



Universiteit  
Leiden  
The Netherlands

## Statistical methods for frailty models: studies on old-age mortality and recurrent events

Böhnstedt, M.

### Citation

Böhnstedt, M. (2021, November 30). *Statistical methods for frailty models: studies on old-age mortality and recurrent events*. Retrieved from <https://hdl.handle.net/1887/3243966>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3243966>

**Note:** To cite this publication please use the final published version (if applicable).

**Statistical Methods for Frailty Models:  
Studies on Old-Age Mortality and Recurrent Events**

Marie Böhnstedt

Cover design: Marie Böhnstedt and adiant druck  
Printing: adiant druck, Roggentin

©2021 Marie Böhnstedt, Rostock, Germany.  
All rights reserved. No part of this publication may be reproduced without prior written permission of the author or, where applicable, of the copyright-owning journal or publisher.

ISBN: 978-3-00-070063-7

Statistical Methods for Frailty Models:  
Studies on Old-Age Mortality and Recurrent Events

Proefschrift

ter verkrijging van  
de graad van doctor aan de Universiteit Leiden,  
op gezag van rector magnificus prof.dr.ir. H. Bijl,  
volgens besluit van het college voor promoties  
te verdedigen op dinsdag 30 november 2021  
klokke 11.15 uur

door

Marie Böhnstedt  
geboren te Stralsund, Duitsland,  
in 1990

Promotor: Prof. dr. H. Putter

Co-promotor: Dr. J. Gampe  
· *Max Planck Institute for Demographic Research,  
Rostock*

Leden promotiecommissie: Prof. dr. S. le Cessie  
Dr. N. van Geloven  
Prof. dr. M.J.C. Eijkemans  
· *Universitair Medisch Centrum Utrecht, Utrecht*  
Dr. G.E. Bijwaard  
· *Netherlands Interdisciplinary Demographic Institute,  
The Hague*

The PhD research was conducted at the Max Planck Institute for Demographic Research, Rostock, Germany.

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Frailty modeling of survival data . . . . .	2
1.3	Assessing mortality deceleration . . . . .	3
1.3.1	The gamma-Gompertz model . . . . .	3
1.3.2	Old-age mortality studies: challenges and design . . . . .	5
1.3.3	Statistical techniques for assessing mortality deceleration . . . . .	5
1.4	Joint modeling of recurrent events and a terminal event . . . . .	7
1.4.1	The joint frailty model . . . . .	7
1.4.2	Different observational settings . . . . .	7
1.4.3	Inference in the joint frailty model . . . . .	8
1.5	Outline of the thesis . . . . .	8
<b>I</b>	<b>Assessing mortality deceleration</b>	<b>11</b>
<b>2</b>	<b>Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model</b>	<b>13</b>
2.1	Introduction . . . . .	13
2.2	Preliminaries . . . . .	15
2.3	Asymptotic distribution of maximum likelihood estimator . . . . .	16
2.4	Asymptotics for the likelihood ratio test . . . . .	18
2.5	AIC for the gamma-Gompertz model . . . . .	20
2.6	Discussion . . . . .	21
2.7	Supplementary material . . . . .	21
<b>3</b>	<b>Information measures and design issues in the study of mortality deceleration: Findings for the gamma-Gompertz model</b>	<b>25</b>
3.1	Introduction . . . . .	26
3.2	Framework: Gamma-Gompertz model, Fisher information & study design	28
3.2.1	Gamma-Gompertz model . . . . .	28
3.2.2	The Fisher information . . . . .	29
3.2.3	Optimal design . . . . .	31

3.3	Information measures in the gamma-Gompertz model . . . . .	32
3.3.1	Fisher information in the gamma-Gompertz model . . . . .	32
3.3.2	$D_A$ -optimality in the gamma-Gompertz model . . . . .	33
3.3.3	Interpretation of information measures in a non-standard setting . . . . .	33
3.3.4	Power of the likelihood ratio test . . . . .	34
3.4	Design considerations in assessing mortality deceleration . . . . .	35
3.5	Empirical results . . . . .	36
3.5.1	Effect of the age at left truncation . . . . .	37
3.5.2	Sample size considerations . . . . .	41
3.5.3	Power of the likelihood ratio test . . . . .	42
3.6	A study on old-age mortality among French-Canadians . . . . .	43
3.7	Discussion . . . . .	46
3.8	Supplementary material . . . . .	47
<b>4</b>	<b>Shifting attention to old age: Detecting mortality deceleration using focused model selection</b>	<b>63</b>
4.1	Introduction . . . . .	64
4.2	Mortality at advanced ages . . . . .	67
4.2.1	Hazard models . . . . .	67
4.2.2	Traditional approaches to inference . . . . .	68
4.3	Focused information criterion for mortality deceleration . . . . .	69
4.3.1	Rationale for FIC . . . . .	70
4.3.2	FIC with a parameter on the boundary of the parameter space . . . . .	71
4.3.3	Choice of the focus parameter . . . . .	73
4.3.4	A modified AIC for the gamma-Gompertz model . . . . .	73
4.4	Simulation study . . . . .	74
4.5	Mortality of French-Canadians at high ages . . . . .	77
4.5.1	Data and age validation . . . . .	77
4.5.2	Comparison of information criteria . . . . .	77
4.5.3	Including effects of year of birth . . . . .	79
4.6	Discussion . . . . .	84
4.7	Supplementary material . . . . .	86
<b>II</b>	<b>Joint modeling of recurrent events and death</b>	<b>107</b>
<b>5</b>	<b>Joint modeling of interval counts of recurrent events and death</b>	<b>109</b>
5.1	Introduction . . . . .	110
5.2	Joint frailty model and interval counts . . . . .	112
5.3	Methods . . . . .	114
5.3.1	Estimation of the joint frailty model based on interval counts . . . . .	114
5.3.2	Score test for the association between recurrences and death . . . . .	116
5.4	Simulation study . . . . .	117

5.4.1	Performance of the estimation method . . . . .	117
5.4.2	Performance of the score test . . . . .	122
5.5	Fertility and mortality in <i>Eleutheria dichotoma</i> . . . . .	124
5.6	Discussion . . . . .	127
5.7	Appendix . . . . .	129
5.8	Supplementary material . . . . .	132
<b>6</b>	<b>Incorporating delayed entry into the joint frailty model for recurrent events and a terminal event</b>	<b>151</b>
6.1	Introduction . . . . .	152
6.2	Joint frailty model and left truncation . . . . .	154
6.2.1	Joint frailty model . . . . .	154
6.2.2	Adjusting for left truncation . . . . .	156
6.3	Estimation of the joint frailty model . . . . .	158
6.4	Simulation study . . . . .	160
6.4.1	Settings . . . . .	160
6.4.2	Estimation and results . . . . .	162
6.5	Recurrent infections and mortality in an older population . . . . .	167
6.6	Discussion . . . . .	170
6.7	Supplementary material . . . . .	172
	<b>English summary</b>	<b>183</b>
	<b>Nederlandse samenvatting</b>	<b>187</b>
	<b>List of publications</b>	<b>191</b>
	<b>Acknowledgements</b>	<b>193</b>
	<b>Curriculum vitae</b>	<b>195</b>





# 1

## Introduction

### 1.1 Introduction

The heterogeneity between individuals is a key aspect in the statistical studies of various phenomena, including mortality and recurrent infections or fertility events. The standard approach for incorporating the differences in observed individual characteristics into the analysis are regression models. However, unobserved individual characteristics often add to the variability in the data as well. Therefore, the statistical models also need to take this unobserved heterogeneity between individuals into account.

When studying duration data that describe the time until an event of interest occurs, the heterogeneity between individuals will also lead to changes in the composition of the sample over time. These selection effects can have serious implications for the statistical analysis in different contexts. In mortality studies, for instance, the sample of survivors to a certain age may have different characteristics than the initial population. Similar effects arise in longitudinal studies if the latent characteristics affect not only the longitudinal component, but also the survival time of the individuals. Thus, in settings with delayed entry, the analysis is based on the selective sample of individuals who survived to the point of entering the study. Even if there is no delayed entry, an association between the longitudinal component and the survival time will result in a dependent drop-out mechanism.

