stata*. Stata Press.

Cook, R. J., Lawless, J. F., et al. (2007). The statistical analysis of recurrent events.

Cox, D. R. (1972a). Models and life-tables regression. *JR Stat. Soc. Ser. B*, *34*, 187–220.

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

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*, vol. 21. CRC Press.

Craig, B. A., & Sendi, P. P. (2002). Estimation of the transition matrix of a discrete-time markov chain. *Health economics*, *11*(1), 33–42.

Crowther, M. J. (2019). Multilevel mixed-effects parametric survival analysis: estimation, simulation, and application. *The Stata Journal*, *19*(4), 931–949.

Crowther, M. J., & Lambert, P. C. (2012). Simulating complex survival data. *The Stata Journal*, *12*(4), 674–687.

Crowther, M. J., Look, M. P., & Riley, R. D. (2014). Multilevel mixed effects parametric survival models using adaptive gauss–hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Statistics in medicine*, *33*(22), 3844–3858.

Czepiel, S. A. (2002). Maximum likelihood estimation of logistic regression models: theory and implementation. *Available at czep. net/stat/mlelr. pdf*, *83*.

De Boor, C., & De Boor, C. (1978). *A practical guide to splines*, vol. 27. springer-verlag New York.

De Leeuw, J., & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, *11*(1), 57–85.

De Wreede, L. C., Fiocco, M., & Putter, H. (2010). The mstate package for estimation and prediction in non-and semi-parametric multi-state and competing risks models. *Computer methods and programs in biomedicine*, *99*(3), 261–274.

de Wreede, L. C., Fiocco, M., Putter, H., et al. (2011). mstate: an r package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, *38*(7), 1–30.

Dean, D. O., Bauer, D. J., & Shanahan, M. J. (2014). A discrete-time multiple event process survival mixture (mepsum) model. *Psychological methods*, *19*(2), 251.

Diamond, I. D., McDonald, J. W., & Shah, I. H. (1986). Proportional hazards models for current status data: application to the study of differentials in age at weaning in pakistan. *Demography*, *23*(4), 607–620.

Diamond, P. A., & Hausman, J. A. (1984). Individual retirement and savings behavior. *Journal of Public Economics*, *23*(1-2), 81–114.

Diggle, P. (2002). *Analysis of longitudinal data*. Oxford university press.

Dobra, A., Bärnighausen, T., Vandormael, A., & Tanser, F. (2017). Space-time migration patterns and risk of hiv acquisition in rural south africa. *Aids (London, England)*, *31*(1), 137.

Doytchinov, B., & Irby, R. (2010). Time discretization of markov chains. *Pi Mu Epsilon Journal*, (pp. 69–82).

Duffy, A., Horrocks, J., Doucette, S., Keown-Stoneman, C., McCloskey, S., & Grof, P. (2014). The developmental trajectory of bipolar disorder. *The British Journal of Psychiatry*, *204*(2), 122–128.

Esquível, M. L., Guerreiro, G. R., Oliveira, M. C., & Corte Real, P. (2021). Calibration of transition intensities for a multistate model: application to long-term care. *Risks*, *9*(2), 37.

Eulenburg, C., Schroeder, J., Obi, N., Heinz, J., Seibold, P., Rudolph, A., Chang-Claude, J., & Flesch-Janys, D. (2016). A comprehensive multistate model analyzing associations of various risk factors with the course of breast cancer in a population-based cohort of breast cancer cases. *American journal of epidemiology*, *183*(4), 325–334.

Fahrmeir, L. (1997). Discrete failure time models.

Fahrmeir, L., & Knorr-Held, L. (1996). *Dynamic discrete-time duration models*. Ludwig-Maximilians-Univ., SFB 386.

Fahrmeir, L., & Knorr-Held, L. (1997). Dynamic discrete-time duration models: Estimation via markov chain monte carlo. *Sociological Methodology*, *27*(1), 417–452.

Fahrmeir, L., Tutz, G., Hennevogl, W., & Salem, E. (1994). *Multivariate statistical modelling based on generalized linear models*, vol. 425. Springer.

Fahrmeir, L., & Wagenpfeil, S. (1996). Smoothing hazard functions and time-varying effects in discrete duration and competing risks models. *Journal of the American Statistical Association*, *91*(436), 1584–1594.

Fang, D., Sun, R., & Wilson, J. R. (2018). Joint modeling of correlated binary outcomes: The case of contraceptive use and hiv knowledge in bangladesh. *PloS one*, *13*(1), e0190917.

Ferguson, N., Datta, S., Brock, G., et al. (2012). mssurv, an r package for nonparametric estimation of multistate models. *Journal of Statistical Software*, *50*(14), 1–24.

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, *94*(446), 496–509.

Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (2008). *Longitudinal data analysis*. CRC press.

Gage, A. J. (2013). Association of child marriage with suicidal thoughts and attempts among adolescent girls in ethiopia. *Journal of Adolescent Health*, *52*(5), 654–656.

Gähler, M., & Palmtag, E.-L. (2015). Parental divorce, psychological well-being and educational attainment: Changed experience, unchanged effect among swedes born 1892–1991. *Social Indicators Research*, *123*(2), 601–623.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, *73*(1), 43–56.

Goldstein, H. (2011). *Multilevel statistical models*, vol. 922. John Wiley & Sons.

Gray, R. J. (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, *87*(420), 942–951.

Grilli, L., & Rampichini, C. (2007). A multilevel multinomial logit model for the analysis of graduates' skills. *Statistical Methods and Applications*, *16*(3), 381–393.

Gutierrez, R. (2010). Competing-risks regression. In *StataCorp LP, Boston. Available: http://www. stata. com/meeting/boston10/boston10_gutierrez. pdf [Accessed 3 September 2012]*.

Habyarimana, F., Zewotir, T., & Ramroop, S. (2016). Joint modeling of poverty of households and malnutrition of children under five years from demographic and health survey data: case of rwanda. *Journal of Economics and Behavioral Studies*, *8*(2 (J)), 108–114.

Han, A., & Hausman, J. (1988). Identification of continuous and discrete competing risk models. Tech. rep., mimeo.

Han, A., & Hausman, J. A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of applied Econometrics*, *5*(1), 1–28.

Han, A. K. (1987). Semiparametric estimation of duration and competing risk models.

Han, Y., & Boves, L. (2007). Hierarchical acoustic modeling based on random-effects regression for automatic speech recognition. In *Eighth Annual Conference of the International Speech Communication Association*.

Harrell, F. E., et al. (2001). *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*, vol. 608. Springer.

Hartzel, J., Agresti, A., & Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, *1*(2), 81–102.

Hasselmo, K., Sbarra, D. A., O'Connor, M.-F., & Moreno, F. A. (2015). Psychological distress following marital separation interacts with a polymorphism in the serotonin transporter gene to predict cardiac vagal control in the laboratory. *Psychophysiology*, *52*(6), 736–744.

Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica: Journal of the Econometric Society*, (pp. 271–320).

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in medicine*, *22*(9), 1433–1446.

Helbert, Z. T. (2015). Modeling enrollment at a regional university using a discrete-time markov chain.

Hosegood, V., McGrath, N., & Moultrie, T. (2009). Dispensing with marriage: Marital and partnership trends in rural kwazulu-natal, south africa 2000-2006. *Demographic research*, *20*, 279.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, *73*(2), 387–396.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime data analysis*, *1*(3), 255–273.

Hougaard, P., Myglegaard, P., & Borch-Johnsen, K. (1994). Heterogeneity models of disease susceptibility, with application to diabetic nephropathy. *Biometrics*, (pp. 1178–1188).

Houle, B., Clark, S. J., Gómez-Olivé, F. X., Kahn, K., & Tollman, S. M. (2014). The unfolding counter-transition in rural south africa: mortality and cause of death, 1994–2009. *PLoS One*, *9*(6), e100420.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

Ikamari, L. (2005). The effect of education on the timing of marriage in kenya. *Demographic Research*, *12*, 1–28.

