Joint models for bivariate discrete longitudinal outcome and survival

Shengning Pan

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Statistical Science

University College London

December 14, 2022

I, Shengning Pan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

In analytical studies of longitudinal and time-to-event data, measuring the relationship between longitudinal outcomes and the time of event occurring simultaneously is of interest. It is common in medical statistics to have non-negative integers as longitudinal responses, and there will often be more than one response variable in the data. The main aim of this thesis is to construct the corresponding bivariate joint models to analyse these discrete longitudinal and time-to-event data. We construct two types of joint models, namely the bivariate shared random-effects joint model and the bivariate latent-class joint model. For the longitudinal model, we use extensions of the binomial distribution and the categorical distribution. In addition, to deal with attrition due to death or dementia, we use the exponential hazard model, the Weibull hazard model and the Gompertz hazard model as the survival model. We will assume that the longitudinal model and the survival model are independent of each other conditional on random effects. The joint models are applied to analyse three datasets. The first data is the English Longitudinal Study of Ageing (ELSA). The second is the PAQUID data, whose full title is PAQUID: Longitudinal data on cognitive and physical aging in the elderly. The third is PBC2: Mayo Clinic primary biliary cirrhosis data.

Impact Statement

In medical statistics, data are usually recorded in the form of longitudinal and time-to-event data, and there is often more than one response variable. The bivariate joint models we constructed in this thesis can be wildly used to analyse discrete longitudinal and time-to-event data. Therefore, our research has practical applications.

These joint models we constructed can be divided into bivariate binomial joint models and bivariate categorical joint models, depending on the longitudinal model used. The bivariate binomial joint models are applicable to datasets where the response variables are the sum of scores. It means that the bivariate binomial joint models can be widely applied to data collected through questionnaires or interviews. The data in medical statistics, particularly concerning cognitive abilities, can be analysed using bivariate binomial joint models. For example, in the ELSA data used in this thesis, we concentrate on the relationship between cognitive ability and an individual's risk of death. In the PAQUID data, we focused on the relationship between cognitive ability and the risk of dementia.

The categorical joint models have a more comprehensive application than the binomial joint models. First, we can construct a joint model for the sum of scores data by considering each possible value as a category. Secondly, we can categorise the responses before applying the joint model to discrete data with a wide range of values. Using the example of the MMSE, a test with a range of values from 0 to 30, we can categorise the data into sub-categories using the common analytical criteria for MMSE before constructing the model: 24-30 is normal cognitive ability; while 19-23 is mild, 10-18 is moderate, and 0-9 is severe cognitive impairment. Modelling according to these four sub-categories makes our joint model applicable

and allows the researcher to obtain model estimates that are more directly relevant to the research question. Finally, for continuous data, similar to discrete data with a large range of values, we can use a categorical joint model to analyse responses after categorising them.

To summarise, the joint model proposed in this thesis can be used for a wide range of applications in medical statistics.

UCL Research Paper Declaration Form

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;

- are in submission to a peer-reviewed publication;

- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

---

1. For a research manuscript that has already been published

(if not yet published, please skip to section 2):

---

2. For a research manuscript prepared for publication but that has not yet been published

(if already published, please skip to section 3):

---

a) What is the current title of the manuscript?	Bivariate joint models for survival and change of cognitive function
b) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv; if 'Yes', please give a link or doi):	No

c) **Where is the work intended to be published?** (e.g. journal names)	Statistical Methods in Medical Research
d) **List the manuscript's authors in the intended authorship order:**	Shengning Pan; Ardo van den Hout
e) **Stage of publication** (e.g. in submission):	Under full consideration for publication

3.For multi-authored work, please give a statement of contribution covering all authors:

Methods developped mainly by SP with some supervision by AvdH.

Writing, computer programming, and data analysis are undertaken by SP.

4.In which chapter(s) of your thesis can this material be found?

Chapter 2, 3, 4, 5

5.e-Signatures confirming that the information above is accurate

Candidate:	Shengning Pan
Date:	11/09/2022
Supervisor/ Senior Author:	Ardo van den Hout
Date:	11/09/2022

Acknowledgements

This thesis is the end of this PhD journey and the beginning of the next phase of my life. Upon completion of my PhD thesis, I would like to express my sincere thanks to all the people who have provided me with invaluable assistance during the process.

First of all, I would like to express my sincere and deep gratitude to my supervisor, Dr Ardo van den Hout, for his continuous guidance and support during my PhD study. Over the past four years, he has given me a lot of valuable advice on identifying topics, constructing models, and analysing results. Every piece of his advice has provided me with new ideas. I am honoured to take his lectures and study under his supervision as a PhD student. Also, I am very grateful for his kindness and help during the writing process. I would never finish this thesis so smoothly without his help.

I want to give a big thanks to the examiners of upgrade viva for the advice they gave me during the viva, which are still very useful when writing my thesis. In addition, I should give my heart thanks to all the staff of the Department of Statistical Science for all the help they have given me over the past four years.

I would also like to thank my parents for the endless love, patience and care they have given me over the past 28 years. Their support has given me the motivation and courage to fulfil my dream of continuing my research. In addition, I would like to extend my heartfelt thanks to my sister Fang and my friends Xinyi, Rosie, Jinye and Ju for all the support they have given me over the years.

In this thesis, I use ELSA data. I really appreciate researchers in the University College London, NatCen Social Research, and the Institute for Fiscal Studies,

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

1.1.1 Longitudinal and time-to-event data

Typically, longitudinal and time-to-event data include several longitudinally measured responses and the time until one or more events of particular interest occurs. In medical statistics, such data provide a wealth of information for studies on ageing, health, longevity, and aspects of disease processes leading to death. The data consist of two main components, the first of which is the longitudinal data. Longitudinal data, also called panel data, are obtained by following the same sample at different time points. This sample could include individuals, households, organisations, etc. Longitudinal data allows researchers to measure changes in the outcome over time, measure the duration of an event, and record the timing of various events. For example, suppose unemployment stays high over a long period. The government could use longitudinal data to determine whether the same group of individuals has been unemployed throughout the period or whether different groups of individuals have been unemployed.

Generally, longitudinal data are multi-dimensional data containing measurements that change over time. These data include observations of multiple phenomena obtained over various periods for the same person, household or organisation [Diggle et al., 2002]. For instance, the English Longitudinal Study of Ageing (ELSA) data contain several cognitive tests. The results of these tests are considered

to reflect, to some extent, the cognitive ability of the tested individual. Individuals are tested every few years, and the results are recorded in the data. Based on these results, the researchers attempt to measure something potentially representative of cognitive function. The ELSA data will be used in Chapter 6 and we will cover it in detail in Chapter 5. Because of the characteristics of longitudinal data, longitudinal studies are widely used in medical science, psychology, sociology, economics and other fields [Rushton et al., 2002].

The second component of the longitudinal and time-to-event data is called survival data (time-to-event data). It shows whether a specific event of interest occurred and when that event occurred [Rizopoulos, 2012]. The response variable in the time-to-event data is the time until that event occurred, often referred to as the failure time, survival time, or event time. The time-to-event data can be treated as data which consist of the last observations for each individual in the longitudinal data. We take the Acquired immune deficiency syndrome (AIDS) data as an example [Organization et al., 2000]. The data record the number of CD4 cells changing over time for 467 patients who are infected by advanced human immunodeficiency virus (HIV). Some patients died at the end of the study, and their time of death was recorded in the data. The number of CD4 cells at the end of the study and the corresponding time was recorded for other patients. Individuals in this dataset were treated with two drugs: didanosine (ddI) and zalcitabine (ddC). The corresponding survival study is designed to compare the efficacy and safety of these two drugs and, more specifically, to identify which drugs can effectively extend the lifespan of patients.

The application of time-to-event data is addressed in different areas of research. For example, the event of interest in ageing research often represents death or dementia. In clinical and epidemiologic studies, the event might be death, the appearance of a tumour, the development of certain diseases, recurrence of a disease, conception, or smoking cessation. The time-to-event data are also widely used in other disciplines. For example, in sociology, the event of interest could be the duration of a first marriage, while in marketing, it might be the timing of a newspaper

subscription [Rizopoulos, 2012].

1.1.2 Joint models

The longitudinal and time-to-event data studies often require a separate analysis of the recorded outcomes: the analysis of longitudinal outcomes and that of time-to-event outcomes. Corresponding longitudinal models and survival models are needed, respectively. When we focus on the longitudinal outcomes, we wish to investigate the effect of time and time-dependent covariates on the outcomes [Rizopoulos, 2012]. It needs to be analysed by the corresponding longitudinal model. The attrition due to any event of interest (e.g. death or dementia) is also recorded and cannot be ignored. The survival models could investigate this nonrandom attrition. When time-to-event outcomes are the focus of the research, we also want to analyse how responses change over time when the event of interest has not occurred. The work in Laird and Ware [1982]; Cox [1972] model longitudinal data and time-to-event data separately. However, if the two processes are correlated, this may lead to biased effect size estimates [Ibrahim et al., 2010]. Moreover, there is evidence in the literature that the average age trajectory of physiological variables in individuals with short lifespans differs from the trajectory of physiological variables in individuals with long lifespans [Yashin et al., 2009, 2012; Arbeev et al., 2014]. Therefore, it is reasonable to use a joint model to simultaneously analyse longitudinal data and time-to-event data. Previous research has shown that joint modelling can improve the efficiency of statistical inference and reduce bias, and it can provide significant benefits when designing experiments [Dupuy and Mesbah, 2002; Hogan and Laird, 1998].

The statistical methods designed to analyse time-to-event data jointly with longitudinal measures are known as joint models [Arbeev et al., 2014]. A joint model for longitudinal responses describes the change of responses using a model for longitudinal data and measures the risk of the event using a survival model [Proust-Lima et al., 2009]. The principle of joint models is to link these two models using a common latent variable. This common latent variable captures the association between the processes so that the two processes are conditionally independent given

the latent variable, which is the random effect. We assume that random effects explain all interdependencies [Rizopoulos, 2012], i.e. the longitudinal outcome is conditionally independent of the time-to-event outcome; the repeated measurements in the longitudinal outcome are conditionally independent of each other. Based on this principle, two widely used joint models have been proposed: the shared random-effects joint model [Henderson et al., 2000] and the latent-class joint model [Lin et al., 2000].

Over the past three decades, researchers have contributed considerably to research on the joint modelling of longitudinal data and time-to-event data [Hickey et al., 2016; Henderson et al., 2000]. For the shared random-effects joint model, the seminal articles in this field were written by Faucett and Thomas [1996]; Wulfsohn and Tsiatis [1997] in 1996 and 1997, respectively, and many extensions to the joint model have been proposed by researchers after that [Mauff et al., 2020]. At the beginning of joint modelling research, the focus was mainly on single longitudinal responses and single time-to-event responses, i.e. univariate joint model. The univariate joint model is widely used in the analysis of clinically related data [Touloumi et al., 2004], and many researchers have written mainstream software packages for computing corresponding problems, such as Rizopoulos [2012]; Philipson et al. [2012]; Crowther et al. [2013]. However, as we mentioned above, longitudinal and time-to-event data, especially the longitudinal component, are usually multidimensional and contain observations of multiple phenomena obtained by the same individual over multiple periods [Diggle et al., 2002].

Although univariate joint models allow us to investigate the relationship between a single time-to-event outcome and a single longitudinal outcome, it is often the case that there will be multiple biomarkers associated with the event of interest [Mauff et al., 2020]. The construction of multivariate joint models is therefore necessary. It allows us to consider more information at the same time, thus enabling us to understand better the dynamic complexity of changes in diseases or biological indicators. In recent years, a number of researchers have worked on multivariate joint models. For instance, Ibrahim et al. [2004]; van Boven et al. [2018]; Hatfield

et al. [2011] investigate multivariate joint models for continuous responses. There are also joint models for other data types, e.g. Huang et al. [2001] contemplates multivariate binary data; the model in Rizopoulos and Ghosh [2011] combines continuous responses and binary responses and Wang et al. [2002] focuses ordinal data. Rue et al. [2017] developed a joint multivariate model associated with discrete data. It is worth noting that although the responses in Rue et al. [2017] are discrete variables, the longitudinal data model in the article still uses a normal distribution model and a beta model.

Many packages also provide relevant functions for joint models. Examples include the `merlin` package in `Stata` [Crowther, 2020]; the `JMbayes` package [Rizopoulos et al., 2022], the `JM` package [Rizopoulos, 2012], the `joineR` package [Philipson et al., 2012], the `lcmm` package [Proust-Lima et al., 2017], the `frailtypack` package [Rondeau et al., 2012], and the `rstanarm` package [Goodrich et al., 2022] in `R`. Amongst these commonly used packages, packages `JMbayes`, `JM` and `rstanarm` can be used to construct multivariate joint models. However, in these multivariate joint models, the relationships between the longitudinal responses are captured by the random effects. This means that in addition to the interdependencies explained by the random effects we mentioned above: the longitudinal outcome is conditionally independent of the time-to-event outcome; the repeated measurements in the longitudinal outcome are conditionally independent of each other. Random effects are also used to explain the relationships between multiple longitudinal responses. For the `lcmm` package, we can use a longitudinal IRT model to analyse homogeneous and heterogeneous data. However, the `lcmm` is a package that focuses on latent-class joint models which cannot be used to construct shared random-effects joint models.

In addition to the above, sometimes the observed responses do not fully reflect the heterogeneity of the outcomes, and the population may be composed of 'latent subpopulations' defined by some unobserved characteristics [Arbeev et al., 2014]. A special type of joint model, the latent-class joint model, has been developed to measure this implicit heterogeneity in the population, see Proust-Lima et al. [2014].

Articles Rouanet et al. [2016]; Proust-Lima et al. [2009, 2016] used latent-class joint models to construct multivariate joint models. The advantage of building a latent-class joint model over a shared random-effects joint model is that we no longer need to do numerical integration, which greatly improves the computational efficiency of the model. The limitation of latent-class joint models is that the classification may not be interpretable because we do not know what the different classes represent. Furthermore, we could only calculate the probability of an individual belonging to a specific class. Although we can assign an individual to a specific class, we can never be sure that the individual belongs to that class.

1.1.3 Ageing research

In this thesis, we will use three datasets related to medical statistics. Two of the datasets, the English Longitudinal Study of Ageing (ELSA) data [J. et al., 2021] and the PAQUID: Longitudinal data on cognitive and physical aging in the elderly (PAQUID) data [Rondeau et al., 2008], are relevant to ageing research. In ageing research, various cognitive function tests are often collected as longitudinal data. As with other longitudinal data in medical statistics, longitudinal data on ageing research is also multi-dimensional [Diggle et al., 2002]. For instance, the ELSA data contain several cognitive tests related to literacy, numeracy, memory and information processing [Banks et al., 2006]. Individuals are tested every few years, and the test scores are recorded; these test scores are recorded as non-negative integers. An important area of investigation in ageing research is change of cognitive function over time.

Cognitive function, also known as cognitive ability, refers to an individual's ability to process information. It mainly reflects an individual's ability to learn and solve problems. Researchers use different tests to obtain scores that reflect an individual's level of cognitive functioning. In analysing these test scores, changes in test scores represent changes in cognitive function. A well-known test that measures cognitive function is the Mini-Mental State Examination (MMSE), also known as the Folstein Test [Folstein et al., 1975]. It is a 30-point questionnaire examining several areas, such as verbal ability, attention, memory, etc. The more questions the

participants answered correctly, the higher the score they received, and the higher score indicates a better cognitive function. The MMSE is widely used in clinical and research settings to reflect cognitive functioning.

It is crucial to investigate changes in individual cognitive function in ageing research. The relationship between changes in cognitive function and the risk of death or dementia is also of interest [Hickey et al., 2016]. We could establish the relationship between time-to-event and risk of death or dementia through survival analysis. Researchers analyse the relationship between cognitive decline and ageing to provide advice on whether older people need care [Van Den Hout and Muniz-Terrera, 2016]. For example, suppose an individual has a sudden decrease in cognitive function over a period of time. In that case, he/she may suffer from a condition that causes cognitive impairment or even life-threatening changes, such as Alzheimer's disease. By analysing the data, the researcher can note this cognitive decline trend and thus make recommendations on the specialist medical care he/she needs.

1.2 Motivation

As we mentioned in Section 1.1.2, studies of longitudinal and time-to-event data often need to be analysed by means of corresponding longitudinal models, and attrition due to any event of interest (e.g. death or dementia) cannot be ignored [Rizopoulos, 2012]. The nonrandom attrition could be investigated by the survival models. In this thesis, we will therefore develop joint models to simultaneously measure the longitudinal response and the event of interest at the time of occurrence.

In addition, researchers often use more than one test in the same study to measure longitudinal outcomes (multivariate). For example, the PAQUID data contain five measures of cognitive test: the Mini-Mental State Examination, the Benton Visual Retention Test, the Isaacs Set Test, the score of physical dependency, and the score measuring depressive symptomatology [Proust-Lima et al., 2017]. Each individual attended these five tests, and the results of these tests reflect the cognitive function of the individual under test. Although the scores on all five tests represent

cognitive function, the different tests have different emphases. All of these results may be important to the research question, and as we consider more of the corresponding variables in the research process, we will obtain more comprehensive information about an individual's cognitive abilities. Therefore, it is worthwhile to construct a multivariate model. We build a bivariate joint model to address the potential estimation bias caused by separate modelling and the need to discard longitudinal responses when using a univariate joint model.

Moreover, we have mentioned in Section 1.1.2 that packages `JMbayes` [Rizopoulos et al., 2022], `JM` [Rizopoulos, 2012] and `rstanarm` [Goodrich et al., 2022] can be used to analyse multivariate joint models. In these multivariate joint models, the correlations between the longitudinal response variables are captured with random effects. It is not easy to intuitively discern the relationship between the longitudinal responses through random effects. We address the inability to observe relationships between longitudinal responses directly by using a bivariate extended binomial distribution [Altham and Hankin, 2012].

1.3 Scope of research

In this thesis, the longitudinal and the survival models are linked by sharing random effects, and we assume that these two models are independent given random effects. The random effects can account for unobservable individual-specific or class-specific characteristics, i.e. in medical science, the unobserved susceptibility for morbidity and mortality can be included in the random effect [Rizopoulos, 2012]. We will use two types of random effects in this thesis. The first is the random effect following the bivariate normal distribution, in which we assume the unobserved individual-specific effects are normally distributed. The normal distribution is the standard modelling choice for random effects [Hickey et al., 2016]. We use normally distributed random effects to construct shared random-effects joint models. Moreover, we introduce the second type of random effect, where the random effects are discrete and class-specific. The corresponding joint models with discrete random effects are namely latent-class joint models [Proust-Lima et al., 2009]. The

latent-class joint models avoid the calculation of integrals and increase the model's efficiency.

As we mentioned in Section 1.1.3, test scores for cognitive abilities in ageing studies are usually in the form of non-negative integers. Not only limited to ageing studies but in other datasets in the field of medical statistics, response variables are often recorded as non-negative integers. As mentioned in the Abstract, we will use three datasets in this thesis. The longitudinal responses in these three datasets are all non-negative integers, with the responses for the first two datasets being the sum of the scores and the response for the third dataset being categorical variables. In this thesis, we will collectively refer to the types of responses as discrete variables. In each dataset, we choose two outcomes as responses for the longitudinal model, i.e. we build bivariate joint models in this thesis. The bivariate joint model could address the potential estimation bias caused by separate modelling and avoid the lack of information caused by dropping longitudinal responses when using a univariate joint model.

We use two distributions to model the discrete longitudinal responses. For the first distribution, we assume that the probability of an individual increasing score by one unit for the longitudinal response follows a Bernoulli distribution. The corresponding probability of obtaining a non-negative score for the response variable follows a binomial distribution. The bivariate extension of the binomial distribution proposed by Altham and Hankin [2012] is applied in the longitudinal model. This distribution not only measures overdispersion or underdispersion compared to the standard binomial distribution but also uses a parameter that measures the association between the two longitudinal outcomes. The model allows us to avoid using random effects to capture all the correlations of the joint model. We could also consider each possible value of the responses as a category and model responses using the categorical distribution [Murphy, 2012]. For the survival part, the exponential hazard model, the Weibull hazard model and the Gompertz hazard model are used. In univariate joint models, the link functions contained in both the longitudinal model and the survival model are well defined [Diggle et al., 2008]. However,

we do not know beforehand which link will provide the best result for the bivariate joint model. We will therefore discuss several different link functions.

We apply the models to analyse the English Longitudinal Study of Ageing (ELSA) data [Banks et al., 2006], the PAQUID: Longitudinal data on cognitive and physical aging in the elderly (PAQUID) data [Rondeau et al., 2008] and the Primary biliary cholangitis (PBC2) data [Mayo, 2022]. The ELSA data are related to verbal learning and recall. Individuals are required to learn ten words and recall these words at two different time points (immediate and later); see Van Den Hout and Muniz-Terrera [2019]. The PAQUID data are random samples of cognitive test scores for people aged over 65 years old in two administrative areas of southwestern France. The PBC2 data are from a group of trials on primary biliary cholangitis of the liver (PBC) conducted between 1974 and 1984 [Mayo, 2022].

As we mentioned in Section 1.1, longitudinal and time-to-event data are used in various fields, such as psychology, sociology and medicine. Therefore, the application of the joint models in this thesis is not limited to the analysis of the three datasets used. The base joint models we have constructed could be used to analyse other discrete longitudinal and time-to-event data by modifying some parameters.

In summary, the main contribution of the thesis is the construction of bivariate joint models that can be widely applied to longitudinal and time-to-event data with discrete responses relevant to medical research. In addition to the discrete data we analysed in this thesis, our joint model can also be applied to the analysis of continuous data and discrete data with a wide range of values. For continuous data, we can discretise the data into several categories and use the joint model presented in this thesis for analysis. By categorising the response variables into subcategories, the bivariate joint models we construct can also be used to analyse discrete data with a wide range of values.

1.4 Outline

Chapter 2 provides a brief overview of the models in this thesis: the extension of the bivariate binomial distribution proposed by Altham and Hankin [2012], the categor-

ical distribution, the exponential hazard model, the Weibull hazard model and the Gompertz hazard model. The link between the longitudinal model and the survival model is also discussed in this part. Chapter 3 constructs the marginal log-likelihood function for joint models, discusses the left truncation and proposes the method for computation. In Chapter 4, we use simulation studies to validate the accuracy of the joint model parameter estimates. We introduce the three data used in this thesis in Chapter 5. We analyse these three datasets by fitting the shared random-effects joint models and the latent-class joint models in Chapter 6. The conclusion of this thesis and possible extension options for the joint model are discussed in Chapter 7.

Chapter 2

Models

In this chapter, we will introduce a series of models used to compose the joint model. As mentioned in Section 1.1.2, the joint model uses a common latent variable to link the longitudinal model and the survival model. These two models are independent given the common latent variable [Rizopoulos, 2012]. We will therefore present several different longitudinal models to analyse discrete longitudinal data and several commonly used survival models.

The joint model used in this thesis is given by

$$p\left(\boldsymbol{y}, \boldsymbol{t}, t_{last}\right) = \int p\left(\boldsymbol{y} | \boldsymbol{t}, \boldsymbol{b}\right) p\left(t_{last} | \boldsymbol{b}\right) p(\boldsymbol{b}) d\boldsymbol{b}, \tag{2.1}$$

where $\boldsymbol{y}$ denotes the longitudinal outcomes, $\boldsymbol{t}$ represents the corresponding time points, t_{last} is the time point for the last observation and $\boldsymbol{b}$ represent the random effects. Distribution $p\left(\boldsymbol{y} | \boldsymbol{t}, \boldsymbol{b}\right)$ represents the longitudinal model, $p\left(t_{last} | \boldsymbol{b}\right)$ is the survival model and $p\left(\boldsymbol{b}\right)$ is the distribution of the random effect.

Section 2.1 introduces the extension of the binomial distribution proposed by Altham and Hankin [2012] and the categorical distribution [Murphy, 2012] used for the longitudinal model. Section 2.2 proposes hazard models for analysing events of interest. Section 2.3 defines the link function between the longitudinal model and the survival model, which consists of the common parameters and random effects of these two models. Finally, we define the distribution of random effects in the shared random-effects joint model in Section 2.4. For what follows, assume the

random effects are given by vector $\boldsymbol{b} \in \mathbb{R}^p$, where the dimension p is determined by the model definition.

2.1 Distributions for the longitudinal outcomes

2.1.1 Binomial distribution

Typically, many response variables used to describe individuals' cognitive abilities in longitudinal data are expressed as non-negative integers. For example, the Mini-Mental State Examination (MMSE) test is a 30-point questionnaire in which an individual's score is taken as an integer from 0 to 30 [Folstein et al., 1975]. This test is widely used in clinical and research settings to measure cognitive function. A higher score on the MMSE test means that the individual has a better cognitive ability. For such discrete data, we model the probability of success, i.e. the individual gets the question right and earns a point. We assume that the 30 questions are independent of each other so that the distribution of the total scores obtained by individuals could be treated as following a binomial distribution. It is worth noting that even though each question had a different probability of success, the total score distribution still follows the binomial distribution [McCullagh and Nelder, 1989].

Let Y denote the response and p denote the probability of success. Parameter m denotes the number of trials. Altham [1978] presented the following extension of the binomial distribution.

$$p(Y = y) = \frac{\binom{m}{y} p^y (1-p)^{(m-y)} \theta^{y(m-y)}}{C_{uni}}, \tag{2.2}$$

for $y \in \{0, 1, ..., m\}$, where $0 < p < 1$, $\theta > 0$, and

$$C_{uni} = \sum_{y=0}^{m} \binom{m}{y} p^y (1-p)^{(m-y)} \theta^{y(m-y)}. \tag{2.3}$$

Compared to the standard binomial distribution, this extended distribution adds the parameter θ to change the shape of the distribution, making it more flexible when fitting the data. When $\theta > 1$, distribution (2.2) is under-dispersed compared to the standard binomial distribution; when $0 < \theta < 1$, the distribution is over-dispersed

relative to the standard distribution.

In this thesis, we are committed to simultaneously measuring the two responses appearing in the longitudinal data. Therefore we use the bivariate extension of Equation (2.2) [Altham and Hankin, 2012]. Let Y_1 and Y_2 denote the bivariate responses.

$$p(Y_1 = y^{(1)}, Y_2 = y^{(2)}) = \frac{g(Y_1 = y^{(1)})g(Y_2 = y^{(2)})\phi^{y^{(1)}y^{(2)}}}{C_{biv}}, \qquad (2.4)$$

for $y^{(j)} \in \{0, 1, ..., m_j\}$, $j = 1, 2$.

$$g\left(Y_j = y^{(j)}\right) = \binom{m_j}{y^{(j)}} p_{Y_j}^{y^{(j)}} (1 - p_{Y_j})^{(m_j - y^{(j)})} \theta_{Y_j}^{y^{(j)}(m_j - y^{(j)})}, \qquad (2.5)$$

where p_{Y_j} denote the probability of success for $j = 1, 2$; $0 < p_{Y_1}, p_{Y_2} < 1$, $\theta_{Y_j}, \phi > 0$. Compared to the univariate binomial extension distribution (2.2), the bivariate extension of the binomial distribution has a new parameter ϕ. The parameter ϕ is used to measure the correlation between two responses. When $0 < \phi < 1$, there is a negative correlation between the two responses; when ϕ is greater than 1, the responses are positively correlated. The denominator of the Equation (2.4) is:

$$C_{biv} = \sum_{y^{(1)}=0}^{m_1} \sum_{y^{(2)}=0}^{m_2} g(Y_1 = y^{(1)})g(Y_2 = y^{(2)})\phi^{y^{(1)}y^{(2)}}. \qquad (2.6)$$

The integers m_1 and m_2 define the range of the two responses $y^{(1)}$ and $y^{(2)}$. In this distribution, the ranges of the two responses could be unequal.

The reason for choosing this bivariate extension of the binomial distribution is that the formula of the probability density function is relatively simple, and it belongs to the exponential family [Altham and Hankin, 2012]. Moreover, this distribution allows Y_1 and Y_2 to have separate choices of m_1 and m_2.

Assume that the random effect $\boldsymbol{b}$ is a vector, where $\boldsymbol{b} = (b_{10}, b_{11}, b_{20}, b_{21})^T$. This vector includes random intercepts (b_{10} and b_{20}) and random slopes (b_{11} and b_{21}). We use the parameters b_{j0} and b_{j1} to denote the corresponding random intercept and random slope for the responses $Y_j, j = 1, 2$. The parameter t represents the

time or age of observation for individuals in the longitudinal data.

To investigate the relationship between time (age) and the responses, we use the logistic regression model, which contains random-effects:

$$p_{Y_j} = \frac{\exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t\right)}{1 + \exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t\right)}. \tag{2.7}$$

Parameter η_0s are the fixed intercepts in the model, and η_1s are the fixed slopes for response Y_j. We could include other covariates, such as gender and education level, in model (2.7) to investigate the effects of these covariates on an individual's cognitive function. We use x to represent any possible covariate, and the extended logistic regression model containing covariates (2.7) can be written as follows:

$$p_{Y_j} = \frac{\exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t + \gamma_L^{(j)}x\right)}{1 + \exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t + \gamma_L^{(j)}x\right)}. \tag{2.8}$$

Parameter $\gamma_L^{(j)}$ is the coefficient of the corresponding covariate, and subscript L represents the longitudinal model.

In this thesis, we will investigate restricted models, i.e. the model contains random intercepts or random slopes for the corresponding two responses. We will explain why restricted models are used in Section 2.3.1 and Chapter 6. The restricted model is defined by changing the value of the random effects, for example:

$$b_{10} = b_{20}, \; b_{11} = b_{21}, \tag{2.9}$$

where two responses Y_1 and Y_2 share the same random intercept b_{10} and the same random slope b_{11}.

We could also set

$$b_{11} = b_{21} = 0, \tag{2.10}$$

which means that we only use random intercept for each response. Also, we can

assume

$$b_{10} = b_{20} = 0, \tag{2.11}$$

where responses have different random slope (b_{11} and b_{21}) and do not have corresponding random intercepts.

2.1.2 Categorical distribution

As we mentioned in Chapter 1, some biomarkers are represented by discrete non-negative integers. Therefore, each of the possible outcome values could be considered a category. Let us use Mini-Mental State Examination (MMSE) again as an example [Folstein et al., 1975]. There are 31 possible outcome values for the MMSE scores, i.e. the individual's score could be any integer from 0 to 30. Therefore, these 31 outcome values can be considered as 31 categories. We could use the categorical distribution to analyse these discrete biomarkers.

The categorical distribution is a discrete probability distribution, also known as the generalised Bernoulli distribution and the multinoulli distribution. The distribution describes the possible outcomes of a variable that can be one and only one of K possible categories, with the probability of each category being specified separately [Murphy, 2012]. For example, we need to model the probabilities of 31 outcomes separately when using categorical distribution to analyse the MMSE scores. It is worth noting that a categorical distribution is a special case of a multinomial distribution. Compared with the multinomial distribution, the categorical distribution gives the probability of the potential outcome of one trial rather than multiple trials. Using the example of dice rolling, the categorical distribution measures the probability of getting a particular side on a single roll rather than the number of times a side is rolled on multiple rolls. Numeric labels are often used in the categorical distribution to indicate the corresponding category.

There are two primary types of response outcomes in categorical distributions: nominal and ordinal. For nominal outcomes, the order of labels is not important for the allocation of outcomes. For example, suppose the response is the preferred type

of music (jazz, rock, blues, etc.), we can label any type as category 1 since there is no natural order between these categories. Correspondingly, ordinal outcomes refer to response variables for which there exists an order between categories. Examples include the number of words immediately recalled and the number of delayed recall words in the ELSA data and histology and edema in the PBC2 data used in this thesis. Agresti [2003] has mentioned that the methods used to analyse nominal can also be used to analyse ordinal. In this thesis, we will use baseline-category logit models [Agresti, 2003] to construct the joint model, which uses the categorical model as the longitudinal model.

Let Y_j denote the response, in which $j = 1, 2$ represents the index of responses. Vector $\boldsymbol{p}$ is the parameter vector for the categorical distribution. Parameter k represents the number of categories. Index i represent the ith category. The probability mass function can be written as:

$$P(Y_j|\boldsymbol{p}) = \prod_{i=1}^{k} [y_j = i] \, p_i^{(j)}, \tag{2.12}$$

where $[y = i]$ evaluates to 1 if $y = i$, 0 otherwise.

The probability mass function can also be written as:

$$P(Y_j|\boldsymbol{p}) = \prod_{i=1}^{k} p_i^{(j)y_{ji}}. \tag{2.13}$$

In this formula, the sample space of y can be thought of as a set of k vectors of dimension k. Moreover, these vectors have the property that exactly one element has a value of 1, and the other elements have a value of 0.

Assume that b_{j0} represents the random intercept of the response Y_j, and b_{j1} represents the random intercept of the response Y_j. Let t represent time or age.

We still use the logistic regression model to investigate the distribution of time and responses. Suppose set $\mathbb{A}_j$ is the set of all possible category indexes except the baseline category l for response Y_j. We use i^* to denote the category with random

effects, and $p_{Y_{ji^*}}$ represents the corresponding probability. We have:

$$p_{Y_{ji^*}}(t) = \frac{\exp\left(\boldsymbol{\eta}_{i^*}^{(j)}\right)}{1 + \exp\left(\boldsymbol{\eta}_{i^*}^{(j)}\right) + \sum_{i \in \mathbb{A}_j \setminus \{i^*\}} \exp\left(\eta_{0i}^{(j)} + \eta_{1i}^{(j)} t\right)}, \quad (2.14)$$

where $\boldsymbol{\eta}_{i^*}^{(j)} = (\eta_{0i^*}^{(j)} + b_{j0}) + (\eta_{1i^*}^{(j)} + b_{j1})t$ represents the linear predictor corresponding to category i^* and response Y_j. In constructing a logistic regression, we need a baseline category l, where

$$p_{Y_{jl}}(t) = \frac{1}{1 + \exp\left(\boldsymbol{\eta}_{i^*}^{(j)}\right) + \sum_{i \in \mathbb{A}_j \setminus \{i^*\}} \exp\left(\eta_{0i}^{(j)} + \eta_{1i}^{(j)} t\right)}. \quad (2.15)$$

Similar to Equation (2.8) in Section 2.1.1, we could also include other covariates in the logistic regression models (2.14) and (2.15). We still use x to denote the covariate used and $\gamma_L^{(j)}$ to denote the corresponding parameter of that covariate. The extension of the models ((2.14) and (2.15)) containing the covariate can be written as:

$$p_{Y_{ji^*}}(t) = \frac{\exp\left(\boldsymbol{\eta}_{i^*}^{(j)} + \gamma_L^{(j)} x\right)}{1 + \exp\left(\boldsymbol{\eta}_{i^*}^{(j)} + \gamma_L^{(j)} x\right) + \sum_{i \in \mathbb{A} \setminus \{i^*\}} \exp\left(\eta_{0i}^{(j)} + \eta_{1i}^{(j)} t\right)}, \quad (2.16)$$

and

$$p_{Y_{jl}}(t) = \frac{1}{1 + \exp\left(\boldsymbol{\eta}_{i^*}^{(j)} + \gamma_L^{(j)} x\right) + \sum_{i \in \mathbb{A} \setminus \{i^*\}} \exp\left(\eta_{0i}^{(j)} + \eta_{1i}^{(j)} t\right)}, \quad (2.17)$$

respectively.

The possible restricted models are the same as in Section 2.1.1:

$$b_{10} = b_{20}, \; b_{11} = b_{21}, \quad (2.18)$$

where two responses Y_1 and Y_2 share the same random intercept b_{10} and the same

random slope b_{11}. When

$$b_{11} = b_{21} = 0, \tag{2.19}$$

it means that only the random intercept is added into the logistic regression. Similarly, when

$$b_{10} = b_{20} = 0, \tag{2.20}$$

responses have different random slopes (b_{11} and b_{21}) and do not have corresponding random intercepts.

2.2 Hazard models

We now define the hazard and the corresponding survival models for the analysis of time-to-event data. The hazard model is a parametric regression model with the formula:

$$h(t) = h_0(t) \exp\left(\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, \gamma_S, t)\right). \tag{2.21}$$

The expression $h_0(t)$ represents the baseline hazard function. Function Δ is defined as a link function which refers to an expression included in both the longitudinal model and the survival model. The link function is composed of random effects $\boldsymbol{b}$, additional parameters $\boldsymbol{\alpha}$ and parameters common to both the longitudinal and survival models $\boldsymbol{\eta}$. We will introduce the link function in detail in Section 2.3. The vector $\boldsymbol{\eta} = (\eta_0^{(j)}, \eta_1^{(j)})$ represents the fixed intercept and fixed slope for response Y_j, which we also use in the logistic regression models (2.7) - (2.8) and (2.14) - (2.17). Similar to the Equation (2.8), γ_S is the parameter of the covariate and the subscript S represents the survival model. Time t in the survival model is the last observed time point (or age).

In survival analysis, there are a number of widely used parametric hazard models, such as the exponential hazard model, the Weibull hazard model, the Gompertz

hazard model and the log-logistic hazard model. In this thesis, we use three widely used hazard models to construct the survival component of the joint model: the exponential model, the Weibull hazard model and the Gompertz hazard model.

The exponential hazard model, denoted by $Exponential(\lambda)$, is a simple hazard model with only one parameter. We build the joint model with the exponential hazard model based on the model estimation results in Chapter 6, and we will explain in detail the reasons for using this hazard model in Chapter 6. The Weibull hazard model denotes by $Weibull(\lambda, \tau)$ is a generalisation of the exponential hazard model. Compared to the exponential hazard model, the Weibull hazard model has an additional positive parameter τ, and the model's shape could be modified by varying the value of parameter τ. The Weibull model and the exponential model are equivalent when τ equals 1. Liu [2012] mentions that of all the families of parametric survival distributions, the Weibull model is probably the most widely used parametric function in survival analysis due to its simplicity and flexibility. Therefore, we chose it as one of the survival models used in this thesis.