The modeling of unobserved heterogeneity has greatly advanced in recent decades. A common approach to modeling the latent information is through random effects. In the analysis of time-to-event data, such a random effect is usually referred to as frailty. In

the following section, we give a brief general introduction to frailty modeling of survival data. Then, we present the two specific frailty models that we study in this thesis to investigate recurrent events, such as repeated infections or fertility events, and old-age mortality, respectively. An outline of the thesis is provided in Section 1.5.

## 1.2 Frailty modeling of survival data

Time-to-event data are commonly modeled through the hazard function, which describes the instantaneous probability of the occurrence of the event of interest for individuals who have not yet experienced the event. As these risks of the occurrence of an event can naturally differ between individuals, the statistical models have to allow for heterogeneity. To account for the differences that are attributable to observed characteristics, covariates can be included in the analysis. However, in most cases, not all characteristics that cause the hazards to vary between individuals can be observed. Frailty models provide an appealing approach to accounting for such unobserved heterogeneity. For example, in the class of frailty proportional hazards models, this type of heterogeneity is incorporated through a positive random effect, the frailty, that acts multiplicatively on a common baseline hazard. For univariate survival data, such frailty models have already been proposed by Beard (1959) and Vaupel et al. (1979).

Unobserved heterogeneity can also arise in studies of multivariate survival data. When analyzing clustered survival data, such as survival times of individuals from different families, the risks of experiencing the event may vary across the clusters. Shared frailty models, in which the value of the random effect is common to all individuals in a cluster, not only take into account the heterogeneity between clusters, but also relax the assumption of independence between the individuals within a cluster. Similarly, shared frailty models are applied in recurrent event studies in which the same type of event can occur repeatedly in the same individual. Individual-specific random effects can then accommodate individual differences in the event rates, as well as the dependence between the recurrence times within an individual. An overview of frailty modeling for multivariate survival data has been provided by, among others, Hougaard (2000).

An important feature of survival data is that they are usually subject to censoring or truncation. For example, if we only know that the individual has not yet experienced the event of interest at a certain point in time, the observation is right-censored. If an individual is included in the study only if he or she has not yet experienced the event of interest by some point in time – that is, if the event time is larger than a certain value – the observation is left-truncated. The statistical methods need to be adapted to such observation schemes. But given that even with these adjustments, the loss of information caused by the incomplete observations may pose considerable challenges for the analysis, this issue should be taken into consideration already when the study is designed.

Another central issue that should be taken into account in time-to-event analysis is that selection effects can occur if there is heterogeneity in the event rates. In studies of a single failure event, the individuals who survive up to a certain point in time will tend to

have lower failure risks than the individuals who already had a failure before that point in time. In the context of frailty models, this implies that the frailty distribution in the sample of survivors changes over time. Ignoring such selection effects may have serious consequences for the inference results.

The issues outlined above also play an important role in the study of the two distinct frailty models that are investigated in this thesis. In the first part, we consider a specific proportional hazards frailty model for univariate survival data that is used as a framework for assessing a demographic phenomenon known as mortality deceleration. In the second part, we examine inference in a joint frailty model for recurrent events and a terminal event under different observation schemes. In the following sections, we briefly introduce the two models, and provide some background on the applications and the statistical approaches used in this thesis.

## 1.3 Assessing mortality deceleration

The first part of this thesis was motivated by studies on old-age mortality patterns, which have frequently observed that human death rates slow down at advanced ages. This mortality deceleration can be described through a specific proportional hazards frailty model. However, non-standard conditions for the inference and limited data availability can complicate the statistical analysis. Thus, our aim is to investigate both traditional and new approaches to assessing mortality deceleration in this setting.

### 1.3.1 The gamma-Gompertz model

The observation that human mortality rates increase exponentially over most of the adult lifespan was already made by Gompertz (1825). However, later investigations based on improved vital registration records suggested that death rates slow down at the oldest ages (Thatcher et al., 1998). Such a downward deviation from the exponentially increasing hazard at high ages is called mortality deceleration, and can be explained by selection effects due to heterogeneous mortality risks. As individuals with higher mortality risks tend to die at younger ages, the population of survivors to advanced ages tends to be made up of individuals with lower mortality risks. This, in turn, results in a comparatively lower population hazard at these advanced ages.

Mortality deceleration can therefore be examined in the framework of a frailty proportional hazards model. The individual risks at any age  $x$  are specified in terms of the conditional hazards  $h(x|Z = z) = z \cdot h_0(x)$  for given frailty value  $Z = z$ , with  $h_0(x)$  denoting the baseline hazard. More specifically, we consider the gamma-Gompertz model in which the individual risks are defined by an exponentially increasing Gompertz baseline hazard,  $h_0(x) = ae^{bx}$ , that is multiplied by a gamma distributed frailty. The frailty variance  $\sigma^2$  describes the amount of heterogeneity in mortality risks. In the presence of

heterogeneity, indicated by a positive value of the frailty variance, the marginal population hazard,

$$h(x) = \frac{ae^{bx}}{1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)}, \quad x \geq 0, \quad a, b > 0, \quad \sigma^2 \geq 0,$$

decelerates at older ages, as illustrated in Figure 1.1. If there is no heterogeneity and the frailty variance is zero, the population hazard is of the same exponentially increasing Gompertz form as the individual hazards. Hence, in the setting of the gamma-Gompertz model, the frailty variance determines whether the population hazard does or does not decelerate at older ages, and the statistical assessment of the phenomenon will therefore largely depend on this variance parameter.

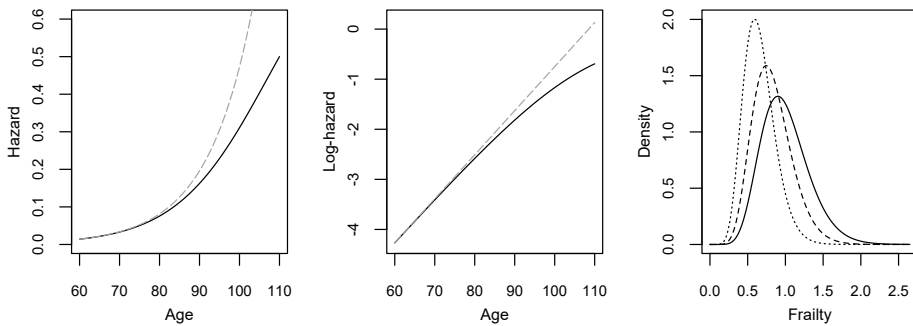


Figure 1.1: Hazard (left) and log-hazard (middle) of the gamma-Gompertz model with parameters  $a = 0.014$ ,  $b = 0.088$ , and  $\sigma^2 = 0.1$  (black-solid) or  $\sigma^2 = 0$  (gray-dashed). Right: Frailty distribution at the starting age of 60 (solid) and among survivors to ages 90 and above (dashed), or to ages 100 and above (dotted), which illustrates that the distribution of frailty among survivors to higher ages is concentrated at lower values, and has smaller variance.

Assuming a fully parametric model for human adult mortality may seem restrictive in light of the advances in semiparametric and nonparametric survival analysis. However, the simple relation of a linear increase of the log mortality rate with age, which is imposed by the Gompertz model, has been repeatedly found to hold across a large part of the age range for human adults in various studies covering different countries and time periods. In addition to its importance in actuarial studies and demography, the Gompertz model has been applied in research on the biology of aging. While the search for an explanation for the exponential increase in mortality with age is still ongoing, it has been shown that models of damage accumulation in specific systems can produce such patterns (Gavrilov and Gavrilova, 2001; Kirkwood, 2015).