Jackson, C. (2007). Multi-state modelling with r: the msm package. *Cambridge, UK*.

Jackson, C. H., et al. (2011). Multi-state models for panel data: the msm package for r. *Journal of Statistical Software*, *38*(8), 1–29.

Jenkins, S. P. (1995). Easy estimation methods for discrete-time duration models. *Oxford bulletin of economics and statistics*, *57*(1), 129–136.

Jenkins, S. P. (1997). Estimation of discrete time (grouped duration data) proportional hazards models: pgmhaz. *Stata Technical Bulletin Reprints, STB*, *17*(7), 109–121.

Jenkins, S. P. (2005). Survival analysis. *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK*, *42*, 54–56.

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, *8*(3), 325–352.

Kandala, N.-B., & Ghilagaber, G. (2006). A geo-additive bayesian discrete-time survival model and its application to spatial analysis of childhood mortality in malawi. *Quality and Quantity*, *40*(6), 935–957.

Kaombe, T. M., & Manda, S. O. (2021). Detecting influential data in multivariate survival models. *Communications in Statistics-Theory and Methods*, (pp. 1–17).

Katz, L. F. (1986). Layoffs, recall and the duration of unemployment.

Kneib, T., & Hennerfeind, A. (2008). Bayesian semi parametric multi-state models. *Statistical Modelling*, *8*(2), 169–198.

Korn, E. L., & Whittemore, A. S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics*, (pp. 795–802).

Kridahl, L., & Silverstein, M. (2017). Parental survival and retirement timing in the swedish population. *Stockholm. https://doi. org/10.17045/sthlmuni*, *5607838*, v1.

Kryscio, R., Schmitt, F., Salazar, J., Mendiondo, M., & Markesbery, W. (2006). Risk factors for transitions from normal to mild cognitive impairment and dementia. *Neurology*, *66*(6), 828–832.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, (pp. 963–974).

Langford, I. H., & Lewis, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *161*(2), 121–160.

Laplace, P. S. (1986). Memoir on the probability of the causes of events. *Statistical science*, *1*(3), 364–378.

Lee, M., Feuer, E. J., & Fine, J. P. (2018). On the analysis of discrete time competing risks data. *Biometrics*, *74*(4), 1468–1481.

Lin, D. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in medicine*, *16*(8), 901–910.

Littell, R. C., & Milliken, G. A. (2006). *SAS for mixed models*.

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in medicine*, *19*(13), 1793–1819.

Little, R. J., & Rubin, D. B. (2014). *Statistical analysis with missing data*, vol. 333. John Wiley & Sons.

Liu, L., & Yu, Z. (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in medicine*, *27*(16), 3105–3124.

Liu, Q., & Pierce, D. A. (1994). A note on gauss—hermite quadrature. *Biometrika*, *81*(3), 624–629.

Ma, X. (2018). *Using classification and regression trees: A practical primer*. IAP.

Mallinckrodt, C. H., Sanger, T. M., Dubé, S., DeBrota, D. J., Molenberghs, G., Carroll, R. J., Potter, W. Z., & Tollefson, G. D. (2003). Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological psychiatry*, *53*(8), 754–760.

Manda, S., & Meyer, R. (2005). Age at first marriage in malawi: a bayesian multilevel analysis using a discrete time-to-event model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *168*(2), 439–455.

Manda, S. O. (2011). A nonparametric frailty model for clustered survival data. *Communications in Statistics—Theory and Methods*, *40*(5), 863–875.

Manda, S. O. M. (1998). Unobserved family and community effects on infant mortality in malawi. *Genus*, (pp. 143–164).

Mantel, N., & Hankey, B. F. (1978). A lofistic rlgression analysis of response-time data where the hazard function is time dependent. *Communications in Statistics-Theory and Methods*, *7*(4), 333–347.

Maqutu, D. (2010). *Statistical methods for longitudinal binary data structure with applications to antiretroviral medication adherence.*. Ph.D. thesis.

Markov, A. (1971). Extension of the limit theorems of probability theory to a sum of variables connected in a chain.

Markov, A. A. (1954). The theory of algorithms. *Trudy Matematicheskogo Instituta Imeni VA Steklova*, *42*, 3–375.

Masyn, K. E. (2009). Discrete-time survival factor mixture analysis for low-frequency recurrent event histories. *Research in Human Development*, *6*(2-3), 165–194.

Matsena Zingoni, Z., Chirwa, T. F., Todd, J., & Musenge, E. (2019). Hiv disease progression among antiretroviral therapy patients in zimbabwe: a multistate markov model. *Frontiers in public health*, *7*, 326.

Matsena Zingoni, Z., Chirwa, T. F., Todd, J., & Musenge, E. (2021). A review of multistate modelling approaches in monitoring disease progression: Bayesian estimation using the kolmogorov-chapman forward equations. *Statistical Methods in Medical Research*, *30*(5), 1373–1392.

Mazroui, Y., Mathoulin-Pélissier, S., MacGrogan, G., Brouste, V., & Rondeau, V. (2013). Multivariate frailty models for two types of recurrent events with a dependent terminal event: application to breast cancer data. *Biometrical Journal*, *55*(6), 866–884.

McCall, B. P. (1996). Unemployment insurance rules, joblessness, and part-time work. *Econometrica: Journal of the Econometric Society*, (pp. 647–682).

McCulloch, C. E., & Searle, S. R. (2004). *Generalized, linear, and mixed models*. John Wiley & Sons.

Mchunu, N. N. (2018). *Modelling CD4 count and Mortality in a cohort of patients initiated on HAART.*. Ph.D. thesis.

Mchunu, N. N., Mwambi, H. G., Reddy, T., Yende-Zuma, N., & Naidoo, K. (2020). Joint modelling of longitudinal and time-to-event data: an illustration using cd4 count and mortality in a cohort of patients initiated on antiretroviral therapy. *BMC infectious diseases*, *20*(1), 1–9.

Meira-Machado, L., de Uña-Álvarez, J., Cadarso-Suarez, C., & Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical methods in medical research*, *18*(2), 195–222.

Mills, M. (2011a). The fundamentals of survival and event history analysis. *Introducing Survival Analysis and Event History Analysis. London: SAGE Publications*, (pp. 1–17).

Mills, M. (2011b). *Introducing survival and event history analysis*. Sage Publications.

Molenberghs, G., & Kenward, M. (2007). *Missing data in clinical studies*, vol. 61. John Wiley & Sons.

Molenberghs, G., & Verbeke, G. (2000). Exploratory data analysis. *Linear Mixed Models for Longitudinal Data*, (pp. 31–40).

Molenberghs, G. V.-G., & Verbeke, G. (2005). Longitudinal and incomplete clinical studies. *Metron-International Journal of Statistics*, *63*(2), 143–176.

Molnar, F. J., Hutton, B., & Fergusson, D. (2008). Does analysis using "last observation carried forward" introduce bias in dementia research? *Cmaj*, *179*(8), 751–753.

Muenz, L. R., & Rubinstein, L. V. (1985). Markov models for covariate dependence of binary sequences. *Biometrics*, (pp. 91–101).

Narendranathan, W., & Stewart, M. B. (1991). Simple methods for testing for the proportionality of cause-specific hazards in competing risk models. *Oxford Bulletin of Economics and Statistics*, *53*(3), 331–340.

Narendranathan, W., & Stewart, M. B. (1993). Modelling the probability of leaving unemployment: competing risks models with flexible base-line hazards. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *42*(1), 63–83.

Nicholson, W. (2013). Dtmcpack: Suite of functions related to discrete-time discrete-state markov chains. *R package version 0.1-2, URL http://CRAN. R-project. org/package= DTMCPack*.

Oppenheimer, V. K. (2003). Cohabiting and marriage during young men's career-development process. *Demography*, *40*(1), 127–149.

Pinheiro, J., & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer science & business media.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, *4*(1), 12–35.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T., & Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, (pp. 541–554).

Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine*, *26*(11), 2389–2430.

Putter, H., et al. (2011). Special issue about competing risks and multi-state models. *Journal of Statistical Software*, *38*(1), 1–4.

Rasbach, J., Steele, F., Browne, W., & Prosser, B. (2003). A user's guide to mlwin version 2.0. *Centre for Multilevel Modelling, Institute of Education, University of London*.

Raudenbush, S., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of education*, (pp. 1–17).

Raudenbush, S. W., & Bryk, A. S. (1988). Chapter 10: Methodological advances in analyzing the effects of schools and classrooms on student learning. *Review of research in education*, *15*(1), 423–475.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, vol. 1. Sage.

Reddy, T., Mwambi, H., & Ndung'u, T. (2011). Modelling hiv progression using multistate markov models. In *Annual Proceedings of the South African Statistical Association Conference*, vol. 2011, (pp. 100–117). South African Statistical Association (SASA).

Regier, M. H. (1968). A two-state markov model for behavioral change. *Journal of the American Statistical Association*, *63*(323), 993–999.

Rice, N., & Leyland, A. (1996). Multilevel models: applications to health data. *Journal of Health Services Research*, *1*(3), 154–164.

Ripatti, S., & Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, *56*(4), 1016–1022.

Rizopoulos, D. (2012). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics & Data Analysis*, *56*(3), 491–501.

Roshani, D., & Ghaderi, E. (2016). Comparing smoothing techniques for fitting the nonlinear effect of covariate in cox models. *Acta Informatica Medica*, *24*(1), 38.

Rubin, D. (1987). Multiple imputation for missing data in sample surveys.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, vol. 81. John Wiley & Sons.

Sampford, M. (1952). The estimation of response-time distributions. ii. multistimulus distributions. *Biometrics*, *8*(4), 307–369.

Sampson, R. J., Laub, J. H., & Wimer, C. (2006). Does marriage reduce crime? a counterfactual approach to within-individual causal effects. *Criminology*, *44*(3), 465–508.

Satty, A., Babikir, A., & Basher, A. (2013). Handling dropouts in longitudinal clinical trials aata: Likelihood-based analysis versus inverse probablity weighting. *Arab Gulf Journal of Scientific Research*, *31*.

Satty, A., & Mwambi, H. (2012). Imputation methods for estimating regression parameters under a monotone missing covariate pattern: A comparative analysis. *South African Statistical Journal*, *46*(2), 327–357.

Satty, A., & Mwambi, H. (2014). A review of methods for handling missing data in the form of dropouts in longitudinal clinical trials. *Bull Pharm Med Sci (BOPAMS)*, *2*, 2310–31.

Satty, A., Mwambi, H., & Molenberghs, G. (2015). Different methods for handling incomplete longitudinal binary outcome due to missing at random dropout. *Statistical Methodology*, *24*, 12–27.

Sbarra, D. A., Law, R. W., Lee, L. A., & Mason, A. E. (2009). Marital dissolution and blood pressure reactivity: Evidence for the specificity of emotional intrusion-hyperarousal and task-rated emotional difficulty. *Psychosomatic Medicine*, *71*(5), 532.

Scheike, T. H., & Jensen, T. K. (1997). A discrete survival model with random effects: an application to time to pregnancy. *Biometrics*, (pp. 318–329).

Scheike, T. H., & Keiding, N. (2006). Design and analysis of time-to-pregnancy. *Statistical methods in medical research*, *15*(2), 127–140.

Schmid, M., & Berger, M. (2020). Competing risks analysis for discrete time-to-event data. *Wiley Interdisciplinary Reviews: Computational Statistics*, (p. e1529).

Schneider, M. (2019). Dealing with heterogeneity in discrete survival analysis using the cure model.

Searle, S. R., Casella, G., & McCulloch, C. E. (2009). *Variance components*. John Wiley & Sons.

Siddiqui, O., & Ali, M. W. (1998). A comparison of the random-effects pattern mixture model with last-observation-carried-forward (locf) analysis in longitudinal clinical trials with dropouts. *Journal of biopharmaceutical statistics*, *8*(4), 545–563.

Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of educational statistics*, *18*(2), 155–195.

Spedicato, G. A., Kang, T. S., Yalamanchi, S. B., & Yadav, D. (2016). The markovchain package: A package for easily handling discrete markov chains in r. *Accessed Dec*.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). Winbugs user manual.

Stata.com (2022). xtmlogit. `https://www.stata.com/manuals/xtxtmlogit.pdf`. [Online; accessed 25 February 2022].

Steele, F. (2005). *Event history analysis*. National Centre for Research Methods.

Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: series A (statistics in society)*, *171*(1), 5–19.

Steele, F. (2011). Multilevel discrete-time event history models with applications to the analysis of recurrent employment transitions. *Australian & New Zealand Journal of Statistics*, *53*(1), 1–20.

Steele, F., Diamond, I., & Amin, S. (1996). Immunization uptake in rural bangladesh: a multilevel analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *159*(2), 289–299.

Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modelling*, *4*(2), 145–159.

Steele, F., Kallis, C., Goldstein, H., & Joshi, H. (2005). The relationship between childbearing and transitions from marriage and cohabitation in britain. *Demography*, *42*(4), 647–673.

Steele, F., Sigle-Rushton, W., & Kravdal, Ø. (2009). Consequences of family disruption on children's educational outcomes in norway. *Demography*, *46*(3), 553–574.

Sutradhar, R., Barbera, L., Seow, H., Howell, D., Husain, A., & Dudgeon, D. (2010). Multistate analysis of interval-censored longitudinal data: application to a cohort study on performance status among patients diagnosed with cancer. *American journal of epidemiology*, *173*(4), 468–475.

Sweeting, M. J., De Angelis, D., & Aalen, O. O. (2005). Bayesian back-calculation using a multi-state model with application to hiv. *Statistics in medicine*, *24*(24), 3991–4007.

Tanser, F., Hosegood, V., Bärnighausen, T., Herbst, K., Nyirenda, M., Muhwava, W., Newell, C., Viljoen, J., Mutevedzi, T., & Newell, M.-L. (2007). Cohort profile: Africa centre demographic information system (acdis) and population-based hiv survey. *International journal of epidemiology*, *37*(5), 956–962.

Therneau, T., Crowson, C., & Atkinson, E. (2018). Multi-state models and competing risks. *CRAN-R (https://cran. r-project. org/web/packages/survival/vignettes/compete. pdf)*.

Tomita, A., Vandormael, A. M., Bärnighausen, T., de Oliveira, T., & Tanser, F. (2017). Social disequilibrium and the risk of hiv acquisition: A multilevel study in rural kwazulu-natal province, south africa. *Journal of acquired immune deficiency syndromes (1999)*, *75*(2), 164.

Tsiatis, A. (1998). Competing risks. *Encyclopedia of biostatistics*.

Tuma, N. B., Hannan, M. T., & Groeneveld, L. P. (1979). Dynamic analysis of event histories. *American journal of Sociology*, *84*(4), 820–854.

Tutz, G. (2011). *Regression for categorical data*, vol. 34. Cambridge University Press.

Tutz, G., & Hennevogl, W. (1996). Random effects in ordinal regression models. *Computational Statistics & Data Analysis*, *22*(5), 537–557.

Tutz, G., Schmid, M., et al. (2016). *Modeling discrete time-to-event data*. Springer.

Uecker, J. E. (2012). Marriage and mental health among young adults. *Journal of Health and Social Behavior*, *53*(1), 67–83.

Verbeke, G. (1997). Linear mixed models for longitudinal data. In *Linear mixed models in practice*, (pp. 63–153). Springer.

Vermunt, J. K. (1996). Log-linear event history analysis. *Series on Work and Organization*.

Waltz, M., Badura, B., Pfaff, H., & Schott, T. (1991). Marriage and the psychological consequences of heart attack: a longitudinal study of adaptation to chronic illness after 3 years. In *Public health*, (pp. 16–36). Springer.