The Gompertz hazard model is the most widely used model within the research field of quantitative description of human mortality and survival [Pollard and Valkovics, 1992; Liu, 2012]. In contrast to the Weibull model, the numerical estimation of the Gompertz model is computationally robust because there is no power function in the Gompertz model. Death is the event of interest for the ELSA data and PBC2 data used in Chapter 6. Given that the event of interest is death, we chose the Gompertz model as one of the survival models. Although we restrict the survival model to the exponential, Weibull and Gompertz models in this thesis, the survival part of our joint model could be easily replaced by other frameworks.

The baseline hazard model for the three models is expressed as:

$$\text{Exponential}: h_0(t) = \lambda \tag{2.22}$$

$$\text{Weibull}: h_0(t) = \lambda \tau t^{\tau-1} \tag{2.23}$$

$$\text{Gompertz}: h_0(t) = \lambda \exp(\xi t) \tag{2.24}$$

where $\lambda, \tau > 0$, and the corresponding survivor functions are:

$$\text{Exponential}: S(t) = \exp(-\lambda t) \tag{2.25}$$

$$\text{Weibull}: S(t) = \exp(-\lambda t^{\tau}) \tag{2.26}$$

$$\text{Gompertz}: S(x) = \exp\left(-\frac{\lambda}{\xi}\left(\exp(\xi t) - 1\right)\right). \tag{2.27}$$

We use the parameterisation $\lambda = \exp(\beta)$ in the joint model for computational reasons to ensure a positive hazard for any $\beta \in \mathbb{R}$.

The range of ξ in the Gompertz model has been defined variously in the literature. For example, some researchers have defined ξ as $\xi \in \mathbb{R}$. In this situation, the hazard grows exponentially with time when $\xi > 0$ and decreases when $\xi < 0$ [Kleinbaum and Klein, 2010]. However, if the parameter ξ is less than 0 when time t tends to infinity, we have the following property of the survival function:

$$\lim_{t \to \infty} S(t) = \exp\left(\frac{\lambda}{\xi}\right) > 0. \tag{2.28}$$

This equation shows that the probability of survival is still greater than zero, even if the time is quite long. It means that the events of interest will not occur in a proportion of the population. The events of interest for the three datasets used in this thesis are death or dementia, and the process from health to the event is irreversible. Therefore we cannot set ξ to $\xi \in \mathbb{R}$. We follow the definition given by Wienke [2010] and restrict ξ to be positive. In this case, the hazard function is equal to λ when $t = 0$ and it increases to ∞ when $t = \infty$. The corresponding survival function is equal to 1 when $t = 0$ and it decreases to 0 when t rises to ∞.

2.3 Link between the models

This section introduces the link function $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, \gamma_S, t)$ of the model, defined as an expression containing the common parameters in both the longitudinal model and the survival model. The link function Δ contains the following parameters: random effects $\boldsymbol{b}$; fixed parameters $\boldsymbol{\eta}$; the covariate x and corresponding parameters γ_S and parameters $\boldsymbol{\alpha}$. Parameters in the link function except $\boldsymbol{\alpha}$ are included in both the

longitudinal and the survival models.

We divide this section into two parts: the link functions used in the shared random-effects joint model and the latent-class joint model. In each part, we will first introduce the expression for the link function and discuss possible restriction models for the link function. After that, we will provide some explanations of the corresponding hazard models based on the link functions. Since in this section we will focus on the structure of the random intercept and random slope, we will simplify the link function $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, \gamma_S, t)$ to $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t)$.

2.3.1   Link functions for shared random-effects models

In the joint model, hazard model (2.21) allows for several specialisations. For univariate joint models, the expression $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t)$ is usually specified as $\alpha(\eta_0 + b + (\eta_1 + b)t)$ [Diggle et al., 2008]. Linear predictors are also commonly used as link functions in many articles relating to multivariate joint models. In our thesis, we also use the linear predictor as the link function. When investigating the bivariate joint model, we do not know in advance which link function will provide the best results. We, therefore, list in this section several possible expressions for the link function and compare the joint models constructed by the different link functions in Chapter 6.

As mentioned in Section 2.1, we restrict models based on Equations (2.10) and (2.11) when constructing link functions. We use restricted models for two main reasons. Suppose we use the full model, i.e. models containing Equation (2.7) or (2.14) for $j = 1, 2$, it would be unclear how to construct the link between the models and how to interpret the impact of the corresponding parameters on the risk of the event. Moreover, we encounter challenges in computing high-dimensional integrals in the shared random-effects joint models. Therefore, we reduce the dimensionality of the integration, i.e. the number of corresponding random effects. For each response variable, we include only one corresponding random effect. We consider a range of definitions of Δ, starting with

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha \left(\eta_0^{(j)} + b_{j0} \right), \tag{2.29}$$

where $\eta_0^{(j)} + b_{j0}$ is the fixed intercept and the random intercept for response Y_j, for $j = 1$ or 2. When the longitudinal model is a categorical distribution, we write the corresponding link function according to Equation (2.14) and (2.15):

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha \left(\eta_{0i*}^{(j)} + b_{j0} \right). \tag{2.30}$$

We can also include the random intercept for both responses Y_1 and Y_2:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \sum_{j=1}^{2} \alpha_j \left(\eta_0^{(j)} + b_{j0} \right), \tag{2.31}$$

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \sum_{j=1}^{2} \alpha_j \left(\eta_{0i*}^{(j)} + b_{j0} \right). \tag{2.32}$$

The advantage of (2.29) - (2.32) is that these link functions does not include time, so they are computationally easier than the link function that contains time. However, changes in longitudinal outcomes over time may have a greater impact on the risk of death or dementia compared to the value of the outcomes at the start of the time scale. Therefore, we try to add random slopes to the link function in the same way:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \sum_{j=1}^{2} \alpha_j \left(\eta_1^{(j)} + b_{j1} \right) t, \tag{2.33}$$

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \sum_{j=1}^{2} \alpha_j \left(\eta_{1i*}^{(j)} + b_{j1} \right) t. \tag{2.34}$$

For the above link function containing random slopes, we also consider the restricted version, i.e. setting one of the α_js to 0:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha \left(\eta_1^{(j)} + b_{j1} \right) t, \tag{2.35}$$

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha \left(\eta_{1i^*}^{(j)} + b_{j1} \right) t. \tag{2.36}$$

Equation (2.33) - (2.36) takes into account the trajectory of Y_1 and Y_2 changing over time by adding the fixed slope and the random slope.

Although we only use restricted models in this thesis, we still discuss the link functions that could be used for the full model. The random intercept and the random slope could be included in the same link function at the same time:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \sum_{j=1}^{2} \alpha_{1j} \left(\eta_0^{(j)} + b_{j0} \right) + \alpha_{2j} \left(\eta_1^{(j)} + b_{j1} \right) t, \tag{2.37}$$

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \sum_{j=1}^{2} \alpha_{1j} \left(\eta_{0i^*}^{(j)} + b_{j0} \right) + \alpha_{2j} \left(\eta_{1i^*}^{(j)} + b_{j1} \right) t. \tag{2.38}$$

In other literature investigating multivariate joint models, there are many common forms of link functions in addition to those mentioned above. For instance, in many joint models that use a linear-mixed model as their longitudinal model, the conditional expectation of the response is commonly used as the link function:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha_j \left(\eta_0^{(j)} + b_{j0} + (\eta_1^{(j)} + b_{j1})t \right). \tag{2.39}$$

Equation (2.39) is a common form of using expectation $E(Y_j)$ in the construction of the link function. We will not use this link function in our thesis. The reason is that in our distribution, the expectation of the response is not a linear predictor. The expectation in our joint model is a function of the range m_j of the response and the probability of success p_{Y_j}.

Alternatively, we could further extend the link function using an expression of probabilities. We choose the probability in the categorical longitudinal model as an example. This probability is constructed from the last observed survival time t_s, i.e. $(p_{Y_{ji^*}}(t_s))$. When we add random intercepts into the i^*th category in the longitudinal

model, the corresponding link function can be written as:

$$\Delta(\boldsymbol{\alpha},\boldsymbol{\eta},\boldsymbol{b},t) = \sum_{j=1}^{2} \alpha_j p_{Y_{ji*}}(t_s). \tag{2.40}$$

When random slopes are included in the i^*th category in the longitudinal model, we write the link function as

$$\Delta(\boldsymbol{\alpha},\boldsymbol{\eta},\boldsymbol{b},t) = \sum_{j=1}^{2} \alpha_j p_{Y_{ji*}}(t_s)t. \tag{2.41}$$

When we combine the above Δ expressions into the baseline hazard models $h_0(t)$, we need to be careful with the parameters λ, τ and ξ in the Weibull hazard model and the Gompertz hazard model.

For the Weibull hazard, when random intercepts are contained in the link function Δ, i.e. the expression of Δ is equal to (2.29) - (2.32) or (2.40). The hazard model can be written as:

$$h(t) = \lambda^* \tau t^{(\tau-1)}. \tag{2.42}$$

With Δ equal to (2.31) or (2.40), the scaling parameter λ^* can be rewritten as an individual-specific parameter:

$$\lambda^* = \exp\left(\beta + \sum_{j=1}^{2} \alpha_j \left(\eta_0^{(j)} + b_{j0}\right)\right), \tag{2.43}$$

or

$$\lambda^* = \exp\left(\beta + \sum_{j=1}^{2} \alpha_j p_{Y_{ji*}}(t_s)\right), \tag{2.44}$$

respectively. The Weibull hazard model can also be written as a hazard model with an individual-specific slope parameter:

$$h(t) = \lambda \tau^* t^{(\tau^*-1)}, \tag{2.45}$$

where

$$\tau^* = \exp\left(\sum_{j=1}^{2} \alpha_j \left(\eta_0^{(j)} + b_{j0}\right)\right)\tau, \tag{2.46}$$

or

$$\tau^* = \exp\left(\sum_{j=1}^{2} \alpha_j p_{Y_{ji^*}}(t_s)t\right)\tau, \tag{2.47}$$

respectively. In this thesis, when we use the Weibull hazard model, we write the corresponding expression in the form of Equation (2.42). When expressions for Δs include random slopes ((2.35) - (2.38) and (2.41)) and are combined with the Weibull hazard model, the shape parameter τ becomes a time-dependent parameter. The time-dependent τ violates the definition of the standard Weibull hazard model. When we want to add random slopes to the Weibull hazard model, we have to compute the log cumulative hazard function [Crowther et al., 2012]. This thesis will not discuss using the log cumulative hazard function.

For the Gompertz hazard, we have:

$$h(t) = \lambda^* \exp(\xi^* t). \tag{2.48}$$

In this case, we need to rewrite the part containing the random intercept as λ^* and the part containing the random slope as ξ^*. Take for example a link function equal to $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha_1\left(\eta_0^{(j)} + b_{j0}\right) + \alpha_2\left(\eta_1^{(j)} + b_{j1}\right)t$, the corresponding parameter λ^* and ξ^* should be written as:

$$\begin{aligned} \lambda^* &= \exp\left(\beta + \alpha_1(\eta_0^{(j)} + b_{j0})\right), \\ \xi^* &= \alpha_2(\eta_1^{(j)} + b_{j1}) + \xi. \end{aligned} \tag{2.49}$$

Both λ^* and ξ^* are parameterised for individuals. The Gompertz hazard model can easily handle the function Δ with random slopes. The reason is that a model conditional on random effects is still a Gompertz hazard model.

2.3.2 Link functions for latent-class models

The latent-class model, also known as the non-parametric frailty model [Van Den Hout, 2017], assumes that the population is heterogeneous and can be divided into a limited number of homogeneous classes [Proust-Lima et al., 2009]. Each class is distinguished by a specific response trajectory and the specific risk of the event. In the latent-class joint model, random effects are discretely distributed. We assume that individuals in the data can be divided into G unobserved classes. For each individual i, if i belongs to class g, it is expressed in the formula as $c_i = g$, $(g = 1, \ldots, G)$.

For the latent-class joint model, we need to use the class-specific random effects b_{g0} and b_{g1} instead of the response-specific random effect b_{j0} and b_{j1} in the shared random-effects joint model. (It is worth noting that the response-specific random effects we refer to here are actually individually specific since in the shared random-effects model, the random effects follow a distribution rather than a definite value.)

In contrast to the widely used latent-class joint model [Proust-Lima et al., 2009, 2014, 2016], we use a similar expression to that of the shared random-effects joint model, i.e. there is a link function between the longitudinal and survival models containing the parameters α as well as other shared parameters. There are two reasons for us to construct the latent-class joint model in this way: 1. We can compare the constructed shared random-effects joint model and the latent-class joint model to some extent by using similar expressions. We will use both models to analyse the PAQUID data and compare the fitted results in Section 6.1.4. 2. Similar to the reason for including the link function in the shared random-effects joint model, we can verify whether there is a correlation between the longitudinal model and the survival model by using the estimated α, i.e. whether it is necessary for us to construct a joint model to analyse the data. The choice of the link between the longitudinal model and the survival model is quite flexible in the joint model. For example,

considering the impact of the random effect on the hazard model directly:

$$\Delta(\boldsymbol{\alpha},\boldsymbol{\eta},\boldsymbol{b},t) = \alpha b_{g0}. \tag{2.50}$$

In this case, the expression of the hazard model (2.24) can be written as (take the Gompertz model as an example):

$$h_g(t) = \lambda^* \exp(\xi t) = \exp(\beta + \alpha b_{g0} + \xi t), \tag{2.51}$$

where

$$\lambda^* = \exp(\beta + \alpha b_{g0}). \tag{2.52}$$

In contrast to how the longitudinal model and the survival model share parameters in the shared random-effects joint model, the grouping probability π_{ig} is usually used in the latent-class joint model to link the two models [Proust-Lima et al., 2009].

We have introduced the link function $\Delta(\boldsymbol{\alpha},\boldsymbol{\eta},\boldsymbol{b},t)$ between the longitudinal model and the survival model for the shared random-effects model in Section 2.3.1. In order to have similar link functions Δ between the latent-class joint model and the shared random-effects model, we use similar expressions as the link function (2.29) - (2.38).

$$\Delta(\boldsymbol{\alpha},\boldsymbol{\eta},\boldsymbol{b},t) = \alpha \left(\eta_0^{(j)} + b_{g0} \right), \tag{2.53}$$

which is the random intercept for response Y_1 or Y_2. We can also plug in the random intercept for both Y_1 and Y_2:

$$\Delta(\boldsymbol{\alpha},\boldsymbol{\eta},\boldsymbol{b},t) = \sum_{j=1}^{2} \alpha_j \left(\eta_0^{(j)} + b_{g0} \right). \tag{2.54}$$

For (2.54), responses Y_1 and Y_2 share the same random intercept b_{g0}, and we use b_{g0} for two times.

To analyse the impact of cognitive decline on the risk of death or dementia, we plug in the random slopes in the same way:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha \left(\eta_1^{(j)} + b_{g1} \right) t, \tag{2.55}$$

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \sum_{j=1}^{2} \alpha_j \left(\eta_1^{(j)} + b_{g1} \right) t. \tag{2.56}$$

As in Section 2.3.1, we discuss the link functions that can be used when using the full model. In the link function corresponding to the full model we can use the random intercept and the random slope at the same time:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha \left(\eta_0^{(j)} + b_{g0} + (\eta_1^{(j)} + b_{g1})t \right), \tag{2.57}$$

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \sum_{j=1}^{2} \alpha_j \left(\eta_0^{(j)} + b_{g0} + (\eta_1^{(j)} + b_{g1})t \right), \tag{2.58}$$

in which the random effects are class-specific, and responses Y_1 and Y_2 share the same random effects.

It is worth noticing that for different classes g, some model parameters do not change. For example, parameters α and $\eta_0^{(j)}$ in (2.53) among different classes are the same. In this case, the model is quite close to the shared random-effects model, in which the parameters do not change across the individual specific models. We can compare the computational efficiency between the shared random-effects model and the latent-class joint model.

2.4 Distribution of random effects

As we mentioned at the beginning of Chapter 2, we use $p(\boldsymbol{b})$ to represent the distribution of random effects. In this thesis, we use the bivariate normal distribution as

the distribution of random effects, i.e.

$$\boldsymbol{b} \sim \left(\boldsymbol{0}, \begin{pmatrix} \sigma_{b1}^2 & \rho\sigma_{b1}\sigma_{b2} \\ \rho\sigma_{b1}\sigma_{b2} & \sigma_{b2}^2 \end{pmatrix}\right). \tag{2.59}$$

Parameter ρ represents the correlation between random effects. The parameters σ_{b1} and σ_{b2} denote the corresponding random effects' standard deviation.

When constructing multivariate joint models, the multivariate normal distribution is the most commonly used distribution for random effects [Hickey et al., 2016]. One advantage of using the multivariate normal distribution is that it allows the use of Gauss-Hermite quadrature in the numerical optimisation of marginal likelihoods. Moreover, it is worth noting that even if we set the random effects distribution before fitting the model, the estimates of joint models will not have a significant general bias, according to the conclusions in Pantazis and Touloumi [2007].

Chapter 3

Left truncation and maximum likelihood estimation

The estimation of parameters is undertaken by maximising the log-likelihood function. In this chapter, we use the same notation in Chapter 2. For individual i, $i = 1, \ldots, N$, the corresponding longitudinal responses are $\boldsymbol{y}_i = (\boldsymbol{y}_i^{(1)}, \boldsymbol{y}_i^{(2)})$. Response $\boldsymbol{y}_i^{(j)}$ is a vector $\left(\boldsymbol{y}_i^{(j)} = (y_{i1}^{(j)}, \ldots, y_{in_i}^{(j)}) \right)$ at age $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})$, where $j = 1, 2$ is the jth response, n_i is the number of observations for each individual. Random effects explain the inclusion of interdependencies in the joint model [Rizopoulos, 2012].

We will use three datasets in this thesis: the ELSA data [Taylor et al., 2007], the PAQUID data [Rondeau et al., 2008] and the PBC2 data [Mayo, 2022]. The time in the ELSA data and the PAQUID data is measured with the age of the individual when the observations are recorded. Therefore we need to consider the left truncation when constructing likelihood functions. For the PBC2 data, since the time variable is the time from the start of the study to the time of recording the observations, there is no need to consider left truncation. We will introduce these datasets in detail in Chapter 5.

In this chapter, we will first introduce the left truncation. After that, the shared random-effects joint model and the latent-class joint model are introduced separately. In the sections corresponding to these two joint models, we will present the likelihood functions with and without left truncation for each type of joint model.

3.1 Left truncation

Left truncation, also called *delayed entry*, occurs when individuals are already at risk before they enter the study[Wienke, 2010]. For example, when the event of interest is death, individuals who die before the start of the experiment will not be included in the data. In this case the data analysis is based on the assumption that individuals who experienced the event before the study are not considered. In other words, the data analysis in this case assumes that individuals will not experience the event before the start of the study. Therefore, it is necessary to deal with the left truncation.

We take the ELSA data [Taylor et al., 2007] used in Chapter 6 as an example. In the ELSA data, the main time scale is the age of the individual when the observations are recorded. Since the event of interest in this data is death, individuals can be included in the data only if they have not experienced the event prior to entering the study. If we do not deal with the left truncation, the estimation is based on the assumption that individuals were not at risk of dying before the start of the study. Therefore, the left truncation should be considered in the joint model.

3.2 Likelihood functions for shared random-effects joint models

A shared random-effects model, also known as the parametric frailty model [Van Den Hout, 2017], can be defined as a joint model constructed by incorporating a mixed model containing random effects as covariates in the survival model [Proust-Lima et al., 2009]. The random effects in the shared random-effects models follow continuous distributions, such as normal distribution, log-normal distribution and gamma distribution. Therefore, the shared random-effects joint model provides a flexible way of building relationships between longitudinal and survival models basis on individuals[Van Den Hout, 2017].

3.2.1   Likelihood function without left truncation

Let $\boldsymbol{\omega}$ represent all the parameters in the joint model except the random effects. Assuming independence between responses given the random effect, the marginal likelihood function is:

$$p(\boldsymbol{y}_i, \boldsymbol{t}_i | \boldsymbol{\omega}) = \int p(\boldsymbol{y}_i | \boldsymbol{t}_i, \boldsymbol{b}_i, \boldsymbol{\omega}) \, p(t_{in_i} | \boldsymbol{b}_i, \delta_i, \boldsymbol{\omega}) \, p(\boldsymbol{b}_i | \boldsymbol{\omega}) \, d\boldsymbol{b}_i, \qquad (3.1)$$

where t_{in_i} is the time corresponding to the last observation. Parameter $\delta_i = 0$ indicates that the event did not occur at the last observation, and $\delta_i = 1$ indicates that the event occurred during the study. Distribution $p(\boldsymbol{y}_i | \boldsymbol{t}_i, \boldsymbol{b}_i, \boldsymbol{\omega})$ is defined by the longitudinal model and $p(t_{in_i} | \boldsymbol{b}_i, \delta_i, \boldsymbol{\omega})$ by the survival model:

$$p(t_{in_i} | \boldsymbol{b}_i, \delta_i, \boldsymbol{\omega}) = h(t_{in_i} | \boldsymbol{b}_i, \boldsymbol{\omega})^{\delta_i} \, p(T \geq t_{in_i} | \boldsymbol{b}_i, \boldsymbol{\omega}). \qquad (3.2)$$

We define the random effect $\boldsymbol{b}_i \in \mathbb{R}^p$ by $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a $p \times p$ covariance matrix.

The corresponding marginal log-likelihood function for the shared random-effects model is:

$$\begin{aligned}
\boldsymbol{L}(\boldsymbol{y}, \boldsymbol{t}_{last} | \boldsymbol{t}, \boldsymbol{\delta}) &= \log \sum_{i=1}^{N} p(\boldsymbol{y}_i, \boldsymbol{t}_i | \boldsymbol{\omega}) \\
&= \log \sum_{i=1}^{N} \int p(\boldsymbol{y}_i | \boldsymbol{t}_i, \boldsymbol{b}_i, \boldsymbol{\omega}) \, p(t_{in_i} | \boldsymbol{b}_i, \boldsymbol{\omega}) \, p(\boldsymbol{b}_i | \boldsymbol{\omega}) \, d\boldsymbol{b}_i \qquad (3.3) \\
&= \log \sum_{i=1}^{N} \int \left[\prod_{k=1}^{n_i} p(y_{ik}^{(1)}, y_{ik}^{(2)} | t_{ik}, b_{ji}, \boldsymbol{\omega}) \right] p(t_{in_i} | \boldsymbol{b}_i, \delta_i, \boldsymbol{\omega}) \, p(\boldsymbol{b}_i | \boldsymbol{\omega}) \, d\boldsymbol{b}_i.
\end{aligned}$$

3.2.2   Likelihood function with left truncation

When we need to consider the left truncation in the shared random-effects joint model, we still use $\boldsymbol{\omega}$ to represent all the parameters in the joint model except the random effects. Let t_{i1} denote the time corresponding to the first observation of the individual, i.e. the truncation time. The likelihood contribution of individual i

conditionally on truncation time t_{i1} is:

$$
\begin{aligned}
L_i(\boldsymbol{\omega}|\boldsymbol{y}_i, \boldsymbol{t}_i, T \geq t_{i1}) &= p(\boldsymbol{y}_i, \boldsymbol{t}_i | T \geq t_{i1}, \boldsymbol{\omega}) \\
&= \frac{p(\boldsymbol{y}_i, \boldsymbol{t}_i | \boldsymbol{\omega})}{p(T \geq t_{i1} | \boldsymbol{\omega})},
\end{aligned}
\tag{3.4}
$$

where $p(T \geq t_{i1}|\boldsymbol{\omega})$ is the survival function evaluated at t_{i1}. For the shared random-effects model, the denominator in (3.4) can be written as:

$$
p(T \geq t_{i1}|\boldsymbol{\omega}) = \int p(T \geq t_{i1}|\boldsymbol{b}_i, \boldsymbol{\omega}) p(\boldsymbol{b}_i|\boldsymbol{\omega}) d\boldsymbol{b}_i.
\tag{3.5}
$$

Assuming independence between responses given the random effect, the numerator in (3.4) is the marginal likelihood function (3.1).

The shared random-effects joint model has several drawbacks. Firstly, because of the high-dimensional integration required in the computation of the joint log-likelihood, the estimation process is numerically intensive [Proust-Lima et al., 2009]. Secondly, random effects capture all the correlations between the joint models. Examples include correlations between individual observations in longitudinal data and between longitudinal models and survival models. In some commonly used multivariate joint models (such as those mentioned in Rizopoulos [2012]), random effects also capture the correlations between several responses. It is, therefore, difficult to interpret these correlations separately. In this thesis, we try to use the bivariate extension of the binomial distribution (2.4) proposed by Altham and Hankin [2012] to improve this problem via the parameter ϕ. In addition, there may be latent heterogeneity in the population, i.e. potential 'latent subpopulations' defined by some unobserved characteristics [Arbeev et al., 2014]. When these heterogeneities are discrete, the shared random-effects joint models with continuous random effects may not describe them well.

3.3 Likelihood functions for latent-class joint models

In order to avoid the above limitations in the shared random-effects joint model, we also construct the latent-class joint model in this thesis. We have introduced this

model in Section 2.3.2, which the model is also called the non-parametric frailty mode. The basic assumption of this model is that the population can be divided into a limited number of homogeneous classes [Proust-Lima et al., 2009]. In this model, the random effects are discretely distributed.

We use the same notation as in Section 2.3.2. We assume that individuals in the data can be divided into G unobserved classes. For each individual i, the individual belongs to class g is expressed in the formula as $c_i = g$, $(g = 1, \ldots, G)$.

3.3.1 Likelihood function without left truncation

Compared with the shared random-effects marginal log-likelihood function (3.1), the latent-class joint model uses summation rather than integral to calculate the marginal likelihood function. Both the longitudinal model and the survival model are class-specific in this model.

Let $\boldsymbol{\omega}$ represent all the parameters in the joint model except the random effects. Assuming independence between responses given the random effect, the marginal likelihood function is:

$$p(\boldsymbol{y}_i, \boldsymbol{t}_i | \boldsymbol{\omega}) = \sum_{g=1}^{G} p(\boldsymbol{y}_i | \boldsymbol{t}_i, c_i = g, \boldsymbol{\omega}) \, p(t_{in_i} | c_i = g, \delta_i, \boldsymbol{\omega}) \, \pi_{ig}, \qquad (3.6)$$

where t_{in_i} is also the corresponding time for the last observation. The parameter $\delta_i = 0$, as we mentioned in Section 3.2.1, represents that the event did not occur at the end of the study ($\delta_i = 0$) and the event occurred ($\delta_i = 1$), respectively. Distribution $p(\boldsymbol{y}_i | \boldsymbol{t}_i, c_i = g, \boldsymbol{\omega})$ represents the class-specific longitudinal model and $p(t_{in_i} | c_i = g, \delta_i, \boldsymbol{\omega})$ is the class-specific survival model, where

$$p(t_{in_i} | c_i = g, \delta_i, \boldsymbol{\omega}) = h_g(t_{in_i} | c_i = g, \boldsymbol{\omega})^{\delta_i} \, p(T \geq t_{in_i} | c_i = g, \boldsymbol{\omega}). \qquad (3.7)$$

Unlike the shared random-effects model, in the latent-class joint model we no longer use the individual-specific random effect $\boldsymbol{b}_i$, but use the class-specific random effect $\boldsymbol{b}_g = (b_{g0}, b_{g1})$. The linear predictor with corresponding random effects in Equation

(2.7) changes to:

$$\boldsymbol{\eta}_g^{(j)} = (\eta_0^{(j)} + b_{g0}) + (\eta_1^{(j)} + b_{g1})t. \tag{3.8}$$

The corresponding logistic regression model between time (age) and the responses is changed to:

$$p_{Y_{jg}} = \frac{\exp\left((\eta_0^{(j)} + b_{g0}) + (\eta_1^{(j)} + b_{g1})t\right)}{1 + \exp\left((\eta_0^{(j)} + b_{g0}) + (\eta_1^{(j)} + b_{g1})t\right)}. \tag{3.9}$$

The random effects b_{g0} and b_{g1} are discrete class-specific random effects. Function $p\left(t_{in_i} | c_i = g, \delta_i, \boldsymbol{\omega}\right)$ is the class-specific survival model, where

$$p\left(t_{in_i} | c_i = g, \delta_i, \boldsymbol{\omega}\right) = h_g\left(t_{in_i} | c_i = g, \boldsymbol{\omega}\right)^{\delta_i} P(T \geq t_{in_i} | c_i = g, \boldsymbol{\omega}). \tag{3.10}$$

Probability $\pi_{ig} = P(c_i = g)$ is the probability for an individual belongs to the class g, and $\sum_{g=1}^{G} \pi_{ig} = 1$. For class g,

$$\begin{aligned}
\pi_{ig} &= P(c_i = g | \boldsymbol{x}_i) \\
&= \frac{\exp(\mu_{0g} + \boldsymbol{x}_i^T \boldsymbol{\mu}_{1g})}{\sum_{l=1}^{G} \exp(\mu_{0l} + \boldsymbol{x}_i^T \boldsymbol{\mu}_{1l})},
\end{aligned} \tag{3.11}$$

where μ_{0g} is the intercept for class g and $\boldsymbol{\mu}_{1g}$ is the vector of class-specific parameters associated with the vector of time-independent covariates $\boldsymbol{x}_i$ [Proust-Lima et al., 2009]. In order to enforce $\sum_{g=1}^{G} \pi_{ig} = 1$, when $g = 1$, we set $\mu_{01} = 0$ and $\mu_{11} = 0$:

$$\begin{aligned}
\pi_{i1} &= P(c_i = 1 | \boldsymbol{x}_i) \\
&= \frac{1}{\sum_{l=1}^{G} \exp\left(\mu_{0l} + \boldsymbol{x}_i^T \boldsymbol{\mu}_{1l}\right)}.
\end{aligned} \tag{3.12}$$

The corresponding marginal log-likelihood function for the latent-class model

is:

$$L(\mathbf{y}, \mathbf{t}_{last}, \mathbf{c} | \mathbf{t}, \boldsymbol{\delta}) = \log \sum_{i=1}^{N} p(\mathbf{y}_i, \mathbf{t}_i | \boldsymbol{\omega})$$

$$= \log \sum_{i=1}^{N} \sum_{g=1}^{G} p(\mathbf{y}_i | \mathbf{t}_i, c_i = g, \boldsymbol{\omega}) \, p(t_{in_i} | c_i = g, \delta_i, \boldsymbol{\omega}) \, \pi_{ig}$$

$$= \log \sum_{i=1}^{N} \sum_{g=1}^{G} \left[\prod_{k=1}^{n_i} p(y_{ik}^{(1)}, y_{ik}^{(2)} | t_{ik}, c_i = g, \boldsymbol{\omega}) \right] p(t_{in_i} | c_i = g, \delta_i, \boldsymbol{\omega}) \, \pi_{ig}.$$

$$(3.13)$$

3.3.2 Likelihood function with left truncation

When using the latent-class joint model, the denominator of (3.4) is the marginal survival function across the classes. It can be defined as:

$$p(T \geq t_{i1} | \boldsymbol{\omega}) = \sum_{g=1}^{G} p(T \geq t_{i1} | c_i = g, \boldsymbol{\omega}) \pi_{ig}. \qquad (3.14)$$

3.4 Numerical integration and optimisation

We code the marginal log-likelihood function in the R software [R Core Team, 2013]. The corresponding parameters are estimated using the `ucminf` function in package `ucminf` [Nielsen and Mortensen, 2016].

For the shared random-effects model (3.1), where the random effect $\boldsymbol{b}$ follows a bivariate normal distribution, we use the Gaussian-Hermite method with 11 nodes to do numerical integration.

It is worth mentioning that the bivariate Gaussian quadrature can be undertaken using two univariate normal distributions (see, e.g. [Van Den Hout, 2017]). The bivariate normal distribution $f(z, x)$ can be expressed as:

$$f_{Z|X}(z | X = x) f_X(x), \qquad (3.15)$$

where $Z \sim N(\mu_Z, \sigma_Z^2)$, $X \sim N(\mu_X, \sigma_X^2)$ and $Z|X \sim N(\mu_Z - \rho(\sigma_Z/\sigma_X)x - \mu_X, \sigma_Z^2(1 - \rho^2))$. Parameter ρ represents the correlation between Z and X.

Chapter 4

Simulation study

In this chapter, we conduct three small simulation studies to investigate the parameter estimation for the bivariate shared random-effects joint models and the bivariate latent-class joint models. The joint model used in the first simulation study is built based on the bivariate extension of the binomial distribution (2.4) and the Gompertz hazard model (2.24). In the second simulation study, we use the bivariate extended binomial longitudinal combined with the Gompertz hazard model to construct the latent-class joint model. In the third simulation study, we use the categorical model (2.12) and the Gompertz model to construct the shared random effect joint model.

The simulation study design in this chapter refers to the articles by Van Den Hout and Muniz-Terrera [2016] and Morris et al. [2019]. We used the ADEMP structure to plan the simulation study [Morris et al., 2019]. The ADEMP structure includes *Aims*, *Data-generating mechanisms*, *Methods*, *Estimands*, and *Performance measures*.

4.1 Bivariate binomial shared random-effects joint model

The main *Aims* of our simulation study are:

1. To check the accuracy of the algebra and code to ensure there are no significant errors.

2. To check the estimation method and investigate the joint model's performance

given various sample sizes.

3. To confirm that the correlation parameter ϕ of the bivariate joint model can be estimated well.

4. To ensure that the relevant arguement settings are reasonable. The arguments refer to the sample size settings for the simulations, the number of nodes for the Gaussian integration settings, and the number of simulations.

5. To confirm that we can obtain accurate estimates when using the results obtained in the Section 6 as the true values in the simulation study. In other words, we can use simulation to generate a dataset similar to the data used in the 6 Section.

Date-generating mechanism: In this section, we will perform a simulation study based on the bivariate extension of the binomial joint model. We use the joint model with minimum AIC in Table 6.2 as a reference to build the simulation study. The data-generating mechanism is to create a bivariate shared random-effects joint model based on the extension of the bivariate binomial distribution (2.4) and the Gompertz hazard model (2.27). When investigating data generation mechanisms, it is common to vary the sample size of a simulated dataset, as performance tends to change with the sample size [Morris et al., 2019]. Therefore, we also try to use various sample sizes when designing our simulation study, and for the sample size setting, we refer to the settings in Van Den Hout and Muniz-Terrera [2016]. For the purpose of investigating small sample bias, we set the sample sizes $N = 100, 400, 1000$ for the joint model, respectively.

In the bivariate extension of the binomial joint model, the logistic regression model is defined by (2.7) with $b_{11} = b_{21} = 0$, i. e. the model contains two random effects, and both random effects are random intercepts. The random effects follow a bivariate normal distribution with expectations of 0. The distribution is specified by parameters σ_{b1}, σ_{b2} and ρ. The corresponding link function is equal to (2.29) for $j = 1$. This joint model is the best performing model in the Binomial joint models Section; see Section 6.1.

The true values of the parameters, i.e. the values used when generating the data, are listed as "Value" in Table 4.1 - 4.3. It is worth noting that we set α to -0.4 in the model. The parameter α is the relevant parameter of the link function $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t)$, if α is significantly different from 0, it means that there is an association between the longitudinal model and the survival model, and it is necessary for us to build the joint model. We select three sets of true values for our simulation experiments. In Table 4.1, we set the correlation parameter ϕ for the two responses to 1. In this situation, the correlation between the two responses is captured by random effects, the same as the multivariate joint model mentioned in Rizopoulos [2012]. In Table 4.2, we set the value of ϕ to 0.8, in which case the correlation between the two responses is measured using the parameter ϕ. When the value of ϕ is between 0 and 1, the two responses are negatively correlated. In addition, a set of simulations is conducted with the estimated values in Section 6.1. The corresponding simulation results are listed in Table 4.3

Estimands: In this simulation study, the estimates of interest are the model parameters in the N_{sim} iteration.

Methods: In generating the data, we set the observation time for all individuals to start at time 0 in order to eliminate the impact of the left truncation. We set the time range from 0 to 24, i.e. if an individual has not dropped out at the end of the study, the corresponding time point of his/her last observation is 24. We set the follow-up interval to three years to ensure that we have adequate information about changes in response variables over time. When the follow-up interval is fixed at every three years, this means that there are 9 observations for individuals who do not drop out during the follow-up.

Based on the specified parameters and follow-up intervals, we first simulate the corresponding random effects for an individual by generating data from the bivariate normal distribution. Given these random effects, bivariate longitudinal trajectories are simulated using the Metropolis-Hastings algorithm. We refer to the `rMM` function in the `MM` package, which is a function used to generate univariate random samples from the multiplicative multinomial [Altham and Hankin, 2012]. We

write the function to generate data from the extension of the binomial distribution based on the `rMM` function. Afterwards, we use the generated individual random effects again to define the relevant parameter λ of the Gompertz hazard model, i.e. $\lambda^* = \exp\left(\beta + \alpha(\eta_0^{(1)} + b_{10})\right)$, and to derive the survival function $S(t)$ for each follow-up time point. We use the cumulative distribution function $F(t) = 1 - S(t)$ to calculate the point in time when the event of interest occurs by the inversion method. After obtaining the generated data, we use the likelihood function of the joint model as well as function `ucminf` to estimate the parameters [Nielsen and Mortensen, 2016]. In the parameter estimation process, the bivariate integral is calculated by the Gauss–Hermite quadrature method, and we set the number of nodes to 11. The number of iterations of the simulation N_{sim} is set to 100.