### 1.3.2 Old-age mortality studies: challenges and design

Several challenges arise in the study of mortality deceleration. In the setting of the gamma-Gompertz model, the absence of mortality deceleration corresponds to a value of zero for the frailty variance, which lies on the boundary of the parameter space. The presence of such a ‘boundary parameter’ introduces a non-standard condition, such that the standard asymptotic results for likelihood-based inference, which are derived under regularity conditions, will generally not apply (Self and Liang, 1987). Therefore, the statistical techniques for assessing mortality deceleration need to be adapted to the non-standard setting.

Empirical studies of mortality deceleration often face the challenge that the available data are limited. Even if the studies are based on large populations, relatively few deaths will be observed at the advanced ages at which the deceleration is most pronounced. This is an inherent feature when studying phenomena that occur at the tail of a distribution, where the data naturally become sparse. Moreover, scientific validation of the ages at death is often considered necessary in order to avoid biased mortality estimates due to age misreporting. Because of the time and the costs involved in verifying individuals’ vital records, mortality data are usually validated for only parts of the population of interest. When studying old-age mortality, samples are often restricted to include only those individuals who survived beyond some advanced age, such as survivors beyond age 80. In the setting of the gamma-Gompertz model, the sample of survivors to ages 80 and above will be less heterogeneous in their mortality risks than the sample of survivors to, say, ages 70 and above, due to selection effects (see the right panel of Figure 1.1 for an illustration). Hence, the left truncation of the observations will affect the inference in this model, and, thus, the ability to detect a slowing down of the death rates.

For the purposes of quantifying the effects of the age of left truncation and the sample size on the assessment of mortality deceleration, information measures can be derived from the Fisher information matrix of the gamma-Gompertz model. This approach draws on the concept of the Fisher information as a means of measuring the amount of information the data carry about the model parameters (Lehmann, 1999). Applying ideas from the theory of optimal design (Atkinson, 1988), study designs can be evaluated based on different optimality criteria, which are defined in terms of scalar functions of the information matrix. As a result, we will be able to gauge, for instance, the amount of information that can be gained by adding the observations on the deaths at ages 85-89 to an existing data set of survivors to ages 90 and above.

### 1.3.3 Statistical techniques for assessing mortality deceleration

The non-standard condition of the boundary parameter and the aspects of the study design laid out above will have an impact on the statistical methods used in the study of mortality deceleration.

In the framework of the gamma-Gompertz model, one approach for assessing the phenomenon is a likelihood ratio test for a zero frailty variance. Taking into account the non-

standard condition that the frailty variance lies on the boundary of the parameter space under the null hypothesis is essential for deriving asymptotic properties of the test. Under the null hypothesis, the test statistic is not asymptotically chi-squared distributed with one degree of freedom, but instead asymptotically follows a 50:50-mixture distribution of a point mass at zero and a chi-squared distribution with one degree of freedom. We also determine a formula for the local asymptotic power of the likelihood ratio test, which involves one of the information measures derived from the Fisher information. Calculations for specific scenarios will illustrate that the test's power to detect mortality deceleration is low if the data are left-truncated at relatively high ages.

A drawback of the hypothesis testing approach is that it is designed to control for the probability of committing a type I error, which in this case is the probability of falsely rejecting the hypothesis of no mortality deceleration. However, the probability of a type II error – that is, of failing to reject the hypothesis of no mortality deceleration although mortality deceleration is present – is not directly controlled, and may be relatively large. This asymmetry of hypothesis testing suggests that alternative methods should be considered here.

A different approach to assessing mortality deceleration is model selection based on information criteria (Burnham and Anderson, 2002). The gamma-Gompertz model is simplified to the Gompertz model if the frailty variance takes the value of zero and there is no mortality deceleration. Thus, we have to choose between two nested models that differ by one parameter only. However, as the selection concerns a boundary parameter, the standard formulas for the information criteria may not be applicable. In Chapter 2, we show that the standard version of the popular Akaike information criterion (AIC; Akaike, 1974) is biased in the setting of the gamma-Gompertz model due to this non-standard condition. Moreover, even after adding a correction term to reduce the bias, model selection based on the AIC will not enable us to reliably detect mortality deceleration for samples that include only survivors to advanced ages.

The AIC was constructed to evaluate overall model performance, whereas in certain applications, model performance for a specific parameter of interest might be more relevant. Claeskens and Hjort (2003) introduced a focused information criterion (FIC) for selecting a model that performs best for a given focus parameter. From a set of nested candidate models, the model with the smallest limiting risk of the estimator of the focus parameter is chosen. This approach is appealing when studying mortality deceleration, because attention can be directed to the quantities that determine old-age mortality, such as the frailty variance or the log-hazard function at some advanced age. As the original FIC was developed under standard regularity conditions, the non-standard condition of the boundary parameter in the present application requires us to derive a new version of the FIC. The performance of this new criterion for detecting mortality deceleration will be assessed in comparison with the performance of the modified AIC.

## 1.4 Joint modeling of recurrent events and a terminal event

The second part of this thesis is devoted to the study of recurrent events in the presence of a terminal event. In particular, methods for inference in a joint frailty model are developed under two specific observation schemes.

### 1.4.1 The joint frailty model

Recurrent events are found in various areas of application, including medical studies on tumor occurrences, demographic studies on fertility, and actuarial studies on insurance claims. Hence, recurrent events have received considerable attention in the recent years; an overview is provided by Cook and Lawless (2007).

In many cases, a terminal event such as death might stop the repeated occurrences of the event of interest in an individual. This potentially dependent censoring of the recurrence process requires that the two event processes are modeled jointly. In an extension of the shared frailty model, Liu et al. (2004) proposed a joint frailty model for recurrent events and a terminal event. As in the shared frailty model for recurrent events, the individual-specific frailties induce heterogeneity in the recurrence rates, and account for the dependence between the recurrences within one individual. But in the joint model, the frailties also affect the hazard of the terminal event, thereby introducing dependence between the recurrence process and the terminal event. The direction and the strength of the association can be modified through an additional dependence parameter.

Consequently, the model not only provides a suitable framework for handling the dependent censoring of the recurrent event process; but it also allows us to assess how the recurrent events and the terminal event are related. The question of whether individuals with a higher recurrence rate simultaneously have a higher or even a lower risk of experiencing the terminal event is of relevance in several applications. This issue is illustrated with a study on the fertility and mortality of a marine organism in Chapter 5. Moreover, analyses based on the joint frailty model can provide insights into how the risks of experiencing recurrent events or the terminal event evolve with time.

### 1.4.2 Different observational settings

The joint frailty model is usually studied in a setting in which each individual is observed from the start of the event processes and the exact times of event recurrence are available. In this thesis, we consider two variations of this observation scheme. First, we deal with a situation in which the recurrence process is observed intermittently, so that only the numbers of recurrent events that occurred between successive observation times are known. This scheme of interval counts of recurrent events can arise in different contexts, such as in a medical study in which the number of epileptic seizures of a patient is recorded at scheduled visits to the doctor, or in a laboratory experiment in which rats are examined for newly developed tumors at fixed inspection times.



Second, we consider a situation in which the recurrence process is continuously observed, but the observation does not begin until sometime after the start of the event processes. Such cases of delayed entry can, for instance, occur in studies of certain diseases in which time is measured from the diagnosis onwards, but patients are enrolled at various points in time after their diagnosis. Another example is that of studies in which age is the main time scale, but individuals enter the study at different ages. In these settings, individuals are included in the study only if they are still at risk of experiencing the terminal event; that is, if the event has not yet occurred for them. The resulting left-truncated study sample may not be representative of the underlying population, as it could consist of individuals who tend to have a lower hazard of experiencing the terminal event. These selection effects can also cause the frailty distribution in the study sample to differ from the initial distribution of frailties in the target population. Hence, inference methods that take the left truncation properly into account must be developed.