Wang, L., Seelig, A., Wadsworth, S. M., McMaster, H., Alcaraz, J. E., & Crum-Cianflone, N. F. (2015). Associations of military divorce with mental, behavioral, and physical health outcomes. *BMC psychiatry*, *15*(1), 128.

Wang, Y. (2011). *Smoothing splines: methods and applications*. CRC press.

Welchowski, T., & Schmid, M. (2015). discsurv: Discrete time survival analysis. *R package version*, *1*(1), 1.

Willekens, F. (2014). *Multistate analysis of life histories with R*. Springer.

Willekens, F. J., Shah, I., Shah, J. M., & Ramachandran, P. (1982). Multi-state analysis of marital status life tables: Theory and application. *Population Studies*, *36*(1), 129–144.

Wolkewitz, M., Vonberg, R. P., Grundmann, H., Beyersmann, J., Gastmeier, P., Bärwolff, S., Geffers, C., Behnke, M., Rüden, H., & Schumacher, M. (2008). Risk factors for the development of nosocomial pneumonia and mortality on intensive care units: application of competing risks models. *Critical Care*, *12*(2), 1–9.

Wood, S. N., & Augustin, N. H. (2002). Gams with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological modelling*, *157*(2-3), 157–177.

Zhang, W., & Steele, F. (2004). A semiparametric multilevel survival model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *53*(2), 387–404.

APPENDIX A - PUBLICATIONS.

frontiers in Psychology
ORIGINAL RESEARCH
published: 18 February 2020
doi: 10.3389/fpsyg.2020.00154

Discrete Survival Time Constructions for Studying Marital Formation and Dissolution in Rural South Africa

Jesca M. Batidzirai[1]*, *Samuel O. M. Manda*[1,2,3], *Henry G. Mwambi*[1] and *Frank Tanser*[4,5,6]

[1] School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa, [2] Biostatistics Unit, South African Medical Research Council, Pretoria, South Africa, [3] Department of Statistics, University of Pretoria, Pretoria, South Africa, [4] Africa Health Research Institute, KwaZulu-Natal, South Africa, [5] Lincoln Institute for Health, University of Lincoln, Lincoln, United Kingdom, [6] School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa

Introduction: Marriage formation and dissolution are important life-course events which impact psychological well-being and health of adults and children experiencing the events. Family studies have usually concentrated on analyzing single transitions including *Never Married* to *Married* and *Married* to *Divorced*. This does not allow understanding and interrogation of dynamics of these life changing events and their effects on individuals and their families. The objective of this study was to assess determinants associated with transitions between and within marital states in South Africa.

Methods: The population-based data available for this study consists of over 55,000 subjects representing over 340,000 person-years exposure from the Africa Health Research Institute (AHRI) in rural KwaZulu-Natal, South Africa. It was collected from 1 January 2004 to 31 December 2016. Multilevel multinomial, binary and competing risks regression models were used to model marital state occupation, transitions between marital states as well as investigate determinants of marital dissolution, respectively.

Results: Between the years 2006 and 2007, a subject was more likely to be married than never married when compared to years 2004 – 2005. After 2007, subjects were less likely to be married than never married and the trend reduced over the years up to 2016 [with *OR*=0.86, *CI*=(0.78; 0.94), *OR*=0.71, *CI*=(0.64; 0.78), *OR*=0.60, *CI*=(0.54; 0.67), *OR*=0.50, *CI*=(0.44; 0.56), and $OR = 0.43, CI = (0.38; 0.48)$] for periods 2008 – 2009, 2010 – 2011, 2012 – 2013, 2014 – 2015, and 2016, respectively. In 2008 – 2009, subjects were more likely to experience a marital dissolution than in the period 2004 – 2005 and the trend slightly reduces from 2010 until 2013 [*OR*=24.49, *CI*=(5.53; 108.37)]. Raising age at first sexual debut was found to be inversely associated with a marital dissolution [$OR = 0.97$; $CI = (0.95; 0.99)$]. Highly educated subjects were more likely to stay in one marital state than those who never went to school [*OR*=6.43, *CI*=(4.89; 8.47), *OR*=18.86, *CI*=(1.14; 53.31), and OR=2.96, CI=(1.96; 4.46) for being married, separated and widowed, respectively, among subjects with tertiary education]. As the age at first marriage increased, subjects became less likely to experience a marital separation [$OR = 0.06, CI = (0.00; 1.11), OR = 0.05, CI = (0.00; 0.91)$, and $OR = 0.04, CI = (0.00; 0.76)$ for subjects who entered a first marriage at ages 18 – 22, 23 – 29, and 30 – 40, respectively].

OPEN ACCESS

Edited by:
Jim Todd, London School of Hygiene and Tropical Medicine, University of London, United Kingdom

Reviewed by:
Mwita Wambura, National Institute of Medical Research, Tanzania
Asungushe Bonaventura Kayombo, London School of Hygiene and Tropical Medicine, University of London, United Kingdom

***Correspondence:**
Jesca M. Batidzirai
batidzirai@ukzn.ac.za

Specialty section:
This article was submitted to Health Psychology, a section of the journal Frontiers in Psychology

Received: *17 May 2019*
Accepted: *21 January 2020*
Published: *18 February 2020*

Citation:
Batidzirai JM, Manda SOM, Mwambi HG and Tanser F (2020) Discrete Survival Time Constructions for Studying Marital Formation and Dissolution in Rural South Africa. Front. Psychol. 11:154. doi: 10.3389/fpsyg.2020.00154

APPENDIX A- ABSTRACT OF THE MANUSCRIPT SUBMITTED TO THE JOURNAL OF APPLIED STATISTICS

ARTICLE TEMPLATE

A Multivariate Discrete Time-to-Event Model for Multiple Recurring Events

Jesca. M. Batidzirai[a], Samuel. O. M. Manda [b], Henry. G. Mwambi[a], and Frank Tanser[c]

[a]School of Mathematics, Statistics and Computer Science, University of KwaZulu- Natal, Pietermaritzburg, South Africa;
[b]Department of Statistics, University of Pretoria, Hatfield, South Africa;
[c]Research Department of Infection & Population Health, University College London, London,UK

ARTICLE HISTORY
Compiled November 19, 2022

ABSTRACT
Recent developments in multi-state models have often considered discrete time in the modeling transition intensities. These models have included univariate multilevel models to account for possible dependence among events that are recurrent in the same subject. We propose a multivariate discrete-time survival model with multiple state transitions where each specific transition has its own separate random effect and the interest is on measuring dependence both between-and within-transitions. Multivariate Normal model for the random effects is suggested. The model parameters are estimated using maximum likelihood methods with non-adaptive Gaussian quadratures numerical integration. The proposed methodology was applied to a real-life data set from an ongoing longitudinal study based on life course history on marriage formation and dissolution events in rural KwaZulu-Natal of South Africa. The five events under consideration in the multi-state process were transition into a first marriage, exiting a marriage through separation, exiting a marriage through death of a partner, re-marriage after a separation and remarriage from widowhood. The model produced smaller standard errors and narrower confidence intervals compared to those from univariate models. Results showed the presence of very small unobserved subject-to-subject heterogeneity for each transition and a weak positive correlation between events.