Performance measures: The results of the simulation are analysed by assessing the means, bias, relative bias (%), empirical standard error (SE), Monte Carlo standard error of bias (4.1) (denoted by MC.Bias) and Monte Carlo standard error of empirical standard error (4.2) (denoted by MC.SE). The Monte Carlo standard errors are statistics that quantify the uncertainty of a finite N_{sim} simulation study [Morris et al., 2019]. Bias and relative bias are the key performance indicators we are interested in.

$$
\text{MC.Bias} = \sqrt{\frac{1}{N_{sim}\left(N_{sim}-1\right)} \sum_{i=1}^{N_{sim}} \left(\hat{\omega}_i - \bar{\omega}\right)^2}, \tag{4.1}
$$

$$
\text{MC.SE} = \frac{\hat{SE}}{\sqrt{2\left(N_{sim}-1\right)}}. \tag{4.2}
$$

Table 4.1, Table 4.2 and Table 4.3 show the simulation results for different starting values when the sample size is equal to $N = 100, 400, 1000$, with follow-up intervals equal to 3. From Table 4.1 we can observe that in 100 iterations, the mean of the estimated values does not deviate much from the true values we set. For most parameters, the relative biases between the true values we set and the mean of the

Table 4.1: Simulation study for the bivariate binomial shared random-effects joint model, where $\phi = 1$. The follow-up interval is 3.

Parameter	Value	N = 100					N = 400					N = 1000				
		Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias
$\eta_0^{(1)}$	1.0	0.995	-0.005	0.5	0.331	0.033	0.973	-0.027	2.7	0.171	0.017	1.010	0.010	1.0	0.123	0.012
$\eta_1^{(1)}$	-0.2	-0.205	-0.005	2.3	0.038	0.004	-0.200	0.000	0.2	0.020	0.002	-0.201	-0.001	0.4	0.014	0.001
θ_{γ_1}	1.0	1.004	0.004	0.4	0.031	0.003	1.001	0.001	0.1	0.016	0.002	1.001	0.001	0.1	0.010	0.001
$\eta_0^{(2)}$	0.8	0.817	0.017	2.1	0.346	0.035	0.780	-0.020	2.5	0.182	0.018	0.803	0.003	0.4	0.128	0.013
$\eta_1^{(2)}$	-0.2	-0.204	-0.004	2.0	0.042	0.004	-0.201	-0.001	0.3	0.019	0.002	-0.199	0.001	0.5	0.014	0.001
θ_{γ_2}	1.0	1.008	0.008	0.8	0.044	0.004	1.002	0.002	0.2	0.016	0.002	1.000	0.000	0.0	0.012	0.001
ϕ	1.0	1.004	0.004	0.4	0.046	0.005	1.004	0.004	0.4	0.023	0.002	0.999	-0.001	0.1	0.016	0.002
ξ	0.1	0.120	0.020	19.8	0.040	0.004	0.103	0.003	3.0	0.016	0.002	0.100	0.000	0.2	0.010	0.001
σ_{b1}	0.4	0.397	-0.003	0.7	0.129	0.013	0.391	-0.009	2.3	0.057	0.006	0.404	0.004	1.0	0.040	0.004
σ_{b2}	0.6	0.613	0.013	2.1	0.172	0.017	0.602	0.002	0.3	0.071	0.007	0.601	0.001	0.1	0.046	0.005
ρ	0.3	0.333	0.033	11.2	0.360	0.036	0.291	-0.009	3.1	0.164	0.016	0.309	0.009	3.1	0.107	0.011
β	-1.5	-1.588	-0.088	5.9	0.799	0.081	-1.488	0.012	0.8	0.266	0.027	-1.507	-0.007	0.4	0.141	0.014
α	-0.4	-0.438	-0.038	9.6	0.969	0.098	-0.442	-0.042	10.5	0.279	0.028	-0.399	0.001	0.1	0.139	0.014

Table 4.2: Simulation study for the bivariate binomial shared random-effects joint model, where $\phi = 0.8$. The follow-up interval is 3.

Parameter	Value	N = 100					N = 400					N = 1000				
		Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias
$\eta_0^{(1)}$	1.0	1.020	0.020	2.0	0.282	0.029	1.011	0.011	1.1	0.128	0.013	1.008	0.008	0.8	0.084	0.008
$\eta_1^{(1)}$	-0.2	-0.207	-0.007	3.5	0.046	0.005	-0.201	-0.001	0.7	0.020	0.002	-0.200	0.000	0.2	0.013	0.001
θ_{Y_1}	1.0	1.009	0.009	0.9	0.041	0.004	1.002	0.002	0.2	0.019	0.002	1.001	0.001	0.1	0.011	0.001
$\eta_0^{(2)}$	0.8	0.821	0.021	2.6	0.295	0.030	0.817	0.017	2.1	0.142	0.014	0.809	0.009	1.2	0.090	0.009
$\eta_1^{(2)}$	-0.2	-0.202	-0.002	1.2	0.044	0.004	-0.201	-0.001	0.4	0.020	0.002	-0.200	0.000	0.2	0.013	0.001
θ_{Y_2}	1.0	1.003	0.003	0.3	0.038	0.004	1.000	0.000	0.0	0.016	0.002	1.000	0.000	0.0	0.011	0.001
ϕ	0.8	0.798	-0.002	0.2	0.048	0.005	0.797	-0.003	0.3	0.022	0.002	0.798	-0.002	0.2	0.015	0.001
ξ	0.1	0.118	0.018	17.6	0.042	0.004	0.102	0.002	2.2	0.018	0.002	0.100	0.000	0.3	0.010	0.001
σ_{b1}	0.4	0.424	0.024	6.0	0.148	0.015	0.404	0.004	0.9	0.063	0.006	0.405	0.005	1.4	0.040	0.004
σ_{b2}	0.6	0.610	0.010	1.7	0.149	0.015	0.604	0.004	0.6	0.074	0.007	0.601	0.001	0.2	0.044	0.004
ρ	0.3	0.248	-0.052	17.5	0.393	0.040	0.308	0.008	2.6	0.173	0.017	0.315	0.015	5.0	0.113	0.011
β	-1.5	-1.500	0.000	0.0	0.603	0.062	-1.504	-0.004	0.3	0.254	0.025	-1.505	-0.005	0.4	0.145	0.015
α	-0.4	-0.453	-0.053	13.2	0.653	0.067	-0.410	-0.010	2.6	0.263	0.026	-0.399	0.001	0.1	0.146	0.015

Table 4.3: Simulation study for the bivariate binomial shared random-effects joint model, where true values equal to results in Section 6.1. The follow-up interval is 3.

Parameter	Value	N = 100					N = 400					N = 1000				
		Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias
$\eta_0^{(1)}$	-0.591	-0.620	-0.029	5.0	0.163	0.016	-0.596	-0.005	0.9	0.071	0.007	-0.601	-0.010	1.7	0.050	0.005
$\eta_1^{(1)}$	-0.023	-0.023	0.000	0.5	0.006	0.001	-0.023	0.000	1.2	0.003	0.000	-0.023	0.000	0.2	0.002	0.000
θ_{Y_1}	1.205	1.211	0.006	0.5	0.026	0.003	1.209	0.004	0.4	0.013	0.001	1.207	0.002	0.2	0.009	0.001
$\eta_0^{(2)}$	-2.026	-2.065	-0.039	1.9	0.208	0.021	-2.046	-0.020	1.0	0.109	0.011	-2.048	-0.022	1.1	0.064	0.006
$\eta_1^{(2)}$	-0.028	-0.028	0.000	0.7	0.005	0.000	-0.027	0.001	2.3	0.002	0.000	-0.027	0.001	2.6	0.001	0.000
θ_{Y_2}	1.121	1.120	-0.001	0.1	0.019	0.002	1.117	-0.004	0.3	0.010	0.001	1.117	-0.004	0.3	0.006	0.001
ϕ	1.416	1.427	0.011	0.8	0.045	0.005	1.419	0.003	0.2	0.021	0.002	1.420	0.004	0.3	0.014	0.001
ξ	0.117	0.121	0.004	3.5	0.035	0.004	0.116	-0.001	0.6	0.017	0.002	0.118	0.001	0.6	0.011	0.001
σ_{b1}	0.450	0.452	0.002	0.5	0.071	0.007	0.453	0.003	0.7	0.032	0.003	0.453	0.003	0.7	0.019	0.002
σ_{b2}	0.646	0.617	-0.029	4.5	0.063	0.006	0.608	-0.038	5.9	0.028	0.003	0.605	-0.041	6.3	0.017	0.002
ρ	0.364	0.344	-0.020	5.6	0.158	0.016	0.349	-0.015	4.2	0.068	0.007	0.344	-0.020	5.5	0.044	0.004
β	-7.061	-7.214	-0.153	2.2	0.871	0.088	-7.066	-0.005	0.1	0.398	0.040	-7.068	-0.007	0.1	0.258	0.026
α	-1.078	-1.056	0.022	2.1	0.701	0.070	-1.074	0.004	0.4	0.346	0.035	-1.055	0.023	2.1	0.210	0.021

final estimates are less than 5%. It shows that our modelling estimates are accurate for most of the parameters. It is worth noting that when the sample size is 100, the relative biases of parameters ξ of the Gompertz hazard model, the correlation coefficient ρ of the random effects, the parameter for the baseline hazard β and the corresponding parameter α of the link function $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t)$ are relatively large. The values of their relative biases are 19.8%, 11.2%, 5.9% and 9.6%, respectively. Although the estimates of these parameters have non-negligible biases when the sample size is small ($N = 100$, these relative biases decrease rapidly as the sample size increases. When the sample size is expanded to 1000, the relative biases of all parameters are fairly close to 0, which is the same as the sample size of the ELSA data we used in Section 6.1. In general, as the sample size N increases, the mean of the estimates will be closer to the set value. As the sample size increases, other relevant estimates such as bias, relative bias and standard error will also improve significantly.

The difference between the true value used in Table 4.2 and the true value in Table 4.1 is that we set ϕ to 0.8 instead of 1. It means that for the joint model using this set of true values, the correlation between the two responses is measured by the parameter ϕ rather than by random effects. From Table 4.2 we can observe that, even with the small sample size, the mean of the ϕ estimates is quite close to our set value. It shows that our joint model can estimate well the correlation coefficients ϕ of the two responses. When the sample size is small ($N = 100$), we encounter the same problem as when analysing the results in Table 4.1, i.e. large relative biases for the parameters ξ, ρ and α. This problem can be well resolved by increasing the sample size.

The true values we use in Table 4.3 are obtained by fitting the ELSA data in Section 6.1. The results from this set of simulations are close to the true values and, to some extent, indicate the parameter estimates we obtained in Section 6.1 are accurate. It is worth noting that the parameter estimates in Section 6.1 are based on the bivariate extended binomial joint model with left truncation, whereas in this section, the left truncation is not incorporated.
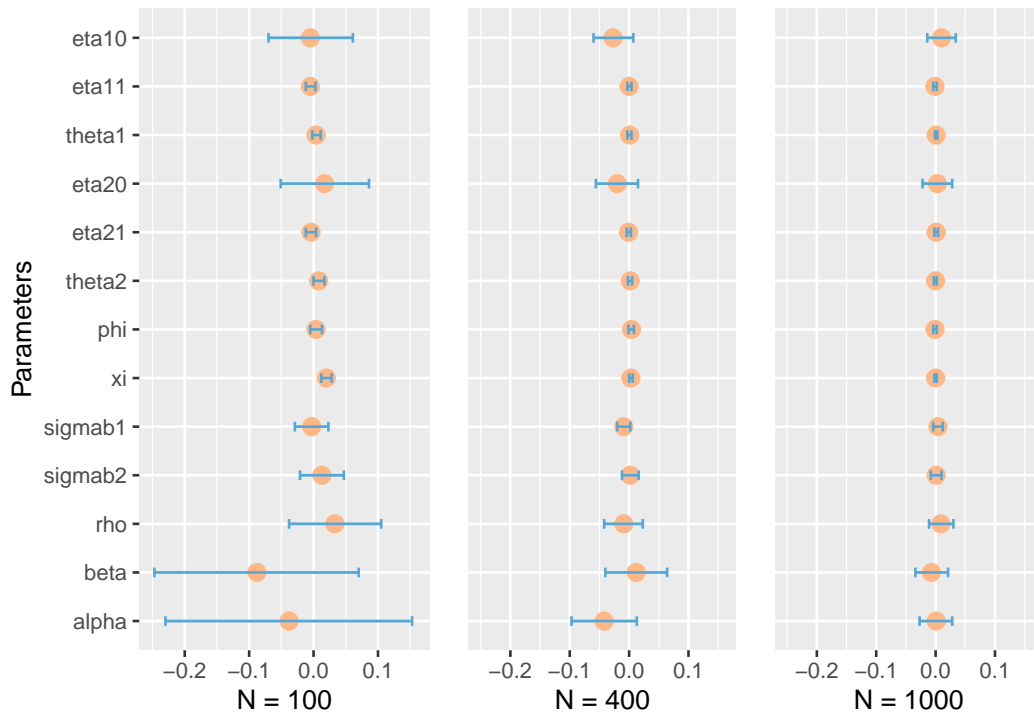
Figure 4.1: Bias and corresponding Monte Carlo 95% confidence interval, based on the results in Table 4.1. Large dots represent biases, and geometry bars represent Monte Carlo 95% confidence interval.

We plot the values of the biases in Table 4.1 and the corresponding Monte Carlo standard error of deviation (MC.Bias) values in Figure 4.1. From this figure, we can see that 0 is included in the 95% Monte Carlo confidence interval for most parameters. As the sample size increases, i.e. from $N = 100$ to $N = 1000$, the corresponding confidence intervals for the biases of the parameters show a significant decrease. We can deduce that the larger the sample size, the closer the average value of the estimated parameters obtained from the simulation based on these 100 iterations is to the true value, which is consistent with our previous discussion related to Table 4.1. We have drawn similar plots A.1 and A.2 in the Appendix A based on the results in Tables 4.2 and Table 4.3, respectively. Both plots demonstrate that the larger the sample size, the closer our simulation results are to the set true value. We also present Figure A.3 - Figure A.5 in the Appendix. The figures show box plots of the corresponding estimates for the simulations in Table 4.1 - Table 4.3.

This simulation shows that

1. The parameters used to generate the data can be reproduced by marginal like-

lihood estimation.

2. The parameter estimation method we use gives accurate estimates, i.e. the function `ucminf` combined with Gauss-Hermite. Moreover, the scenarios in the simulation study are reasonable, e.g. in the Gauss-Hermite method, the nodes are set to 11, and in the simulations, the number of iterations is set to 100. It is worth noting that although in this section it is reasonable for us to set the number of nodes to 11 by simulation, in other cases, e.g. when there is considerable heterogeneity between subjects, the Gauss-Hermite method may require more nodes.

3. As the sample size increases, the bias in the parameter estimates is significantly reduced. In particular, when the sample size is $N = 1000$, the 95% confidence interval for the bias of all parameter estimates includes 0. This result also justifies our choice of a sample size of 1000 in Section 6.1.

4.2 Misspecification

In Section 4.1, we constructed two sets of simulations in Tables 4.1 and Table 4.2. These two sets of simulation results demonstrate that our joint model can accurately estimate the correlation between the two longitudinal response variables, i.e. the parameter ϕ.

Our *Aims* in this section is to demonstrate that it is worthwhile to use the bivariate extension of the binomial distribution to construct joint models, i.e. that the correlation coefficient ϕ is a necessary parameter when constructing the bivariate joint model.

Data-generating mechanism: We will construct two sets of simulations, $\mathscr{A}$ and $\mathscr{B}$, to prove the need for the existence of the parameter ϕ. We can observe from the simulations in Section 4.1 that we can obtain more accurate parameter estimates when the sample size reaches $N = 400$. Therefore, simulations in this section set the sample size equal to 400. In both sets of simulations, we use Gompertz hazard as the survival model, and the corresponding link function is Equation (2.31). The logistic regression model is defined by (2.7) with $b_{11} = b_{21} = 0$, i. e. the model

contains two random effects, and both are random intercepts. The random effects follow a bivariate normal distribution with expectations of 0. The distribution is specified by parameters σ_{b1}, σ_{b2} and ρ.

In simulation $\mathscr{A}$, we generate data from a shared random-effects joint model containing two univariate extended binomial distributions. In this data, the correlation between the two longitudinal responses is captured by random effects. The true value of the parameter ρ is set to 0.3. After obtaining the data, we fit the data with a joint model containing the extension of the bivariate binomial distribution. In this simulation, we expect the estimated ϕ to be equal to 1, i.e. the two longitudinal responses are independent of each other given random effects. Moreover, ϕ equals 1 means that the fact that random effects capture the correlation between the two longitudinal responses can be reflected by the parameter estimation results.

In simulation $\mathscr{B}$, we use the joint model containing the extension of the bivariate binomial distribution to generate data. We set the true value of the parameter ϕ to 0.8. Afterwards, we fit the data using the bivariate joint model containing two univariate extended binomial distributions. In theory, the expectation is that all parameter estimates, except for parameter ρ will be accurate. The parameter ρ will not be consistent with the true value because it captures the correlation between the two longitudinal responses, whereas, in the data part of this correlation is captured by the parameter ϕ.

Estimands: The estimates of interest are the model parameters in the N_{sim} iteration.

Methods: The process of generating data is similar to that in Section 4.1. We first generate individual-specific random effects based on parameters σ_{b1}, σ_{b2} and ρ. Given generated random effects, we use the Metropolis-Hasting method to simulate bivariate longitudinal trajectories. It is worth noting that in simulation $\mathscr{A}$, we generate trajectories from two univariate extended binomial distributions; in simulation $\mathscr{B}$, we generate trajectories from the bivariate extended binomial distribution. For the survival part, we use the cumulative distribution function to generate the event occurrence time. The parameter estimation is undertaken by the function `ucminf`

in R [Nielsen and Mortensen, 2016]. We use the Gauss-Hermite method with 11 nodes to calculate the bivariate integral. The number of iterations of the simulation N_{sim} is set to 100.

Performance measures: We still use the means, bias, relative bias (%), empirical standard error (SE) and Monte Carlo standard error of bias (4.1) (denoted by MC.Bias) to evaluate the parameter estimates.

We present the results of simulation $\mathscr{A}$ and $\mathscr{B}$ in Table 4.4. The corresponding results for the parameter ϕ in simulation $\mathscr{A}$ (bias, SE, etc.) are obtained by assuming the true $\phi = 1$. The estimated values of simulation $\mathscr{A}$ are similar to the true values. The relative biases for all parameters except for α_2 are below 5%. It shows that the estimates can well reflect the distribution of the data. Moreover, there is no parameter ϕ in the true value, and the true value of parameter ρ is equal to 0.3. Our bivariate joint model successfully estimates these two parameters, with ϕ close to 1 and parameter ρ close to 0.3. The estimation results of these two parameters indicate that even if longitudinal responses are independent of each other given random effects, the joint model we propose in this thesis can still fit the parameters accurately.

In simulation $\mathscr{B}$, we can observe that when the parameter ϕ exists in the data-generating model but not in the fitted model, the estimated values of all parameters differ significantly from the true values.

Combining simulation $\mathscr{A}$ as well as simulation $\mathscr{B}$, we can conclude that it is necessary to use the bivariate extension of the binomial distribution to construct the joint model. Joint models that only use random effects to capture the correlation between responses do not always reflect the distribution of the data well.

4.3 Bivariate binomial latent-class joint model

The *Aims* of this simulation are similar to those in Section 4.1. The first aim is to check whether we can get accurate parameter estimates for the binomial latent-class joint model across various sample sizes. The second aim is to verify that the relevant arguments, such as the sample sizes and the number of iterations we use,

Table 4.4: Simulation study for the shared random-effects joint model. The follow-up interval is 3.

Parameter	𝒜						𝓑					
	Value	Mean	Bias	% bias	SE	MC.Bias	Value	Mean	Bias	% bias	SE	MC.Bias
$\eta_0^{(1)}$	1.0	1.020	0.020	2.0	0.163	0.043	1.0	0.045	-0.955	95.5	0.037	0.004
$\eta_1^{(1)}$	-0.2	-0.203	-0.003	1.4	0.019	0.005	-0.2	-0.116	0.084	41.9	0.010	0.001
θ_{γ_1}	1.0	1.003	0.003	0.3	0.016	0.004	1.0	0.959	-0.041	4.1	0.010	0.001
$\eta_0^{(2)}$	0.8	0.809	0.009	1.1	0.180	0.048	0.8	-0.269	-1.069	133.7	0.042	0.004
$\eta_1^{(2)}$	-0.2	-0.201	-0.001	0.4	0.021	0.006	-0.2	-0.112	0.088	43.8	0.011	0.001
θ_{γ_2}	1.0	1.001	0.001	0.1	0.019	0.005	1.0	0.956	-0.044	4.4	0.012	0.001
ϕ	/	1.000	0.000	0.0	0.022	0.006	0.8	/	/	/	/	/
ξ	0.1	0.105	0.005	4.5	0.016	0.004	0.1	0.213	0.113	113.0	0.071	0.007
σ_{b1}	0.4	0.412	0.012	3.1	0.067	0.018	0.4	0.462	0.062	15.4	0.046	0.005
σ_{b2}	0.6	0.596	-0.004	0.7	0.076	0.020	0.6	0.624	0.024	4.0	0.057	0.006
ρ	0.3	0.304	0.004	1.4	0.178	0.048	0.3	-0.932	-1.232	410.8	0.037	0.004
β	-1.5	-1.513	-0.013	0.9	0.237	0.063	-1.5	-3.663	-2.163	144.2	0.831	0.084
α_1	-0.4	-0.385	0.015	3.7	0.264	0.071	-0.4	-6.506	-6.106	1526.4	4.145	0.417
α_2	-0.4	-0.425	-0.025	6.2	0.157	0.042	-0.4	-5.066	-4.666	1166.5	3.140	0.316

are feasible.

Date-generating mechanism: We use the joint model with link function (2.53) to build the simulation study. The data-generating mechanism is to create a bivariate latent-class joint model based on the extension of the bivariate binomial distribution (2.4) and the Gompertz hazard model (2.27). The sample sizes are also set to $N = 100, 400, 1000$, respectively in order to investigate the small sample bias.

For the marginal likelihood function (3.13), we divide individuals into two groups ($G = 2, g \in \{1, 2\}$). We set the class-specific random effect $b_1 = -b_2$ and define the probability for an individual belonging to the class g (represented by π_{ig}) as

$$\pi_{i1} = \frac{\exp(\mu)}{1 + \exp(\mu)}. \tag{4.3}$$

For the logistic regression model in the longitudinal model, we have:

$$p_{Y_{jg}} = \frac{\exp\left((\eta_0^{(j)} + b_{g0}) + \eta_1^{(j)}t\right)}{1 + \exp\left((\eta_0^{(j)} + b_{g0}) + \eta_1^{(j)}t\right)}. \tag{4.4}$$

For the Gompertz hazard model with the link function Δ equal to (2.53), and the corresponding hazard model is written as:

$$h_g(t) = \exp(\beta + \alpha(\eta_0^{(j)} + b_{g0}) + \xi t). \tag{4.5}$$

The corresponding simulation results are listed in Table 4.5. The true values of the parameters, i.e. the values used when generating the data, are listed as "Value" in the table.

Estimands: In this simulation study, the estimates of interest are the model parameters in the N_{sim} iteration.

Methods: In generating the data, we set the observation time for all individuals to start at time 0 to eliminate the impact of the left truncation. We still set the time range from 0 to 24, i.e. if an individual has not dropped out at the end of the study, the corresponding time point of his/her last observation is 24. We set the follow-up

interval to three years to ensure that we have adequate information about changes in response variables over time. When the follow-up interval is fixed every three years, this means that there are 9 observations for individuals who do not drop out during the follow-up.

Similar to Section 4.1, we first simulate the bivariate longitudinal trajectories based on the corresponding class-specific random effect b_g using the Metropolis-Hastings algorithm. Afterwards, we use the class-specific random effects again to define the relevant parameter λ of the Gompertz hazard model, i.e. $\lambda^* = \exp\left(\beta + \alpha(\eta_0^{(j)} + b_{g0})\right)$ for each follow-up time point. We use the formula mentioned in Bender et al. [2005] to generate the event times rather than using the cumulative distribution function and the inversion method:

$$T = \frac{1}{\xi} \log\left(1 - \frac{\xi \log(u)}{\lambda^*}\right), \tag{4.6}$$

where $u \sim U(0,1)$ and β is the vector of regression coefficients.

After obtaining the generated data, we use the joint model's likelihood function as well as function `ucminf` to estimate the parameters [Nielsen and Mortensen, 2016]. The number of iterations of the simulation N_{sim} is set to 100.

Performance measures: We compute the mean of estimated parameters over N_{sim} iterations. We still assess means, bias, relative bias (%), empirical standard error (SE) and Monte Carlo standard error of bias (MC.Bias) [Morris et al., 2019].

Table 4.5 shows the simulation results when the sample size equals to $N = 100, 400, 1000$, with follow-up intervals equal to 3. From the table, we can observe some deviation between the mean of the estimated values and the true values we set for the 100 iterations. Especially the relative biases of the mean value of parameters ξ of the Gompertz hazard model and the corresponding parameter α of the link function $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t)$ are relatively large. For most of the remaining parameters, the relative biases between the true values we set and the mean of the final estimates are less than 5%. It shows that our modelling estimates are relatively accurate for most parameters. Although the estimates of parameters have non-negligible biases when the sample size is small ($N = 100$), these relative biases decrease rapidly as

Table 4.5: Simulation study for the bivariate binomial latent-class joint model. The follow-up interval is 3.

Parameter	Value	N = 100					N = 400					N = 1000				
		Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias
$\eta_0^{(1)}$	-0.59	-0.595	-0.005	0.8	0.145	0.015	-0.593	-0.003	0.5	0.074	0.007	-0.590	0.000	0.0	0.044	0.004
$\eta_1^{(1)}$	-0.02	-0.020	0.000	1.4	0.004	0.000	-0.020	0.000	1.2	0.003	0.000	-0.020	0.000	0.7	0.002	0.000
θ_{Y_1}	1.21	1.212	0.002	0.2	0.022	0.002	1.209	-0.001	0.1	0.013	0.001	1.210	0.000	0.0	0.008	0.001
$\eta_0^{(2)}$	-2.03	-2.030	0.000	0.0	0.198	0.020	-2.015	0.015	0.7	0.091	0.009	-2.023	0.007	0.3	0.061	0.006
$\eta_1^{(2)}$	-0.03	-0.030	0.000	1.5	0.004	0.000	-0.030	0.000	1.1	0.002	0.000	-0.030	0.000	0.3	0.001	0.000
θ_{Y_2}	1.12	1.125	0.005	0.5	0.017	0.002	1.122	0.002	0.2	0.009	0.001	1.120	0.000	0.0	0.006	0.001
ϕ	1.42	1.420	0.000	0.0	0.041	0.004	1.418	-0.002	0.1	0.021	0.002	1.419	-0.001	0.1	0.013	0.001
b_1	0.30	0.310	0.010	3.3	0.033	0.003	0.303	0.003	0.9	0.016	0.002	0.300	0.000	0.0	0.010	0.001
ξ	0.12	0.163	0.043	35.8	0.070	0.007	0.155	0.035	29.2	0.032	0.003	0.154	0.034	28.6	0.018	0.002
μ	2.00	2.032	0.032	1.6	0.331	0.034	2.014	0.014	0.7	0.156	0.016	2.010	0.010	0.5	0.104	0.010
β	-7.06	-7.107	-0.047	0.7	1.730	0.176	-7.194	-0.134	1.9	0.582	0.058	-7.165	-0.105	1.5	0.311	0.031
α	-1.08	-0.083	0.997	92.3	4.406	0.447	-0.950	0.130	12.1	0.685	0.068	-0.958	0.122	11.3	0.344	0.034

the sample size increases. In general, as the sample size *N* increases, the mean of the estimates will be closer to the set value. As the sample size increases, other relevant estimates, such as bias, relative bias and standard error will also improve significantly.
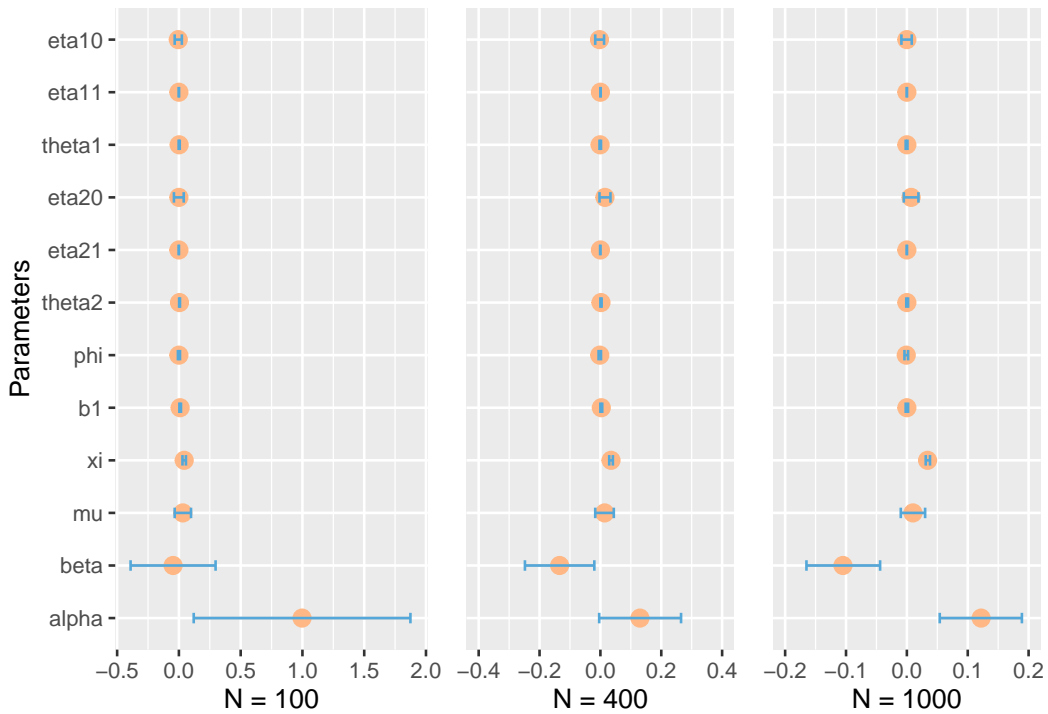
Figure 4.2: Bias and corresponding Monte Carlo 95% confidence interval, based on the results in Table 4.5. Large dots represent biases, and geometry bars represent Monte Carlo 95% confidence interval.

We plot the values of the biases in Table 4.5 and the corresponding Monte Carlo standard error of deviation (MC.Bias) values in Figure 4.2. This figure shows that 0 is included in the 95% Monte Carlo confidence interval for most parameters. As the sample size increases, i.e. from $N = 100$ to $N = 1000$, the corresponding confidence intervals for the biases of the parameters show a significant decrease. We can deduce that the larger the sample size, the closer the average value of the estimated parameters obtained from the simulation based on these 100 iterations is to the true value, which is consistent with our previous discussion related to Table 4.5. We also present Figure A.6 in the Appendix. The figures show box plots of the corresponding estimates for the simulations in Table 4.5.

This simulation shows that: with marginal likelihood estimation, it is possible

to reproduce the parameter values used to generate the data; the parameter estimation method we use gives accurate estimates, i.e. the function `ucminf`. In addition, the arguments set in the simulation study are reasonable - for example, the number of iterations, the time interval and the sample size. As the sample size increases, the bias in the parameter estimates is significantly reduced.

4.4 Bivariate categorical joint model

The *Aims* of this simulation are similar to those in Section 4.1. The first aim is to check whether we can get accurate parameter estimates for the categorical joint model across various sample sizes. The second aim is to verify that the relevant arguments, such as the number of nodes and the number of iterations we use, are feasible.

Date-generating mechanism: The mechanism is based on the categorical distribution (2.12) and the Gompertz hazard model (2.27). We also set the sample size $N = 100, 400, 1000$ respectively in order to investigate the small sample bias. For the longitudinal model, we use the categorical distribution (2.12), in which the logistic regression model is defined by (2.14) and (2.15) with $b_{11} = b_{21} = 0$. We use two univariate categorical distributions, with the first distribution having four categories and the second having three. We set the baseline category to 1, i.e. for both categorical distributions, we set $l = 1$ in Equation (2.15) when constructing the corresponding logistic regression. The values of i^* in Equation (2.14) are set to 4 and 2, respectively. For the survival model, we use the Gompertz hazard model (2.27). The corresponding link function is given by (2.31). The random-effects distribution for (b_{10}, b_{20}) is specified by σ_{b1}, σ_{b2} and ρ. The joint model we use in this section is based on the model that obtained the best results in Section 6.2.2.

The true values of the parameters are listed as "Value" in Table 4.6. We set αs to -0.4 in the model. The parameter α is the relevant parameter of the link function $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t)$, if αs are significantly different from 0, it means that there is an association between the longitudinal model and the survival model, and it is necessary for us to build the joint model.

Estimands: In this simulation study, the estimands of interest are also the model parameters over the N_{sim} iterations.

Methods: We set the observation time for all individuals to start from zero so that we do not need to consider left truncation in this joint model. If an individual does not drop out, the follow-up ends at 20 years. The simulation study in this section is less computationally expensive than that we constructed in Section 4.1, so we can add more observations to improve the accuracy of the simulation study. Follow-up intervals are set to one year to ensure we have enough information about changes in responses over time.

Given the specified parameters and follow-up intervals, data for an individual are firstly simulated by drawing random effects. Based on these effects, we simulate two longitudinal trajectories using `sample` function separately R Core Team [2013]. Afterwards, random effects are used again to define the Gompertz parameter $\lambda^* = \exp\left(\beta + \alpha_1(\eta_0^{(1)} + b_{10}) + \alpha_2(\eta_0^{(2)} + b_{20})\right)$ for the joint model, and to calculate the survival function $S(t)$ for each follow-up time point. In this simulation study, we use the formula mentioned in Bender et al. [2005] to generate the event times, rather than using the cumulative distribution function and the inversion method:

$$T = \frac{1}{\xi} \log\left(1 - \frac{\xi \log(u)}{\lambda^*}\right), \tag{4.7}$$

where $u \sim U(0,1)$ and β is the vector of regression coefficients.

It is worth noting that we could not generate data that follows the categorical distribution by using the function `rcat` in `LaplacesDemon` package [Statisticat and LLC., 2021]. The distribution of the data generated using this `rcat` function does not match the categorical distribution we expect. We have verified this by using `rcat` to generate the data and then fitting the data with the corresponding categorical distribution.

When estimating the parameters, we still use the Gauss-Hermite method to calculate the double integral. We choose 11 nodes for the Gauss–Hermite quadrature.

The corresponding parameters are estimated using the `ucminf` function in package `ucminf` [Nielsen and Mortensen, 2016]. The number of iterations is 100.

Performance measures: As with the previous simulation, we compute the mean of estimated parameters over N_{sim} iterations. We still assess means, bias, relative bias (%), empirical standard error (SE), Monte Carlo standard error of bias (MC.Bias) and Monte Carlo standard error of empirical standard error (MC.SE) Morris et al. [2019].

As expected, the mean of the estimated value will be closer to the set value when N is increased. The relative bias shows a large difference (11.1%) between the mean of the estimated values and the true value of α_1 when the sample size is 100. We can reduce this bias by increasing the sample size. We could already achieve a relatively accurate estimate when the sample size is expanded to 400. This sample size is close to the PBC2 data we used in Section 6.2.2. Therefore, we demonstrate to some extent that the parameter estimates obtained in Chapter 6 are reliable. As in Section 4.1, we also produced plots of the corresponding bias and Monte Carlo 95% confidence intervals based on the results in Table 4.6. From Figure 4.3 we could also conclude that the estimated parameters of the joint model we constructed fit the data well.

Table 4.6: Simulation study for the bivariate categorical shared random-effects joint model. The follow-up interval is 1.