### 1.4.3 Inference in the joint frailty model

Several estimation procedures for the joint frailty model have been proposed for settings in which the recurrence process is observed from the start, and the exact recurrence times are known. The methods are frequently based on the marginal likelihood, which can be approximated using Gaussian quadrature. As the likelihood can be easily adjusted to different observation schemes, we can also adopt this approach for estimating the joint frailty model in the above situations with intermittent observations or delayed entry. More precisely, we build on the work of Liu and Huang (2008), who used Gauss-Hermite quadrature to approximate the likelihood and specified the event rates as piecewise-constant functions.

Given the complexity of the joint frailty model, it may be of interest to test a priori whether the recurrence process and the terminal event are associated. For that purpose, Balan et al. (2016) developed a score test for the association between the two event processes in the setting in which recurrence times are exactly observed. We adapt this test to the situation in which there are interval counts of the recurrent events.

## 1.5 Outline of the thesis

The two parts of this thesis are structured as follows. The first part encompasses three chapters that cover the different aspects of assessing mortality deceleration in the framework of the gamma-Gompertz model.

In Chapter 2, we present asymptotic results for likelihood-based inference in the gamma-Gompertz model in a local misspecification setting. We determine the limiting distribution of the maximum likelihood estimator and an approximation of the local asymptotic power of the likelihood ratio test for a zero frailty variance under the non-standard condition of the boundary parameter. In addition, we derive the bias of the standard AIC in this model with a boundary parameter.

To address questions relevant to empirical studies of mortality deceleration, we investigate in Chapter 3 how the sample size and the age range covered by a data set affect the assessment of mortality deceleration. The corresponding information measures based on the Fisher information matrix of the gamma-Gompertz model are described. One of the criteria is used in calculating the approximate local power of the likelihood ratio test according to the formula derived in Chapter 2. We demonstrate that samples that are relatively small in size or that cover only the most advanced ages can make drawing reliable inferences difficult.

Chapter 4 discusses the approach of using model selection to assess mortality deceleration. The new version of the FIC is introduced and its performance is evaluated, especially for more demanding study designs. Comparisons between this criterion and a modified AIC with a bias correction term, which is motivated by the result from Chapter 2, are also included.

The second part of the thesis is comprised of two chapters about the inference in the joint frailty model for recurrent events and a terminal event under two different observational schemes. Chapter 5 focuses on the situation in which only interval counts of recurrent events are available. We present the method for estimating the joint frailty model by direct maximization of the approximate marginal likelihood. The question of whether the recurrence process and the terminal event process are associated is addressed by means of a score test.

In Chapter 6, the joint frailty model is studied in the situation with delayed entry. We outline the construction of the likelihood, and propose an estimation procedure based on an approximated marginal likelihood. We demonstrate the importance of taking into account the selection effects on the frailty distribution in the left-truncated sample.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.
- Atkinson, A. C. (1988). Recent developments in the methods of optimum and related experimental designs. *International Statistical Review* 56(2), 99–115.
- Balan, T. A., S. E. Boonk, M. H. Vermeer, and H. Putter (2016). Score test for association between recurrent events and a terminal event. *Statistics in Medicine* 35, 3037–3048.
- Beard, R. E. (1959). Note on some mathematical mortality models. In G. E. W. Wolstenholme and M. O'Connor (Eds.), *The Lifespan of Animals*, Ciba Foundation Colloquia on Ageing, pp. 302–311. Boston: Little, Brown.
- Burnham, K. P. and D. R. Anderson (2002). *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*. Springer.

- Claeskens, G. and N. L. Hjort (2003). The focused information criterion. *Journal of the American Statistical Association* 98, 900–916.
- Cook, R. J. and J. F. Lawless (2007). *The Statistical Analysis of Recurrent Events*. Statistics for Biology and Health. New York: Springer.
- Gavrilov, L. A. and N. S. Gavrilova (2001). The reliability theory of aging and longevity. *Journal of Theoretical Biology* 213, 527 – 545.
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* 115, 513–583.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Statistics for Biology and Health. New York: Springer.
- Kirkwood, T. B. L. (2015). Deciphering death: a commentary on Gompertz (1825) ‘On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies’. *Philosophical Transactions of the Royal Society of London B* 370. doi: 10.1098/rstb.2014.0379.
- Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer texts in statistics. New York: Springer.
- Liu, L. and X. Huang (2008). The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine* 27, 2665–2683.
- Liu, L., R. A. Wolfe, and X. Huang (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* 60, 747–756.
- Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 82(398), 605–610.
- Thatcher, A. R., V. Kannisto, and J. W. Vaupel (1998). *The force of mortality at ages 80 to 120*. Odense Monographs on Population Aging 5. Odense University Press.
- Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3), 439–454.

## **Part I**

# **Assessing mortality deceleration**



# 2

## Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model

### Abstract

We study the asymptotic properties of the maximum likelihood estimator and the likelihood ratio test in the gamma-Gompertz model for local alternatives. We also show that the standard AIC is biased in this model due to the boundary parameter.

### 2.1 Introduction

Benjamin Gompertz (1825) pioneered human mortality research by demonstrating that death rates for adults increase exponentially with age. Since then, the Gompertz distribution has been widely used to model adult lifespans. It was not until more and better data at high ages became available that the overall validity of the Gompertz distribution was called into question. Downward deviations from an exponentially increasing hazard (*mortality deceleration*) were observed when data for the oldest-old were analyzed

---

This chapter has been published as: M. Böhnstedt and J. Gampe (2019). Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model. *Statistics and Probability Letters* 150, 68–73, Copyright Elsevier B.V. 2019.

(Thatcher et al., 1998). This kind of mortality deceleration results if hazards differ between individuals, and a proportional hazards (PH) frailty model is the standard approach used to model such heterogeneity (Vaupel et al., 1979; Wienke, 2010).

If we denote by  $Y$  the continuous random variable that describes adult lifespans (typically above age 30) and its hazard by

$$h(y) = \lim_{\Delta y \searrow 0} P(y \leq Y < y + \Delta y | Y \geq y) / \Delta y, \quad (2.1)$$

a PH frailty model is of the form  $h(y|Z = z) = z \cdot h_0(y)$ , where  $Z$  is a positive random effect (*'frailty'*),  $h_0(y)$  is a baseline hazard, and  $h(y|Z = z)$  denotes the conditional hazard of an individual at age  $y$ , given that his or her frailty is  $Z = z$ . If the baseline hazard is an exponential function,  $h_0(y) = ae^{by}$  with parameters  $a > 0, b > 0$ , and the random effect  $Z$  has a gamma distribution with mean one and variance  $\sigma^2$ , we obtain the so-called gamma-Gompertz model. The variance parameter  $\sigma^2$  describes the heterogeneity in the risks of death: i.e., individuals with higher frailty values tend to die earlier, while more robust individuals tend to survive. While all individual hazards  $h(y|Z)$  are exponentially increasing, the resulting marginal hazard,

$$h(y) = \frac{ae^{by}}{1 + \sigma^2 \frac{a}{b} (e^{by} - 1)}, \quad (2.2)$$

shows a downward deviation from the exponential trajectory if  $\sigma^2 > 0$ . The deviation is discernible at high ages, when differential mortality has played out sufficiently. If  $\sigma^2 = 0$  – i.e., if there is no heterogeneity – model (2.2) reduces to a plain Gompertz model with hazard  $h(y) = ae^{by}$ . The choice of the gamma distribution for the frailty  $Z$  is both mathematically convenient and theoretically justified (Abbring and van den Berg, 2007).

The question of whether mortality deceleration is present or an exponential hazard fits even at advanced ages has been repeatedly discussed (e.g., Gavrilova and Gavrilov, 2015). For the gamma-Gompertz model, the question is reduced to whether  $\sigma^2 > 0$  or  $\sigma^2 = 0$ . In the latter case, the parameter  $\sigma^2$  is on the boundary of the parameter space, which violates the standard assumptions that underlie the asymptotic properties of likelihood-based inference. Thus, asymptotic results need to be derived for this setting.