KEYWORDS
Discrete-time-to-event, Multi-state models, Random effects, Multivariate Competing risks, Gauss-Hermite

1. Introduction

Multistate models have been proposed and extensively used in biostatistics, demography and economics [1, 19, 50, 56, and the references therein] and may be used to inform how subjects evolve through the states in a stochastic process. Multistate models are important in providing the general trajectories through intermediate states [30]. A common case scenario is observed where time is treated as continuous, such that the exact time of event occurrence is known, since it is modeled at every

CONTACT J. M. Batidzirai. Email: Batidzirai@ukzn.ac.za

APPENDIX B- MAP OF SOUTH AFRICA

(a) Africa **(b)** South Africa **(c)** KZN Province

Figure (7.1) Data Collection Area

APPENDIX C-DATA PREPARATION

Data preparation for Chapter 3

Depending on the time scale of interest, either age or calendar time may be used. If age is used as the time scale,then the effect of age as a covariate will no longer have to modeled because its effect will now be represented in the baseline. Age is then considered as a discrete random variable throughout this chapter. We consider subjects between ages 17 and 65, which are the ages for subjects being able to make independent marital decisions in society. For this analysis, a subject's marital status is assumed to change not more than once within a period of 2 years, hence we chose the size of the intervals to be 2-year intervals, $[17, 19), [19, 21), ...[64, 65]$ years. This step of grouping of durations was found to have minimal loss of precision (Diamond et al., 1986). By so doing, the size of the data set was reduced to fewer records which had an added advantage of manageability. Yearly intervals are common in family formation and dissolution studies as subjects rarely change marital statuses more than once in a single year (Sampson et al., 2006; Oppenheimer, 2003), so are 2- year intervals. If shorter intervals are considered, one risks enlarging the data set while no little or no transitions occurring in most of the intervals. Steele et al. (2004) went on to point out that where the hazard function and covariate values are constant within an interval, grouping will not necessarily lead to information loss as long as the intervals are weighted by exposure time, which (in this case) is 24 months. This is the number of months during that interval for which a subject is under the exposure of risk of observing an event (Steele et al., 2004).

Data Arrangement for Chapter 3

One important variable which must be created (if not already there) is a unique identifier variable for each subject. In the data snapshot in Table 7.1 below, this variable is labeled *ID*. It shows all observations which are made for each single individual. For discrete-time analysis, the data should be in the long format, where it is expanded in such a manner that for every subject, there is a response for each age interval (a person-period data set (Davis et al, 1992)). The expanded dataset for the subjects considered on yearly intervals was rearranged as shown in the snapshot in Table 7.1 above. Subject 19, for example, was followed up for 6 years (2009-2014) under which she was never married until she got lost to follow-up when she was now 74 years old. On the other hand, subject 23 entered the study in 2004 when he was 32 years old but only entered a first marriage in his 8^{th} year of follow-up in 2011, at the age

of 39. Any subject who experiences the event is then taken out of the risk set.

Table (7.1) Snapshot of Data: Discrete-time Survival Model

ID	Year	Marital Status	ε_{it}	Age	Gender	Income	Employed
19	2009	Never Married	0	69	Female	Yes	No
19	2010	Never Married	0	70	Female	Yes	No
19	2011	Never Married	0	71	Female	No	Yes
19	2012	Never Married	0	72	Female	No	No
19	2013	Never Married	0	73	Female	No	No
19	2014	Never Married	0	74	Female	No	No
23	2004	Never Married	0	32	Male	No	No
23	2005	Never Married	0	33	Male	No	No
23	2006	Never Married	0	34	Male	No	No
23	2007	Never Married	0	35	Male	No	No
23	2008	Never Married	0	36	Male	No	No
23	2009	Never Married	0	37	Male	No	No
23	2010	Never Married	0	38	Male	No	Yes
23	2011	Married	1	39	Male	No	No
42	2005	Never Married	0	29	Female	No	No
42	2006	Never Married	0	30	Female	No	No
42	2007	Married	1	31	Female	No	No
45	2010	Never Married	0	32	Male	No	Yes
45	2011	Never Married	0	33	Male	No	No

Data Preparation for Chapter 4

For a discrete-time analysis of recurrent events, the data must be expanded and re-arranged in an expanded person-episode-period structure (Davis et al, 1992). There must be a record for each discrete-time interval for every episode. Table 7.2 gives a snapshot of how the data must be restructured before the analysis for a multinomial response for state occupation, ε_{ktj}. Variables for subject, episode and time are *ID, Episode* and *Period* respectively. Covariates are also displayed and these include *Age, Gender, Income* and *Employed*.

Table (7.2) Snapshot of Data: Multi-state Model

ID	Year	Episode	Period	Marital Status	ε_{ktj}	Age	Gender	Income	Employed
16	2009	1	1	Widowed	4	56	Female	No	Yes
16	2010	1	2	Widowed	4	57	Female	No	Yes
16	2011	1	3	Widowed	4	58	Female	No	Yes
16	2012	1	4	Widowed	4	59	Female	Yes	No
17	2005	1	1	Married	2	36	Female	No	No
17	2006	1	2	Married	2	37	Female	No	No
17	2007	2	1	Widowed	4	38	Female	Yes	No
17	2008	2	2	Widowed	4	39	Female	Yes	No
17	2009	2	3	Widowed	4	40	Female	Yes	No
17	2010	2	4	Widowed	4	41	Female	No	No
17	2011	2	5	Widowed	4	42	Female	No	Yes
17	2012	3	1	Married	2	43	Female	No	No
17	2013	3	2	Married	2	44	Female	No	No
17	2014	4	1	Widowed	4	45	Female	No	No
25	2004	1	1	Never Married	1	25	Male	No	No
25	2005	1	2	Never Married	1	26	Male	No	No
25	2006	1	3	Never Married	1	27	Male	No	Yes
25	2007	1	4	Never Married	1	28	Male	No	Yes
25	2008	1	5	Never Married	1	29	Male	No	No
25	2009	1	6	Never Married	1	30	Male	No	No
25	2010	1	7	Never Married	1	31	Male	No	No
25	2011	1	8	Never Married	1	32	Male	No	No
25	2012	1	9	Never Married	1	33	Male	No	No
25	2013	1	10	Never Married	1	34	Male	No	No
25	2014	1	11	Never Married	1	35	Male	No	No
25	2015	1	12	Never Married	1	36	Male	No	No
25	2016	1	13	Never Married	1	37	Male	No	No
27	2011	1	1	Married	1	35	Female	No	Yes
28	2004	1	1	Never Married	1	36	Male	No	No
28	2005	1	2	Never Married	1	37	Male	No	Yes

It is clear from Table 7.2 that subject 16 only experienced 1 episode which lasted for a period of 4 years. She entered the study in 2009 when she was already a 56-year-old widow and stayed widowed until she left the study in 2012 at the age of 59. Subject 17, however, entered in 2005 when she was married and the first episode lasted for 2 years. In 2007, she started her second episode in the *Widowed* state which lasted for a period of 5 years before she remarried and started a third episode in 2012. She experienced 4^{th} episode after the death of the partner again in 2014 which was the last observation made on her before she left the study. Subject 25, on the other hand had only 1 episode which lasted for all the 13 years that he was under study. Another visual illustration of the paths taken by a typical subject (17) can also be represented graphically as in Figure 7.2 below:

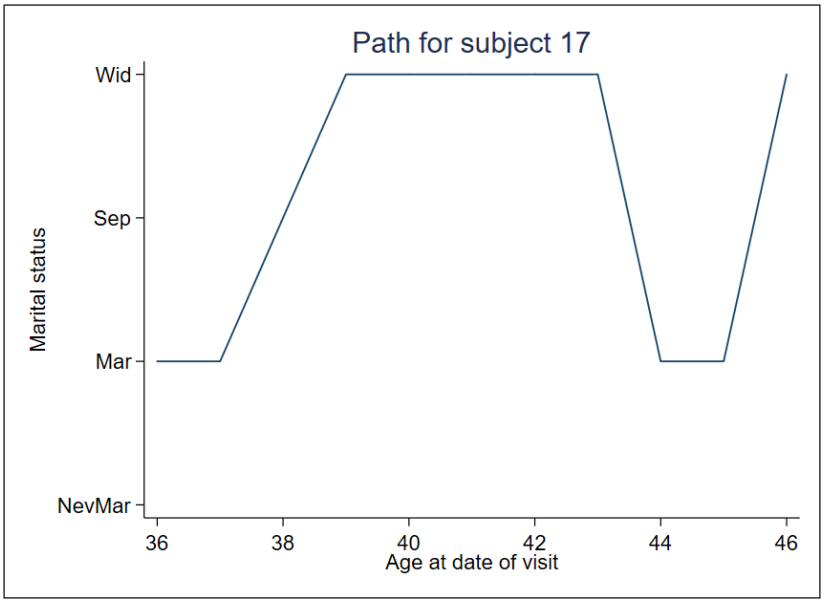


Figure (7.2) Typical path for a subject k

APPENDIX D- STATA CODES FOR CHAPTERS 3 AND 4

D1- Discrete Survival

```
use "C:\Users\batidzirai\Desktop\PhD in SAS\Chpt2.dta",clear

*************************************************************
            *Table 2.2
*************************************************************

ta Gender Y_nm if Duration==1 & Age>16 & Age<=65,row
ta Income Y_nm if Duration==1 & Age>16 & Age<=65,row
ta Employed Y_nm if Duration==1 & Age>16 & Age<=65,row
ta Educationn Y_nm if Duration==1 & Age>16 & Age<=65,row
ta Employed Y_nm if Duration==1 & Age>16 & Age<=65,row


*************************************************************
            *Table 2.4
*************************************************************
/*FE Model with a Quartic Baseline */
logit Y_nm AgeNM AgeNM_Quad AgeNM_Cubic AgeNM_Quartic i.Gender i.Income i.Employed i.Educationn
AgeAtFirstSex1/*FE Model with Quartic Baseline */
logit,or /*Odds Ratios*/
estat ic
estimates store LogitNM /* Store the estimates as LogitNM*/



/*RE Model with a Quartic Baseline, Gauss-Hermite Quad */
*Declare data to be panel data
xtset ID Year
xtlogit Y_nm AgeNM AgeNM_Quad AgeNM_Cubic AgeNM_Quartic  i.Gender i.Income i.Employed i.Educationn
 , re
xtlogit, or

melogit Y_nm AgeNM AgeNM_Quad AgeNM_Cubic AgeNM_Quartic i.Gender i.Income i.Employed i.Educationn
AgeAtFirstSex1|| ID:,nolog /*RE Model with 2 Levels*/
melogit, or
estimates store MeLogitNM2 /* Store the estimates as MeLogitNM*/

melogit Y_nm AgeNM AgeNM_Quad AgeNM_Cubic AgeNM_Quartic i.Gender i.Income i.Employed i.Educationn
AgeAtFirstSex1 || HIntId: || ID:,nolog /*RE Model with 3 Levels*/
melogit, or
estimates store MeLogitNM3 /* Store the estimates as MeLogitNM*/
estat ic

lrtest LogitNM MeLogitNM2, force /* Likelihood Ratio Test*/
lrtest LogitNM MeLogitNM3, force /* Likelihood Ratio Test*/

**************Printing the Output**************
findit esttab /*Instal st0085_1  */
esttab LogitNM MeLogitNM2 MeLogitNM3, eform b(2) aic  se /*tabulate the Exponentiated estimates
correct to 2.dp, all the 3models models (with Standard Errors in the parenthesis)*/
esttab LogitNM MeLogitNM2 MeLogitNM3, eform b(2) aic  ci /*tabulate the Exponentiated estimates
correct to 2.dp, all the 3models models (with COnfidence Intervals in the parenthesis)*/

```

D1- Multinomial for State Occupation

```
use "C:\Users\batidzirai\Desktop\PhD in SAS\Chpt2.dta",clear

*********************************************************
            *Table 3.2
*********************************************************

ta Gender MaritalStatus if Duration==1 & Age>16 & Age<=65,row
ta Income MaritalStatus if Duration==1 & Age>16 & Age<=65,row
ta Employed MaritalStatus if Duration==1 & Age>16 & Age<=65,row
ta Educationn MaritalStatus if Duration==1 & Age>16 & Age<=65,row

su AgeAtFirstSex1 if Duration ==1 & Age>16 & Age<=65


*********************************************************
*Table 3.3 Determine the number of ij transitions over time
*********************************************************
bysort ID: ge Previous_MStatus = MaritalStatus[_n-1]
bysort ID: replace Previous_MStatus =MaritalStatus if Previous_MStatus ==.
replace MaritalStatus=1 if ID==24597 & Previous_MStatus==1
la define Previous_MStatus1 1"Never Married" 2"Married" 3"Seperated" 4"Widowed"
la values Previous_MStatus Previous_MStatus1

ta Previous_MStatus, missing

ta Previous_MStatus MaritalStatus if Age>16 & Age<=65,row

*********************************************************
            *Table 3.4
*********************************************************
mlogit MaritalStatus i.Gender i.Income i.Employed i.Educationn AgeAtFirstSex1, cluster(ID) /*with
SE's corrected at ID level*/
mlogit,rr


```

D1- Competing Risks

```
use "C:\Users\batidzirai\Desktop\PhD in SAS\Chpt2.dta",clear

***********************************************************
            *Table 3.5
***********************************************************
melogit Y_ms i.Gender i.Income i.Employed i.Educationn AgeAtFirstSex1 || ID:,nolog/*RE Model with CONSTANT Baseline */
melogit,or /*Odds Ratios*/
estimates store MeLogitMS


melogit Y_mw AgeMW AgeMW_Quad AgeMW_Cubic i.Gender i.Income i.Employed i.Educationn AgeAtFirstSex1 || ID:,nolog/*RE Model with CUBIC Baseline */
melogit,or /*Odds Ratios*/
estimates store MeLogitMW

melogit Y_sm  AgeMS i.Gender i.Income i.Employed i.Educationn AgeAtFirstSex1 || ID:,nolog/*RE Model with LINEAR Baseline */
melogit,or /*Odds Ratios*/
estimates store MeLogitSM

melogit Y_wm AgeMW i.Gender i.Income i.Employed i.Educationn AgeAtFirstSex1 || ID:,nolog/*RE Model with LINEAR Baseline */
melogit,or /*Odds Ratios*/
estimates store MeLogitWM

esttab MeLogitMS MeLogitMW MeLogitSM MeLogitWM, eform b(2) aic  ci



***********************************************************
            *Table 3.7 and 3.8
***********************************************************
*Generates the Variable: Yijk from Married- without censorig competing events)*
bysort ID Episode: ge Yijt_CR = . /* M_SW= 1 if subject HASN'T made a transition into rither S or W*/
bysort ID: replace Yijt_CR = 0 if MaritalStatus==2
bysort ID: replace Yijt_CR  = 1 if MaritalStatus==3 & MaritalStatus[_n-1]==2 & MaritalStatus  ~=. /* M_SW= 1 if subject made a transition into S*/
bysort ID: replace Yijt_CR  = 2 if MaritalStatus==4  & MaritalStatus[_n-1]==2 & MaritalStatus ~=. /* M_SW= 2 if subject made a transition into W*/
la var Yijt_CR "Transition OUT of Married"
label define Yijt_CR 0"Married" 1 "Separated" 2"Widowed"
label values Yijt_CR Yijt_CR
ta Yijt_CR


*Competing Risks Model: MARGINAL FIXED MODEL*
mlogit Yijt_CR AgeMD i.Gender i.Income i.Employed i.Educationn AgeAtFirstSex1
mat b=e(b)
mlogit, rrr
estat ic /*AIC &BIC*/

xtmlogit Yijt_CR AgeMD i.Gender i.Income i.Employed i.Educationn AgeAtFirstSex1 /*with independent covariance structure*/
estat ic /*AIC*/
xtmlogit,rr
```