Parameter	Value	N = 100					N = 400					N = 1000				
		Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias	Mean	Bias	% bias	SE	MC.Bias
$\eta_{02}^{(1)}$	0.3	0.292	-0.008	2.7	0.131	0.013	0.290	-0.010	3.3	0.057	0.006	0.300	0.000	0.0	0.042	0.004
$\eta_{12}^{(1)}$	-0.1	-0.098	0.002	1.8	0.012	0.001	-0.099	0.001	1.0	0.005	0.001	-0.100	0.000	0.2	0.004	0.000
$\eta_{03}^{(1)}$	0.3	0.297	-0.003	1.0	0.130	0.013	0.292	-0.008	2.7	0.069	0.007	0.299	-0.001	0.3	0.043	0.004
$\eta_{13}^{(1)}$	-0.1	-0.099	0.001	0.9	0.012	0.001	-0.099	0.001	0.7	0.006	0.001	-0.100	0.000	0.0	0.004	0.000
$\eta_{04}^{(1)}$	0.3	0.283	-0.017	5.6	0.158	0.016	0.294	-0.006	1.9	0.086	0.009	0.301	0.001	0.4	0.051	0.005
$\eta_{14}^{(1)}$	-0.1	-0.098	0.002	1.9	0.012	0.001	-0.099	0.001	1.1	0.006	0.001	-0.099	0.001	0.7	0.004	0.000
$\eta_{02}^{(2)}$	0.3	0.318	0.018	5.8	0.103	0.010	0.296	-0.004	1.2	0.051	0.005	0.294	-0.006	2.1	0.036	0.004
$\eta_{12}^{(2)}$	-0.1	-0.102	-0.002	1.7	0.011	0.001	-0.100	0.000	0.2	0.005	0.000	-0.100	0.000	0.4	0.003	0.000
$\eta_{03}^{(2)}$	0.3	0.325	0.025	8.4	0.153	0.015	0.307	0.007	2.2	0.079	0.008	0.300	0.000	0.0	0.048	0.005
$\eta_{13}^{(2)}$	-0.1	-0.101	-0.001	1.4	0.012	0.001	-0.100	0.000	0.0	0.006	0.001	-0.099	0.001	0.5	0.003	0.000
ξ	0.1	0.106	0.006	6.2	0.033	0.003	0.104	0.004	4.3	0.017	0.002	0.101	0.001	0.6	0.010	0.001
σ_{b1}	1.0	0.993	-0.007	0.7	0.107	0.011	0.996	-0.004	0.4	0.055	0.005	0.991	-0.009	0.9	0.036	0.004
σ_{b2}	1.0	0.996	-0.004	0.4	0.106	0.011	0.977	-0.023	2.3	0.047	0.005	0.979	-0.021	2.1	0.033	0.003
ρ	0.3	0.302	0.002	0.8	0.139	0.014	0.298	-0.002	0.7	0.062	0.006	0.301	0.001	0.2	0.038	0.004
β	-5.0	-5.124	-0.124	2.5	0.487	0.049	-5.058	-0.058	1.2	0.242	0.024	-5.019	-0.019	0.4	0.141	0.014
α_1	-0.4	-0.444	-0.044	11.1	0.366	0.037	-0.408	-0.008	2.0	0.149	0.015	-0.396	0.004	1.0	0.092	0.009
α_2	-0.4	-0.418	-0.018	4.5	0.241	0.024	-0.399	0.001	0.4	0.131	0.013	-0.407	-0.007	1.9	0.091	0.009
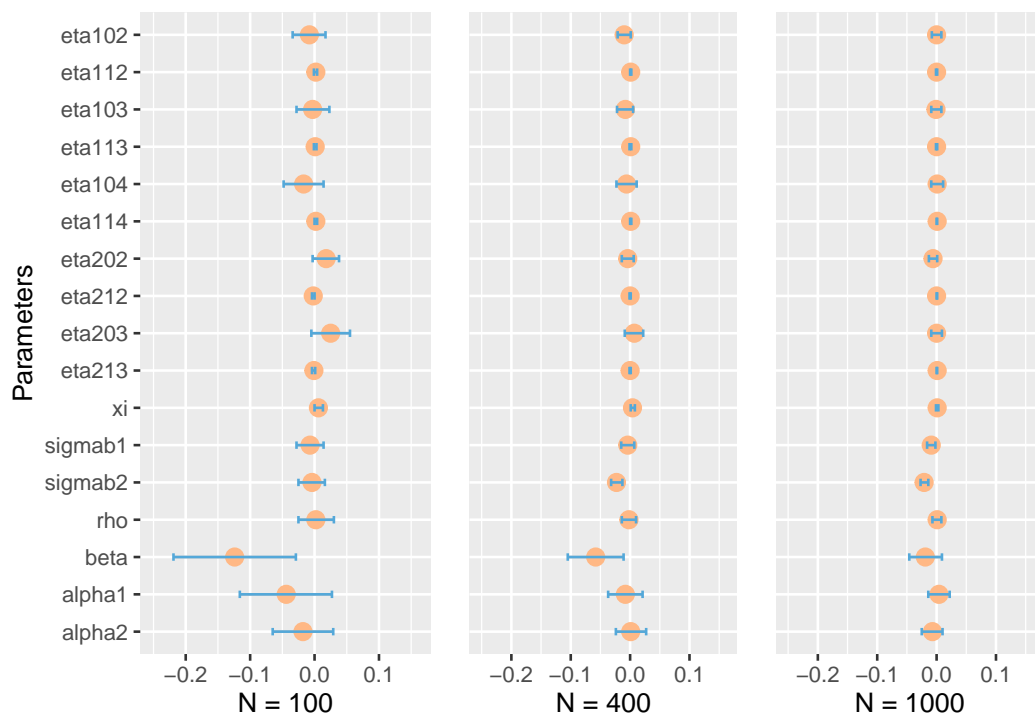
Figure 4.3: Bias and corresponding Monte Carlo 95% confidence interval, based on the results in Table 4.6. Large dots represent biases, and geometry bars represent Monte Carlo 95% confidence interval.

Chapter 5

Data

In this chapter, we will introduce the data used in Section 6: the English Longitudinal Study of Ageing (ELSA); Longitudinal data on cognitive and physical aging in the elderly (PAQUID); the primary biliary cholangitis (PBC2).

5.1   English Longitudinal Study of Ageing (ELSA)

The English Longitudinal Study of Ageing (ELSA) is a repository of information on changes in the health, social, welfare and economic status of the UK population aged 50+ over time. This study was established in 2002. The study's original sample was drawn from households that had responded to the Health Survey for England (HSE) between 1998 and 2001 [J. et al., 2021]. The same group of individuals was interviewed every two years to measure changes in their health, economic and social status. ELSA data are available via the Economic and Social Data Service (www.esds.ac.uk) [J. et al., 2021]. The information collected provides data on household and individual demographics, health, social care, work and pensions, income and assets, housing, cognitive function, social participation, effort and reward, expectations, walking speed and weight for individuals [Taylor et al., 2007].

In this thesis, we focus on a test with two responses in the ELSA data [Banks et al., 2006]. This test requires individuals to read aloud and remember 10 words in the same interview. The first response in the test, which to some extent represents an individual's short-term memory capacity, is the number of words the individual recalls immediately (immediate-recall). An individual's long-term memory capacity

is indicated by recording the number of words he/she can remember after 5 minutes (delayed-recall) [Banks et al., 2006]. The number of immediate-recall words is denoted by Y_1, and the number of delayed-recall words is Y_2. In our data analysis, we use age as the time scale. On this basis, when we analyse the ELSA data, we need to take into account the left truncation. For confidentiality reasons, the corresponding ages in ELSA data are given in integers [Taylor et al., 2007].

From 1998 to 2009, ELSA had 11932 interviewees [Van Den Hout, 2017]. The collected data are pre-processed following four aspects: excluding individuals with only one interview and missing data; excluding individuals without birth year information; excluding individuals younger than 50 years at first observation; and excluding individuals with censored first observed age [Van Den Hout and Muniz-Terrera, 2019]. The sample size of the pre-processed data is $N = 10852$. Of these 10852 individuals, 1884 die during the follow-up period, and the attrition rate is 17.36%. The attrition rate is too large to be ignored. In the joint model, this attrition is modelled by using the age of death as the time-to-event. There are 5946 women and 4906 men in the data. In the ELSA data, the highest education refers to a National Vocational Qualification level 2 or a General Certificate of Education equivalent to O level or above. Individuals with any of these qualifications are considered to have a higher education level, which is represented with value 1, otherwise 0 [Banks et al., 2006]. There are 4699 individuals with high educational qualifications in the data.

The main purpose of this thesis is to illustrate the methodology rather than a study of the ELSA data. Moreover, we need to compute high-dimensional integrals for the shared random-effects joint model, and the large sample size will lengthen the computation time. Therefore, we use a subset of 1000 individuals in this thesis. This subset is randomly drawn from the full data under two constraints: individuals are younger than 90 years at the time of the first observation and have no censored ages during the follow-up period; and there are at least two records for each individual, which should include two observations or one observation and one time of death.

This subset was also used in Van Den Hout and Muniz-Terrera [2019]. There are 4038 observations in the subset, and the number of individuals with 2, 3, 4, 5 and 6 observations (including death) is 174, 146, 156, 516 and 8, respectively. The number of deaths is 195, where the attrition rate is close to the full data. The ratio of women to men in the data is 540 : 460, which is also close to that ratio for the full data. There are 445 individuals with higher education levels selected in the data. The minimum recorded age is 50, and the maximum is 89 years old. The average first recorded immediate-recall words for all individuals is 5.60, and that for the delayed recall is 4.19. This phenomenon indicates that individuals can recall more words through short-term memory than through long-term memory, i.e. short-term memory is better than long-term memory to some extent. Other details for this subset can be found in Van Den Hout and Muniz-Terrera [2019].
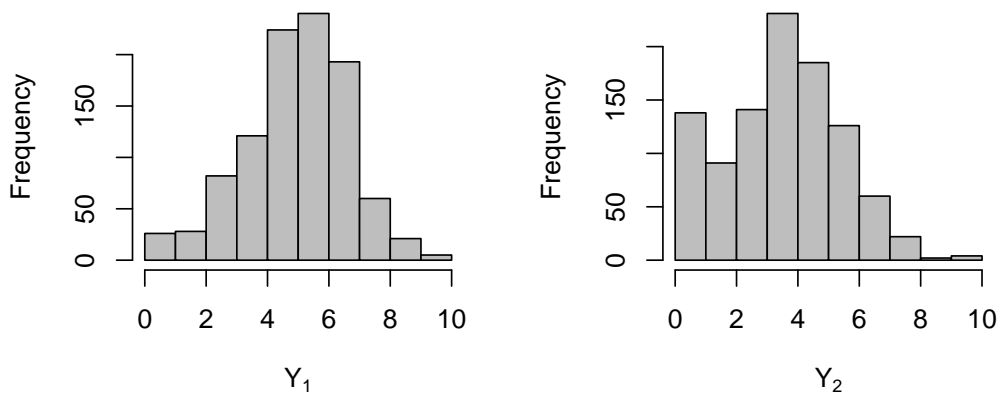
Figure 5.1: Frequency distribution of recalled words at the first interview. Y_1 represents immediate-recall words; Y_2 represents delayed-recall words.

Figure 5.1 presents the frequency distribution of recalled words at the first interview. The distribution of delayed-recall words is clustered to the left compared to immediate-recall words. This phenomenon could also reflect that individuals have better short-term memory than long-term memory, which is in line with the conclusion we reached in the previous paragraph by analysing the average of the first responses. In Figure 5.2, we select thirty individuals and plot their response trajec-
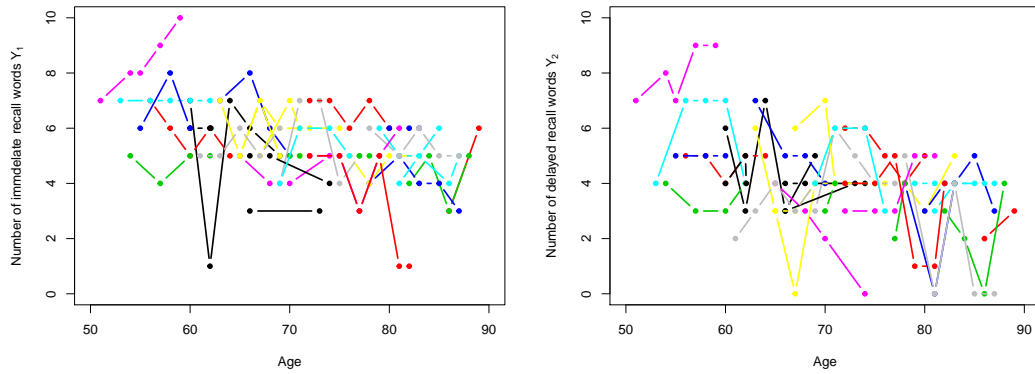
Figure 5.2: Recalled words trajectories. Left-hand side: immediate-recall words Y_1 trajectories. Right-hand side: delayed-recall words Y_2 trajectories. Individuals are represented by same colour in both plots.

tories. We can observe from the graph that individuals' ability to memorise words (cognitive ability) fluctuates and that, generally, the number of words memorised decreases with ageing.

In Chapter 6, we have rescaled the age by subtracting 49 years, which means that 50 years old is expressed as 1, and the maximum age is expressed as 40.

5.2 PAQUID data

Both responses in the ELSA data are integers range from 0 to 10, i.e. $m_1 = m_2 = 10$ in Equation (2.5). To demonstrate the flexibility of the method, where the range of the two responses could be different, we will also use PAQUID: Longitudinal data on cognitive and physical aging in the elderly [Proust-Lima et al., 2017] in Chapter 6.

The PAQUID (or Paquid) prospective cohort is a group of 3,777 people aged 65 or over. The study began in 1988 and lasted until 2004 [Rondeau et al., 2008; Letenneur et al., 1994]. Researchers selected this group from more than 91 different regions in southwest France. This study was designed to investigate the effects of different environmental, behavioural and social vectors of age-related medical conditions and diseases. One of the main goals of the study is to identify some of the causes of dementia and Alzheimer's disease [Nejjari et al., 1997].

Five repeated measures of cognition are included in the data: the Mini-Mental State Examination (MMSE), a psychometric test of overall cognitive function (integer in the range $0 - 30$); the Benton Visual Retention Test (BVRT), a psychometric test of spatial memory (integer in the range $0 - 15$); the Isaacs Set Test (IST), with a 15-second cutoff, is a test of verbal memory (integer in the range $0 - 40$); the physical dependency; and the CES-D, a brief self-report scale designed to measure depressive symptoms in the general population (integer in the range $0 - 52$) [Letenneur et al., 1994]. These measurements are collected together along with dementia information for up to 20 years. Dementia information includes age at dementia diagnosis and dementia diagnosis information. In addition, the data provide sociodemographic information such as gender and age.

This thesis uses the MMSE and the BVRT as the bivariate responses Y_1 and Y_2.

5.2.1 Mini-Mental State Examination (MMSE)

The Mini-Mental State Examination (MMSE) or Folstein test, initially developed by Folstein et al. [1975], is a 30-point questionnaire. This questionnaire is widely used in clinical and research settings to measure cognitive impairment [Pangman et al., 2000]. The test was originally proposed to distinguish between patients with organic and functional psychiatric disorders [Folstein et al., 1975]. It is currently used in medicine and healthcare to detect dementia. It also allows estimating the severity and progression of a patient's cognitive impairment and monitoring their cognitive processes over time. Therefore, it has become an effective recording of an individual's response to treatment. The test lasts $5 - 10$ minutes and covers recording, attention and calculation, memory, language use, the ability to follow simple commands and sense of direction [Tuijl et al., 2012].

In our model, we refer to the MMSE as Y_1, and the range of the MMSE is $Y_1 \in \{0, \dots, 30\}$. Typically researchers consider individuals who score 24 or more out of 30 on the test to have normal cognitive abilities. Individuals who score below this may be considered to have mild $(19 - 23)$, moderate $(10 - 18)$ or severe (≤ 9) cognitive impairment [Vertesi et al., 2001].

The advantages of the MMSE are that no special equipment or training is re-

quired to use this test; it is reliable and valid for diagnosing and longitudinal assessment of Alzheimer's disease [Harrell et al., 2000]. The disadvantages of using the MMSE are that it is influenced by demographic factors, such as age and education; it lacks sensitivity to mild cognitive impairment and does not accurately distinguish between patients with mild Alzheimer's disease and normal patients; the MMSE test lacks an adequate examination of visuospatial cognitive abilities [Tombaugh and McIntyre, 1992].

To compensate for the shortcomings of the MMSE in testing visuospatial cognitive abilities, we used the BVRT as the second response.

5.2.2 Benton Visual Retention Test (BVRT)

The Benton Visual Retention Test (referred to as the Benton Test or BVRT) is a separately administered test that measures visual perception and visual memory in individuals aged eight years to adulthood [Walsh and Betz, 1995; Lockwood et al., 2011]. The test is designed to provide a brief assessment of immediate nonverbal memory, i.e., it is used to assess visual memory, perception, and construction abilities [Benton, 1945; Tamkin and Kunce, 1985; Lockwood et al., 2011]. It can be used to help researchers identify individuals who may have learning disabilities and other disorders that may affect memory. Participants are shown 10 designs, one at a time, and are asked to reproduce each design as accurately as possible from memory on plain paper [Lockwood et al., 2011]. The test has no time limit, and the results are professionally scored by form, shape, pattern and arrangement on the paper. The test score is a non-negative integer between $0 - 15$, with a higher score indicating greater cognitive function. The BVRT is represented as Y_2 in the model, in which $Y_2 \in \{0, \ldots, 15\}$.

The PAQUID data consist of a sub-sample of the PAQUID prospective cohort study. It can be obtained from `LCMM` package in `R` [Proust-Lima et al., 2017]. The dataset contains 500 individuals, with 2250 observations and 12 variables. We remove all missing observations to obtain a subset of 500 individuals and 1950 corresponding observations. Of these individuals, 128 got dementia disease.
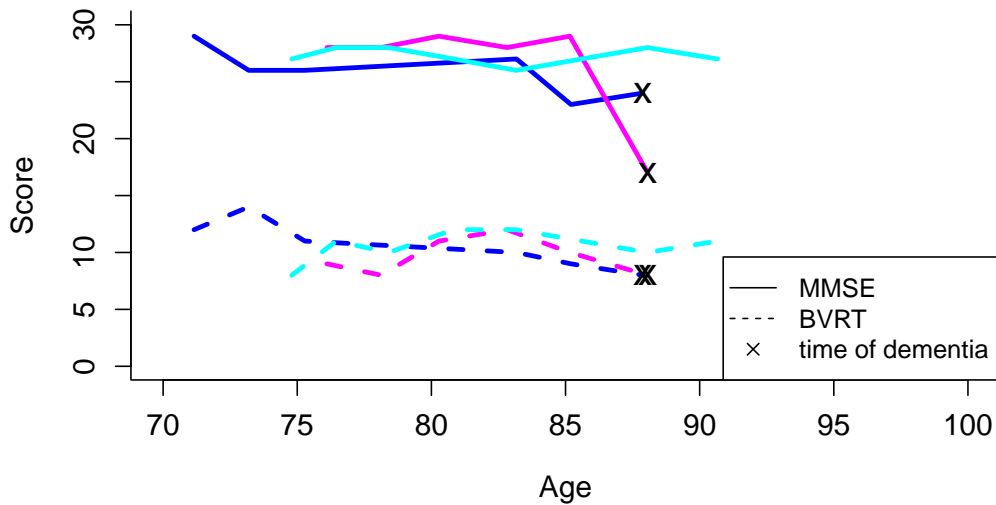
Figure 5.3: Trajectories of three individuals in the PAQUID data

5.3 Primary biliary cholangitis (PBC2)

The PBC2: Mayo Clinic primary biliary cirrhosis data (PBC2) is the third dataset we use in Chapter 6. Primary biliary cholangitis (PBC), formerly known as primary biliary cirrhosis, is a chronic progressive cholestatic liver disease [Mayo, 2022; Lammert et al., 2013]. The disease is caused by an autoimmune attack on the small bile ducts. As a result, the bile ducts in the patient's liver are slowly destroyed [Kaplan, 1996]. Its pathogenesis is largely unknown, and some researchers have suggested that there may be a complex interaction between environmental and genetic factors and susceptibility to this disease [Kaplan, 1996; Lazaridis and Talwalkar, 2007]. The PBC occurs almost exclusively in women between the ages of 35 and 70 [Lammert et al., 2013]. We access the PBC2 dataset via the JM package in R [Rizopoulos, 2010; R Core Team, 2013].

The PBC2 data are from the Mayo Clinic trial of hepatic primary biliary cholangitis (PBC) conducted between 1974 and 1984 [Rizopoulos, 2010]. During this decade, 424 patients with PBC were referred to the Mayo Clinic and met the eligibility criteria for a randomised placebo-controlled trial of the drug D-penicillamine, but only the first 312 cases in the dataset were enrolled in the

randomised trial [Hickey et al., 2018]. Therefore, there are 312 individuals in this data. There are 1945 observations in the data, and time in the data is defined as the number of years between registration and the time of death, transplantation or study analysis (whichever is earlier). Our event of interest is death for the PBC2 data. The ratio of individuals surviving to die in this dataset is 172:140, which is too large to ignore the effect of death on the data analysis process. In addition, we mentioned above that PBC almost exclusively occurs in women, and this feature is also reflected in the gender ratio in the dataset: 36:276 men to women.

We focus on two responses in the PBC2 dataset. The first response, represented by Y_1 is the histologic stage of liver biopsy. We used the classification criteria for the histological stages given in the dataset [Rizopoulos, 2010]. Assume $Y_i \in \{1, 2, 3, 4\}$. The corresponding stages (categories) indicate the following states [Rizopoulos, 2010]:

- Stage 1: Inflammation, abnormal connective tissue, or both, confined to the portal areas.

- Stage 2: Inflammation, fibrosis, or both, confined to the portal and periportal areas.

- Stage 3: Bridging fibrosis.

- Stage 4: Cirrhosis.

An individual with PBC may be asymptomatic for up to 10 years. If an individual has PBC in the early stages (stage 1 or stage 2), their life expectancy is normal. There is no cure for cirrhosis, but eliminating the cause of the disease can slow it down. If the damage is not too severe, the liver can heal itself over time [Mayo, 2022]. However, PBC usually progresses to an advanced stage within 15 to 20 years. If an individual with PBC develops advanced symptoms, the average life expectancy is reduced to 10-15 years [Mayo, 2022].

The second response is the edema conditions, represented by Y_2. Edema is one of the signs and symptoms of advanced cirrhosis [Mayo, 2022]. There are three categories in this response in Rizopoulos [2010]:

- "No edema" = 1

- "edema no diuretics" = 2

- "edema despite diuretics" = 3

We made a cross tabulation of these two responses in Table 5.1. We can ob-

Table 5.1: PBC2: Cross tabulation of two responses in PBC2.

		Edema conditions		
		1	2	3
Histologic stages of liver biopsy	1	90	5	0
	2	234	29	3
	3	471	107	34
	4	606	238	128

serve from Table 5.1 that when $Y_1 = 1$ and $Y_2 = 3$, there are no corresponding observations in the dataset. The reason for this phenomenon may be the one we mentioned above: edema is a possible symptom of advanced primary biliary cholangitis [Mayo, 2022] that does not appear in the early stages of the disease.

Chapter 6

Data Analysis

In this chapter, we will use different joint models to analyse the three datasets mentioned in Chapter 5. We divide this chapter into two main sections based on the longitudinal model used in joint models: binomial joint models and categorical joint models. These longitudinal distributions are mentioned in Section 2.1. Section 6.1 on binomial joint models, we construct shared random-effects models and latent-class models to analyse ELSA and PAQUID data. We will use the ELSA and PBC2 data in the Categorical joint models Section; see Section 6.2. The Gompertz hazard model, the Weibull hazard model and the exponential hazard model mentioned in Section 2.2 are used for the survival model. The longitudinal and the survival models are joined together, given the random effects. The random effects for shared random-effects joint models follow a bivariate normal distribution (2.59) with a mean of 0.

6.1 Binomial joint models

Longitudinal parts of all joint models presented in this section are based on the extension of the binomial distribution (2.4). We first present the bivariate extension shared random-effects joint models based on the univariate model proposed by Van Den Hout and Muniz-Terrera [2016]. Afterwards, we present bivariate shared random-effects and latent-class joint models, respectively. These joint models are constructed on the basis of the link functions we discuss in Section 2.3.

6.1.1 Bivariate extended shared random-effects joint model based on previous paper

Van Den Hout and Muniz-Terrera [2016] constructed univariate shared random-effects joint models with the binomial and beta-binomial distributions. We want to build a bivariate joint model using a similar approach, the shared random-effects method. Notwithstanding that we will use different binomial distributions and have a different number of responses, we still think it is worthwhile to construct a bivariate joint model based on the univariate joint model proposed in Van Den Hout and Muniz-Terrera [2016] because the same shared random-effects approach is applied. In constructing the bivariate joint model, we use the same format as the random intercept and random slope presented in Van Den Hout and Muniz-Terrera [2016] to model one of the responses, and for the other response, we use the fixed effects in the corresponding logistic regression model.

In Van Den Hout and Muniz-Terrera [2016], they use a similar structure to the link function Δ we discussed in Section 2.3. For ease of understanding, we represent this structure as follows:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha_1(\eta_0 + b_1) + \alpha_2(\eta_1 + b_2)t. \tag{6.1}$$

The estimates of the link function's parameter αs in the univariate joint model are negative. These estimates illustrate that: (i) there is a relationship between the longitudinal and survival models, i.e. it is reasonable to model the joint model rather than modelling the longitudinal data and time-to-event data separately; (ii) take random slope as an example, the parameter α of the random slope is negative indicating that the risk of death of an individual increase as time t increases.

If the estimates of αs for the bivariate joint model are similar to those of the univariate joint model, i.e. both αs are negative, we could infer that our approach is consistent with the analysis based on the univariate joint model, i.e. the idea of constructing bivariate joint models in this thesis is feasible. Since the response in the univariate joint model is delayed-recall Y_2, we also include the random effects

to the linear predictor associated with Y_2. We assume that the logistic regression model for delayed-recall words p_{Y_2} is equal to (2.7), where $j = 2$. For immediate recall Y_1, we have the logistic regression model p_{Y_1}:

$$p_{Y_1} = \frac{\exp(\eta_0^{(1)} + \eta_1^{(1)}t)}{1 + \exp(\eta_0^{(1)} + \eta_1^{(1)}t)}. \tag{6.2}$$

To build the joint model, we need to add expressions that include random effects in the survival model, i.e. the link function. The corresponding link function, similar to Equation (6.1), is

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha_1 \left(\eta_0^{(2)} + b_{20} \right) + \alpha_2 \left(\eta_1^{(2)} + b_{21} \right) t, \tag{6.3}$$

where the random effects b_{20} and b_{21} follow a bivariate normal distribution with a mean zero (2.59). The link function associated with delayed recall is included in the hazard model. We use the Gompertz hazard model in this section, which equals to (2.48). Parameter λ^* and ξ^* in the Gompertz hazard model equal to (2.49) for $j = 2$. The estimated results are shown in Table 6.1.

Table 6.1: ELSA: Parameter estimates for the bivariate extension based on the joint model in [Van Den Hout and Muniz-Terrera, 2016]. Values in brackets are the standard error of the corresponding parameters.

$\eta_0^{(1)}$	-0.857 (0.053)	θ_{Y_1}	1.122 (0.009)	σ_{b1}	1.116 (0.073)	β	-6.800 (0.497)
$\eta_0^{(2)}$	-1.969 (0.094)	θ_{Y_2}	1.167 (0.011)	σ_{b2}	0.055 (0.005)	ξ	0.097 (0.012)
$\eta_1^{(1)}$	-0.014 (0.002)	ϕ	1.432 (0.016)	ρ	-0.682 (0.190)		
$\eta_1^{(2)}$	-0.038 (0.004)			α_1	-0.400 (0.168)	α_2	-0.193 (0.086)

The estimates of both parameters αs for the link function in the joint model in Table 6.1 align with our expectations. The estimate of the parameter α_1 associated with the random intercept is negative and significantly different from zero ($\hat{\alpha}_1 = -0.400$), implying that the risk of death is relatively low when individuals have a better cognitive function at baseline age ($t = 0$). Meanwhile, the parameter α_2 associated with the random slope is also negative ($\hat{\alpha}_2 = -0.193$), which means that

the risk of death increases over time when b_{21} is equal to 0 on average.

We can calculate the probability of remembering a word at baseline age ($t = 0$) utilising the fixed intercepts $\hat{\eta}_0^{(j)}, j = 1, 2$, given that the random effects b_{20} and b_{21} are both equal to their mean 0. For the number of immediate-recall words, we have the fixed intercept $\hat{\eta}_0^{(1)} = -0.857$, with a corresponding probability equal to 0.298; for the number of delayed-recall words, the corresponding probability drops to 0.122 with $\hat{\eta}_0^{(2)} = -1.969$. At baseline age ($t = 0$), the probability of recalling a word immediately is higher than the probability of recalling it later. It means that the number expected to be remembered immediately is greater than the number of words expected to be delayed. The fixed slopes of both responses $\hat{\eta}_1^{(1)}$ and $\hat{\eta}_1^{(2)}$ are less than 0, indicating that individuals' memory or cognitive function declines with age. Furthermore, since $\hat{\eta}_1^{(1)} > \hat{\eta}_1^{(2)}$, we can infer that for each individual, temporary memory decays more slowly with age than long-term memory.

The estimates of the parameters $\hat{\theta}_{Y_1}$ and $\hat{\theta}_{Y_2}$ are greater than 1, meaning that for two responses of the dataset, their corresponding distributions are under-dispersed compared to the standard binomial distribution. Moreover, $\hat{\phi} > 1$ means a positive correlation between the responses.

6.1.2 Discussion of bivariate shared random-effects joint models with different link functions Δ

After verifying the feasibility of the joint model in Section 6.1.1, we will now construct the shared random-effects joint models based on the link function mentioned in Section 3.2. In our joint models, we use the bivariate extension of the binomial distribution (2.4) as the longitudinal model and the Gompertz hazard or the Weibull hazard as the survival model. Gender is added to the model as a covariate (0 for women, 1 for men). The corresponding logistic regression model is:

$$p_{Y_j} = \frac{\exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t + \gamma_L^{(j)} sex\right)}{1 + \exp\left((\eta_0^{(j)} + b_{j0}) + (\eta_1^{(j)} + b_{j1})t + \gamma_L^{(j)} sex\right)}, \tag{6.4}$$

where $j = 1$ or 2.

In the link function we add the covariate gender and the associated parameter γ_S, i.e. the function is written as $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, \gamma_S, t)$. Taking Equation (2.31) as an example, in this section, we write it as follows:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, \gamma_S, t) = \sum_{j=1}^{2} \alpha_j \left(\eta_0^{(j)} + b_{j0} \right) + \gamma_S sex. \tag{6.5}$$

As mentioned in Section 2.1 and Section 2.3, we will investigate restricted joint models. We restrict models by adding only one random effect for each response for two reasons. Firstly, we want to be able to distinguish the effect of changes in the two responses overtime on the risk of death. Secondly, because the shared random-effects joint models require the calculation of high-dimensional integrals, we want to reduce the computational difficulty. It means that we will only discuss the situations where the expressions associated with the two responses both contain random intercepts or random slopes. Furthermore, in these joint models, four parameters could be used to measure the correlation between two responses: ϕ in the bivariate extension of the binomial model (2.4); correlation ρ between corresponding random effects for two responses and link function parameters α_1 and α_2. Therefore, we will discuss the case in which α_1, α_2 and ρ are equal to 0, respectively.

We use *I* to denote the joint model with a random intercept in the link function, i.e. link function equals to (2.29) or (2.31). In this situation we restrict the random effects in the longitudinal model to $b_{11} = b_{21} = 0$ (Equation (2.10)). The letter *S* is used to denote a model in which the link function contains a random slope. In this situation the corresponding link function is equal to either (2.33) or (2.35), and the corresponding longitudinal is restricted to Equation (2.11). The letters *G* and *W* denote joint models containing the Gompertz hazard model and the Weibull hazard model, respectively.

Table 6.2 lists the AIC values and the corresponding parameters for the link function Δ. Model IW_1 - IW_4 and Model IG_1 - IG_4 in Table 6.2 represent joint models with random intercepts. For Model SG_1 - SG_4, we construct the joint model

with random slopes.

When individuals have a good cognitive function at baseline age ($t = 0$) or have a slow trajectory of cognitive decline over time, we expect the risk of death to be relatively low. In Table 6.2, the αs estimates are less than 0 for all models except the Model SG_1. For Models IW_1 - IW_4 and Models IG_1 - IG_4, αs estimates less than 0 imply a relatively low risk of death when individuals have a better cognitive function at time 0. For models that include a random slope (Models SG_2 - SG_4), a negative α means that the risk of death is lower when individuals experience a relatively slow decline in cognitive function.

For Model SG_1, when the link function has random slopes for both responses (2.33), the model estimate of α_1 is inconsistent with our expectations of the process. The estimated parameter associated with the random slope $\hat{\alpha}_1 = 0.062$ in the Model SG_1 is positive. This positive α_1 affects an individual's risk of death. It represents that affected by the random slope and the fixed slope of immediate recall, an individual's risk of death becomes smaller over time. This result is contrary to our expectation: the risk of death decreases instead of increasing over time.

Generally, the AIC value for the model with random intercept (e.g. Model IW_2) is smaller than that for the corresponding model with random slope (e.g. Model SW_2). It means that the heterogeneity in slope between individuals is small, so the increase in model likelihood is negligible compared to the increase in complexity.

6.1.2.1 Random intercept for both Y_1 and Y_2

The joint Model IG_2, using the Gompertz survival model with link function (2.29) for $j = 1$, has the smallest AIC value of all joint models that include random intercepts. We consider this model the best fit among all models that include random intercepts. We list the parameter estimates in Table 6.3.

In Table 6.3, we first analyse the corresponding parameters $\eta_0^{(1)}$ and $\eta_0^{(2)}$ for the fixed intercepts of the two responses. we have $\hat{\eta}_0^{(1)} = -0.580$ greater than parameter for random intercept Y_2 ($\hat{\eta}_0^{(2)} = -2.014$), and we can use these two parameters to calculate the corresponding probability of remembering a word at time 0. With $b = 0$, the probability of recalling a word immediately at time 0 is 0.359,

Table 6.2: ELSA Data: AIC and estimated αs for shared random-effects joint models $-2LL$ represents the -2loglikelihood.

MODEL	py_j	Link function	Hazard distribution	$-2LL$	AIC	α	α_1	α_2	ρ
IW_1		(2.31)	Weibull	100985.1	101019.1		-0.480	-0.031	0.316
IW_2		(2.29) $j=1$	Weibull	100985.2	101017.2	-0.508			0.318
IW_3		(2.29) $j=2$	Weibull	100989.2	101021.2	-0.154			0.327
IW_4	(2.10)	(2.31)	Weibull	101001.2	101033.2		-0.507	-0.075	
IG_1		(2.31)	Gompertz	100965.5	100999.5		-0.609	-0.007	0.315
IG_2		(2.29) $j=1$	Gompertz	100965.5	100997.5	-0.616			0.316
IG_3		(2.29) $j=2$	Gompertz	100971.2	101003.2	-0.168			0.329
IG_4		(2.31)	Gompertz	100981.5	101013.5		-0.632	-0.065	
SG_1		(2.33)	Gompertz	103373.2	103407.2		0.062	-0.015	0.474
SG_2	(2.11)	(2.35) $j=1$	Gompertz	101170.6	101202.6	-0.617			0.436
SG_3		(2.35) $j=2$	Gompertz	101179.1	101211.1	-0.117			0.431
SG_4		(2.33)	Gompertz	101197.5	101229.5		-0.583	-0.069	

Table 6.3: ELSA: Parameter estimates for shared random-effects Model IG_2 with (2.29) $j = 1$. Values in brackets are the standard error of the corresponding parameters.

$\eta_0^{(1)}$	-0.580 (0.080)	θ_{Y_1}	1.206 (0.035)	σ_{b1}	0.447 (0.101)	β	-7.043 (0.350)
$\eta_0^{(2)}$	-2.014 (0.108)	θ_{Y_2}	1.123 (0.028)	σ_{b2}	0.649 (0.094)	ξ	0.122 (0.009)
$\eta_1^{(1)}$	-0.024 (0.002)	ϕ	1.425 (0.061)	ρ	0.316 (0.082)		
$\eta_1^{(2)}$	-0.029 (0.003)					α	-0.616 (0.227)
$\gamma_L^{(1)}$	-0.063 (0.042)	$\gamma_L^{(2)}$	-0.130 (0.051)	γ_S	0.370 (0.148)		

which is higher than the probability of recalling a word later (0.118). A high probability of remembering words means better expected cognitive function. The number of words expected to be recalled immediately is greater for an individual than the number of words for delayed recall.

Meanwhile, the estimate of the parameter α included in the link function is -0.616. A negative estimate of the parameter α indicates that the risk of death is relatively low when the individual has a better cognitive function. We could calculate that the estimate of α is significantly different from 0 based on the standard error. This phenomenon indicates that there is a relationship between the longitudinal process and the survival process.

Furthermore, both fixed slope $\hat{\eta}_1^{(1)}$ and $\hat{\eta}_1^{(2)}$ are less than 0, which implies that the probability of recalling a word decreases with age. We also notice that random effects b_1 and b_2 are positively correlated with $\hat{\rho} = 0.370$. This situation suggests that if individuals have good short-term memory (immediate recall ability) at the beginning of the study, corresponding long-term memory (delayed recall ability) will be slightly positively affected. The parameters $\hat{\theta}_{Y_1}$ and $\hat{\theta}_{Y_2}$ are greater than 1, implying that the estimated distribution has a more top-shaped distribution than the standard bivariate binomial distribution. The two gender-related parameters $\hat{\gamma}_L^{(1)}$ and $\hat{\gamma}_L^{(2)}$ in the longitudinal model are both less than 0. This result implies that all other parameters being equal, men are expected to have lower cognitive function than women. Parameter $\hat{\gamma}_S = 0.370$ represents a higher risk of death for men than

for women, given the same conditions.

6.1.2.2 Additional remarks

Compared to the probability of remembering words at time 0 calculated from $\eta_0^{(1)}$ and $\eta_0^{(2)}$, we would expect changes in cognitive ability over time to have a greater impact on the risk of death.

For the Gompertz hazard model, we rewrite the hazard model based on (2.48):

$$
\begin{aligned}
h(t) &= \exp\left(\beta + \eta_0^{(j)} + \alpha(\eta_1^{(j)} + b_{j1})t + \xi t\right) \\
&= \exp\left\{\beta + \eta_0^{(j)} + \left(\alpha(\eta_1^{(j)} + b_{j1}) + \xi\right)t\right\}.
\end{aligned}
\tag{6.6}
$$

In this case, the parameter ξ^* for the Gompertz model is:

$$
\xi^* = \alpha(\eta_1^{(j)} + b_{j1}) + \xi.
\tag{6.7}
$$

In Table 6.2, we constructed four models that use random slopes as link functions (Model SG_1 - SG_4). The Model SG_2, using the Gompertz survival model with link function (2.35) for $j = 1$, has the best fit among the four models.