In this chapter, we consider the asymptotic distribution of the likelihood ratio test (LRT) statistic for  $H_0: \sigma^2 = 0$  in the gamma-Gompertz model. As we are interested in deriving large-sample approximations of the power of the LRT to detect  $\sigma^2 > 0$ , we will work within a framework of local alternatives (Lehmann, 1999). We will also consider model selection based on the Akaike information criterion (AIC, Akaike, 1974). In the gamma-Gompertz model, the standard AIC is not an asymptotically unbiased estimator of the Akaike information. We derive the bias using a local misspecification framework (Hjort and Claeskens, 2003).

The rest of the chapter is structured as follows. In Section 2.2, we introduce some further notations and assumptions. In Section 2.3, we derive the asymptotic distribution of the maximum likelihood estimator (MLE) in the gamma-Gompertz model. In Section 2.4,

we establish the asymptotic distribution of the likelihood ratio test statistic, and provide a large-sample approximation to the power of the LRT to detect a positive  $\sigma^2$ . In Section 2.5, we show that the standard AIC is a biased estimator of the Akaike information in the gamma-Gompertz model. We conclude with a discussion in Section 2.6.

## 2.2 Preliminaries

We consider a sample of  $n$  iid lifespans  $y$  from a gamma-Gompertz density  $f(y, \eta)$  where the parameter vector  $\eta = (a, b, \sigma^2)^\top$  consists of the elements  $\theta = (a, b)^\top$ , resulting from the Gompertz baseline, and the gamma-variance  $\sigma^2$ . The density is

$$f(y, \eta) = \begin{cases} ae^{by} [1 + \sigma^2 \frac{a}{b} (e^{by} - 1)]^{-(1 + \frac{1}{\sigma^2})} & \text{for } \sigma^2 > 0 \\ ae^{by} \exp\{-\frac{a}{b}(e^{by} - 1)\} & \text{for } \sigma^2 = 0. \end{cases} \quad (2.3)$$

The framework of local alternatives assumes that the observations  $y$  are generated from a density

$$f_{\text{true}}(y) = f(y, \theta_0, \gamma_0 + \delta/\sqrt{n}), \quad (2.4)$$

where  $\theta_0$  is a  $p$ -dimensional parameter vector, and  $\gamma = \gamma_0 + \delta/\sqrt{n}$  is a  $q$ -dimensional parameter vector perturbed around a (known)  $\gamma_0$  in the direction of  $\delta$ . In the gamma-Gompertz model (2.2), we have  $\theta = (a, b)^\top$ , so  $p = 2$ , and  $\gamma$  is the single ( $q = 1$ ) boundary parameter  $\gamma = \sigma^2$ , with  $\gamma_0 = 0$  and  $\delta = \sqrt{n}\sigma^2 \geq 0$ .

The asymptotic distribution of the MLEs in the general framework (2.4) was derived by Hjort and Claeskens (2003) under the usual regularity conditions, which require that the true parameter is an inner point of the parameter space. Hjort (1994) considered a boundary parameter, but exclusively dealt with the  $t$ -distribution as an extension of the normal model. Self and Liang (1987) presented asymptotic distributions of MLEs and likelihood ratio test statistics in the presence of boundary parameters, but while assuming a fixed true model, rather than allowing for a local specification as in (2.4).

To state our main results, we introduce some further notations, closely following Hjort and Claeskens (2003). We denote by  $U(y) = \partial \ln f(y, \theta_0, \gamma_0)/\partial \theta$  and  $V(y) = \partial \ln f(y, \theta_0, \gamma_0)/\partial \gamma$  the score functions with respect to  $\theta$  and  $\gamma$ , respectively, of the log-likelihood of a single observation  $y$  from  $f_{\text{true}}$ , with both evaluated at the point  $(\theta, \gamma) = (\theta_0, \gamma_0)$ . Let  $J_{\text{full}}$  be the corresponding information matrix; i.e.,  $J_{\text{full}}$  is the variance-covariance matrix of the score vector  $(U(y)^\top, V(y)^\top)^\top$ , with blocks  $J_{00}$ ,  $J_{01}$ ,  $J_{10}$ , and  $J_{11}$ .

In the following,  $Y = (Y_1, \dots, Y_n)$  denotes a random sample of size  $n$  from density (2.4), and we abbreviate  $\eta = (\theta^\top, \gamma)^\top$  and  $\eta_0 = (\theta_0^\top, \gamma_0)^\top$ . We can express the log-likelihood as  $\ell_n(\eta, Y) = \sum_{i=1}^n \ln f(Y_i, \theta, \gamma)$ . Similarly, the averages of the score functions are denoted by  $\bar{U}_n = n^{-1} \sum_{i=1}^n U(Y_i)$  and  $\bar{V}_n = n^{-1} \sum_{i=1}^n V(Y_i)$ , with the shorthand notation  $\bar{W}_n = (\bar{U}_n^\top, \bar{V}_n)^\top$ .

When  $\gamma$  could lie on the boundary of the parameter space, the derivatives of the log-likelihood have to be taken from the appropriate side. It is also important to note that



these derivatives need to exist, and that they have to be bounded on intersections of neighborhoods of the true parameter value and the parameter space (see Self and Liang, 1987).

The derivation of the asymptotic distribution of the MLEs is based on the following result showing the weak convergence of the averaged score vector. It was given as Lemma 3.1 in Hjort and Claeskens (2003). Their proof carries over to the boundary setting considered here.

**Lemma 2.1.** *Under the sequence of local alternatives (2.4), the score vector is asymptotically normally distributed,*

$$\begin{pmatrix} \sqrt{n}\bar{U}_n \\ \sqrt{n}\bar{V}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} J_{01}\delta \\ J_{11}\delta \end{pmatrix} + \begin{pmatrix} M \\ N \end{pmatrix}, \quad \text{with} \quad \begin{pmatrix} M \\ N \end{pmatrix} \sim \mathcal{N}_{p+1}(0, J_{\text{full}}).$$

### 2.3 Asymptotic distribution of maximum likelihood estimator

In the setting with one boundary parameter, the MLE is asymptotically distributed as the projection of a normal random vector onto the subspace of admissible parameter values.

**Theorem 2.2.** *Under the sequence of models (2.4) and with a normally distributed random vector  $(A^\top, B)^\top \sim \mathcal{N}(0, J_{\text{full}}^{-1})$ , it holds that*

$$\begin{pmatrix} \sqrt{n}(\hat{\theta} - \theta_0) \\ \sqrt{n}(\hat{\gamma} - \gamma_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} A \\ B + \delta \end{pmatrix} \cdot \mathbb{1}_{\{B+\delta>0\}} + \begin{pmatrix} A + J_{00}^{-1}J_{01}(B + \delta) \\ 0 \end{pmatrix} \cdot \mathbb{1}_{\{B+\delta\leq 0\}},$$

where  $\mathbb{1}$  is the indicator function.

**Remark 2.1.** Theorem 2.2 applies not only to the gamma-Gompertz model, but more generally to parametric models with one boundary parameter. In particular, the theorem holds for other gamma-PH models if the parameters of the baseline hazard are not boundary parameters. This includes the gamma-exponential and the gamma-Weibull model.