SAS CODE FOR DISCRETE-TIME MULTIVARIATE SURVIVAL MODEL

############################## **Chapter 5** ##############################

```
/* Import Never Married to Married Data*/
PROC IMPORT OUT= WORK.NM
DATAFILE= "C:\Users\batidzirai\Desktop\PhD in SAS\NM1.dta"
DBMS=STATA REPLACE;
RUN;


/* Import Married to Separated Data*/
PROC IMPORT OUT= WORK.MS
DATAFILE= "C:\Users\batidzirai\Desktop\PhD in SAS\MS1.dta"
DBMS=STATA REPLACE;
RUN;

/* Import Married to Widowed Data*/
PROC IMPORT OUT= WORK.MW
DATAFILE= "C:\Users\batidzirai\Desktop\PhD in SAS\MW1.dta"
DBMS=STATA REPLACE;
RUN;

/* Import Separated to Married Data*/
PROC IMPORT OUT= WORK.SM
DATAFILE= "C:\Users\batidzirai\Desktop\PhD in SAS\SM1.dta"
DBMS=STATA REPLACE;
RUN;

/* Import Widowed to Married Data*/
PROC IMPORT OUT= WORK.WM
DATAFILE= "C:\Users\batidzirai\Desktop\PhD in SAS\WM1.dta"
DBMS=STATA REPLACE;
RUN;
```

```
/*COMBINING THE DATA WHICH WAS CLEANED IN STATA*/

LIBNAME ANALYSES "C:\USERS\BATIDZIRAI\DESKTOP\PHD IN SAS";

DATA ANALYSES.NM;
SET NM;
RUN;

DATA ANALYSES.MS;
SET MS;
RUN;

DATA ANALYSES.MW;
SET MW;
RUN;

DATA ANALYSES.SM;
SET SM;
RUN;

DATA ANALYSES.WM;
SET WM;
RUN;

DATA JOINT;
SET NM MS MW SM WM;
/*MERGING_VARIABLE=1;*/
IF TRANS=1 THEN EVENT=Y_NM;
ELSE IF TRANS=2 THEN EVENT=Y_MS;
ELSE IF TRANS=3 THEN EVENT=Y_MW;
ELSE IF TRANS=4 THEN EVENT=Y_SM;
ELSE IF TRANS=5 THEN EVENT=Y_WM;
RUN;

PROC SORT DATA=JOINT OUT=ANALYSES.JOINT_FINAL;
BY ID;
RUN;
```

```
/*THE MULTIVARIATE JOINT MODEL*/

PROC NLMIXED DATA=ANALYSES.JOINT_FINAL NOAD QPOINTS=7 MAXITER=500;
PARMS

NM01=1 NM02=1 NM03=1 NM04=1 NM05=1 NM06=1 NM07=1 NM08=1 NM09=1 NM10=1 NM11=1
MS01=1 MS02=1 MS03=1 MS04=1 MS05=1 MS06=1 MS07=1 MS08=1 MS MS09=1 MS10=1 MS1
MW01=1 MW02=1 MW03=1 MW04=1 MW05=1 MW06=1 MW07=1 MW08=1 MW09=1 MW10=1 MW11=1
SM01=1 SM02=1 SM03=1 SM04=1 SM05=1 SM06=1 SM07=1 SM08=1 SM09=1 SM10=1 SM11=1
WM01=1 WM02=1 WM03=1 WM04=1 WM05=1 WM06=1 WM07=1 WM08=1 WM09=1 WM10=1 WM11=1

NMBETA1=-0.1 NMBETA2=-0.5 NMBETA3=-0.8 NMBETA41=-0.8 NMBETA42=-0.8 NMBETA43=
MSBETA1=-0.6 MSBETA2=-0.1 MSBETA3=-0.8 MSBETA41=0.4 MSBETA42=0.4 MSBETA43=0.4
MWBETA1=-1.7 MWBETA2=-0.9 MWBETA3=0.4 MWBETA41=0.4 MWBETA42=0.4 MWBETA43=0.4
SMBETA1=-0.4 SMBETA2=-0.1 SMBETA3=-0.1 SMBETA41=-0.1 SMBETA42=-0.1 SMBETA43=
WMBETA1=0.1 WMBETA2=-0.3 WMBETA3=0.1 WMBETA41=0.2 WMBETA42=0.2 WMBETA43=0.2 W
SDU1=0 SDU2=0 COV12=0;


BOUNDS
NM01 MS01 MW01 SM01 WM01 SDU1 SDU2
NNM01 NM02 NM03 NM04 NM05 NM06 NM07 NM08 NM09 NM10 NM11 NM12 NM13
MS01 MS02 MS03 MS04 MS05 MS06 MS07 MS08 MS09 MS10 MS11 MS12 MS13
MW01 MW02 MW03 MW04 MW05 MW06 MW07 MW08 MW09 MW10 MW11 MW12 MW13
SM01 SM02 SM03 SM04 SM05 SM06 SM07 SM08 SM09 SM10 SM11 SM12 SM13
WM01 WM02 WM03 WM04 WM05 WM06 WM07 WM08 WM09 WM10 WM11 WM12 WM13
SDU1 SDU2 > 0;


IF EVENT=1 THEN DO;
/*ESTIMATING THE DURATION IN EACH AGE INTERVAL FOR NM*/
BASELINEHZNM =  NM01*Indicator_NM1 + NM02* Indicator_NM2 + NM03* Indicator_NM

MU_NM = BASELINEHZNM + GENDER*NMBETA1 + INCOME*NMBETA2 + EMPLOYED*NMBETA3 +
LAMBDA = EXP(MU_NM)/ (1+ EXP(MU_NM));
```

```
END;


ELSE IF EVENT=2 THEN DO;
/*MODELLING THE BASELINE HAZARD FOR THE MS TRANSITION*/
BASELINEHZMS= MS01* Indicator_MS1 + MS02*Indicator_MS2 + MS03* Indicator_MS3
+ MS08*Indicator_MS8 + MS09*Indicator_MS9 + MS10*Indicator_MS10 + MS11*Indica

MU_MS = BASELINEHZMS + GENDER*MSBETA1 + INCOME*MSBETA2 + EMPLOYED*MSBETA3 +
LAMBDA = EXP(MU_MS)/ (1+ EXP(MU_MS));
END;


ELSE IF EVENT=3 THEN DO;
/*MODELLING THE BASELINE HAZARD FOR THE MW TRANSITION*/
BASELINEHZMW= MW01*Indicator_MW1 + MW02*Indicator_MW2 + MW03* Indicator_MW3 +

MU_MW = BASELINEHZMW + GENDER*MWBETA1 + INCOME*MWBETA2 + EMPLOYED*MWBETA3 +
LAMBDA = EXP(MU_MW)/ (1+ EXP(MU_MW));
END;


ELSE IF EVENT=4 THEN DO;
/*MODELLING THE BASELINE HAZARD FOR THE SM TRANSITION*/
BASELINEHZSM = SM01*Indicator_SM1 +SM02*Indicator_SM2 +SM03*Indicator_SM3+ SM

MU_SM = BASELINEHZSM + GENDER*SMBETA1 + INCOME*SMBETA2 + EMPLOYED*SMBETA3 +
LAMBDA = EXP(MU_SM)/ (1+ EXP(MU_SM));
END;


ELSE IF EVENT=5 THEN DO;
/*MODELLING THE BASELINE HAZARD FOR THE WM TRANSITION*/
BASELINEHZWM= WM01* Indicator_WM1 + WM02*Indicator_WM2 + WM03*Indicator_WM3 +

MU_WM = BASELINEHZWM + GENDER*WMBETA1 + INCOME*WMBETA2 + EMPLOYED*WMBETA3 +
LAMBDA = EXP(MU_WM)/ (1+ EXP(MU_WM));
END;
```

```
RHO12=(EXP(2*COV12)-1)/(EXP(2*COV12)+1);


MODEL EVENT ~ BINARY(LAMBDA);
RANDOM U1 U2 U3 U4 U5 ~ NORMAL([0, 0, 0, 0, 0], [EXP(2*SDU1),
RHO12*EXP(SDU1+SDU2), EXP(2*SDU1), RHO12*EXP(SDU1+SDU2),
RHO12*EXP(SDU1+SDU2), EXP(2*SDU1), RHO12*EXP(SDU1+SDU2),
RHO12*EXP(SDU1+SDU2), RHO12*EXP(SDU1+SDU2), EXP(2*SDU1), RHO12*EXP(SDU1+SDU2
SUBJECT= ID OUT=EB;
ESTIMATE 'NM' EXP(2*SDU1);
ESTIMATE 'NM_MS' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'MS' EXP(2*SDU1);
ESTIMATE 'NM_MW' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'MS_MW' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'MW' EXP(2*SDU1);
ESTIMATE 'NM_SM' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'MS_SM' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'MW_SM' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'SM' EXP(2*SDU1);
ESTIMATE 'NM_WM' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'MS_WM' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'MW_WM' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'SM_WM' RHO12*EXP(SDU1+SDU2);
ESTIMATE 'WM' EXP(2*SDU1);
RUN;
```