In Model SG_2, the estimation of $\hat{\eta}_0^{(1)} = -0.631$ is higher than $\hat{\eta}_0^{(2)} = -1.956$, representing the higher probability of remembering a word. An individual's expected number of immediate-recall words is greater than the delayed recall. The parameter $\hat{\alpha}$ associated with the random slope and the fixed slope for the immediate recall has a negative estimate (-0.617), and the estimate of the fixed slope is negative, indicating that when an individual's cognitive function declines slowly, he/she has a relatively low risk of death. Estimated $\hat{\eta}_1^{(1)}$ and $\hat{\eta}_1^{(2)}$ are below 0, indicating that the probability of remembering a word decreases with age. As with the results presented in Table 6.3, the correlation between random effects b_1 and b_2 is positive $(\hat{\rho} = 0.436)$. In this model, with the inclusion of a random slope, this result shows that if an individual's instant memory declines rapidly with age, his/her long-term memory may also decline fast. The estimation results for Model SG_2 are presented in Table B.1 in Appendix B.

In Table 6.2, we have $\hat{\alpha}_1 > 0$ when the random effects for both Y_1 and Y_2 are

included in the survival model, i.e. Model SG_1. A positive estimate of the parameter α_1 means that when an individual's cognitive ability declines relatively slowly per unit of time, his/her risk of death becomes higher. Even if the AIC value of Model SG_1 is not the smallest of all joint models that include random slopes, we still think it is worthwhile to analyse the estimated value of α_1. This result may be caused by the correlation between responses Y_1 and Y_2. In constructing the bivariate joint model, we capture the correlation between the two responses several times through the parameters ϕ, ρ and αs. Overfitting of correlations between responses may lead to this estimate.

6.1.2.3   Model fit

In order to investigate the fit of the model, in this section, we will first calculate the corresponding individual-specific random effects using the maximum a posterior (MAP) method. After that, we will calculate the mean of these individual-specific random effects and plot the fitted distribution based on the mean. Finally, we will discuss the accuracy of the model predictions by comparing the predicted survival curves with the observed survival curves (Kaplan-Meier survival curves).

We use $\widetilde{y}_i$ to represent the response of individual i. The parameter $\widetilde{t}_{in_i}$ indicates the corresponding age to the last observation of the individual, and $\widetilde{t}_{i1}$ denotes the age corresponding to the first observed age. When an individual has only one observation, we set $\widetilde{t}_{in_i}$ and $\widetilde{t}_{i1}$ to be equal. We need the corresponding values of the random effects to predict individual trajectories given the observed data. The MAP estimation is used to derive the most probable value of the random effect:

$$
\begin{aligned}
&p(\boldsymbol{b}_i|\widetilde{t}_{i1},\widetilde{t}_{in_i},\delta_i=0,\widetilde{\boldsymbol{y}}_i;\boldsymbol{\omega}=\hat{\omega})\\
&\propto p(\widetilde{t}_{in_i}|\delta_i=0,\boldsymbol{b}_i;\boldsymbol{\omega}=\hat{\omega})p(\widetilde{\boldsymbol{y}}_i|\boldsymbol{b}_i;\boldsymbol{\omega}=\hat{\omega})p(\boldsymbol{b}_i|\boldsymbol{\omega}=\hat{\omega}).
\end{aligned}
\tag{6.8}
$$

In the shared random-effects joint model, the distribution of random effects is set before fitting the model. This thesis sets the random effects are set to follow a bivariate normal distribution (2.59) with a mean of 0. After obtaining the corresponding random effects from Equation (6.8), we can verify whether the estimated random effects are the same as the distribution we set. Using the model fitted in Sec-

tion 6.1.1 as an example, the estimated random effects follow the bivariate normal

distribution $\boldsymbol{b}_i \sim N\left(\begin{pmatrix} -0.017 \\ 0.002 \end{pmatrix}, \begin{pmatrix} 0.306 & -0.345 \\ -0.345 & 0.001 \end{pmatrix}\right)$.

The mean values of random effects are close to what we set. However, the esti-

mated variances and correlation are smaller than the result in Table 6.1. According

to Laird [2004], it is common for the variance of the estimates to be smaller than the

set value. Therefore the distribution of the random effects we calculated from the

MAP is consistent with our expectations. Both survival and longitudinal responses

can be predicted, given estimated random effects and assumed ages.
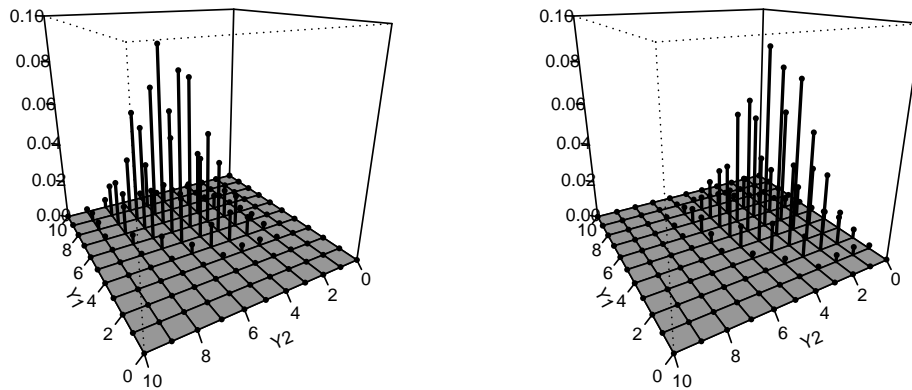
Figure 6.1: Fitted bivariate binomial distribution for immediate recall (Y_1) and delayed re-
call (Y_2) based on Table 6.1.
Left hand side: conditional on age 50; Right hand side: conditional on age 80.

We plot the three-dimensional fitted images of the bivariate extended binomial

distribution (Figure 6.1) based on Table 6.1. Random effects $\hat{b}_i$s are equal to the

mean of $\boldsymbol{b}_i$ estimated in (6.8) in Figure 6.1. The density plot on the left is conditional

on age 50, and the density on the right is conditional on age 80. We can observe

from the plot that the number of words recalled by an individual is likely to decrease

with age. At age 50, the number of immediate-recall words is concentrated between

4 and 10. When age reaches 80, the number of immediate-recall words decreases

to $2-8$. This change is more pronounced in the number of delayed-recall words than than in immediate-recall words. At the age of 50, the number of delayed-recall words is concentrated at $3-8$. However, when the individual reaches the age of 80, the number of delayed-recall words drops dramatically to $0-5$ for most people. Numerically, we can confirm this finding with the corresponding expectations: $E(Y_1, Y_2 | t = 50) = (6.8, 5.6)$; $E(Y_1, Y_2 | t = 80) = (4.6, 3.0)$. Figure B.1 - Figure B.4 in Appendix B are the corresponding three-dimensional plots for Models IG_1 - IG_4 in Table 6.2. We list the expectation values of Models IG_1 - IG_4 in Appendix B as well.
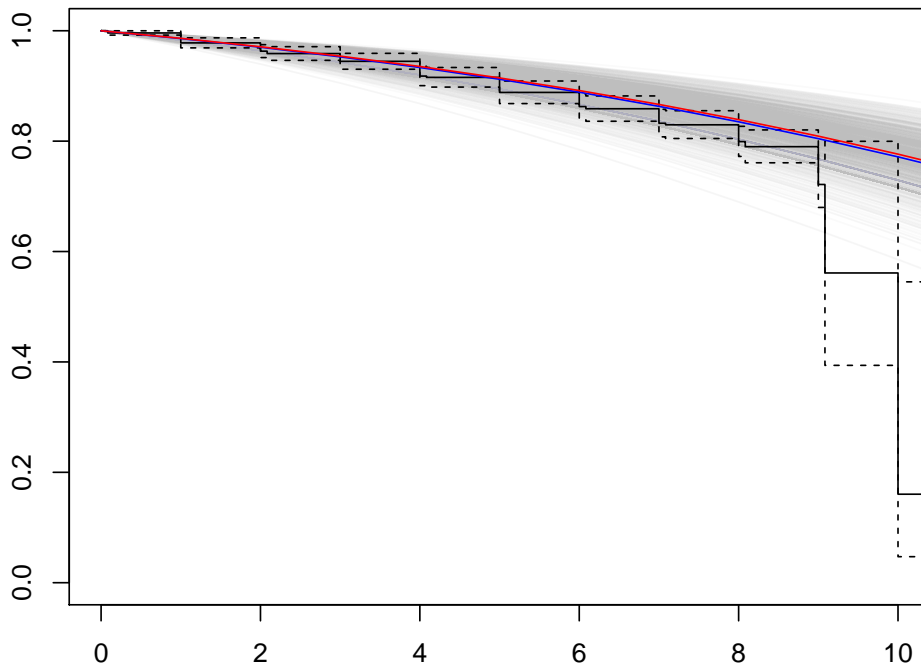
Figure 6.2: Comparison of K-M survival curves: predicted survival curves (grey lines), the mean of those survival curves (blue line) and the median of survival curves (red line).

We expect our joint model could predict the risk of death given the observed age and the corresponding responses. To demonstrate this, we plot the predicted survival curves based on the calculated random effects and compare these curves

with the observed survival curves (i.e. the Kaplan-Meier survival curves). Figure 6.2 shows the comparison of Kaplan-Meier survival curves and predicted model survival curves based on the model fitted in Section 6.1.1. The blue line shows the mean of predicted survival curves, and the red line shows the median of predicted survival curves. Only the baseline test score is used in Figure 6.2. The comparative plots illustrate our joint model's relatively good predictive power for individual risk of death.

6.1.3 Bivariate extended latent-class joint models

The shared random-effects joint model is computationally intensive because we need to calculate the integral associated with the random effects when fitting the model. Moreover, as we mentioned in Section 1.1.2, sometimes the observed responses do not fully reflect the heterogeneity of the results, and the population may be composed of discrete 'latent subpopulations' defined by some unobserved characteristics [Arbeev et al., 2014]. A shared random-effects joint model with continuous random effects may not fit these 'latent subpopulations' well. Therefore, instead of constructing shared random-effects joint models with individual-specific random effects, in this section, we build latent-class joint models with class-specific random effects.

For the marginal likelihood function (3.13), we divide individuals into two groups ($G = 2, g \in \{1, 2\}$). We set the class-specific random effect $b_1 = -b_2$ and define the probability for an individual belonging to the class g (represented by π_{ig}) with the gender and education level of the individual, where

$$\pi_{i1} = \frac{\exp(\mu_{01} + \mu_{11}\mathrm{sex}_i + \mu_{21}\mathrm{educ}_i)}{1 + \exp(\mu_{01} + \mu_{11}\mathrm{sex}_i + \mu_{21}\mathrm{educ}_i)}. \tag{6.9}$$

We use the same longitudinal model and survival models as in the shared random-effects joint models in Section 6.1.2 to construct latent-class joint models. The only class-specific element in the latent-class joint model is the random effect b_g. Taking Equation (3.9) as an example, in this logistic regression model, only the random effect b_{g0} and b_{g1} vary according to the grouping, while all other

parameters are response-specific. For logistic regression model (3.9) we have: (i) Assuming $b_{g1} = 0$ in (3.9), i.e. random intercepts are included in the linear predictor,

$$p_{Y_{jg}} = \frac{\exp\left(\left(\eta_0^{(j)} + b_{g0}\right) + \eta_1^{(j)}t\right)}{1 + \exp\left(\left(\eta_0^{(j)} + b_{g0}\right) + \eta_1^{(j)}t\right)}. \tag{6.10}$$

The corresponding link functions in this case are Equation (2.53) and Equation (2.54). (ii) Assuming $b_{g0} = 0$ in (3.9), i.e. random slopes are included in the linear predictor,

$$p_{Y_{jg}} = \frac{\exp\left(\eta_0^{(j)} + (\eta_1^{(j)} + b_{g1})t\right)}{1 + \exp\left(\eta_0^{(j)} + (\eta_1^{(j)} + b_{g1})t\right)}. \tag{6.11}$$

In this situation the corresponding link functions are Equation (2.55) and Equation (2.56).

According to Table 6.2, the joint model containing the Gompertz hazard has a lower AIC value than the Weibull hazard under the same conditions. Therefore, we could infer that the joint model containing the Gompertz hazard is more suitable for analysing ELSA data than the joint model containing the Weibull hazard. Therefore, we only use the Gompertz hazard model for latent-class joint models.

When we use the Gompertz hazard model, the only class-specific element is the random effect b_g. We take the example of the link function Δ equal to (2.53), and the corresponding hazard model is written as:

$$h_g(t) = \exp(\beta + \alpha(\eta_0^{(j)} + b_{g0}) + \xi t). \tag{6.12}$$

As with the shared random-effects models, we divide the joint models into several types according to the expression of the logistic regression models (3.9). We still use *I* to represent joint models with a random intercept in the link function, i.e. where the link function is equal to (2.53) or (2.54). In this case for the longitudinal model, we restrict $b_{g1} = 0$ (Equation (6.10)). We use *S* to denote the model with a

random slope in the link function. In this situation, the corresponding link function is equal to (2.55) or (2.56) and the corresponding longitudinal model is restricted to Equation (6.11). The letter G represents a model containing the Gompertz hazard model. The AIC values and the corresponding parameters of link function Δ are presented in Table 6.4.

Table 6.4: ELSA Data: AIC and estimated αs for latent-class joint models
$-2LL$ represents the -2loglikelihood

MODEL	p_{Y_j}	Link function	$-2LL$	AIC	α	α_1	α_2
IG_1		(2.53) $j=1$	29335.34	29363.34	-2.096		
IG_2	(6.10)	(2.53) $j=2$	29372.58	29400.58	-1.670		
IG_3		(2.54)	29629.06	29659.06		-1.086	1.364
SG_1		(2.55) $j=1$	29496.99	29524.99	-1.604		
SG_2	(6.11)	(2.55) $j=2$	29476.84	29504.84	-1.138		
SG_3		(2.56)	29534.6	29564.6		-4.149	2.833

The AIC values for the joint models with link functions for one response ((2.53) and (2.55)) present better fit than models with link functions for two responses ((2.54) and (2.56)). The AIC values of these models in Table 6.4 are pretty close. We choose Model IG_1 with the smallest AIC for further inference.

Table 6.5: ELSA: Parameter estimates for latent-class Model IG_1 with (2.53) $j=1$. Values in brackets are the standard error of the corresponding parameters.

$\eta_0^{(1)}$	-0.529 (0.060)	θ_{Y_1}	1.153 (0.009)	$b1$	0.260 (0.010)	β	-6.838 (0.372)
$\eta_0^{(2)}$	-1.780 (0.077)	θ_{Y_2}	1.038 (0.006)	$b2$	-0.260 (0.010)	ξ	0.097(0.009)
$\eta_1^{(1)}$	-0.022 (0.002)	ϕ	1.372 (0.015)				
$\eta_1^{(2)}$	-0.024 (0.002)					α	-2.096 (0.377)
μ_{01}	-0.757 (0.170)	μ_{11}	-0.778 (0.184)	μ_{21}	2.042(0.190)		

Similar to the discussion for the shared random-effects models, the probability of remembering a word at baseline age ($t=0$) can be obtained via parameter $\hat{\eta}_0^{(1)} =$

-0.529 and $\hat{\eta}_0^{(2)} = -1.780$. The probability of recalling a word immediately is higher than recalling it later. It means that the expected immediate recall number is larger than the expected number of the later recall. Both fixed slopes for two responses $\hat{\eta}_1^{(1)}$ and $\hat{\eta}_1^{(2)}$ are smaller than 0, which shows that individuals' memory or cognitive function declines with age. Moreover, since $\eta_1^{(1)} > \eta_1^{(2)}$, we can infer that for each individual, the transient memory decays more slowly than long-term memory with age. The estimation of $\hat{\alpha} = -2.096$ is negative, which means that the risk of death is relatively low when the individual has a better cognitive function at baseline age.

The probability for an individual belongs to the first class is negatively related to gender ($\hat{\mu}_{11} = -0.778$) and positively correlated with education level ($\hat{\mu}_{21} = 2.042$). It indicates that highly educated females are more likely to be grouped into the first class. It is worth noting that although we used the level of higher education as one of the criteria for grouping individuals into two classes, the results obtained through such grouping criteria are not necessarily accurate. The reason is that an individual's level of education is to some extent related to his/her age, i.e. the education level contains different information in the different calendar years. For example, an individual born in 1950 is less likely to be in higher education than a person born in 2000.

Based on the estimated $\hat{\alpha}$s for Model IG_3 and Model SG_3, we still have the parameter estimation problem, in which the estimations of αs are still positive after converting the model from the shared random-effects joint model to the latent-class joint model. Moreover, we have verified in Section 4.1 that our joint model estimates for the bivariate binomial shared random-effects joint models are accurate. Therefore, we can infer that this problem is not caused by the limitation of the shared random-effects joint model.

6.1.4 Bivariate binomial joint models using PAQUID data

In the previous section, we have used the ELSA data, where the range of the two responses is the same, i.e. $m_1 = m_2 = 10$ in Equation (2.4) - (2.5). In fact, the bivariate extension of the binomial distribution (2.4) allows us to have two responses

with different ranges. To verify the distribution's flexibility and compare the shared random-effects model and the latent-class model, we use the PAQUID data mentioned in Section 5.2.

We use the same model in Section 6.1.2 and Section 6.1.3 to analyse the PAQUID data. For the longitudinal model, the discrete ranges of the two responses are $0-30$ and $0-15$, respectively. Based on the results in Table 6.2, under the same conditions, the AIC value of the model using the Gompertz hazard is smaller than the model with the Weibull hazard. Therefore, for the survival model part, we still only use the Gompertz hazard model.

We add the letters S and L to the model numbers to denote the shared random-effects joint models and the latent-class joint models, respectively. The rest of the model numbers are the same as that used in Tables 6.2 and 6.4. For example, Model SIG_1 represents a random intercept included in the shared random-effects joint model, and the survival model uses the Gompertz hazard model. Moreover, we provide the running time for the corresponding joint models to compare the fitting efficiency for these two types of models. Table 6.6 summarises the AIC values, estimated αs and the running times.

Although the running time of these models may be affected by the computer configuration and the number of programs running simultaneously, we could clearly see that the running speed of the latent-class model is considerably faster than the shared random-effects model. In the shared random-effects model, we use two-dimensional integration, which increases the computational challenge. Since the shared random-effects model is more directly aimed at the individual, it is worthwhile to use the shared random-effects joint model to estimate individual-specific quantities [Proust-Lima et al., 2009]. Compared to the shared random-effects model, the latent-class model does not involve any numerical integration. However, the latent-class model also has the notable drawback that it is difficult to define what the classes in the model represent.

In Table 6.6, the αs are estimated to be less than 0. It means that the risk of dementia is relatively low when the individual has a better cognitive function at the

Table 6.6: PAQUID Data: AIC and estimated αs

SREM represents the shared random-effects model; LCJM represents the latent-class joint model

	MODEL	p_{Y_j}	Link function	$-2LL$	AIC	α	α_1	α_2	Time(secs)
SREM	SIG_1		(2.29) $j=1$	12520.49	12546.49	-0.019			153292
	SIG_2	(2.10)	(2.29) $j=2$	12493.57	12519.56	-1.739			148378
	SIG_3		(2.31)	12456.43	12484.44		-0.116	-1.496	1 22811
	SSG_1		(2.35) $j=1$	12724.85	12750.85	-1.285			87332
	SSG_2	(2.11)	(2.35) $j=2$	13229.67	13255.67	-2.270			108520
	SSG_3		(2.33)	12739.06	12767.06		-0.196	-1.761	100170
LCJM	LIG_1		(2.53) $j=1$	12899.66	12927.66	-2.176			6248
	LIG_2	(6.10)	(2.53) $j=2$	12897.56	12925.56	-1.619			5640
	LIG_3		(2.54)	12961.83	12991.83		-0.834	-1.018	6188
	LSG_1		(2.55) $j=1$	13047.04	13075.04	-1.759			4939
	LSG_2	(6.11)	(2.55) $j=2$	13051.86	13079.86	-1.147			4262
	LSG_3		(2.56)	13079.62	13109.62		-0.138	-1.003	4798

baseline age; the risk of dementia is relatively low when the cognitive declines in individuals are relatively slow. Comparing the estimation in Table 6.4 with Table 6.6, we obtain different results when we apply the same model with random intercept/slope for both Y_1 and Y_2 to different data. We take the joint model with random intercepts for both Y_1 and Y_2 as an example. For the ELSA data (Model IG_3 in Table 6.4), the estimation of α_2 is positive, representing higher risk of death when the number of later recall words are relatively larger at baseline age. This result is opposite to what we expected. By contrast, for the PAQUID data (Model SIG_3 in Table 6.6), the estimations of $\hat{\alpha}$s are negative, representing a lower risk of dementia when individuals have greater test scores at baseline age. Therefore we can infer that the unexpected estimation for joint models with two α parameters is caused by the specific feature of the ELSA data.

6.2 Categorical joint models

We will now use the categorical distribution (2.12) instead of the extension of the binomial distribution (2.2) and (2.4) to construct the joint model. We consider each possible value of the response as a category. One advantage of using the categorical distribution rather than the extension of the binomial distribution is that we could model each category of the response separately.

In this section, we will first construct a univariate joint model to analyse the ELSA data, after which we will construct two different bivariate shared random-effects joint models to analyse the PBC2 data. In each subsection, we compare the categorical distribution with the extension of the binomial distribution to verify whether the categorical distribution can fit the data better than the extension of the binomial distribution.

6.2.1 Univariate shared random-effects categorical joint models

We still focus on the recall test in the ELSA data in this subsection. We refer to the ELSA data in Section 5.1, in which the test consists of remembering a list of 10 words and recalling them immediately and later. The test score is the number of words remembered by the individual. We will start by comparing the fit of the

categorical distribution and the extension of the binomial distribution. After that, we will construct the univariate categorical shared random-effects joint model.

In this univariate shared random-effects joint model, we will use the number of delayed-recall words as the response variable. When using the categorical distribution to analyse the number of delayed-recall words, we treat the 11 possible outcomes (from 0 to 10) as 11 categories. The probabilities for each category will be modelled separately.

To make the notation consistent with the previous sections, we also use Y_1 and Y_2 to denote the number of immediate-recall words and the number of delayed-recall words separately. The corresponding time scale in ELSA data is age.

6.2.1.1 Comparison of categorical distribution and binomial distribution

As mentioned in Chapter 1 and Section 6.1.1, many researchers assume that integer biomarkers follow a binomial distribution. The models proposed in these articles fit the data well. In particular, we have analysed ELSA data using the extension of the binomial distribution in Section 6.1.2 and verified in the simulation Section 4.1 that the bivariate binomial joint model fits the data well. We would like to validate whether it is worth replacing the extension of the binomial distribution with a categorical distribution before fitting the categorical shared random-effects joint model. Therefore, we will use the first interview data for each individual and compare fitting the categorical distribution (2.13) and the extension of the binomial distribution (2.2) proposed by Altham and Hankin [2012].

The first interview refers to the first observation of an individual in the study. We draw the histogram of immediate-recall words at the first interview. Data from the first interview are fitted with the categorical distribution and the extension of the binomial distribution, respectively.

Based on Figure 6.3, the histogram of categorical distribution is closer to that for the first interview data than the histogram of the extension of the binomial distribution. More precisely, the histogram of the categorical distribution is the same as the histogram of the data. This phenomenon is especially obvious when the data
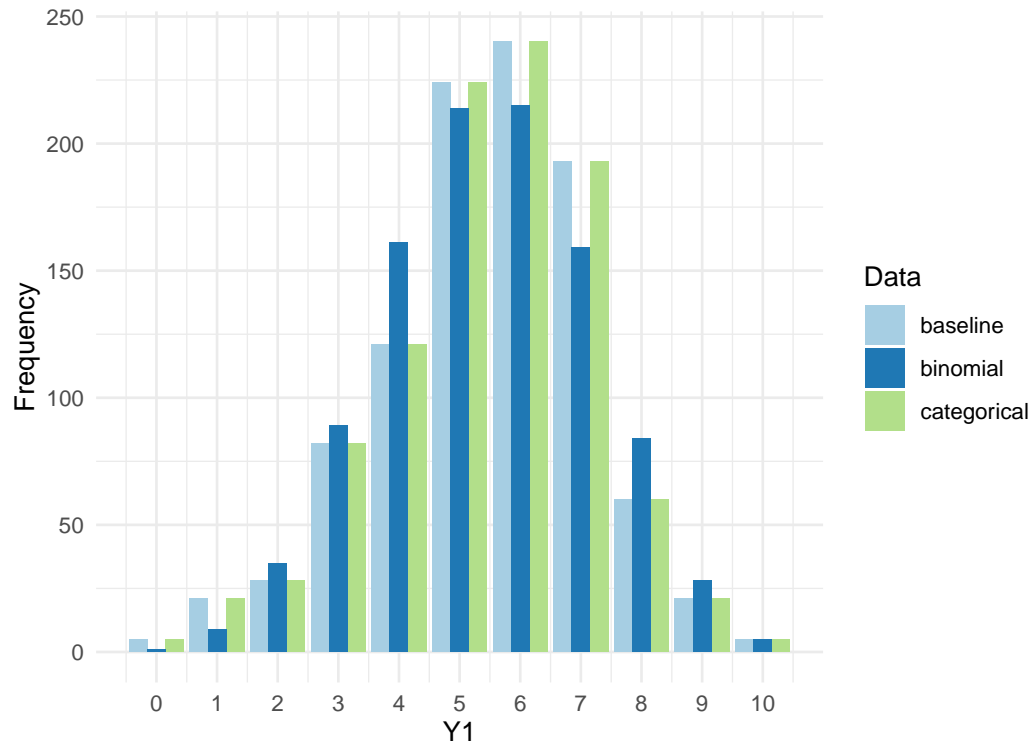
Figure 6.3: Comparison of immediate-recall words at the first interview: histogram of the first interview data (light blue), fitted extension of the binomial distribution (blue) and fitted categorical distribution (green bar).

equals 4 or 8. Therefore, we could expect a better fitting effect of the categorical distribution than the extension of binomial distribution based on this histogram.

We also draw the comparison plot for the delayed-recall words at the first interview (Figure 6.4). Compared with the distribution of immediate recall words, the distribution of delayed-recall words is more concentrated to the left. Figure 6.4 illustrates more clearly the large deviations between the extension of the binomial distribution and the first observations. In contrast, the corresponding histograms obtained from the categorical distribution are the same as those for the first observations.

Since individuals take the test for the first time at different ages, we use the linear predictors and corresponding logistic regression to describe the relationship between time (age) and the responses. The corresponding expressions are quite similar to Equations (2.7), (2.14) and (2.15). For the extension of the binomial
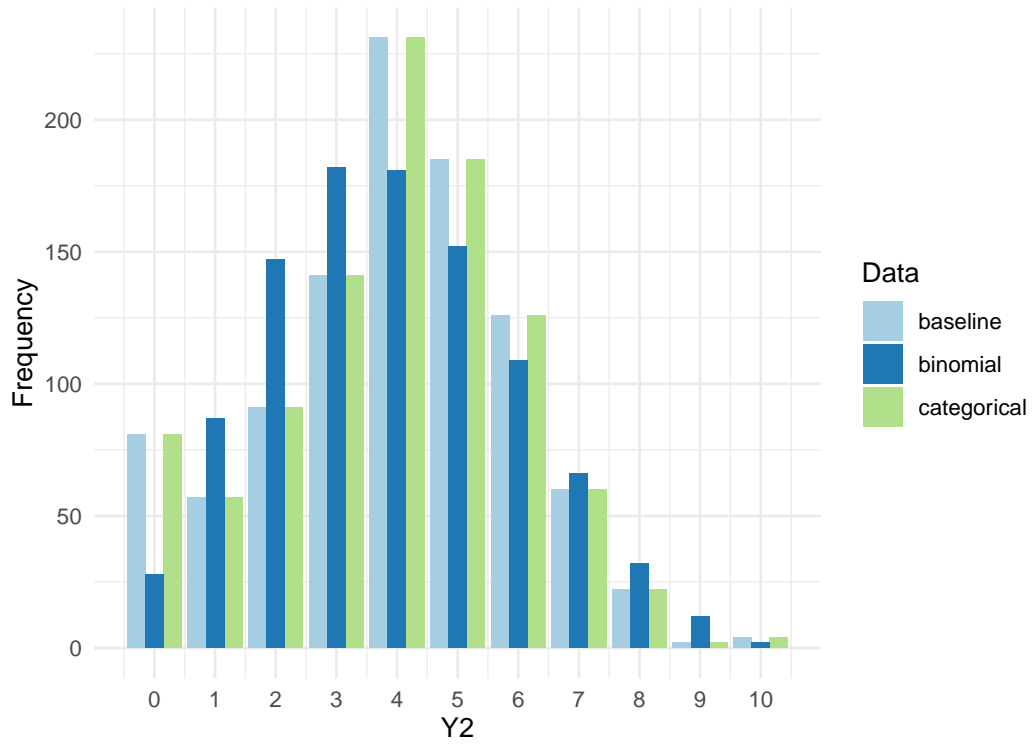
Figure 6.4: Comparison of delayed-recall words at the first interview: histogram of the first interview data (light blue), fitted extension of the binomial distribution (blue) and fitted categorical distribution (green bar)

distribution, we have:

$$\boldsymbol{\eta}^{(j)} = \eta_0^{(j)} + \eta_1^{(j)} t. \tag{6.13}$$

The corresponding probability can be written as:

$$p_{Y_j} = \frac{\exp\left(\boldsymbol{\eta}^{(j)}\right)}{1 + \exp\left(\boldsymbol{\eta}^{(j)}\right)}. \tag{6.14}$$

For the categorical distribution, let i denote the number of recalled words, where $i \in \mathbb{A}$, $\mathbb{A} = \{0, ..., 9\}$:

$$\boldsymbol{\eta}_i^{(j)} = \eta_{0i}^{(j)} + \eta_{1i}^{(j)} t. \tag{6.15}$$

Therefore, the corresponding probability can be written as:

$$p_{Y_{ji}} = \frac{\exp\left(\boldsymbol{\eta}_i^{(j)}\right)}{1 + \sum_{i \in \mathbb{A}} \exp\left(\boldsymbol{\eta}_i^{(j)}\right)}, \text{ for } i \in \mathbb{A}, \tag{6.16}$$

$$p_{Y_{j10}} = \frac{1}{1 + \sum_{i \in \mathbb{A}} \exp\left(\boldsymbol{\eta}_i^{(j)}\right)}. \tag{6.17}$$

Equations (6.13) - (6.17) are applied to fit the corresponding extension of the binomial distributions and categorical distributions with/without time effects, respectively. The models without time effects are defined by setting $\eta_1^{(j)}$ and $\eta_{1i}^{(j)}$ to 0 in Equations (6.13) - (6.17). The AIC values for the models are presented in Table 6.7.

Table 6.7: ELSA: Model comparison of binomial and categorical models, based on AIC values for models.

	Without time		
Immediate recall		Delayed recall	
Extended Binomial	Categorical	Extended Binomial	Categorical
3935.694	3903.018	4274.507	4150.822
	With time		
Immediate recall		Delayed recall	
Binomial	Categorical	Binomial	Categorical
3762.569	3713.046	4093.111	3975.365

From Table 6.7, we can observe that the AIC of the categorical distribution is smaller than that of the binomial distribution for the same response. From Table 6.7 and Figures 6.3 - 6.4, we can infer that using the categorical distribution as a longitudinal model may give a better fit than the extension of the binomial distribution. Therefore, it is reasonable to construct a joint model with the categorical distribution as an alternative to the extension of the binomial distribution.

6.2.1.2 Discussion of univariate joint models with different link functions Δ

For the extension of the binomial joint model, there is only one corresponding probability for each response. Thus we only need to discuss how to include random effects in the corresponding logistic regression model. When constructing a joint model using the categorical distribution, since we will model the probability of each response value separately, there are more options for link functions in categorical joint models than in the extension of the binomial joint models.

Taking a univariate joint distribution as an example, in the extension of the binomial distribution in the case of using logistic regression, only one probability p needs to be modelled. Whereas in the categorical distribution, since we model the probability of each category separately, we need to discuss which probability we will model using random effects across the k categories.

In this section, we use the delayed-recall number as the response variable. For longitudinal model (2.12), the probability with a random effect is written as Equations (2.14) and (2.15) for $k = 11$. Index $j = 2$, representing the response is the number of relayed-recall words. We use the Gompertz hazard as the survival model. The corresponding link function is (2.30) when the random effects are random intercepts. We assume that the random effect b follows a normal distribution $N(0, \sigma^2)$. We discuss univariate joint models given left truncation (3.4), where random effects are included in the models for the selected probabilities.

The AIC values for these models are presented in Table 6.8. We use the letter U to indicate the univariate joint model and C to indicate the categorical joint model, and the subscript indicates the corresponding categories containing random effects. We use the AIC value to select models. The selected model is Model UC_0 in which the random intercept is included in the corresponding probability of 0 words being recalled (Equation (2.14) for $i^* = 0$). The estimated parameters of this model are presented in Table 6.9.

From Table 6.9 we can observe that parameter $\hat{\alpha}$ of the link function equals to 0.201. It means that the higher the probability of an individual getting a score of 0,

Table 6.8: ELSA Data: AIC and estimated αs for the univariate categorical shared random-effects joint models
$-2LL$ represents the -2loglikelihood.

MODEL	p_{Y_j}	Link function	$-2LL$	AIC	α
UC_0	(2.14) $i^* = 0$		16738.57	16786.57	0.201
UC_1	(2.14) $i^* = 1$		16966.17	17014.17	1.726
UC_2	(2.14) $i^* = 2$		16966.17	17014.17	1.728
UC_3	(2.14) $i^* = 3$		16944.91	16992.91	0.033
UC_4	(2.14) $i^* = 4$	(2.30)	16939.02	16987.02	-0.479
UC_5	(2.14) $i^* = 5$		16936.81	16984.81	-0.413
UC_6	(2.14) $i^* = 6$		16914.24	16962.24	-0.066
UC_7	(2.14) $i^* = 7$		16921.02	16969.02	-0.158
UC_8	(2.14) $i^* = 8$		16930.61	16978.61	-0.007
UC_9	(2.14) $i^* = 9$		16966.17	17014.17	0.118

Table 6.9: ELSA: Parameter estimates for univariate categorical joint model with random intercept for the number of delayed-recall words is 0. Values in brackets are the standard error of the corresponding parameters.

|							
$\eta_{00}^{(2)}$	-3.463 (0.664)	$\eta_{10}^{(2)}$	0.215 (0.031)	$\eta_{01}^{(2)}$	-0.372 (0.505)	$\eta_{11}^{(2)}$	0.126 (0.028)
$\eta_{02}^{(2)}$	0.584 (0.482)	$\eta_{12}^{(2)}$	0.103 (0.028)	$\eta_{03}^{(2)}$	1.675 (0.465)	$\eta_{13}^{(2)}$	0.083 (0.027)
$\eta_{04}^{(2)}$	2.652 (0.458)	$\eta_{14}^{(2)}$	0.058 (0.027)	$\eta_{05}^{(2)}$	3.035 (0.458)	$\eta_{15}^{(2)}$	0.030 (0.027)
$\eta_{06}^{(2)}$	3.169 (0.459)	$\eta_{16}^{(2)}$	0.006 (0.027)	$\eta_{07}^{(2)}$	2.877 (0.467)	$\eta_{17}^{(2)}$	-0.023 (0.028)
$\eta_{08}^{(2)}$	2.095 (0.490)	$\eta_{18}^{(2)}$	-0.034 (0.030)	$\eta_{09}^{(2)}$	1.389 (0.568)	$\eta_{19}^{(2)}$	-0.078 (0.038)
σ	2.458 (0.894)						
ξ	0.119 (0.087)						
β	-5.773 (0.303)	α	0.201 (0.047)				

i.e. the higher the probability of not recalling any word, the higher the risk of death for him/her. This conclusion is consistent with our expectation that individuals with poorer cognitive abilities may have a higher risk of death. Estimates of fixed intercepts $\eta_{0i}^{(2)}$, $i \in \{0, \ldots, 10\}$ are also worth noting. At baseline age ($t = 0$) conditional on random effects being equal to its mean 0 ($b = 0$), an individual's probability of remembering i words increases and then decreases as i increases. For example, for category 0, i.e. the number of words recalled is 0, we have $\hat{\eta}_{00}^{(2)} = -3.463$ and the corresponding probability equals to 0.0003; for category 6, the corresponding probability increases to 0.244 with $\eta_{06}^{(2)} = 3.169$; after that the probability continues to decrease and drops to 0.010 at $i = 10$.

Moreover, estimates of fixed slopes $\eta_{1i}^{(2)}$ decrease as the number of recalled words i increases. To give an example, for $i = 0$, $\hat{\eta}_{10}^{(2)} = 0.215$ is greater than 0, which leads to an increase in the probability of the corresponding category with age. It means that the probability of an individual not being able to remember any word increases with age. For $i = 9$, $\hat{\eta}_{19}^{(2)} = -0.078$ is negative, which means that the probability of an individual being able to remember 9 words decreases with age. We can infer from the analysis above that as people get older, their ability to recall words will decrease.

We have represented all relevant probabilities in Figure 6.5. In this figure, we plot a comparison of the probabilities of different categories when the random effect takes the value of its expectation. We have calculated three corresponding probabilities for each category around ages 50, 70 and 90, respectively.