*Proof.* Two cases need to be distinguished: namely, that the log-likelihood  $\ell_n$  is maximized at  $\hat{\gamma} > \gamma_0$  or at  $\hat{\gamma} = \gamma_0$  (cf. Hjort, 1994). In the first case, we have  $\partial\ell_n(\hat{\eta})/\partial\theta = 0$  and  $\partial\ell_n(\hat{\eta})/\partial\gamma = 0$ . If we apply, first, the usual Taylor arguments and, second, the result of Lemma 2.1 on the limiting distribution of the score vector, we find that for a maximum at  $\hat{\gamma} > \gamma_0$ ,

$$\begin{pmatrix} \sqrt{n}(\hat{\theta} - \theta_0) \\ \sqrt{n}(\hat{\gamma} - \gamma_0) \end{pmatrix} \stackrel{d}{=} J_{\text{full}}^{-1} \begin{pmatrix} \sqrt{n}\bar{U}_n \\ \sqrt{n}\bar{V}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 0 \\ \delta \end{pmatrix} + J_{\text{full}}^{-1} \begin{pmatrix} M \\ N \end{pmatrix} = \begin{pmatrix} A \\ B + \delta \end{pmatrix}. \quad (2.5)$$

In the second case, when the maximum occurs at  $\hat{\gamma} = \gamma_0$ , we have  $\partial\ell_n(\hat{\eta})/\partial\theta = 0$ , but  $\partial\ell_n(\hat{\eta})/\partial\gamma \leq 0$  and  $\hat{\gamma} = \gamma_0$ , such that

$$\begin{pmatrix} \sqrt{n}(\hat{\theta} - \theta_0) \\ \sqrt{n}(\hat{\gamma} - \gamma_0) \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} J_{00}^{-1}\sqrt{n}\bar{U}_n \\ 0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} J_{00}^{-1}J_{01}\delta + J_{00}^{-1}M \\ 0 \end{pmatrix}. \quad (2.6)$$

Direct calculations show that  $J_{00}^{-1}J_{01}\delta + J_{00}^{-1}M$  can be expressed as  $A + J_{00}^{-1}J_{01}(B + \delta)$ . If the last component of  $J_{\text{full}}^{-1}(\sqrt{n}\bar{U}_n^\top, \sqrt{n}\bar{V}_n^\top)^\top$  in (2.5) is denoted by  $\Delta_n$ , and if  $\Omega_n$  denotes the set of samples for which the estimate  $\hat{\gamma} > \gamma_0$ , then  $\Delta_n$ , and  $\Omega_n$  are (asymptotically) equivalent in the following sense:

$$\mathbb{1}_{\Omega_n} - \mathbb{1}_{\{\Delta_n > 0\}} \xrightarrow{P} 0. \quad (2.7)$$

This convergence is verified in Section 2.7.

The following corollary explicitly gives the asymptotic distribution of the MLEs for the gamma-Gompertz model.

**Corollary 2.3.** *Under the sequence of true gamma-Gompertz models (2.4) with  $\theta_0 = (a, b)^\top$ ,  $\gamma_0 = 0$ , and  $\delta = \sqrt{n}\sigma^2$ , the MLEs  $\hat{a}$ ,  $\hat{b}$ , and  $\hat{\sigma}^2$  asymptotically follow mixture distributions given by*

$$\begin{aligned} \sqrt{n}(\hat{a} - a) &\xrightarrow{d} \Phi\left(\frac{\delta}{\kappa}\right) \mathcal{TN}_1(\mu_\delta, \Sigma_1, s, t) + \Phi\left(-\frac{\delta}{\kappa}\right) \mathcal{N}\left(-\frac{c_{13}}{c_{33}}\delta, c_{11} - \frac{c_{13}^2}{c_{33}}\right) \\ \sqrt{n}(\hat{b} - b) &\xrightarrow{d} \Phi\left(\frac{\delta}{\kappa}\right) \mathcal{TN}_1(\mu_\delta, \Sigma_2, s, t) + \Phi\left(-\frac{\delta}{\kappa}\right) \mathcal{N}\left(-\frac{c_{23}}{c_{33}}\delta, c_{22} - \frac{c_{23}^2}{c_{33}}\right) \\ \sqrt{n}\hat{\sigma}^2 &\xrightarrow{d} \Phi\left(\frac{\delta}{\kappa}\right) \mathcal{TN}(\delta, \kappa^2, 0, \infty) + \Phi\left(-\frac{\delta}{\kappa}\right) \chi_0^2. \end{aligned}$$

Here, the  $c_{kl}$  are the elements of  $J_{\text{full}}^{-1}$ , with  $\kappa^2 = c_{33}$ ,  $\Phi(\cdot)$  is the cdf of the standard normal distribution and  $\chi_0^2$  denotes a point mass at zero.  $\mathcal{TN}_1(\mu, \Sigma, s, t)$  is the marginal distribution of the first component of a truncated bivariate normal distribution with mean vector  $\mu$ , covariance matrix  $\Sigma$ , and the lower and upper truncation limits  $s = (-\infty, 0)^\top$  and  $t = (\infty, \infty)^\top$ , respectively; where in particular,  $\mu_\delta = (0, \delta)^\top$ ,  $\Sigma_1 = \begin{pmatrix} c_{11} & c_{13} \\ c_{13} & c_{33} \end{pmatrix}$ , and  $\Sigma_2 = \begin{pmatrix} c_{22} & c_{23} \\ c_{23} & c_{33} \end{pmatrix}$ .

**Remark 2.2.** Two special cases of Corollary 2.3 are of particular interest: First, if  $\delta = \sqrt{n}\sigma^2$  is sufficiently large, we obtain the standard result of asymptotic normality of the MLEs. Second, if  $\delta = 0$  – i.e., if the true model is a pure Gompertz model – we obtain the typical mixture weights of 0.5.

*Proof.* Define  $Z = (Z_1, Z_2, Z_3)^\top \sim \mathcal{N}((0, 0, \delta)^\top, J_{\text{full}}^{-1})$ , which is distributed as  $(A^\top, B + \delta)^\top$ , and the projection matrix

$$T = \begin{pmatrix} I & J_{00}^{-1}J_{01} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -c_{13}/c_{33} \\ 0 & 1 & -c_{23}/c_{33} \\ 0 & 0 & 0 \end{pmatrix},$$

to express the result of Theorem 2.2 as

$$\sqrt{n}(\hat{\eta} - \eta_0) \xrightarrow{d} \hat{\eta}_* = Z \cdot \mathbb{1}_{\{Z_3 > 0\}} + TZ \cdot \mathbb{1}_{\{Z_3 \leq 0\}}. \quad (2.8)$$

As the components of  $TZ$  are independent of  $Z_3 \sim \mathcal{N}(\delta, \kappa^2)$ , it follows that

$$\begin{aligned} \mathbb{P}[\hat{\eta}_* \leq z] &= \mathbb{P}[\hat{\eta}_* \leq z | Z_3 > 0] \mathbb{P}[Z_3 > 0] + \mathbb{P}[\hat{\eta}_* \leq z | Z_3 \leq 0] \mathbb{P}[Z_3 \leq 0] \\ &= \mathbb{P}[Z \leq z | Z_3 > 0] \Phi\left(\frac{\delta}{\kappa}\right) + \mathbb{P}[TZ \leq z] \Phi\left(-\frac{\delta}{\kappa}\right). \end{aligned}$$

Now for  $\hat{\eta}_n = (\hat{a}, \hat{b}, \hat{\sigma}^2)^\top$  and  $\eta_0 = (a, b, 0)^\top$ , a large-sample approximation to  $\mathbb{P}[\sqrt{n}(\hat{\eta}_n - \eta_0) \leq u]$  based on (2.8) is given by

$$\mathbb{P}[\hat{\eta}_* \leq u] = \Phi\left(\frac{\delta}{\kappa}\right) \mathbb{P}[Z \leq u | Z_3 > 0] + \Phi\left(-\frac{\delta}{\kappa}\right) \mathbb{P}[TZ \leq u],$$

where  $TZ$  is normally distributed with mean  $T(0, 0, \delta)^\top = (-\delta c_{13}/c_{33}, -\delta c_{23}/c_{33}, 0)^\top$  and covariance matrix

$$TJ_{\text{full}}^{-1}T^\top = \begin{pmatrix} c_{11} - c_{13}^2/c_{33} & c_{12} - c_{13}c_{23}/c_{33} & 0 \\ c_{12} - c_{13}c_{23}/c_{33} & c_{22} - c_{23}^2/c_{33} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The result of Corollary 2.3 follows by determining the marginal distributions.