SAS CODE FOR DISCRETE-TIME MULTIVARIATE COMPETING RISKS MODEL

############################## **Chapter** 6 ##############################

```
/* Import From Never Married Data*/
PROC IMPORT OUT= WORK.Xn
DATAFILE= "C:\Users\batidzirai\Desktop\JointCompet\Xn.dta"
DBMS=STATA REPLACE;
RUN;


/* Import From Married Data*/
PROC IMPORT OUT= WORK.Xm
DATAFILE= "C:\Users\batidzirai\Desktop\JointCompet\Xm.dta"
DBMS=STATA REPLACE;
RUN;

/* Import From Separated Data*/
PROC IMPORT OUT= WORK.Xs
DATAFILE= "C:\Users\batidzirai\Desktop\JointCompet\Xs.dta"
DBMS=STATA REPLACE;
RUN;

/* Import Widowed to Married Data*/
PROC IMPORT OUT= WORK.Xw
DATAFILE= "C:\Users\batidzirai\Desktop\JointCompet\Xw.dta"
DBMS=STATA REPLACE;
RUN;



/*COMBINING THE DATA WHICH WAS CLEANED IN STATA*/
```

```
LIBNAME ANALYSES "C:\USERS\BATIDZIRAI\DESKTOP\PHD IN SAS";

DATA ANALYSES.Xn;
SET Xn;
RUN;

DATA ANALYSES.Xm;
SET Xm;
RUN;

DATA ANALYSES.Xs;
SET Xs;
RUN;

DATA ANALYSES.Xw;
SET Xw;
RUN;

DATA JOINT;
SET Xn Xm Xs Xw;
/*MERGING_VARIABLE=1;*/
IF TransCR=1 THEN EVENT=Xn;
ELSE IF TransCR=2 THEN EVENT=Xm;
ELSE IF TransCR=3 THEN EVENT=Xs;
ELSE IF TransCR=4 THEN EVENT=Xw;
RUN;

PROC SORT DATA=JOINT OUT=ANALYSES.JOINT_CR;
BY ID;
RUN;



/*THE MULTIVARIATE JOINT MODEL*/

PROC NLMIXED DATA=ANALYSES.JOINT_CR QPOINTS=3 MAXITER=50;
PARMS
```

```
NM01=1 MS01=1 MW01=1 SM01=1 WM01=1
NMBETA1=-0.1 NMBETA2=-0.5 NMBETA3=-0.8 NMBETA41=-0.8 NMBETA42=-0.8 NMBETA43=
MSBETA1=-0.6 MSBETA2=-0.1 MSBETA3=-0.8 MSBETA41=0.4 MSBETA42=0.4 MSBETA43=0.4
MWBETA1=-1.7 MWBETA2=-0.9 MWBETA3=0.4 MWBETA41=0.4 MWBETA42=0.4 MWBETA43=0.4
SMBETA1=-0.4 SMBETA2=-0.1 SMBETA3=-0.1 SMBETA41=-0.1 SMBETA42=-0.1 SMBETA43=
SMBETA5=0
WMBETA1=0.1 WMBETA2=-0.3 WMBETA3=0.1 WMBETA41=0.2 WMBETA42=0.2 WMBETA43=0.2 W
LOGSDU1=0
Z12=0;

BOUNDS NM01 MS01 MW01 SM01 WM01 LOGSDU1 >=0;

/*ESTIMATING THE LIKELIHOOD OUT OF THE MARRIED STATE*/
IF TransCR=2 THEN DO;
BASELINEHZMS = MS01 ;
BASELINEHZMW = MW01 ;

MU1=0;
MU2 =MS01 + GENDER*MSBETA1 + INCOME*MSBETA2 + EMPLOYED*MSBETA3 + (EDUCATIONN

MU3 = MW01 + GENDER*MWBETA1 + INCOME*MWBETA2 + EMPLOYED*MWBETA3 +(EDUCATIONN


exp_MU1 = 1;
exp_MU2 = exp(MU2);
exp_MU3 = exp(MU3);
DEN = exp_MU1 + exp_MU2 + exp_MU3;


IF EVENT=0 THEN LOG_LIK= exp_MU1/DEN;
IF EVENT=2 THEN LOG_LIK= exp_MU2/DEN;
IF EVENT=3 THEN LOG_LIK= exp_MU3/DEN;
END;


/*ESTIMATING THE LIKELIHOOD FOR NM*/
ELSE IF TransCR=1 THEN DO;
BASELINEHZNM = NM01 ;
MU_NM = BASELINEHZNM + GENDER*NMBETA1 + INCOME*NMBETA2 + EMPLOYED*NMBETA3 +(E
```

```
LAMBDA = EXP(MU_NM)/ (1+ EXP(MU_NM));
LOG_LIK= X_n*LOG(LAMBDA) + (1-X_n)*LOG(1-LAMBDA);
END;

/*ESTIMATING THE LIKELIHOOD FOR SM*/
ELSE IF TransCR=3 THEN DO;
/*MODELLING THE BASELINE HAZARD FOR THE SM TRANSITION*/
BASELINEHZSM = SM01;
MU_SM = BASELINEHZSM + GENDER*SMBETA1 + INCOME*SMBETA2 + EMPLOYED*SMBETA3 +(E
LAMBDA = EXP(MU_SM)/ (1+ EXP(MU_SM)) ;
LOG_LIK= X_s*LOG(LAMBDA) + (1-X_s)*LOG(1-LAMBDA);

END;


ELSE IF TransCR=3 THEN DO;
/*MODELLING THE BASELINE HAZARD FOR THE WM TRANSITION*/
BASELINEHZWM= WM01;

MU_WM = BASELINEHZWM + GENDER*WMBETA1 + INCOME*WMBETA2 + EMPLOYED*WMBETA3 +(E

LAMBDA = EXP(MU_WM)/ (1+ EXP(MU_WM));
LOG_LIK= X_w*LOG(LAMBDA) + (1-X_w)*LOG(1-LAMBDA);
END;

RHO12=(EXP(2*Z12)-1)/(EXP(2*Z12)+1);


MODEL EVENT ~ GENERAL(LOG_LIK);
RANDOM U11 U12 U2 U3 U4 ~ NORMAL([0, 0, 0, 0, 0], [EXP(2*LOGSDU1),
RHO12*EXP(LOGSDU1+LOGSDU1), EXP(2*SDU1), RHO12*EXP(LOGSDU1+LOGSDU1),
RHO12*EXP(LOGSDU1+LOGSDU1), EXP(2*SDU1), RHO12*EXP(LOGSDU1+LOGSDU1),
RHO12*EXP(LOGSDU1+LOGSDU1), RHO12*EXP(LOGSDU1+LOGSDU1), EXP(2*SDU1),
RHO12*EXP(LOGSDU1+LOGSDU1),RHO12*EXP(LOGSDU1+LOGSDU1), RHO12*EXP(LOGSDU1+LOGS
SUBJECT= ID OUT=EB;
ESTIMATE 'VAR' EXP(2*LOGSDU1);
ESTIMATE 'COV' RHO12*EXP(LOGSDU1+LOGSDU1);
RUN;
```