From Figure 6.5, we can see that the corresponding probabilities for each category increase and then decrease when $b = 0$ and the age of the individuals is not that large, i.e. 50 and 70. We can conclude that at baseline age given $b = 0$, the probability of an individual remembering 0 words is smaller than the probability of remembering 1 word. Among all possible categories, an individual has the highest probability of scoring 6 when he/she is 50 years old, i.e. remembering 6 words at baseline age given $b = 0$. When the individual's age increased to 70 years, his/her most likely score decreased to 4. As age increases further, reaching 90 years, we can
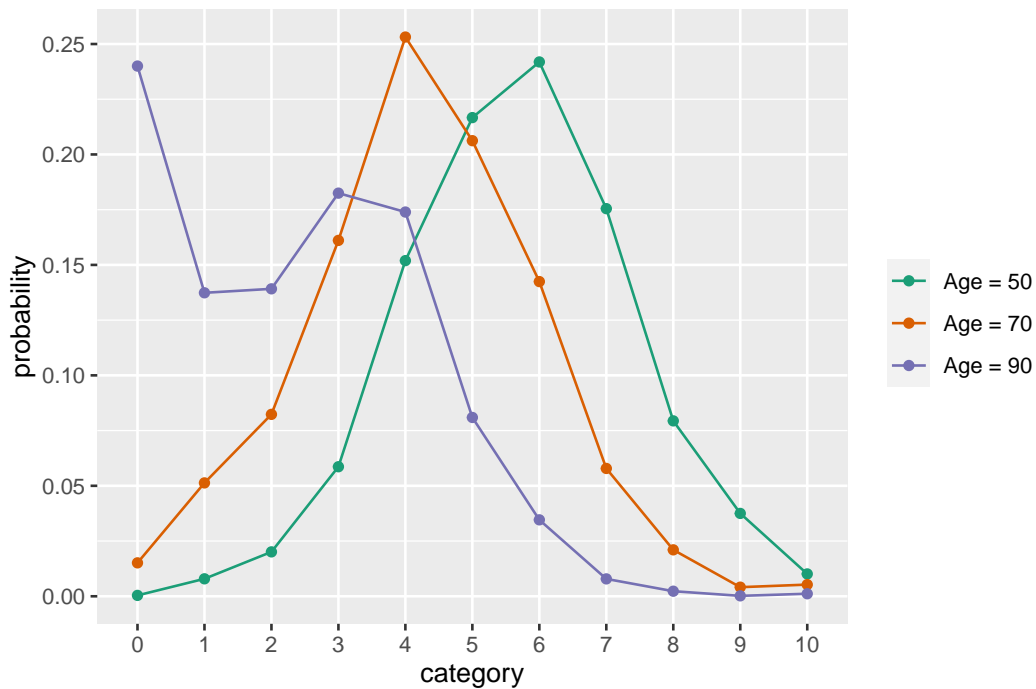
Figure 6.5: Comparison of probabilities between different categories: given that random effects *b*s equal to the mean, i.e. 0.

observe an inevitable decline in the individuals' long-term memory ability. When they are asked to recall some words they saw a few minutes ago, there is a high probability that they will not be able to recall any of them. These conclusions are in line with our expectations. Individuals' memory declines with age.

We also plot the distribution of the observed data versus the estimated probabilities, see Figure 6.6. Although the corresponding bars and lines do not overlap precisely, we can see a consistent trend in the distribution of the observed data and the corresponding estimated probabilities with increasing age. For each category in the figure, we have a limited number of observations, which may be one of the reasons cause the difference between the histogram and the line.

6.2.2   Bivariate shared random-effects categorical joint models

We now extend the univariate categorical joint model into a bivariate categorical joint model. Two variables mentioned in Section 5.3 from the PBC2 data are treated as response variables in this section: histologic stages of liver biopsy Y_1 and edema conditions Y_2.
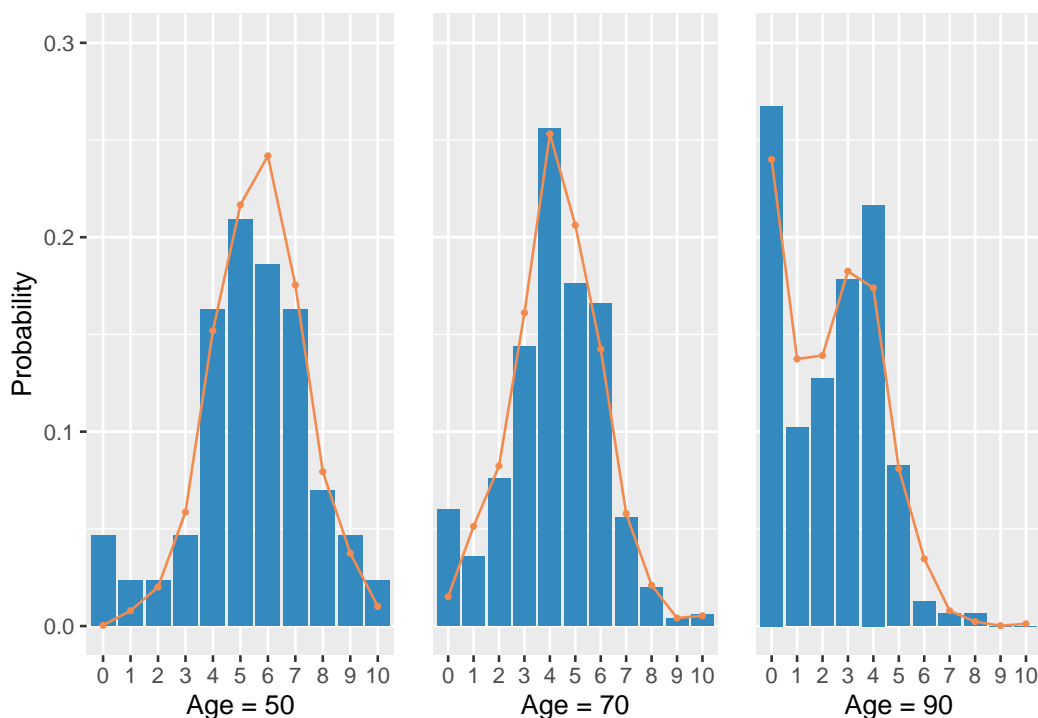
Figure 6.6: Comparison between estimated probabilities and data distributions with increasing age: given that random effects *b*s equal to the mean, i.e. 0. The histograms represent the distribution of the corresponding data and the line represent the estimated probabilities of the different categories.

In the PBC2 data, time is defined as the number of years between the start of the study and the date of the current visit, so we use the marginal likelihood function (3.1) when constructing the corresponding joint model, i.e. the marginal likelihood that does not take into account the left truncation.

In this section, we will first compare the fit of the categorical distribution (2.12) with that of the extension of the binomial distribution (2.2), as we did in Section 6.2.1, to show that the categorical distribution fits the data better than the extension of the binomial distribution.

After that, we will construct two types of bivariate joint models based on the categorical distribution. We use two univariate categorical distributions in the first joint model to model the corresponding responses separately. In contrast to Section 6.1, the correlation between the two outcomes in the first joint model is captured by the random effects rather than measured by the longitudinal model. This method of capturing correlation between outcomes is the same as that proposed for the

multivariate joint model in Rizopoulos [2012]. We construct the two responses in the second joint model as a table. Using the PBC2 data as an example, the two responses in this data actually define a table. We have converted two univariate categorical models with 4 and 3 categories into a categorical model with 12 categories. We will explain this approach in detail in Section 6.2.2.3.

6.2.2.1 Comparison of categorical distribution and the extended binomial distribution

Before fitting the bivariate model, we compare the fit effect of the categorical distribution (2.13) with the extension of the bivariate distribution (2.2) proposed by Altham and Hankin [2012], as in Section 6.2.1.

We still choose the first interview data as the sample to test the fit of the two distributions. We draw the histogram of histologic stages at the first interview. The first interview data are fitted with the extension of the binomial distribution and the categorical distribution, respectively.

As illustrated by Figure 6.7, the histogram for the categorical distribution is the same as the histogram for the first interview data, while the histogram for the extended binomial distribution deviates from the first interview data. The reason is that the categorical distribution allows us to model the probabilities of each category separately. Therefore, the fit of the categorical distribution is better than that of the extension of the binomial distribution.

We also draw the comparison plot for the edema conditions at the first interview (Figure 6.8). Again, the categorical distribution is the same as the histogram of the data.

The AIC values for the above four models are presented in Table 6.10. With this table, we can observe how much the extended binomial distribution deviates from the categorical distribution. Furthermore, since in the categorical distribution, we model each category separately, and the histogram of the first interview data and the histogram of the categorical data in Figures 6.7 and 6.8 are the same, we can assume that the categorical distribution is the distribution that can perfectly first observations for this non-negative longitudinal data. Therefore we can also say
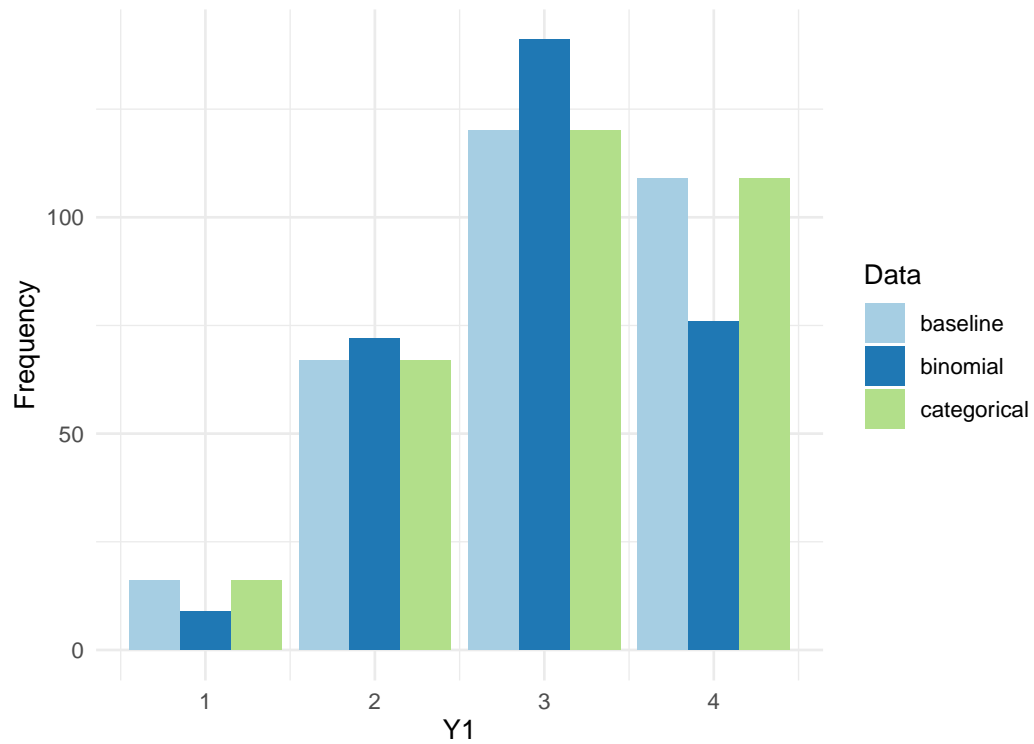
Figure 6.7: Comparison of the histologic stages at the first interview: histogram of the first interview data (light blue), fitted extension of the binomial distribution (blue) and fitted categorical distribution (green).

that this table shows how much deviation there is between the extended binomial distribution and the best model.

It is clear that for the first observations, the categorical distribution has a better fit than the extension of the binomial distribution. Therefore, for the PBC2 data, we assume that using categorical distribution to construct a bivariate joint model can improve the fit effect of the model.

Table 6.10: PBC2: Model comparison of binomial and categorical models, using first observations, and based on AIC values for models.

Immediate recall		Delayed recall	
Extended Binomial	Categorical	Extended Binomial	Categorical
809.3079	765.7705	537.0429	405.1175
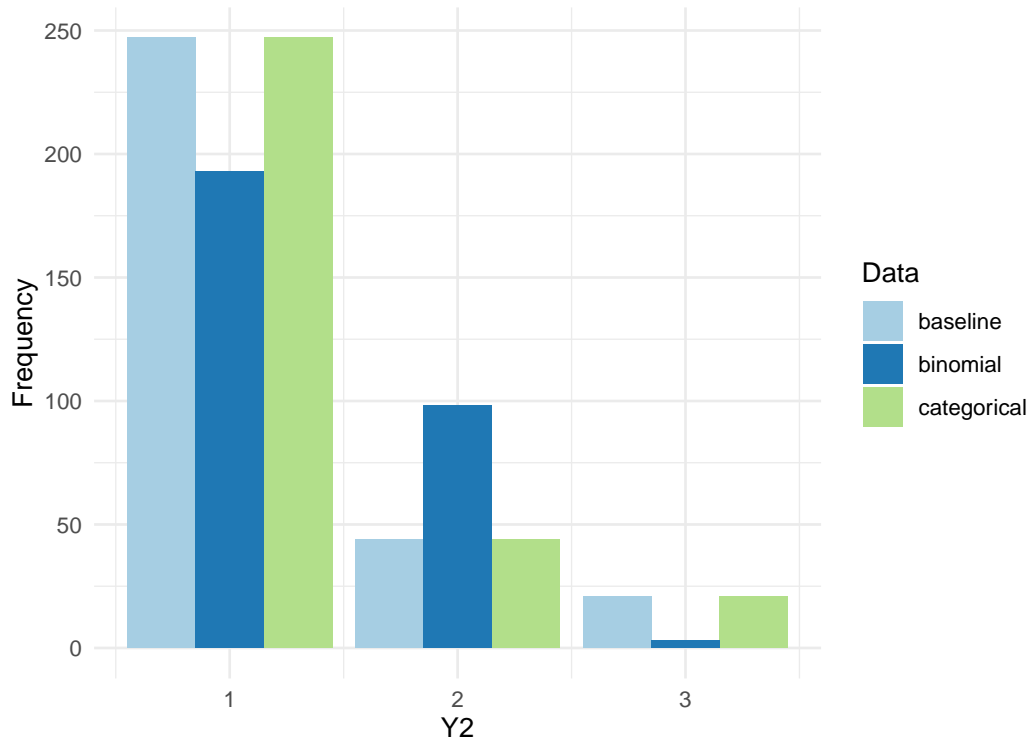
Figure 6.8: Comparison of the edema conditions at the first interview: histogram of the first interview data (light blue), fitted extension of the binomial distribution (blue) and fitted categorical distribution (green bar).

6.2.2.2 Discussion of bivariate joint models with different link functions Δ

In this section, we use two variables mentioned in Section 5.3 from the PBC2 dataset as responses: histologic stages of liver biopsy Y_1 and edema conditions Y_2. For these two response variables, their longitudinal distribution (2.12) has k equal to 4 and 3, respectively. The probability with random effects is written as Equation (2.14 - 2.15) for $\mathbb{A}_1 = \{2,3,4\}$ and $\mathbb{A}_2 = \{2,3\}$. We use the exponential hazard, the Weibull hazard and the Gompertz hazard as survival models for joint models. We assume that in each joint model, there are two correlated random effects, which follow a bivariate normal distribution, i.e. Equation (2.59):

$$\boldsymbol{b} \sim \left(\boldsymbol{0}, \begin{pmatrix} \sigma_{b1}^2 & \rho\,\sigma_{b1}\sigma_{b2} \\ \rho\,\sigma_{b1}\sigma_{b2} & \sigma_{b2}^2 \end{pmatrix} \right).$$

The corresponding link function is (2.32) or (2.34) when the random effects are random intercept and random slope respectively. For these bivariate categorical joint models, the correlation between the two responses is captured by random effects. It is worth noting that for the PBC2 data, since the time t for each individual starts at 0, we do not need to consider left truncation.

We have already discussed in Section 6.2.1.2 that there are more possible link functions for the categorical joint model than for the extended binomial joint model. In a bivariate categorical joint model, the possible scenarios are more complex. The main purpose of this thesis is to illustrate the method rather than an in-depth clinical-relevant study of the dataset. Therefore, we select a few possible scenarios and analyse potential models.

We use I to denote the joint model containing random intercept in the link function, i.e. link function equals to (2.32). In this case for the longitudinal model we restrict $b_{11} = b_{21} = 0$ (Equation (2.19)). We use S to represent models with the random slope in the link function. In this situation the corresponding link function is equal to Equation (2.34), and the corresponding longitudinal are restricted to Equation (2.20). The model named after *PROB* means that the corresponding probabilities are included in the link function (Equation (2.40)). The model with *sex* in the name means that we have used individuals' gender as the covariate in this joint model. The corresponding logistic regression models are given by:

$$p_{Y_{ji^*}}(t) \;=\; \frac{\exp\left(\boldsymbol{\eta}_{i^*}^{(j)} + \gamma_L^{(j)}sex\right)}{1 + \exp\left(\boldsymbol{\eta}_{i^*}^{(j)} + \gamma_L^{(j)}sex\right) + \sum_{i \in \mathbb{A}_j \setminus \{i^*\}} \exp\left(\eta_{0i}^{(j)} + \eta_{1i}^{(j)}t\right)}, \quad (6.18)$$

$$p_{Y_{jl}}(t) \;=\; \frac{1}{1 + \exp\left(\boldsymbol{\eta}_{i^*}^{(j)} + \gamma_L^{(j)}sex\right) + \sum_{i \in \mathbb{A}_j \setminus \{i^*\}} \exp\left(\eta_{0i}^{(j)} + \eta_{1i}^{(j)}t\right)}. \quad (6.19)$$

The gender variable is also added to the link function $\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, \gamma_S, t)$. Taking Equa-

tion (2.31) as an example, in this section we write Equation (2.31) as:

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, \gamma_S, t) = \sum_{j=1}^{2} \alpha_j \left(\eta_0^{(j)} + b_{j0} \right) + \gamma_S sex, \tag{6.20}$$

where $\gamma_L^{(j)}$ and γ_S are the corresponding parameters of the covariates defined in Chapter 2. The reason we include covariates in only one of the categories is similar to the reason we include random effects in only one of the categories. We assume that it is possible for a category that is relatively important across all categories. That category is something we want to investigate. We will investigate this choice later in this section.

The letters *G*, *E* and *W* denote models containing the Gompertz hazard model, the exponential hazard model and the Weibull hazard model, respectively. The two numbers in the model name indicate the number of categories where the random effect is included, i.e. the values of i^* in Equation (2.14). For example, IG_{12} means that the joint model uses the Gompertz hazard model with two random intercepts; the first random intercept is included in the probability of the first category of Y_1, and the second random intercept is included in the second category probability of Y_2.

The AIC values for these models are presented in Table 6.11 and Table 6.12. By comparing the AIC values of Model IG_{32} and Model $IG_{32}.sex$, we can see that using *sex* as a covariate can make the model fit better to some extent. When we use random intercepts to model the same categories, changing the type of hazard model does not significantly change the AIC values. For example, the differences in the corresponding AIC values for models IG_{32}, IW_{32} and IE_{32} are quite small.

The model that has the smallest AIC value in Table 6.11 is IG_{42}, i.e. the joint model uses the Gompertz hazard model, where the random intercept of the first response is included in the corresponding probability of stage 4 for histologic stages of liver biopsy and the probability of stage 2 for edema conditions. This result is to some extent in line with our expectations, as both stage 4 of Y_1 and stage 2 of Y_2 are stages where disease symptoms are pronounced, and we assume that more

Table 6.11: PBC2 Data: AIC and estimated αs for bivariate categorical shared random-effects joint models.
With random intercepts included in link function.
$-2LL$ represents the -2loglikelihood.

MODEL	py_j	Link function	Hazard distribution	$-2LL$	AIC	α_1	α_2	ρ
IG_{22}	(2.32)	Gompertz	7413.805	7447.805	-0.174	0.190	-0.279
IG_{23}		Gompertz	7458.399	7492.399	-0.089	0.463	-0.283
IG_{31}		Gompertz	6938.313	6972.313	-0.109	-0.304	0.208
IG_{321}	(2.30) $j=1$	Gompertz	7383.954	7415.954	-0.163		-0.199
IG_{322}	(2.30) $j=2$	Gompertz	7364.652	7396.652		0.249	-0.166
IG_{32}		Gompertz	7359.829	7393.829	-0.144	0.261	-0.202
$IG_{32.sex}$	(2.32)	Gompertz	7351.479	7387.479	-0.125	0.270	-0.203
$IG_{32.PROB}$	(2.40)	Gompertz	7390.473	7424.473	-0.221	0.717	-2.716
IG_{33}	(2.19)	Gompertz	7408.944	7442.944	-0.037	0.475	-0.297
IG_{42}		Gompertz	6725.215	6759.215	0.116	0.186	0.468
IG_{43}		Gompertz	6741.243	6775.243	0.033	0.456	0.726
IW_{32}		Weibull	7360.226	7394.226	-0.144	0.257	-0.201
IW_{42}	(2.32)	Weibull	6730.435	6764.435	0.114	0.251	0.655
IW_{43}		Weibull	6743.426	6777.426	0.037	0.459	0.726
IE_{32}		Exponential	7364.091	7396.091	-0.130	0.236	-0.197
IE_{42}		Exponential	6736.962	6768.962	0.113	0.179	0.572
IE_{43}		Exponential	6778.924	6810.924	0.029	0.324	0.726

Table 6.12: PBC2 Data: AIC and estimated αs for bivariate categorical shared random-effects joint models. With random slopes included in link function. $-2LL$ represents the -2loglikelihood.

MODEL	py_j	Link function	Hazard distribution	$-2LL$	AIC	α_1	α_2	ρ
SG_{32}			Gompertz	7494.949	7528.949	-0.057	0.180	0.369
SG_{43}			Gompertz	7169.757	7203.757	0.026	0.331	0.373
SE_{32}	(2.20)	(2.34)	Exponential	7494.949	7526.949	-0.057	0.180	0.369
SE_{42}			Exponential	7092.354	7124.354	0.050	0.080	0.335
SE_{43}			Exponential	7169.757	7201.757	0.026	0.331	-3.048

accurate model estimates will be obtained when the corresponding random effects are included in these stages. We will analyse the estimation results of this model in detail later. The estimated parameters of this model are presented in Table 6.13.

By observing the AIC values of the models in Table 6.12, we could obtain similar conclusions to those in Table 6.11: when we include the random effects into the same categories, the corresponding AIC values, as well as parameter estimates, are quite similar whether we use the Gompertz hazard or the exponential hazard. Moreover, when we use the Gompertz hazard to build the model, including random slopes, the corresponding parameter ξ is estimated to be 0. Although this is inconsistent with our defined parameter ξ in Section 2, Van Den Hout and Muniz-Terrera [2016] have mentioned that when ξ is equal to 0, the corresponding Gompertz hazard model is equal to the exponential hazard model. We infer from the above that in the joint model containing random slopes, the longitudinal part of the model already captures most of the time dependence and adding parameter ξ does not affect the fit effect of the joint model.

The model with the lowest AIC value in Table 6.12 is SE_{42}. This joint model uses the exponential hazard model in which the random slope of the first response Y_1 is included in the corresponding probability of category 4 and the random slope of the second response Y_2 is included in the corresponding probability of category 2. Estimated parameters for this model are shown in Table 6.14.

Table 6.13: PBC2: Parameter estimates for Model IG_{42}. Values in brackets are the standard error of the corresponding parameters.

$\eta_{02}^{(1)}$	1.170 (0.162)	$\eta_{12}^{(1)}$	-0.066 (0.049)	$\eta_{03}^{(1)}$	1.748 (0.151)	$\eta_{13}^{(1)}$	0.046 (0.044)
$\eta_{04}^{(1)}$	1.083 (0.221)	$\eta_{14}^{(1)}$	0.799 (0.071)				
$\eta_{02}^{(2)}$	-2.350 (0.169)	$\eta_{12}^{(2)}$	0.182 (0.028)	$\eta_{03}^{(2)}$	-2.710 (0.157)	$\eta_{13}^{(2)}$	0.157 (0.024)
σ_{b1}	6.934 (0.713)	σ_{b2}	1.779 (0.076)				
ρ	0.468 (0.019)	ξ	0.089 (0.326)				
β	-2.786 (0.219)	α_1	0.116 (0.022)	α_2	0.186 (0.066)		

From Table 6.13 we can observe that parameter $\hat{\alpha}_1 = 0.116$ and $\hat{\alpha}_2 = 0.186$

Table 6.14: PBC2: Parameter estimates for Model SE_{42}. Values in brackets are the standard error of the corresponding parameters.

$\eta_{02}^{(1)}$	1.149 (0.156)	$\eta_{12}^{(1)}$	-0.056 (0.045)	$\eta_{03}^{(1)}$	1.768 (0.145)	$\eta_{13}^{(1)}$	0.038 (0.040)
$\eta_{04}^{(1)}$	1.180 (0.160)	$\eta_{14}^{(1)}$	0.883 (0.079)				
$\eta_{02}^{(2)}$	-1.712 (0.098)	$\eta_{12}^{(2)}$	0.139 (0.028)	$\eta_{03}^{(2)}$	-2.695 (0.131)	$\eta_{13}^{(2)}$	0.153 (0.024)
σ_{b1}	1.675 (0.133)	σ_{b2}	0.626 (0.083)				
ρ	0.335 (0.025)						
β	-2.860 (0.110)	α_1	0.050 (0.014)	α_2	0.080 (0.036)		

of the link function are both positive. This means that the higher the probability of an individual being in stage 4 of Y_1 and stage 2 of Y_2 at the baseline, the higher the probability of mortality.

Estimates of fixed intercept $\eta_{03}^{(1)}$ is the largest estimated intercept in Y_1, and based on this result we could calculate the probability of the corresponding stage 3 at the beginning of the study ($t = 0$), i.e. $p_{Y_{13}}(0) = 0.445$ given $b = 0$. This phenomenon indicates that at the start of the study individuals have a nearly 50% probability of being in histologic stage 3 of liver biopsy. For Y_2, we could use the same method to calculate the probability of the corresponding category at the beginning of the research equal to 0.861, 0.082 and 0.057, respectively.

The probability of an individual being at stage 1 is highest at $t = 0$, meaning that the majority of individuals do not have symptoms of edema at the beginning of the research. Estimates of fixed slopes $\eta_{1i}^{(j)}$ increases as the stage increases. For histologic stages of liver biopsy Y_1, $\hat{\eta}_{12}^{(1)} = -0.066$, the probability of the patient remaining at stage 2 decreases as time increases. The probability of the patient's condition being at stage 3 and stage 4 increases with time. The estimated $\eta_{14}^{(1)} = 0.799$ is much larger than $\hat{\eta}_{13}^{(1)} = 0.046$, which means that the probability of a patient's condition deteriorating to stage 4 is higher than the probability of deterioration to stage 3 per unit time. For edema conditions Y_2, we can obtain similar conclusions: The probability of the patient's condition being at stage 2 and stage 3 increases with time, and the risk of reaching stage 3 increases faster per unit time than stage 2.

By analysing the results of the parameter estimates in Table 6.14, we can draw similar conclusions to the analysis in Table 6.13. The probability of an individual being in the histologic stage of liver biopsy $1 - 4$ at the beginning of the experiment is 0.075, 0.238, 0.442 and 0.245, respectively. The probability of an individual not showing signs of edema is 0.801.

We plot line charts of the corresponding probabilities over time for each category of the two responses given $b = 0$ (Figure 6.9). The histological stage of the liver biopsy Y_1, stages of the individuals at the beginning of the study ($t = 0$) are relatively dispersed. The probability of being in stages 2, 3 and 4 is relatively even. Two years after the start of the study ($t = 2$), the probability of being in the 1, 2 and 3 stages decreased rapidly. The probability of an individual being in stage 4 increased from 25% to approximately 65%. The probability of an individual being in the histologic stage 3 of liver biopsy decreases from 0.442 to less than 0.25. After that, the probability decreases rapidly to close to 0. From the fourth year of the research, individuals have a quite high probability (close to 1) of progressing to stage 4. This change in probability is consistent with what we mentioned in Chapter 5, i.e., cirrhosis is incurable, and if the patient is in a more advanced histological stage, the disease will worsen over time [Mayo, 2022].

For the second response Y_2, at the beginning of the study, there is a high probability that the patient does not experience edema (nearly 80%). However, the probability of having edema gradually increases over time, and there is a slight increase in the probability of severe edema. In the 8th year of the study, the probability of patients having no edema decreased from 80% to 55%. In contrast, the probability of having edema doubled (from 15% to 30%) and the probability of patients having severe edema, i.e., failure to eliminate it with diuretics, increased from 5% to 15%.

6.2.2.3 Bivariate categorical joint models based on a univariate categorical distribution in table form

As we mentioned at the beginning of this subsection, we will represent the two responses as a table and construct the corresponding joint model using a univariate categorical distribution. We are using PBC2 data in this subsection, and the two
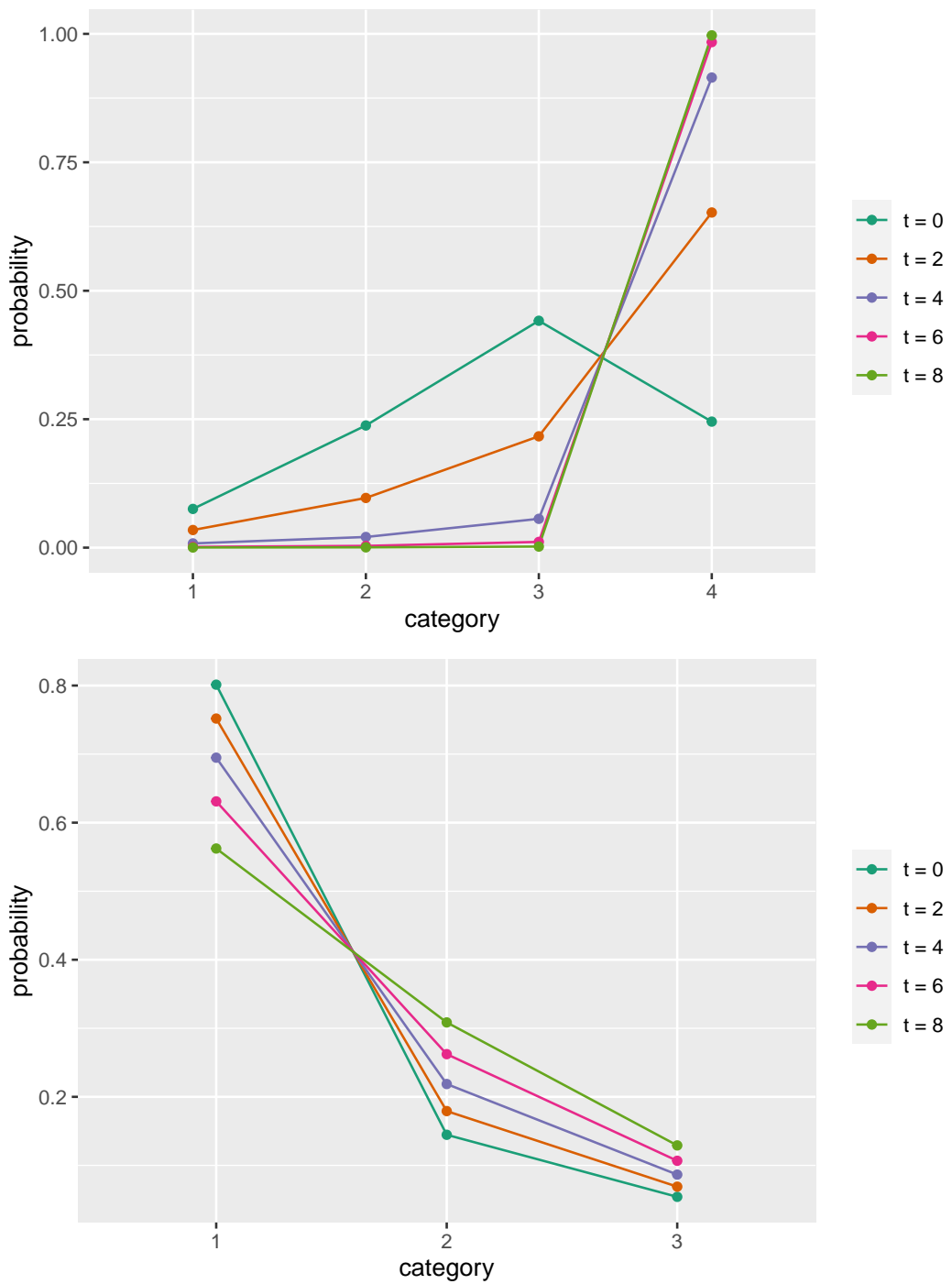
Figure 6.9: Line graph of probability for change over time for each category. The first graph: the histologic stages of liver biopsy Y_1 trajectories. The second graph: the edema conditions Y_2 trajectories.

responses in PBC2 have four and three categories, respectively. Therefore 12 categories will be present in the matrix form we have constructed. The corresponding probabilities for twelve categories are listed in Table 6.15.

Table 6.15: PBC2: Probabilities for the corresponding 12 categories.

		Edema conditions		
		1	2	3
Histologic stages of liver biopsy	1	p_{11}	p_{12}	p_{13}
	2	p_{21}	p_{22}	p_{23}
	3	p_{31}	p_{32}	p_{33}
	4	p_{41}	p_{42}	p_{43}

Each probability in Table 6.15 contains two numbers, the first representing the histology stage of liver biopsy and the second representing the edema condition. For example, p_{21} represents the probability that an individual is in histologic stage 2 of liver biopsy and has no edema. In this section, we will build categorical joint models based on the 12 categories in Table 6.15.

We set p_{11} as the probability of baseline category. Similar to Equation (2.14) and Equation (2.15), we use $p_{i^*}(t)$ to represent the logistic regression of the corresponding categorical probabilities, where i^* represents the number of the category index which has corresponding random effects.

Suppose the set $\mathbb{A}$ is the set of all possible category indices except for the baseline category. The categories in Table 6.15 are the elements contained in the set $\mathbb{A}$, i.e. $\mathbb{A} \in \{12, 13, 21, 22, 23, 31, 32, 33, 41, 42, 43\}$. We have:

$$p_{i^*}(t) = \frac{\exp(\boldsymbol{\eta}_{i^*})}{1 + \exp(\boldsymbol{\eta}_{i^*}) + \sum_{i \in \mathbb{A} \setminus \{i^*\}} \exp(\eta_{0i} + \eta_{1i}t)}, \qquad (6.21)$$

where $\boldsymbol{\eta}_{i^*} = (\eta_{0i^*} + b_0) + (\eta_{1i^*} + b_1)t$ represents the linear predictor corresponding to category i^* and response Y_j. We set category 11 as the baseline category, where

$$p_{11}(t) = \frac{1}{1 + \exp(\boldsymbol{\eta}_{i^*}) + \sum_{i \in \mathbb{A} \setminus \{i^*\}} \exp(\eta_{0i} + \eta_{1i}t)}. \qquad (6.22)$$

The corresponding link function is

$$\Delta(\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{b}, t) = \alpha_1 \left(\eta_{0i*} + b_0 \right) + \alpha_2 \left(\eta_{1i*} + b_1 \right) t. \tag{6.23}$$

However, the joint models we built based on the above set $\mathbb{A}$ and Equations (6.21) - (6.23) encounter problems in estimating the parameters associated with p_{13}. The standard error of the parameters associated with p_{13}, i.e. the corresponding fixed intercept and fixed slope, is `NaN`. According to Table 5.1, we can see that no observation corresponds to p_{13} in PBC2, which might be one of the reasons for this parameter estimation problem. The empty observation on p_{13} might be a sampling zero or a structural zero [Agresti, 2003]. During our review of information on PBC, we learned that edema is a symptom of advanced cirrhosis Mayo [2022]. We presume that in the early stages of the disease, the patient will not develop severe oedema that diuretics cannot eliminate. Therefore, we assume the missing observation for p_{13} is structural zero. To deal with the structural zero, we set the p_{13} probability to 0 in the model. The elements contained in the set $\mathbb{A}$ modified to $\{12, 21, 22, 23, 31, 32, 33, 41, 42, 43\}$.

We propose one potential way to verify whether the empty observation of category 13 is sampling zero or structural zero. We have verified in Section 4.4 that our code can generate data from the joint model we set up and that the bivariate categorical joint model estimates are accurate. We could use the results obtained by Section 6.2 as true values to generate a dataset large enough to represent PBC2. By verifying the existence of observations of category 13 in this dataset, we may be able to know the type of this empty observation. There are two main problems with this approach. Firstly, we cannot guarantee that the PBC2 dataset with 312 individuals captures the main information about the PBC patient population. It means that even if our joint model describes the PBC2 data properly, there may be a bias between the dataset we generate from the joint model and the real data. Secondly, we generate data randomly from the joint model; even if the empty observation of category 13 is structural zero, we may still get the corresponding observation in the generated data. Therefore, we need to develop criteria for determining the outcome

as structural zero/sampling zero. We will not validate this method in this thesis.

We present in Table 6.16 the AIC values of the constructed joint model and the estimated values of the parameter αs. The letters G and E represent the Gompertz hazard and the exponential hazard in the joint model, respectively. The numbers in the model names represent the category containing the random effects. For example model G_{12} represents a joint model using Gompertz hazard as the survival model and $i^* = 12$ in Equation (6.21) - (6.23).

We choose the model with the smallest AIC value in Table 6.16 to analyse, i.e. Model E_{43}. The parameter estimation results for Model E_{43} are presented in Table 6.17. From Table 6.13 we can observe that parameter related to random intercept of the link function $\hat{\alpha}_1 = 0.246$ is positive. This positive $\hat{\alpha}_1$ means that the higher the probability of category 43, the higher the probability of death of the individual. Parameter related random slope of the link function $\hat{\alpha}_2 = 0.195$ is also positive. The estimated $\hat{\alpha}_2$ means that when time increases, the risk of death for the individual increases.

Estimates of fixed intercept $\hat{\eta}_{031}$ is the largest estimated intercept, and based on this result we could calculate the corresponding probability at the beginning of the study ($t = 0$), i.e. $p_{31}(0) = 0.321$ given random effects $b = 0$. This largest intercept means that at the start of the study, individuals have 30% probability of being in histologic stage 3 of liver biopsy and without edema. The estimated value of fixed slope η_{142} is the largest among all fixed slopes, which means that as time increases, the corresponding probability of an individual being in category 42 becomes higher. Probability p_{42} increases per unit of time the fastest among all categories.