## 2.4 Asymptotics for the likelihood ratio test

This section deals with the asymptotics of the LRT for  $H_0: \sigma^2 = 0$  vs.  $H_1: \sigma^2 > 0$  in the gamma-Gompertz model in the framework of local alternatives (2.4). After presenting the limiting distribution of the likelihood ratio test statistic, we will provide a formula that approximates the power of the test to detect  $\sigma^2 > 0$  in large samples.

The likelihood ratio test statistic based on a sample  $Y = (Y_1, \dots, Y_n)$  is defined as

$$-2 \ln \lambda_n(Y) = 2[\ell_n(\hat{\eta}, Y) - \ell_n(\hat{\eta}_G, Y)],$$

where  $\hat{\eta}_G = (\hat{a}_G, \hat{b}_G, 0)^\top$  with the MLE  $(\hat{a}_G, \hat{b}_G)^\top$  for the Gompertz model. Under the non-standard condition of testing the boundary parameter  $\sigma^2$ , the test statistic does not converge to the standard chi-squared distribution, but rather to a mixture; see

**Theorem 2.4.** *Under the sequence of local alternatives (2.4), the test statistic of a LRT for  $H_0: \sigma^2 = 0$  is asymptotically distributed as*

$$-2 \ln \lambda_n(Y) \xrightarrow{d} \left\{ \max\left(0, \frac{Z_3}{\kappa}\right) \right\}^2 \quad \text{with } Z_3 \sim \mathcal{N}(\delta, \kappa^2).$$

**Corollary 2.5.** *In terms of the cumulative distribution function, we find that  $P[-2 \ln \lambda_n(Y) \leq z]$  asymptotically equals*

$$\frac{\Phi\left(\sqrt{z} - \frac{\delta}{\kappa}\right) - \Phi\left(-\frac{\delta}{\kappa}\right)}{\Phi\left(\sqrt{z} - \frac{\delta}{\kappa}\right) - \Phi\left(-\sqrt{z} - \frac{\delta}{\kappa}\right)} F_{\chi_1^2} \left(z, \text{ncp} = \frac{\delta^2}{\kappa^2}\right) + \Phi\left(-\frac{\delta}{\kappa}\right) F_{\chi_0^2}(z),$$

which is a combination of a non-central chi-squared distribution with one degree of freedom and non-centrality parameter  $\delta^2/\kappa^2$ , and a point mass at zero.

**Remark 2.3.** Setting  $\delta = 0$  – i.e.,  $\sigma^2 = 0$  – we obtain the asymptotic distribution of the LRT statistic under  $H_0$  as  $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_0^2$ , which is a 50:50 mixture of a chi-squared distribution with one degree of freedom and a point mass at zero.

**Remark 2.4.** Theorem 2.4 is also valid for other gamma-PH models, such as the gamma-Weibull model. The limiting distribution will depend on the respective model via the element  $\kappa^2$  of the inverse information matrix.

*Proof* (of Theorem 2.4). We again distinguish two cases depending on whether the estimate  $\hat{\sigma}^2$  does or does not lie on the boundary of the parameter space. If the likelihood is maximized at  $\hat{\sigma}^2 = 0$ , we have  $\hat{\eta} = \hat{\eta}_G$ , and, hence,  $-2 \ln \lambda_n(Y) = 0$ . If the likelihood is maximized at some point with  $\hat{\sigma}^2 > 0$ , a second-order Taylor expansion yields (see also Hjort and Claeskens, 2003, Section 3.3)

$$2[\ell_n(\hat{\eta}, Y) - \ell_n(\eta_0, Y)] \stackrel{d}{=} \begin{pmatrix} \sqrt{n}\bar{U}_n \\ \sqrt{n}\bar{V}_n \end{pmatrix}^\top J_{\text{full}}^{-1} \begin{pmatrix} \sqrt{n}\bar{U}_n \\ \sqrt{n}\bar{V}_n \end{pmatrix} \text{ and} \\ 2[\ell_n(\hat{\eta}_G, Y) - \ell_n(\eta_0, Y)] \stackrel{d}{=} \sqrt{n}\bar{U}_n^\top J_{00}^{-1} \sqrt{n}\bar{U}_n.$$

The quadratic form of the full score vector  $\sqrt{n}(\bar{U}_n^\top, \bar{V}_n)^\top$  can be calculated as  $n[\bar{U}_n^\top J_{00}^{-1} \bar{U}_n + \kappa^2(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n)^\top(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n)]$ . Consequently, if  $\hat{\sigma}^2 > 0$ , the likelihood ratio test statistic,  $-2 \ln \lambda_n(Y)$ , fulfills

$$-2 \ln \lambda_n(Y) \stackrel{d}{=} n\kappa^2(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n)^\top(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n). \quad (2.9)$$

From the asymptotic normality of  $\sqrt{n}(\bar{U}_n^\top, \bar{V}_n)^\top$  given in Lemma 2.1, we obtain that  $\sqrt{n}(\bar{V}_n - J_{10}J_{00}^{-1}\bar{U}_n)$  is asymptotically normally distributed with mean  $\delta/\kappa^2$  and variance  $\kappa^{-2}$ . Hence, (2.9) is distributed as the square of a normal random variable with mean  $\delta/\kappa$  and unit variance. The proof is completed by noting that  $Z_3/\kappa \sim \mathcal{N}(\delta/\kappa, 1)$  and recalling from the proof of Corollary 2.3 that  $\hat{\sigma}^2 > 0$  if  $Z_3 > 0$  and  $\hat{\sigma}^2 = 0$  if  $Z_3 \leq 0$ .

One reason for working under the framework in (2.4) is that it allows us to derive large-sample approximations of the power of the LRT to detect  $\sigma^2 > 0$ .

**Lemma 2.6.** *Under the sequence of local alternatives (2.4), the power  $\beta_n$  of the LRT for  $H_0 : \sigma^2 = 0$  at level  $\alpha$  based on a gamma-Gompertz sample of size  $n$  is approximately equal to*

$$\beta_n(\delta) \approx 1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{\delta}{\kappa}\right) = 1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{\sqrt{n}\sigma^2}{\kappa}\right).$$

The proof is deferred to Section 2.7.

## 2.5 AIC for the gamma-Gompertz model

In this section, we study the AIC for the gamma-Gompertz model. The AIC is a popular tool for model selection. In the standard setting, the AIC is defined as

$$\text{AIC} = -2 \ell_n(\hat{\eta}, Y) + 2(p + 1). \quad (2.10)$$

The AIC aims to give an unbiased estimate of the Akaike information, the expected relative Kullback-Leibler distance between the true data-generating density and the best parametric approximation  $f(y, \hat{\eta})$ , which is determined by inserting the MLE  $\hat{\eta}$  for  $\eta$ . To obtain such an unbiased estimate, the term  $-2 \ell_n(\hat{\eta}, Y)$  is penalized by twice the number of parameters in definition (2.10), which here is  $2(p + 1) = 2 \cdot 3$ . This results from asymptotic considerations that, again, require regularity conditions that do not hold for the gamma-Gompertz model. We find that formula (2.10) needs to be modified, as stated in

**Theorem 2.7.** *Under the sequence of local alternatives (2.4), an asymptotically unbiased estimator of the Akaike information of the gamma-Gompertz model is given by*

$$\text{AIC}^* = -2 \ell_n(\hat{\eta}, Y) + 2 \cdot 3 - 2 \Phi \left( -\frac{\delta}{\kappa} \right).$$

*Proof.* Here we give only an outline of the proof, and provide further details in Section 2.7.