We plot line charts of the corresponding probabilities over time for each category given $b = 0$ (Figure 6.10). From Figure 6.10 we can observe that when $t = 0$, the probability of an individual being in category 31 is highest (32%), and the probability of an individual being in category 41 is also high (27%). As time increases, the probability of an individual being in category 21 and category 31 continues to decrease, with the probability of category 21 decreasing from 18% at $t = 0$ to 6% at $t = 8$, and the probability of category 31 decreasing by half at $t = 8$ compared to

Table 6.16: PBC2 Data: AIC and estimated αs for categorical shared random-effects joint models with 12 categories −2LL represents the -2loglikelihood.

MODEL	p_{Y_j}	Link function	Hazard distribution	−2LL	AIC	α_1	α_2	ρ
G_{32}			Gompertz	7944.351	7998.351	0.038	-0.057	-0.507
G_{42}			Gompertz	7780.718	7834.718	0.246	0.195	0.143
G_{43}	(6.21)	(6.23)	Gompertz	7739.571	7793.571	0.307	0.220	0.680
E_{32}			Exponential	7944.528	7996.528	0.028	-0.094	-0.502
E_{42}			Exponential	7780.718	7832.718	0.246	0.195	0.143
E_{43}			Exponential	7730.429	7782.429	0.332	0.264	0.137

Table 6.17: PBC2: Parameter estimates for Model E_{43} with 12 categories. Values in brackets are the standard error of the corresponding parameters.

η_{012}	-2.889 (0.622)	η_{112}	-0.001 (0.178)	η_{021}	1.074 (0.167)	η_{121}	-0.054 (0.049)
η_{022}	-0.645 (0.273)	η_{122}	-0.296 (0.129)	η_{023}	-3.598 (0.829)	η_{123}	0.075 (0.202)
η_{031}	1.643 (0.156)	η_{131}	0.005 (0.044)	η_{032}	-0.048 (0.198)	η_{132}	0.083 (0.052)
η_{033}	-1.856 (0.322)	η_{133}	0.254 (0.063)	η_{041}	1.478 (0.155)	η_{141}	0.146 (0.043)
η_{042}	-1.597 (0.365)	η_{142}	0.377 (0.064)	η_{043}	-0.379 (0.203)	η_{143}	0.221 (0.048)
σ_{b_0}	2.307 (0.224)	σ_{b_1}	0.471 (0.098)				
ρ	0.143 (0.182)	β	-2.649 (0.158)				
α_1	0.246 (0.057)	α_2	0.195 (0.061)				

Figure 6.10: Line graph of probability for change over time for each category for Model E_{43} in Table 6.17.

$t = 0$. In contrast, the probability of an individual being in category 41 continues to increase, from 27% ($t = 0$) to 42% ($t = 8$). The probability of an individual being in category 42 and category 43 also increases with time. The trend in these categories related to stage 4 for the histologic of liver biopsy illustrates that the probability of individuals' cirrhosis progressing towards an advanced stage is increasing over time. This trend is consistent with the information we obtained in Mayo [2022] that there is no cure for cirrhosis.

Moreover, the probabilities associated with the category of patients being in the early histologic of the liver biopsy but with edema are small, especially for category 12 and category 23, which are consistently close to 0. This trend also confirms the high probability of individuals not showing symptoms of edema when they are in the early histological stage, in line with our discussion in the previous paragraph.

The advantages of the model we construct in this section over the bivariate joint model constructed in previous sections using two categorical distributions are:

1. The dimensionality of the joint model is reduced. We convert the bivariate joint model to a one-dimensional joint model by pre-processing the data for the model (using two response variables to construct a table).

2. This model allows us to analyse the relationship between the two responses directly rather than capturing the relationship between them via random effects.

3. The presence of structural zero in the data can be taken into account explicitly.

This model also has some drawbacks. Even though we can use this method to analyse data with more than two corresponding variables, the addition of more response variables increases the number of parameters in the joint model rapidly. Moreover, when we use this method to analyse multivariate data, we have a high probability of encountering empty observations in some categories. It is difficult to distinguish whether the empty observation for a category consisting of multiple responses is a structure zero or a sampling zero, and it is also difficult to analyse which responses jointly lead to the empty observation. Therefore, this method is not suitable for

analysing multivariate data. In contrast, a joint model using two categorical distributions can capture more responses by increasing the number of categorical distributions.

Chapter 7

Conclusion

In research of longitudinal and time-to-event data, the simultaneous measurement of trends in longitudinal outcomes over time and the time of occurrence of the event is an important research direction. Many responses for longitudinal data in medical statistics are recorded as discrete data. Furthermore, it is common for researchers to measure more than one biomarker in longitudinal data, i.e. there is more than one response variable in longitudinal data. In this thesis, we build bivariate discrete joint models for such discrete longitudinal and time-to-event data.

The joint model for analysing longitudinal and time-to-event data contains a longitudinal model and a survival model. These two models are linked together by a link function Δ containing random effects $\boldsymbol{b}$. Random effects are used to capture dependence across time and dependence between responses.

For the longitudinal model, we chose two distributions for discrete responses. The first distribution is the bivariate extension of the binomial distribution. For the outcome of a sum of scores, we assume that the probability of an individual's response increasing by one, i.e. the probability of an individual earning a point, follows a Bernoulli distribution. Therefore, the response follows a binomial distribution. We choose a bivariate extension of the binomial distribution rather than the standard binomial distribution. The advantages of using this extended model to build a joint model are: (i) Compared to the standard binomial distribution, this distribution can measure over- or under-dispersion of the data. (ii) This extended model has a separate parameter ϕ that can measure the relationship between the

longitudinal outcomes. In this case, the random effects are not required to capture all the correlations in joint models. This parameter ϕ allows us to discern more clearly the relationship between the two responses. We used the logistic regression model with random effects to model the probabilities in this extended binomial distribution.

The second longitudinal model we use in this thesis is the categorical distribution. We assume that each possible value of the longitudinal response is a category and model the probability that the data for an individual belongs to a specific category. Because the categorical distribution models the probability of each possible value separately, it provides a better description of the information contained in the longitudinal data than the extension of the binomial distribution. We demonstrate this in Chapter 6 by modelling the baseline data using these two distributions separately. The corresponding probabilities for each category are also constructed using the logistic regression model.

We used three parametric hazard models for the survival model: the exponential hazard model, the Weibull hazard model and the Gompertz model. Our joint model can be used for prediction through parametric hazard models.

Joint models in this thesis are estimated by maximising the marginal likelihood. Model comparisons are performed by applying the Akaike information criterion. We use two approaches to construct joint models in this thesis. The first approach, known as the shared random-effects joint model, is a model in which the random effects follow a continuous distribution. This model provides a flexible way to measure the correlation between longitudinal response and the risk of the event occurring, i.e. random effects do not need to be a fixed value. We assume that the random effects follow a bivariate normal distribution, which is the distribution often used in multivariate shared random-effects joint models [Hickey et al., 2016]. In general, even if we set the distribution of random effects in advance, the fixed effects parameter estimates are relatively robust, except for the parameter estimates in the survival model [Pantazis and Touloumi, 2007]. However, for heavily skewed distributed random effects, this may lead to biased estimates if the actual

distribution differs from the normal distribution.

The shared random-effects joint models are used to analyse the ELSA data, PAQUID and PBC2 data. For the ELSA data and the PAQUID data, joint models are constructed with the extension of the binomial distribution. We compare the fit of joint models using different link functions by AIC values. Based on the same expressions of the longitudinal model and survival model, the joint model with a random intercept link function has a smaller AIC than the joint model with a random slope link function. We use the MAP method to make predictions in Section 6.1.2.3 based on the model estimation results. The analysis of the PAQUID data validates the flexibility of our joint model, i.e. the two responses of the bivariate extension of the binomial distribution do not need to have the same range.

For the PBC2 data, we use the categorical distribution to construct the shared random-effects joint models. We treat each possible value of the responses for the PBC2 data as a category. A disadvantage of joint models based on the categorical distribution over joint models based on the bivariate extended binomial distribution is that we cannot model the relationship between the two longitudinal responses by a specific parameter.

As we discussed in Section 1.1.2, the shared random-effects model has two drawbacks: (i) this joint model requires the calculation of multi-dimensional integrals, which greatly slows down the computation; (ii) If the latent heterogeneity of responses is discrete, i.e. there are 'latent subpopulations' among individuals defined according to some characteristics [Arbeev et al., 2014], the shared random-effects model with continuous random effects may not describe the structure of this heterogeneity well. Therefore, we also construct latent-class joint models in our thesis.

The latent-class model connects the longitudinal model and the survival model through a discrete distribution of random effects. It divides individuals into groups. This method does not involve any numerical integration. We use the bivariate extended binomial distribution as the longitudinal model in latent-class joint models. In Section 6.1.3, we analyse the ELSA data using latent-class joint models and

select one of the models for detailed analysis. In Section 6.1.4, we analyse the PAQUID data using both shared random-effects joint models and latent-class joint models. We compare the running times of the two joint models and conclude that the joint model runs considerably more efficiently when the computational integral is removed. The drawback of latent-class joint models is that the divided classes are difficult to explain. Nevertheless, sometimes explaining the meaning of each latent class may not be part of the research.

In the research of this thesis, we encountered several problems. (i) When we attempted to generate data from a categorical distribution in Section 4.4, we are unable to get the data we expect using the function `rcat` in `LaplacesDemon` package [Statisticat and LLC., 2021]. We demonstrated this by comparing the generated data distribution with the probabilities we set in `rcat`. We used the `sample` function to generate data from categorical distributions in the simulation study. (ii) The estimation results of both joint models for analysing the ELSA data encountered the same problem: when the link function is equal to the random slope of the two responses, the estimates suggest that the risk of death is higher in individuals with a slower decline in cognitive function over time. We have shown via simulation study that the estimation results of our joint model are accurate. Moreover, the joint model for other data does not show such unexpected results. Therefore, we speculate that the specific feature of the ELSA data causes this phenomenon, and the true reason for this phenomenon needs further investigation. (iii) We used education level as a covariate when constructing the probabilities for each class in the latent-class model. However, an individual's education level may be related to his birth year. For example, a person born in 2000 may have a higher probability of having a higher education than a person born in 1980. Therefore, using education level as a covariate may lead to bias in the estimation results and make the prediction results inconsistent from the true picture.

There are some aspects that could be used as a starting point for further research.

1. First, one of the binomial distribution assumptions is that the questions used

to collect the data (e.g. words in ELSA data and questions in MMSE) are independent of each other. However, if there is a dependence between questions, then this violates the assumptions of the independent Bernoulli test. Alternatively, we can replace the observed total score with a latent variable that is linked to the score for a particular question through an item response theory model [Fox, 2010].

2. We use linear predictors and logistic regression to model the probabilities in the longitudinal model. In further studies, the linear predictors in logistic regression can be modelled more flexibly by using B-splines or other semi-parametric methods.

3. Although we have greatly improved the computational efficiency of the shared random-effects joint model through code optimisation and the use of some packages in `R` (the running time of the code has been reduced from over 24 hours to 9 hours), the code of the shared random-effects model still runs relatively slow. It is still challenging to compute high-dimensional integrals quickly when we have more random effects. An alternative might be to use a Bayesian approach; for recent development in this area, see the `JMbayes` package Rizopoulos et al. [2022].

4. In this thesis, we construct bivariate joint models, whereas, in longitudinal data, there may be more than two responses. We can construct the corresponding multivariate extended binomial distribution based on Altham and Hankin [2012] and use this distribution in our joint model. We can also construct a high-dimensional joint model by adding a new categorical distribution.

5. As we mentioned in Chapter 2, our framework is flexible in its choice of hazard models, i.e. it is easy to replace the exponential hazard model, the Gompertz hazard model and the Weibull hazard model with other parametric or semi-parametric hazard models. Joint models using other hazard models are worth being investigated. Moreover, in this thesis, we discuss the case of data containing two states, i.e. alive and dead/dementia. Multi-state mod-

els are worth investigating in future research as individuals have more states in many datasets. For example, if individuals have three states of health/illness/death, we can analyse such data by combining the joint model and the multi-state model.

6. In this thesis, we have used two approaches to construct the joint model. However, we can use other methods to construct the joint model in further research. An alternative potential approach is to use a copula model to link the longitudinal and survival models [Rizopoulos et al., 2008]. When applying this method, we need to use the cumulative distribution functions of the longitudinal and survival models.

7. The datasets we use in this thesis are all discrete. When the data to be analysed are discrete with large ranges or continuous data, we believe that our model is still applicable. For discrete data with large ranges, by changing the value of m_j in Equation (2.4), we can use the binomial joint model to analyse the dataset. Alternatively, we can categorise the responses prior to applying the joint model for discrete data with a wide range of values and continuous data.

We use the MMSE test as an example. We have introduced in Chapter 5 that MMSE scores are discrete data that take values in the range of 0 to 30. We could directly take the 31 possible values as 31 categories and analyse them using a categorical joint model. However, the number of parameters required to model the 31 categories is relatively large. In practice, researchers usually categorise the results of MMSE scores into four categories in order to diagnose the cognitive abilities of participants, i.e. 24-30 for normal cognitive ability, 19-23 for mild, 10-18 for moderate, 0-9 for severe cognitive impairment, respectively [Vertesi et al., 2001]. Therefore, we can categorise MMSE scores according to this already widely used categorisation method. After categorising the responses into four categories, we can analyse the data using the binomial joint model and the categorical joint model proposed in this thesis.

Pre-processing the data in the manner of our example not only allows for a wider application of our proposed model but also allows the researcher to reduce the number of parameters when analysing the data and to obtain inferences explicitly targeting the main research questions.

To conclude, in this thesis, we construct bivariate joint models to simultaneously investigate changes in longitudinal outcomes over time and the time of event occurrence. The bivariate joint models are general and can be used in a wide range of bivariate discrete-valued outcomes in the medical statistics. In addition, we discussed link functions which are contained in both the longitudinal model and the survival model. We hope our discussion will provide ideas for constructing other multivariate joint models for data consisting of discrete longitudinal outcomes and time to events.

Appendix A

Appendix for Simulation Section

We conducted simulation studies using three sets of true values in Section 4.1 and plotted bias and corresponding Monte Carlo 95% confidence interval according to Table 4.1. In this section, We plot the images based on Table 4.2 and Table 4.3 as well. We plot box plots based on parameter estimates from the four sets of simulation studies conducted in Section 4.1, Section 4.3 and Section 4.4, and these box plots are also presented in this section.
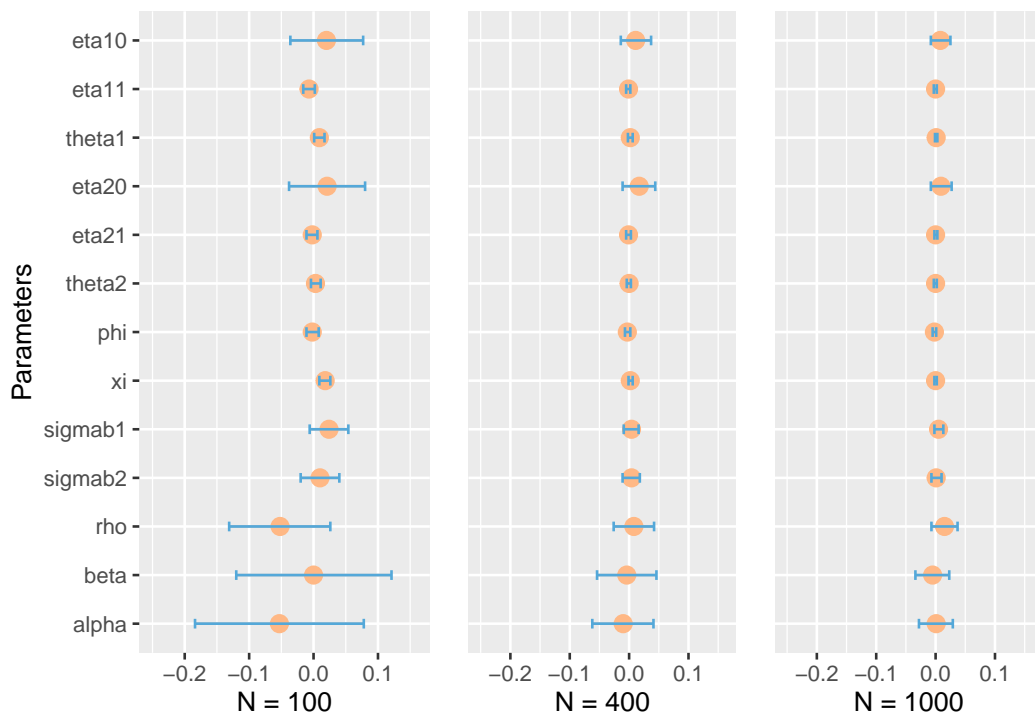
Figure A.1: Bias and corresponding Monte Carlo 95% confidence interval, based on the results in Table 4.2. Large dots represent biases, and geometry bars represent Monte Carlo 95% confidence interval.
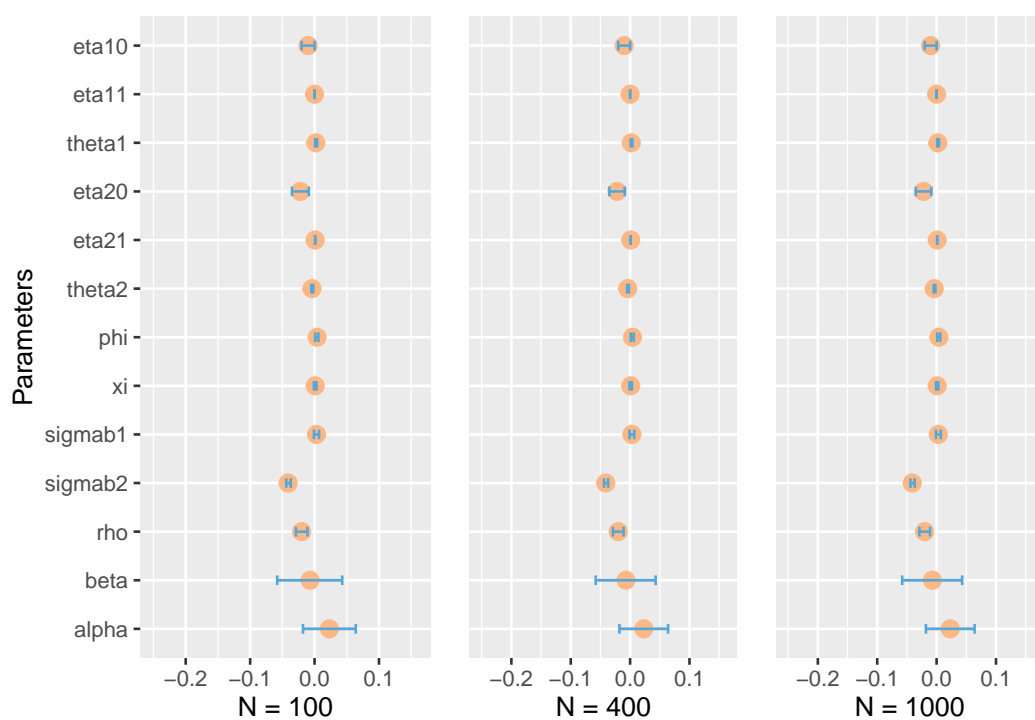
Figure A.2: Bias and corresponding Monte Carlo 95% confidence interval, based on the results in Table 4.3. Large dots represent biases, and geometry bars represent Monte Carlo 95% confidence interval.

N = 100; Follow−up Interval = 3

N = 400; Follow−up Interval = 3

N = 1000; Follow−up Interval = 3
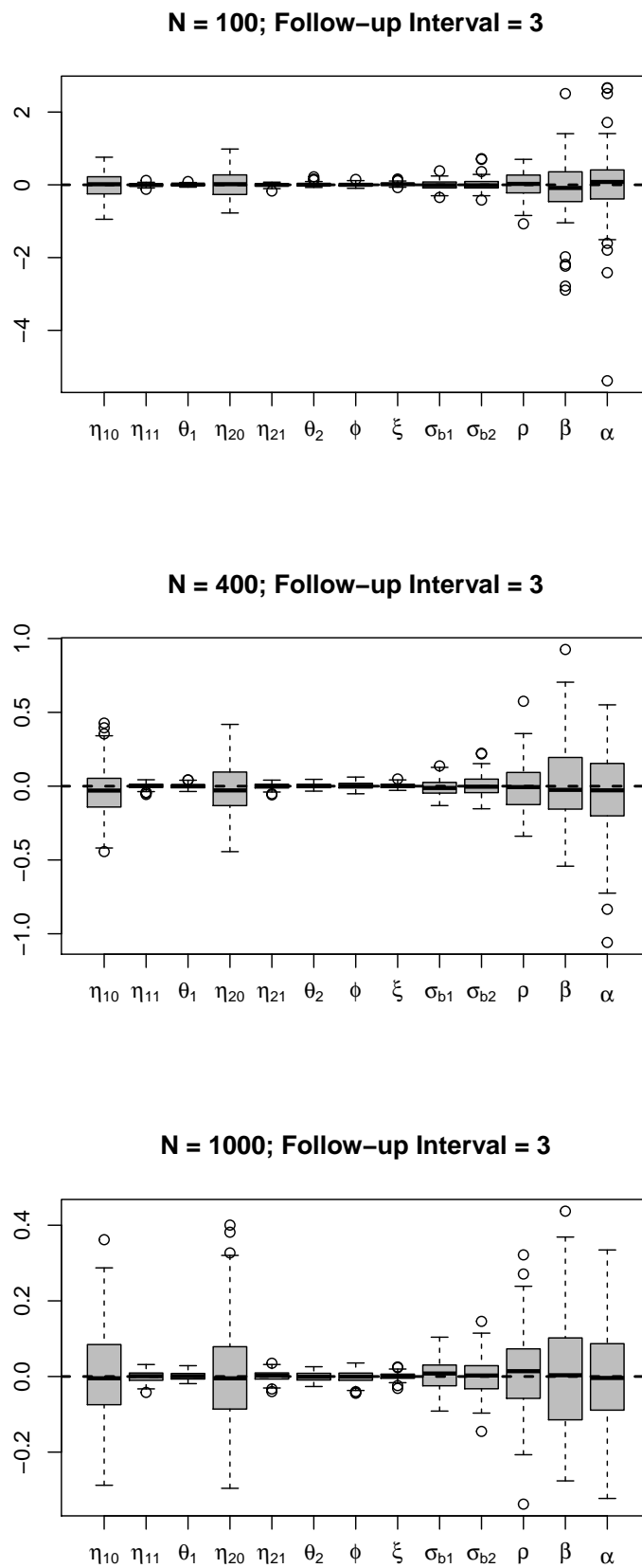
Figure A.3: Additional images related to the simulation studies in Table 4.1: Box plots for estimated parameters.

N = 100; Follow−up Interval = 3

N = 400; Follow−up Interval = 3

N = 1000; Follow−up Interval = 3
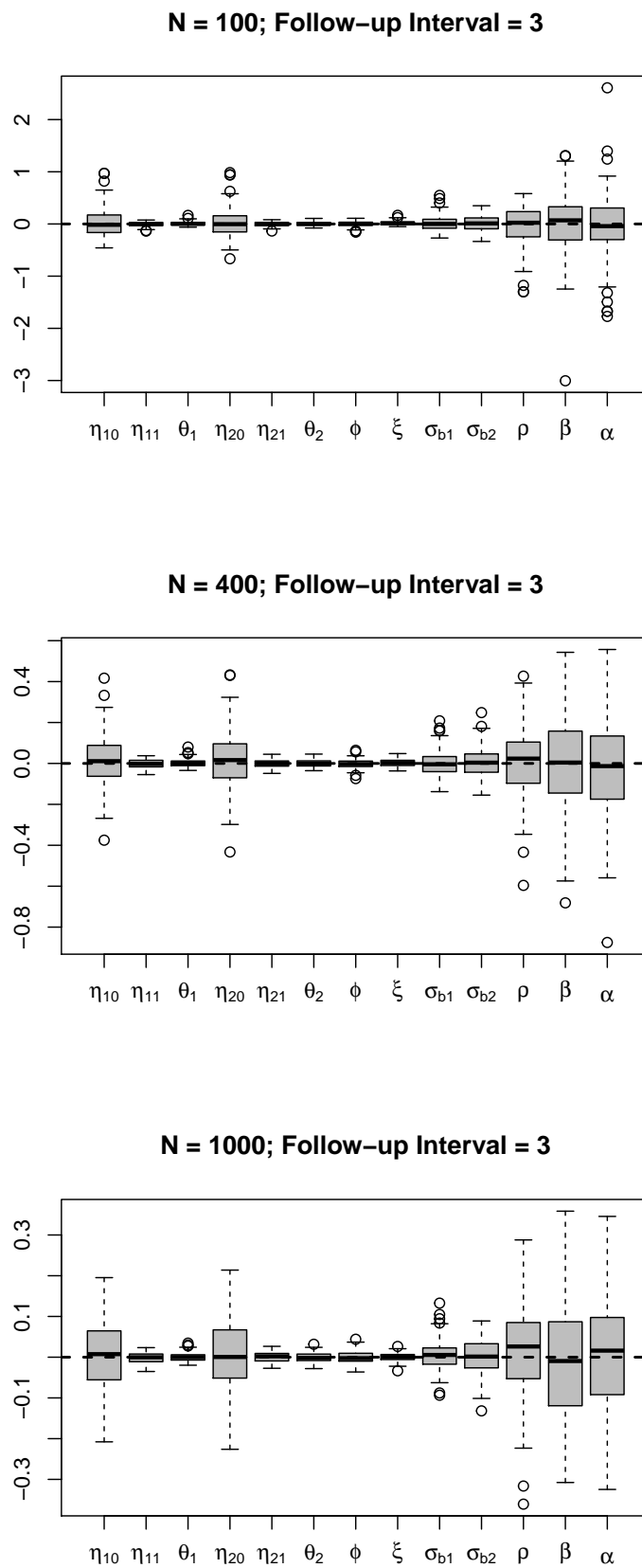
Figure A.4: Additional images related to the simulation studies in Table 4.2: Box plots for estimated parameters.
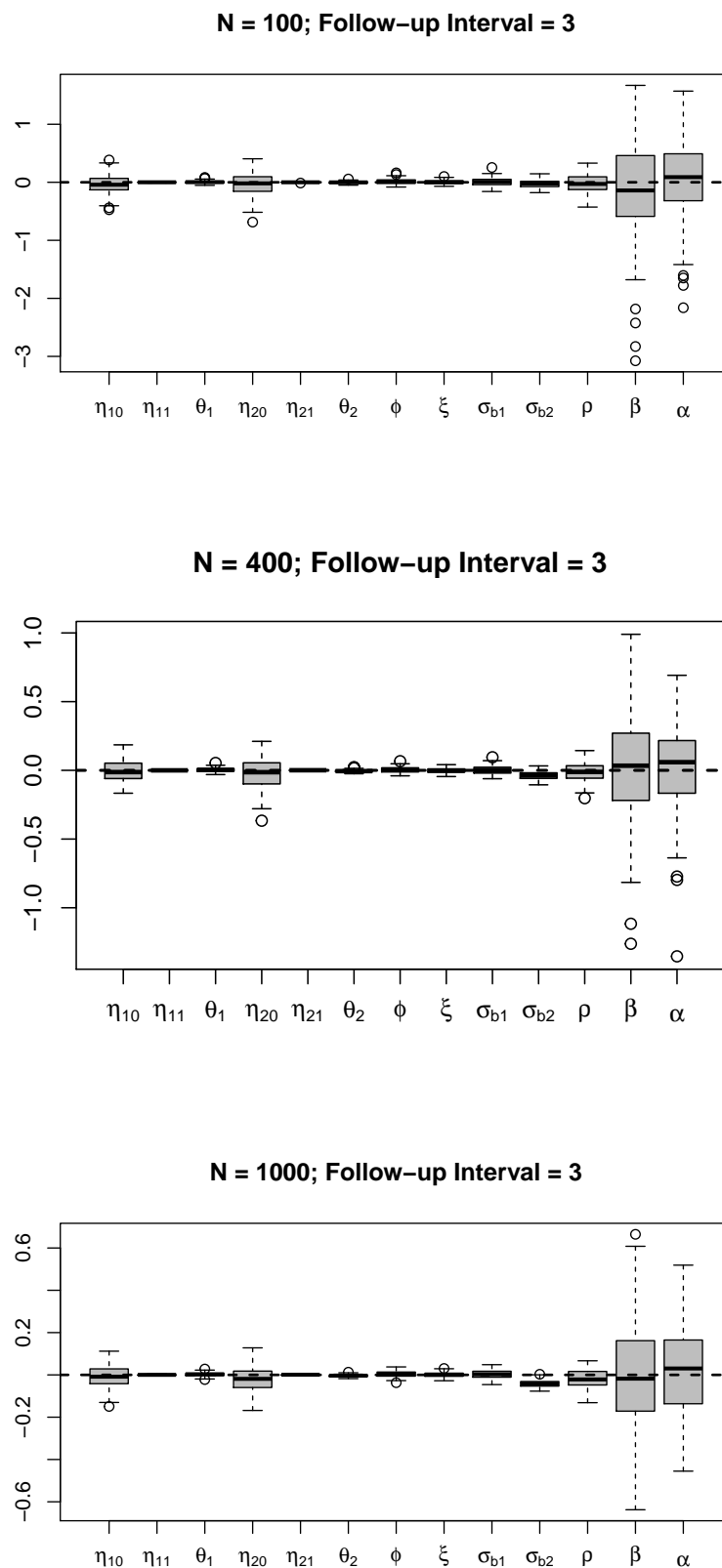
Figure A.5: Additional images related to the simulation studies in Table 4.3: Box plots for estimated parameters.

N = 100; Follow−up Interval = 3

N = 400; Follow−up Interval = 3

N = 1000; Follow−up Interval = 3
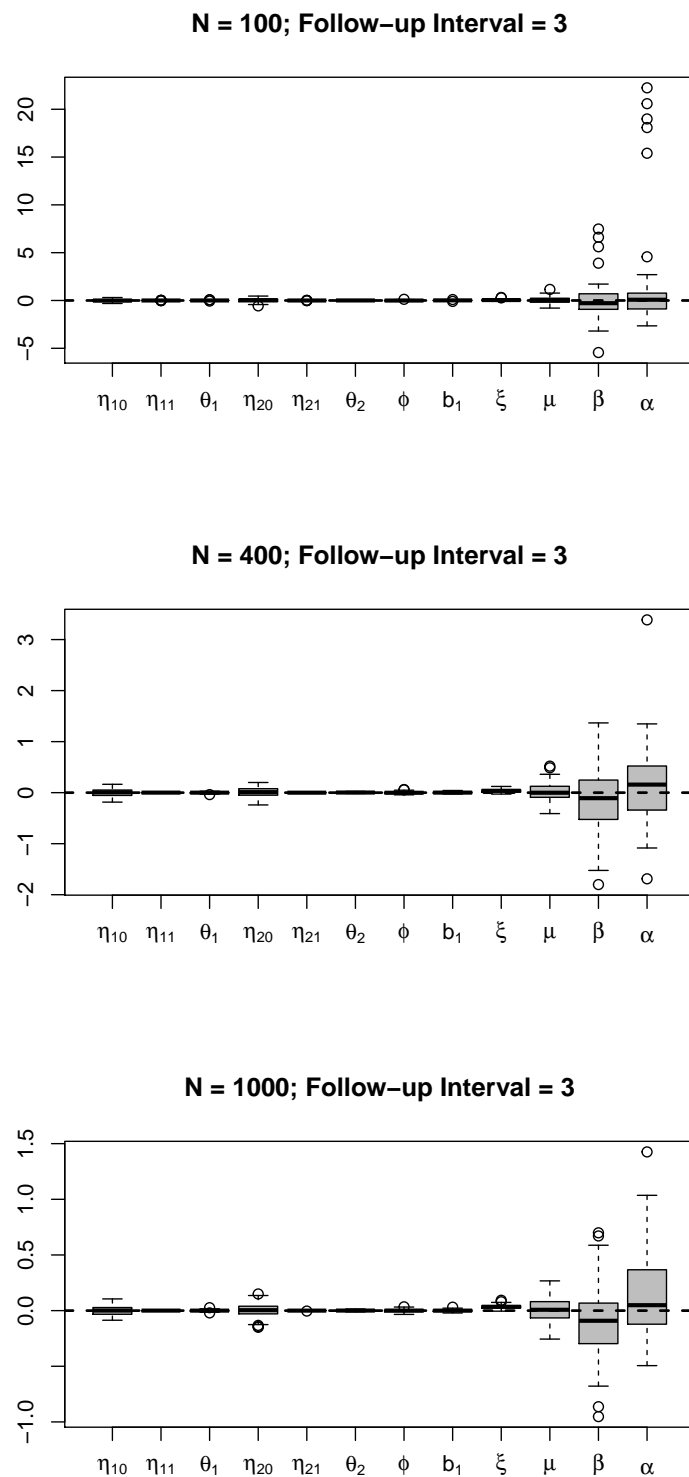
Figure A.6: Additional images related to the simulation studies in Table 4.5: Box plots for estimated parameters.
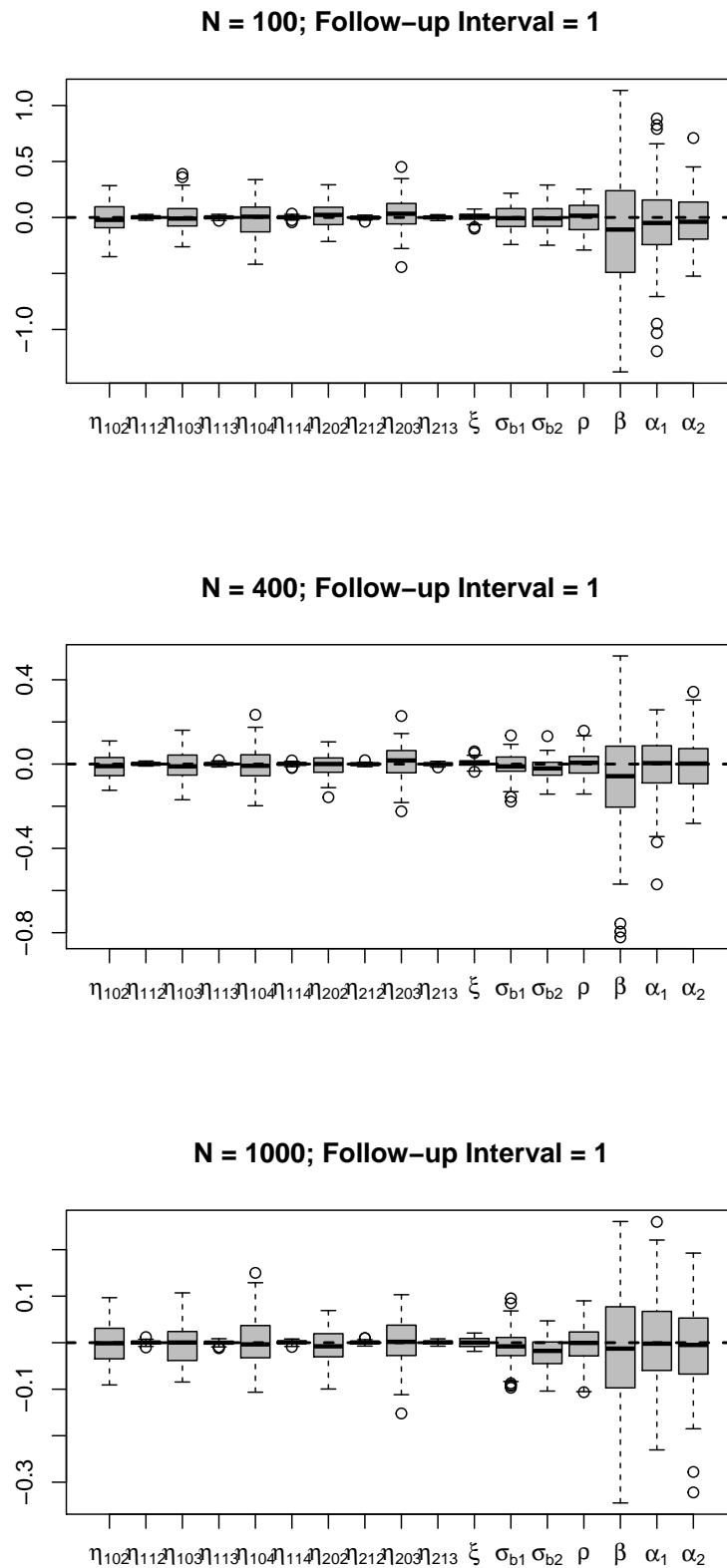
Figure A.7: Additional images related to the simulation studies in Table 4.6: Box plots for estimated parameters.

Appendix B

Appendix for Application Section

In this chapter, we will first present the estimation results of Model SG_2 in Table 6.2. After that, we will plot the three-dimensional images of Models IG_1 - IG_4 in Table 6.2 and present the corresponding expectations.

B.1   Estimation results of Model SG_2

We have analysed the estimation results of Model SG_2 in the Additional remarks in Section 6.1.2. In this section, we list all estimated parameters of the model as supplementary.

Table B.1: ELSA: Parameter estimates for shared random-effects Model SG_2 with (2.35)

$\eta_0^{(1)}$	-0.631 (0.074)	θ_{Y_1}	1.188 (0.032)	σ_{b1}	0.020 (0.011)	β	-6.641 (0.290)
$\eta_0^{(2)}$	-1.956 (0.097)	θ_{Y_2}	1.102 (0.025)	σ_{b2}	0.032 (0.013)	ξ	0.106 (0.010)
$\eta_1^{(1)}$	-0.022 (0.002)	ϕ	1.423 (0.057)	ρ	0.436 (0.087)		
$\eta_1^{(2)}$	-0.027 (0.002)					α	-0.617 (0.201)
$\gamma_L^{(1)}$	-0.041 (0.036)	$\gamma_L^{(2)}$	-0.169 (0.040)	γ_S	0.353 (0.147)		

B.2   Plots and expectations of Model IG_1 - IG_4

In the Model fit in Section 6.1.2, we calculated individual-specific random effects using the MAP method based on the result in Table 6.1. After that, we plotted the fitted distribution based on the mean of the calculated random effects and presented the corresponding expectations. In this section, we list the plots of the three-

dimensional fitted distributions and the corresponding expectations of Models IG_1 - IG_4 in Table 6.2. In addition, we add gender as a covariate in the model in Table 6.2 so that the three-dimensional plots and the corresponding expectations are calculated conditional on age and gender.

Expectations for Model IG_1:

$$E[(Y1,Y2)|t = 50, sex = 0] = (4.755, 3.329) \tag{B.1}$$

$$E[(Y1,Y2)|t = 50, sex = 1] = (4.448, 3.041) \tag{B.2}$$

$$E[(Y1,Y2)|t = 80, sex = 0] = (3.712, 1.092) \tag{B.3}$$

$$E[(Y1,Y2)|t = 80, sex = 1] = (3.429, 0.859) \tag{B.4}$$

Figure B.1: Model IG_1: Fitted bivariate binomial distribution for immediate recall (Y_1) and delayed recall (Y_2).

Expectations for Model *IG$_2$*:

$$E[(Y1, Y2)|t = 50, sex = 0] = (5.548, 3.927) \tag{B.5}$$

$$E[(Y1, Y2)|t = 50, sex = 1] = (5.485, 3.845) \tag{B.6}$$

$$E[(Y1, Y2)|t = 80, sex = 0] = (4.978, 3.301) \tag{B.7}$$

$$E[(Y1, Y2)|t = 80, sex = 1] = (4.916, 3.22) \tag{B.8}$$
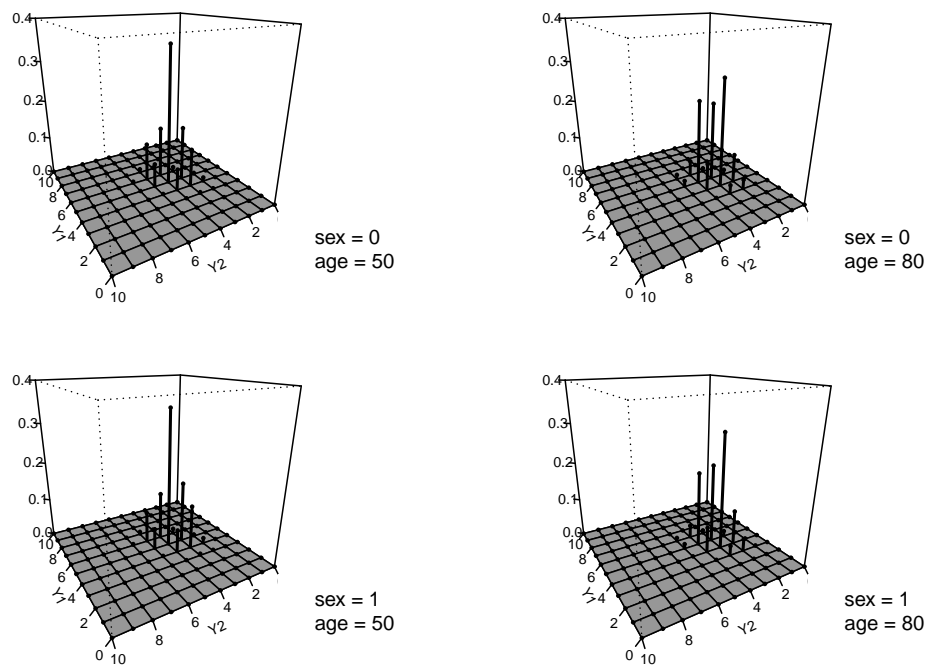
Figure B.2: Model *IG$_2$*: Fitted bivariate binomial distribution for immediate recall (Y_1) and delayed recall (Y_2).

Expectations for Model *IG*$_3$:

$$E[(Y1,Y2)|t = 50, sex = 0] = (4.683, 3.839) \tag{B.9}$$

$$E[(Y1,Y2)|t = 50, sex = 1] = (4.619, 3.755) \tag{B.10}$$

$$E[(Y1,Y2)|t = 80, sex = 0] = (4.126, 3.208) \tag{B.11}$$

$$E[(Y1,Y2)|t = 80, sex = 1] = (4.063, 3.126) \tag{B.12}$$
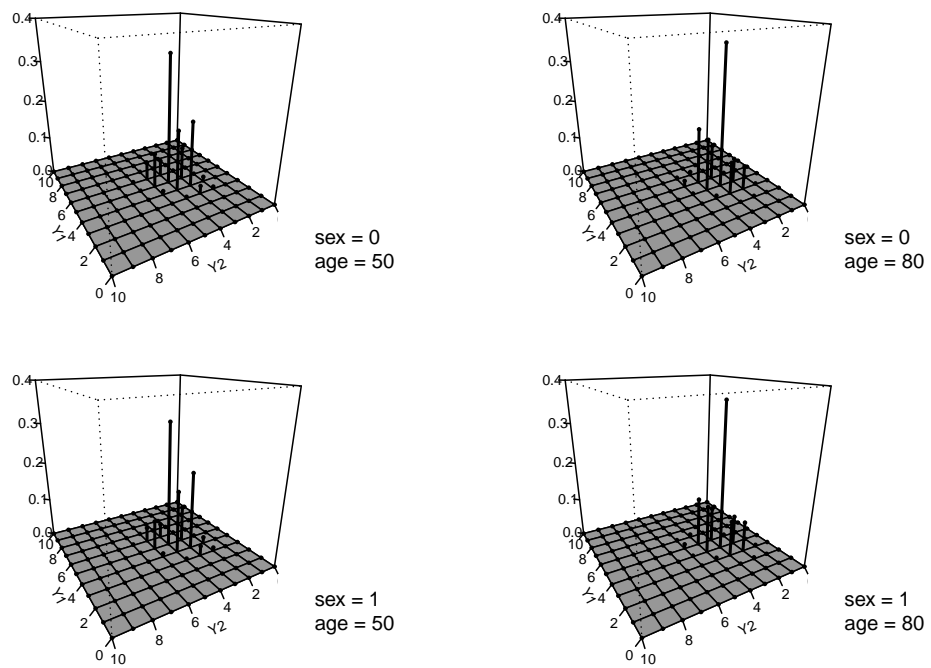
Figure B.3: Model *IG*$_3$: Fitted bivariate binomial distribution for immediate recall (Y_1) and delayed recall (Y_2).

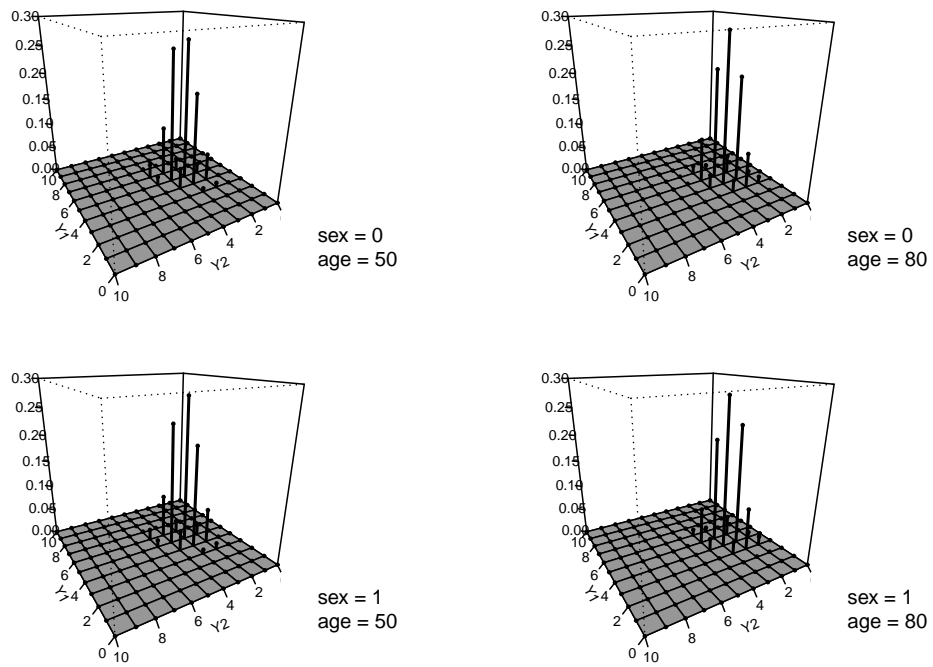Expectations for Model IG_4:

$$E[(Y1,Y2)|t = 50, sex = 0] = (5.121, 3.225) \tag{B.13}$$

$$E[(Y1,Y2)|t = 50, sex = 1] = (5.06, 3.143) \tag{B.14}$$

$$E[(Y1,Y2)|t = 80, sex = 0] = (4.558, 2.604) \tag{B.15}$$

$$E[(Y1,Y2)|t = 80, sex = 1] = (4.498, 2.523) \tag{B.16}$$
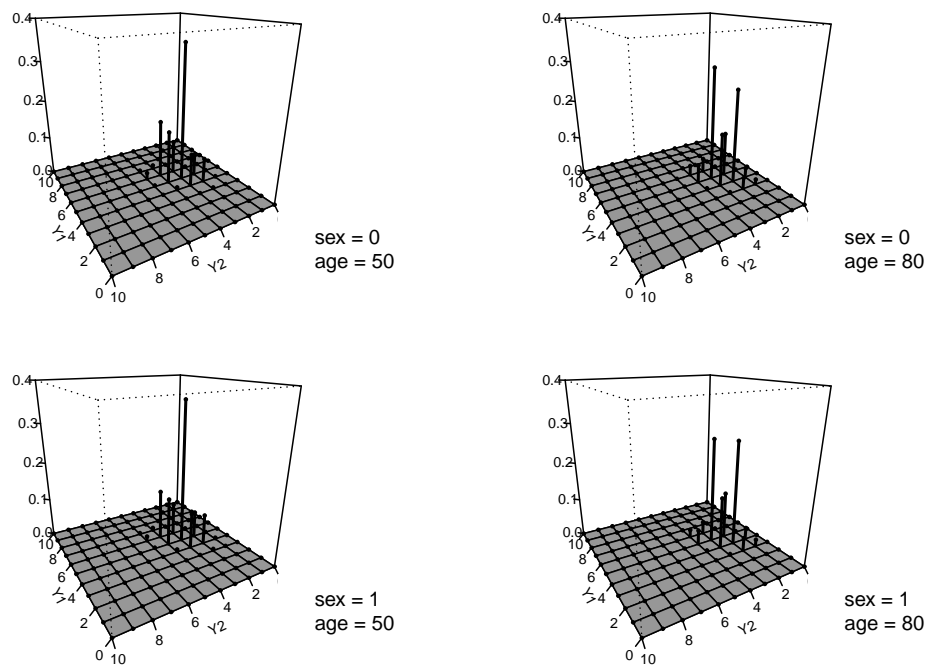
Figure B.4: Model IG_4: Fitted bivariate binomial distribution for immediate recall (Y_1) and delayed recall (Y_2).

Appendix C

Code

We present some of the code written during the research in this chapter. These codes are used in the simulation study in Section 4.1. This chapter consists of three parts. The first part is the functions used in the simulation study; the second part is the code that generates the data from the bivariate joint model; the third part is the corresponding log-likelihood function used for parameter estimation.

C.1   Code: Functions used in the simulation study

The main functions we use in our simulation study are the function to calculate the Gaussian-Hermite integral, the density function of Equation (2.4) `dbin` and the function to generate data from Equation (2.4) `rbivMM`. Function `rbivMM` is inspired by the `rMM` function in `MM` package [Altham and Hankin, 2012], and the `rMM` function is used to generate data from Equation (2.2).

```
# =====================================
# Functions used in the
# bivariate joint model simulation
# =====================================
# Gaussian-Hermite integral:
# -------------------------------------
library(statmod)
# Calculate nodes and weights for Gaussian quadrature
quad    <- gauss.quad(nnodes, "hermite")
nodes   <- quad$nodes
```

```r
weights <- quad$weights
# Compute integral
integrateGH.biv <- function(integrand, mu, sigma){
  approx <- 0
  for(i in 1:nnodes ){
    x <- sqrt(2)*sigma*nodes[i] + mu
    approx <- (1/sqrt(pi))*weights[i]%*%integrand(x)+ approx
  }
  return(approx)
}


# -------------------------------------
# Density function for
# the bivariate extension of the binomial:
# -------------------------------------
dbin <- function(x,size,prob,theta,phi){
  # Define:
  # Range of responses Y1 and Y2
  n  <- size
  # Probability of success
  p1 <- prob[1]
  p2 <- prob[2]
  j1 <- x[1]
  j2 <- x[2]
  theta1 <- theta[1]
  theta2 <- theta[2]
  if (j1<0 || j2<0) {return(0)}
  # Pre-compute binomial coefficients:
  C1.choose   <- rep(NA,size+1)
  part.choose <- rep(NA,size+1)
  for(j in 0:size){
    C1.choose[j+1]   <- choose(size,j)
```

```r
    part.choose[j+1] <- choose(size,j)
  }
  # Model:
  num1 <- choose(n,j1)*(1-p1)^(n-j1)*p1^(j1)*theta1^(j1*(n-j1))
  num2 <- choose(n,j2)*(1-p2)^(n-j2)*p2^(j2)*theta2^(j2*(n-j2))
  num  <- num1*num2*phi^(j1*j2)
  C    <- 0
  for(j in 0:n){
    C1 <-  C1.choose[j+1]**(1-p1)^(n-j)*p1^(j)*theta1^(j*(n-j))
    for(jj in 0:n){
      part <- part.choose[jj+1]*(1-p2)^(n-jj)*p2^(jj)*theta2^(jj*(n-jj))
      C    <-  C + C1*part*phi^(j*jj)
    }
  }
  # Return:
  return(num/C)
}


# -------------------------------------
# Sample from the bivariate extension of the binomial:
# -------------------------------------
rbivMM <- function(size,prob,theta,phi,burn.in = 1000){
  # burn-in = 1000
  X <- numeric(burn.in+1); Y <- numeric(burn.in+1)
  X[1] <- 1; Y[1] <- 1
  for (i in 2:(burn.in+1)) {

    Epsilon1 <- sample(c(1, -1), 1)
    Epsilon2 <- sample(c(-1, 1), 1)

    Xproposal <- X[i - 1] +  Epsilon1
    Yproposal <- Y[i - 1] +  Epsilon2
```

```r
  if (Xproposal > size || Xproposal < 0) {
    Xproposal <- X[i - 1]
  }
  if(Yproposal > size || Yproposal < 0){
    Yproposal <- Y[i - 1]
  }
  # Set probability a
  # With probability a we accept the proposed value:
  # X[i] <- Xproposal; Y[i] <- Yproposal
  # Otherwise we reject it and set:
  # X[i] <- X[i-1]; Y[i] <- Y[i-1]
  a <- min(1,dbin(c(Xproposal, Yproposal), size, prob, theta, phi)/
  dbin(c(X[i-1], Y[i-1]), size, prob, theta, phi))

  if (runif(1) < a) {
    X[i] <- Xproposal; Y[i] <- Yproposal
  } else {
    X[i] <- X[i-1]; Y[i] <- Y[i-1]
  }
}
X <- X[-c(1:burn.in)]; Y <- Y[-c(1:burn.in)]
data <- cbind(X, Y)
return(data)
}
```

C.2   Code: Data generation for the bivariate joint model

```r
# ======================================
# Simulate data for bivariate joint model
# ======================================
# Simulate random effects and data:
```

```r
# -------------------------------------
# Matrix used to record the generated random effects
b <- matrix(NA,Npop,2)
for(i in 1:Npop){

  # Generate random effects from bivariate normal distribution:
  b1.i     <- rnorm(n=1,mean=0,sd=sigma.b1)
  cond.mn  <- 0+rho*(b1.i-0)*sigma.b2/sigma.b1
  cond.var <- sigma.b2^2*(1-rho^2)
  b2.i     <- rnorm(n=1,mean=cond.mn,sd=sqrt(cond.var))
  b[i,]    <- c(b1.i,b2.i)


  # Age grid:
  age.i    <- ages


  # Longitudinal model:
  # ---- Y1 ----
  nu1.i    <- nu1+c(b1.i,0)
  # Linear predictor with random intercept for Y1
  lp1.i    <- cbind(1,age.i)%*%nu1.i
  # Corresponding logistic regression for Y1
  pie1.i   <- exp(lp1.i)/(1+exp(lp1.i))


  # ---- Y1 ----
  nu2.i    <- nu2+c(b2.i,0)
  # Linear predictor with random intercept for Y2
  lp2.i    <- cbind(1,age.i)%*%nu2.i
  # Corresponding logistic regression for Y1
  pie2.i   <- exp(lp2.i)/(1+exp(lp2.i))


  # longitudinal responses:
  y.ij     <- matrix(NA,length(age.i),2)
```

```r
for (j in 1:length(age.i)) {
  # Generate bivariate longitudinal outcomes
  # from the bivariate extension of the binomial distribution
  y.ij[j, ] <- rbivMM(
    # Range of responses Y1 and Y2
    size,
    # Logistic regression for two response
    prob = c(pie1.i[j], pie2.i[j]),
    # theta1 and theta2
    theta = theta,
    phi = phi,
    # warm-up period
    burn.in = 1000
    )
}
y1.i <- y.ij[, 1]
y2.i <- y.ij[, 2]


# Survival model:
lambda.i <- exp(beta[1]+alpha.2D%*%c(nu1.i[1],nu2.i[1]))
S        <- exp(c(-lambda.i*xi^(-1))*(exp(xi*dages)-1))
# Death time:
F      <- 1-S
u      <- runif(n=1,0,1)
# Compute time to events using the inversion method
dage.i <- dages[which(abs(F-u)==min(abs(F-u)))]


# Enforce max age:
status.i <- as.numeric(dage.i <= max.age)
dage.i <- ifelse(dage.i>=max.age,max.age, dage.i)


# Build data:
```

```r
  data.i <- cbind(No.sim = sim, idn=i,
                  y1=y1.i,y2=y2.i,
                  age=age.i,dage=dage.i,bage=age.i[1],
                  calender=calender.year+age.i,
                  status = status.i)
}
data <- as.data.frame(data)


# Remove observations after death
data.split <- split(data,data$idn)

for(i in 1: Npop){
  data.i  <- data.split[[i]]
  idn      <- data.i$idn
  y1       <- data.i$y1
  y2       <- data.i$y2

  age.i    <- data.i$age
  dage.i   <- data.i$dage[1]
  bage.i   <- data.i$age[1]
  No.sim   <- data.i$No.sim
  calender.i <- data.i$calender
  status.i <- data.i$status
  n.i      <- length(which(age.i< dage.i))

  # Death record:
  drecord  <- c(No.sim[1],idn[1],NA, NA,
                dage.i, dage.i ,bage.i, NA, status.i)
  data.spl <- rbind(data.i[1:n.i,],drecord)
  if(i==1){dta <- data.spl}else{dta <- rbind(dta, data.spl)}
}
```

C.3   Code: Log-likelihood function

```r
# ====================================
# Loglikelihood for bivariate joint model
# ====================================
# Pre-compute binomial coefficients:
C1.choose   <- rep(NA,size+1)
part.choose <- rep(NA,size+1)
for(j in 0:size){
  C1.choose[j+1]   <- choose(size,j)
  part.choose[j+1] <- choose(size,j)
}


# Loglikelihood:
loglikelihood <- function(p){
    # Parameters:
    # longitudinal:
    nu1      <- p[1:2]
    theta1   <- exp(p[3])


    nu2      <- p[4:5]
    theta2   <- exp(p[6])
    delta    <- exp(p[7])


    # survival:
    xi       <- p[8]


    # random effects:
    sigma.b1 <- exp(p[9])
    sigma.b2 <- exp(p[10])
    rho <- 2*exp(p[11])/(exp(p[11])+1) - 1


    # shared part:
```

```r
beta    <- p[12]
alpha1  <- p[13]

# Build loglikelihood :
loglik <- 0
for(i in 1:Npop){
  time  <- tt[i]
  death <- dd[i]
  y1    <- yy1[ind1[i]:ind2[i]]
  y2    <- yy2[ind1[i]:ind2[i]]
  x     <- xx[ind1[i]:ind2[i]]
  L     <- length(x)

  integrand <- function(b1, b2){
    # Longitudinal model:
    nu.i1 <- nu1+c(b1,0)
    lp.i1 <- cbind(1,x)%*%nu.i1
    nu.i2 <- nu2+c(b2,0)
    lp.i2 <- cbind(1,x)%*%nu.i2
    p1 <- exp(lp.i1)/(1+exp(lp.i1))
    p2 <- exp(lp.i2)/(1+exp(lp.i2))
    # Data-dependent part of distribution:
    f1 <- 1
    for(ind in 1:L){
      # Bivariate extension of the binomial distribution:
      num1 <- (1-p1[ind])^(size-y1[ind])*p1[ind]^(y1[ind])*
      theta1^(y1[ind]*(size-y1[ind]))
      num2 <- (1-p2[ind])^(size-y2[ind])*p2[ind]^(y2[ind])*
      theta2^(y2[ind]*(size-y2[ind]))
      num  <- num1*num2*delta^(y1[ind]*y2[ind])
      # Normalising constant for distribution:
      # with pre-computed binomial coefficients:
```

```r
    C <- 0
    for(j in 0:size){
      C1 <- C1.choose[j+1]*(1-p1[ind])^(size-j)*p1[ind]^(j)*
      theta1^(j*(size-j))
      for(jj in 0:size){
        part <- part.choose[jj+1]*(1-p2[ind])^(size-jj)*
        p2[ind]^(jj)*theta2^(jj*(size-jj))
        C    <-  C + C1*part*delta^(j*jj)
      }
    }
    f1 <- num/C*f1
  }


  # Survival model:
  lambda.i <- exp(beta+alpha1*nu.i1[1])
  h   <- lambda.i*exp(xi*time)
  S   <- exp(-lambda.i/xi*(exp(xi*time)-1))
  f2 <- death*(h*S)+(1-death)*S
  # Integrand:
  f1*f2
}


# Integrate for likelihood contribution for i:
marg <- integrateGH.biv(Vectorize(function(b2) {
  integrateGH.biv(function(b1) { integrand(b1, b2) },
               rho*(sigma.b1/sigma.b2)*b2,
               sqrt(sigma.b1^2*(1-rho^2)) ) }), 0,sigma.b2)


loglik <- loglik+log(marg)
}


# Return:
```

```
    -loglik
}
```

Bibliography

Alan Agresti. *Categorical data analysis*. John Wiley & Sons, 2003.

Pat M. E. Altham and Robin K. S. Hankin. Multivariate generalizations of the multiplicative binomial distribution: Introducing the MM package. *Journal of Statistical Software*, 46(12):1–23, 2012. doi: 10.18637/jss.v046.i12.

Patricia ME Altham. Two generalizations of the binomial distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 27(2):162–167, 1978.

Konstantin G Arbeev, Igor Akushevich, Alexander M Kulminski, Svetlana V Ukraintseva, and Anatoliy I Yashin. Joint analyses of longitudinal and time-to-event data in research on aging: Implications for predicting health and survival. *Frontiers in public health*, 2:228, 2014.

James Banks, Elizabeth Breeze, Carli Lessof, and James Nazroo. Retirement, health and relationships of the older population in england: The 2004 english longitudinal study of ageing (wave 2), 2006.

Ralf Bender, Thomas Augustin, and Maria Blettner. Generating survival times to simulate cox proportional hazards models. *Statistics in medicine*, 24(11):1713–1723, 2005.

Arthur L Benton. A visual retention test for clinical use. *Archives of Neurology & Psychiatry*, 54(3):212–216, 1945.

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

Michael J Crowther. merlin—a unified modeling framework for data analysis and methods development in stata. *The Stata Journal*, 20(4):763–784, 2020.

Michael J Crowther, Keith R Abrams, and Paul C Lambert. Flexible parametric joint modelling of longitudinal and survival data. *Statistics in Medicine*, 31(30): 4456–4471, 2012.

Michael J Crowther, Keith R Abrams, and Paul C Lambert. Joint modeling of longitudinal and survival data. *The Stata Journal*, 13(1):165–184, 2013.

Peter J Diggle, Patrick Heagerty, Kung-Yee Liang, Patrick J Heagerty, Scott Zeger, et al. *Analysis of Longitudinal Data*. Oxford University Press, Oxford, 2002.

PJ Diggle, Robin Henderson, and Pete Philipson. Random-effects models for joint analysis of repeated-measurement and time-to-event outcomes. *Longitudinal Data Analysis*, pages 349–366, 2008.

Jean-françois Dupuy and Mounir Mesbah. Joint modeling of event time and nonignorable missing longitudinal data. *Lifetime Data Analysis*, 8(2):99–115, 2002.

Cheryl L Faucett and Duncan C Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in medicine*, 15(15):1663–1685, 1996.

Marshal F Folstein, Susan E Folstein, and Paul R McHugh. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, 1975.

Jean-Paul Fox. *Bayesian item response modeling: Theory and applications*. Springer, 2010.

Ben Goodrich, Jonah Gabry, Imad Ali, and Sam Brilleman. rstanarm: Bayesian applied regression modeling via Stan., 2022. URL `https://mc-stan.org/rstanarm/`. R package version 2.21.3.

Lindy E Harrell, Daniel Marson, Anjan Chatterjee, and Jo Ann Parrish. The severe mini-mental state examination: a new neuropsychologic instrument for the bedside assessment of severely impaired patients with alzheimer disease. *Alzheimer Disease & Associated Disorders*, 14(3):168–175, 2000.

Laura A Hatfield, Mark E Boye, and Bradley P Carlin. Joint modeling of multiple longitudinal patient-reported outcomes and survival. *Journal of Biopharmaceutical Statistics*, 21(5):971–991, 2011.

Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

Graeme L Hickey, Pete Philipson, Andrea Jorgensen, and Ruwanthi Kolamunnage-Dona. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, 16(1): 1–15, 2016.

Graeme L Hickey, Pete Philipson, Andrea Jorgensen, and Ruwanthi Kolamunnage-Dona. joinerml: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC medical research methodology*, 18(1):1–14, 2018.

Joseph W Hogan and Nan M Laird. Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. *Statistical Methods in Medical Research*, 7(1):28–48, 1998.

Wenzheng Huang, Scott L Zeger, James C Anthony, and Elizabeth Garrett. Latent variable model for joint analysis of multiple repeated measures and bivariate event times. *Journal of the American Statistical Association*, 96(455):906–914, 2001.

Joseph G Ibrahim, Ming-Hui Chen, and Debajyoti Sinha. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*, pages 863–883, 2004.

Joseph G Ibrahim, Haitao Chu, and Liddy M Chen. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16):2796, 2010.

Banks J., Batty G.D., Breedvelt J.J.F., Coughlin K., Crawford R., Marmot M., Nazroo J., Oldfield Z., Steel N., Steptoe A., Wood Martin., and P. Zaninotto. English longitudinal study of ageing: Waves 0-9, 1998-2019, 2021. URL `http://doi.org/10.5255/UKDA-SN-5050-23`.

Marshall M Kaplan. Primary biliary cirrhosis. *New England Journal of Medicine*, 335(21):1570–1580, 1996.

David G Kleinbaum and Mitchel Klein. *Survival Analysis*, volume 3. Springer, New York, 2010.

Nan Laird. Analysis of longitudinal and cluster-correlated data. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 8:i–155, 2004. ISSN 19355920, 23290978. URL `http://www.jstor.org/stable/4153193`.

Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

Craig Lammert, Douglas L Nguyen, Brian D Juran, Erik Schlicht, Joseph J Larson, Elizabeth J Atkinson, and Konstantinos N Lazaridis. Questionnaire based assessment of risk factors for primary biliary cirrhosis. *Digestive and Liver Disease*, 45(7):589–594, 2013.

Konstantinos N Lazaridis and Jayant A Talwalkar. Clinical epidemiology of primary biliary cirrhosis: incidence, prevalence, and impact of therapy. *Journal of clinical gastroenterology*, 41(5):494–500, 2007.

Luc Letenneur, Daniel Commenges, Jean-François Dartigues, and Pascale Barberger-Gateau. Incidence of dementia and alzheimer's disease in elderly com-

munity residents of south-western france. *International journal of epidemiology*, 23(6):1256–1261, 1994.

Haiqun Lin, Charles E McCulloch, Bruce W Turnbull, Elizabeth H Slate, and Larry C Clark. A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine*, 19(10):1303–1318, 2000.

Xian Liu. *Survival analysis: models and applications*. John Wiley & Sons, 2012.

Courtney A Lockwood, Yael Mansoor, Elizabeth Homer-Smith, and James A Moses Jr. Factor structure of the benton visual retention tests: Dimensionalization of the benton visual retention test, benton visual retention test–multiple choice, and the visual form discrimination test. *The Clinical Neuropsychologist*, 25(1):90–107, 2011.

Katya Mauff, Ewout Steyerberg, Isabella Kardys, Eric Boersma, and Dimitris Rizopoulos. Joint models with multiple longitudinal outcomes and a time-to-event outcome: a corrected two-stage approach. *Statistics and Computing*, 30(4):999–1014, 2020.

Clinic Mayo. Primary biliary cholangitis - symptoms and causes, 2022. URL `https://www.mayoclinic.org/diseases-conditions/primary-biliary-cholangitis/symptoms-causes/syc-20376874`.

P McCullagh and JA Nelder. Generalized linear models, 1989.

Tim P Morris, Ian R White, and Michael J Crowther. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11):2074–2102, 2019.

Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

C Nejjari, JF Tessier, I Baldi, P Barberger-Gateau, JF Dartigues, and R Salamon. Epidemiologic aspects of respiratory aging: contribution of the paquid survey. *Revue d'epidemiologie et de sante publique*, 45(5):417–428, 1997.

Hans Bruun Nielsen and Stig Bousgaard Mortensen. *ucminf: General-Purpose Unconstrained Non-Linear Optimization*, 2016. URL `https://CRAN.R-project.org/package=ucminf`. R package version 1.1-4.

World Health Organization et al. Who report on global surveillance of epidemic-prone infectious diseases. Technical report, World Health Organization, 2000.

Verna C Pangman, Jeff Sloan, and Lorna Guse. An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice. *Applied Nursing Research*, 13(4):209–213, 2000.

Nikos Pantazis and Giota Touloumi. Robustness of a parametric model for informatively censored bivariate longitudinal data under misspecification of its distributional assumptions: A simulation study. *Statistics in Medicine*, 26(30):5473–5485, 2007.

Pete Philipson, Peter Diggle, Ines Sousa, Ruwanthi Kolamunnage-Dona, Paula Williamson, and Robin Henderson. *joineR: Joint modelling of repeated measurements and time-to-event data*, 2012. URL `https://github.com/graemeleehickey/joineR/`. R package version 1.2.6.

John H Pollard and Emil J Valkovics. The gompertz distribution and its applications. *Genus*, pages 15–28, 1992.

Cécile Proust-Lima, Pierre Joly, Jean-François Dartigues, and Hélène Jacqmin-Gadda. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a nonlinear latent class approach. *Computational Statistics & Data Analysis*, 53(4):1142–1154, 2009.

Cécile Proust-Lima, Mbéry Séne, Jeremy MG Taylor, and Hélène Jacqmin-Gadda. Joint latent class models for longitudinal and time-to-event data: a review. *Statistical methods in medical research*, 23(1):74–90, 2014.

Cécile Proust-Lima, Jean-François Dartigues, and Hélène Jacqmin-Gadda. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. *Statistics in Medicine*, 35(3):382–398, 2016.

Cécile Proust-Lima, Viviane Philipps, and Benoit Liquet. Estimation of extended mixed models using latent classes and latent processes: The R package lcmm. *Journal of Statistical Software*, 78(2):1–56, 2017. doi: 10.18637/jss.v078.i02.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`.

Dimitris Rizopoulos. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9):1–33, 2010. URL `https://doi.org/10.18637/jss.v035.i09`.

Dimitris Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data: with Applications in R*. Chapman and Hall/CRC, Boca Raton, 2012.

Dimitris Rizopoulos and Pulak Ghosh. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30(12):1366–1380, 2011.

Dimitris Rizopoulos, Geert Verbeke, and Geert Molenberghs. Shared parameter models under random effects misspecification. *Biometrika*, 95(1):63–74, 2008.

Dimitris Rizopoulos, Grigorios Papageorgiou, and Pedro Miranda Afonso. *JMbayes2: Extended Joint Models for Longitudinal and Time-to-Event Data*, 2022. https://drizopoulos.github.io/JMbayes2/, https://github.com/drizopoulos/JMbayes2.

Virginie Rondeau, Hélène Jacqmin-Gadda, Daniel Commenges, Catherine Helmer, and Jean-François Dartigues. Aluminum and silica in drinking water and the risk

of alzheimer's disease or cognitive decline: findings from 15-year follow-up of the paquid cohort. *American Journal of Epidemiology*, 169(4):489–496, 2008.

Virginie Rondeau, Yassin Mazroui, and Juan R. Gonzalez. frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, 47(4):1–28, 2012. URL `https://www.jstatsoft.org/v47/i04/`.

Anaïs Rouanet, Pierre Joly, Jean-François Dartigues, Cécile Proust-Lima, and Hélène Jacqmin-Gadda. Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia. *Biometrics*, 72(4):1123–1135, 2016.

Montserrat Rue, Eleni-Rosalina Andrinopoulou, Danilo Alvares, Carmen Armero, Anabel Forte, and Lluis Blanch. Bayesian joint modeling of bivariate longitudinal and competing risks data: an application to study patient-ventilator asynchronies in critical care patients. *Biometrical Journal*, 59(6):1184–1203, 2017.

Jerry L Rushton, Michelle Forcier, and Robin M Schectman. Epidemiology of depressive symptoms in the national longitudinal study of adolescent health. *Journal of the American Academy of Child & Adolescent Psychiatry*, 41(2):199–205, 2002.

Statisticat and LLC. *LaplacesDemon: Complete Environment for Bayesian Inference*, 2021. URL `https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software`. R package version 16.1.6.

Arthur S Tamkin and Joseph T Kunce. A comparison of three neuropsychological tests: the weigl, hooper, and benton. *Journal of clinical psychology*, 41(5):660–664, 1985.

Rebecca Taylor, Laura Conway, Lisa Calderwood, Carli Lessof, Hayley Cheshire, Kate Cox, and Shaun Scholes. Health, wealth and lifestyles of the older popula-

tion in england: The 2002 english longitudinal study of ageing technical report. *London: Institute of Fiscal Studies*, 2007.

Tom N Tombaugh and Nancy J McIntyre. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*, 40(9):922–935, 1992.

Giota Touloumi, Nikos Pantazis, Abdel G Babiker, Sarah A Walker, Olga Katsarou, Anastasia Karafoulidou, Angelos Hatzakis, Kholoud Porter, CASCADE Collaboration, et al. Differences in hiv rna levels before the initiation of antiretroviral therapy among 1864 individuals with known hiv-1 seroconversion dates. *Aids*, 18(12):1697–1705, 2004.

Jolien P Tuijl, Evert M Scholte, Anton JM de Craen, and Roos C van der Mast. Screening for cognitive impairment in older general hospital patients: comparison of the six-item cognitive impairment test with the mini-mental state examination. *International Journal of Geriatric Psychiatry*, 27(7):755–762, 2012.

Nick van Boven, Linda C Battes, K Martijn Akkerhuis, Dimitris Rizopoulos, Kadir Caliskan, Sharda S Anroedh, Wisam Yassi, Olivier C Manintveld, Jan-Hein Cornel, Alina A Constantinescu, et al. Toward personalized risk assessment in patients with chronic heart failure: detailed temporal patterns of nt-probnp, troponin t, and crp in the bio-shift study. *American Heart Journal*, 196:36–48, 2018.

Ardo Van Den Hout. *Multi-state survival models for interval-censored data*. CRC Press, Boca Raton, 2017.

Ardo Van Den Hout and Graciela Muniz-Terrera. Joint models for discrete longitudinal outcomes in aging research. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(1):167–186, 2016.

Ardo Van Den Hout and Graciela Muniz-Terrera. Hidden three-state survival model for bivariate longitudinal count data. *Lifetime Data Analysis*, 25(3):529–545, 2019.

Andrea Vertesi, Judith A Lever, D William Molloy, Brett Sanderson, Irene Tuttle, Laura Pokoradi, and Elaine Principi. Standardized mini-mental state examination. use and interpretation. *Canadian Family Physician*, 47(10):2018–2023, 2001.

W Bruce Walsh and Nancy E Betz. *Tests and Assessment*. Prentice-Hall, Inc, New Jersey, 1995.

Chen Wang, Jeffrey Douglas, and Susan Anderson. Item response models for joint analysis of quality of life and survival. *Statistics in Medicine*, 21(1):129–142, 2002.

Andreas Wienke. *Frailty Models in Survival Analysis*. Chapman and Hall/CRC, Boca Raton, 2010.

Michael S Wulfsohn and Anastasios A Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339, 1997.

Anatoli I Yashin, Svetlana V Ukraintseva, Konstantin G Arbeev, Igor Akushevich, Liubov S Arbeeva, and Alexander M Kulminski. Maintaining physiological state for exceptional survival: What is the normal level of blood glucose and does it change with age? *Mechanisms of ageing and development*, 130(9):611–618, 2009.

Anatoliy I Yashin, Konstantin G Arbeev, Svetlana V Ukraintseva, Igor Akushevich, and Alexander Kulminski. Patterns of aging-related changes on the way to 100: an approach to studying aging, mortality, and longevity from longitudinal data. *North American Actuarial Journal*, 16(4):403–433, 2012.