The relative distance between the true underlying distribution  $g$  of  $X$  and an approximating parametric density  $f(\cdot, \eta)$  is measured by  $\mathbb{E}_X[\ln f(X, \eta)]$ . In practice, the unknown parameter  $\eta$  is estimated from a sample  $Y$  generated from  $g$ , and the estimator is denoted by  $\hat{\eta}(Y)$ . The Akaike information, which is to be estimated by (2.10), takes the form (Akaike, 1974)

$$-2 \mathbb{E}_Y [\mathbb{E}_X [\ln f(X, \hat{\eta}(Y))]]. \quad (2.11)$$

With the previous notation and  $\hat{\eta} = \hat{\eta}(Y)$ , an unbiased estimator of (2.11) is given by

$$-2 \ell_n(\hat{\eta}, Y) + \underbrace{2 \mathbb{E}_Y [\ell_n(\hat{\eta}, Y) - \ell_n(\eta_0, Y)]}_{=:S_1} + \underbrace{2 \mathbb{E}_Y [\mathbb{E}_X [\ell_n(\eta_0, X) - \ell_n(\hat{\eta}, X)]]}_{=:S_2}. \quad (2.12)$$

Based on a Taylor expansion of the log-likelihood about  $\eta_0$  and with the limiting distribution of the score given in Lemma 2.1, we can show that  $S_2$  is asymptotically equal to

$$\underbrace{-2 \mathbb{E}_Y [\sqrt{n}(\hat{\eta} - \eta_0)]^\top \begin{pmatrix} J_{01} \delta \\ J_{11} \delta \end{pmatrix}}_{=:S_{21}} + \underbrace{\mathbb{E}_Y [\sqrt{n}(\hat{\eta} - \eta_0)^\top J_{\text{full}} \sqrt{n}(\hat{\eta} - \eta_0)]}_{=:S_{22}}. \quad (2.13)$$

Exploiting the result on the limiting distribution of the MLE (2.8) and the arguments presented above yields

$$S_{21} \approx -2 \left\{ \left[ \frac{\delta}{\kappa} \frac{\phi(-\delta/\kappa)}{1 - \Phi(-\delta/\kappa)} + J_{11} \delta^2 \right] \left[ 1 - \Phi \left( -\frac{\delta}{\kappa} \right) \right] + J_{10} J_{00}^{-1} J_{01} \delta^2 \Phi \left( -\frac{\delta}{\kappa} \right) \right\}. \quad (2.14)$$

With regard to  $S_{22}$ , we have to combine (2.8) with the results on the mean of quadratic forms of random vectors, to arrive at

$$S_{22} \approx \left[ 1 - \Phi\left(-\frac{\delta}{\kappa}\right) \right] \left[ 3 + \delta^2 J_{11} + \frac{\delta}{\kappa} \frac{\phi(-\delta/\kappa)}{[1 - \Phi(-\delta/\kappa)]} \right] + [2 + \delta^2 J_{10} J_{00}^{-1} J_{01}] \Phi\left(-\frac{\delta}{\kappa}\right). \quad (2.15)$$

Finally,  $S_1$  in (2.12) is shown to be asymptotically equivalent to  $S_{22}$  in (2.13); and by inserting (2.14) and (2.15) into (2.12), we obtain the postulated result. All of these steps are laid out in more detail in Section 2.7.

## 2.6 Discussion

In this chapter, we have studied likelihood inference in the gamma-Gompertz model. Standard likelihood theory does not apply in this case because of the boundary parameter  $\sigma^2$ . The limiting distributions of the MLEs and of the likelihood ratio test statistic were found to be mixtures. The results can be extended to other gamma-PH models, in which the baseline hazard differs from the exponential Gompertz hazard. Our findings further indicated that under the non-standard conditions of the gamma-Gompertz model, the common definition of the AIC was biased. We also found, however, that the bias term depended on the unknown parameter  $\sigma^2$ , which complicated immediate bias correction. Future work will investigate alternative model selection tools.

## 2.7 Supplementary material: Technical appendix and proofs

**ad Theorem 2.2:** *Proof of equation (2.7)*

Using a second-order Taylor expansion about  $\eta_0$ , the log-likelihood can be approximated by a parabola in  $\hat{\gamma}$ , which has its maximum at  $\hat{\gamma}$ , satisfying

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = \left[ -\frac{1}{n} \frac{\partial^2 \ell_n}{\partial \gamma^2}(\hat{\theta}, \gamma_0) \right]^{-1} \sqrt{n} \bar{V}_n(\hat{\theta}, \gamma_0), \quad (2.16)$$

if the right-hand side is positive, and at  $\hat{\gamma} = \gamma_0$  otherwise. While the first term,  $-n^{-1} \cdot \partial^2 \ell_n(\hat{\theta}, \gamma_0) / \partial \gamma^2$ , tends to  $J_{11}$ , the second term,  $\sqrt{n} \bar{V}_n$ , is expanded in a first-order Taylor series as

$$\sqrt{n} \bar{V}_n(\hat{\theta}, \gamma_0) \approx \sqrt{n} \bar{V}_n(\theta_0, \gamma_0) + \sqrt{n}(\hat{\theta} - \theta_0)^\top \frac{1}{n} \frac{\partial^2 \ell_n}{\partial \gamma \partial \theta}(\theta_0, \gamma_0).$$

Note that  $-n^{-1} \partial^2 \ell_n(\theta_0, \gamma_0) / \partial \gamma \partial \theta$  tends to  $J_{10}$  and that  $\sqrt{n}(\hat{\theta} - \theta_0)$  is distributed on  $\Omega_n$  as the first components of  $J_{\text{full}}^{-1}(\sqrt{n} \bar{U}_n^\top, \sqrt{n} \bar{V}_n)^\top$  according to (2.5). From this it follows directly that on the set  $\Omega_n$  the right-hand side of (2.16) can be approximated to first-order by  $\Delta_n$ .

**Proof of Lemma 2.6**

The power  $\beta_n$  is the probability to correctly reject  $H_0$ , if  $H_0$  does not hold. For a LRT at level  $\alpha$ , we reject  $H_0$  if the  $p$ -value is smaller than  $\alpha$ . If  $t_{\text{obs}}$  is the value of the test statistic for the observed sample, the  $p$ -value is computed according to Corollary 2.5 as

$$p = \mathbb{P}[-2 \ln \lambda_n(Y) \geq t_{\text{obs}} | H_0] \approx 1 - \left( \frac{1}{2} F_{\chi_1^2}(t_{\text{obs}}) + \frac{1}{2} F_{\chi_0^2}(t_{\text{obs}}) \right).$$

Since  $t_{\text{obs}} \geq 0$  by definition,  $F_{\chi_0^2}(t_{\text{obs}}) = 1$ . Furthermore,

$$F_{\chi_1^2}(t_{\text{obs}}) = \frac{\gamma\left(\frac{1}{2}, \frac{t_{\text{obs}}}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} = \frac{\sqrt{\pi} \operatorname{erf}\left(\sqrt{\frac{t_{\text{obs}}}{2}}\right)}{\sqrt{\pi}} = \operatorname{erf}\left(\sqrt{\frac{t_{\text{obs}}}{2}}\right),$$

using the gamma function  $\Gamma(v) = \int_0^\infty u^{v-1} e^{-u} du$ , the lower incomplete gamma function  $\gamma(v, w) = \int_0^w u^{v-1} e^{-u} du$  and the error function  $\operatorname{erf}(x) = 2/\sqrt{\pi} \int_0^x e^{-u^2} du$ . Moreover,  $1/2 + 1/2 \operatorname{erf}(x/\sqrt{2}) = \Phi(x)$ , and from this  $p = 1 - \Phi(\sqrt{t_{\text{obs}}})$ . Under the true model (2.4), we know from Theorem 2.4 that  $t_{\text{obs}}$  is a realization of a random variable that is asymptotically distributed as  $\max\{0, Z_3/\kappa\}^2$ . Thus,

$$\beta_n(\delta) = \mathbb{P}[p < \alpha | \text{true } \delta] \approx \mathbb{P}[1 - \Phi(\max\{0, Z_3/\kappa\}) < \alpha].$$

Rearranging the terms and exploiting the normality of  $Z_3/\kappa$  proves Lemma 2.6.

**ad Theorem 2.7:**

*Proof of (2.13):*

A Taylor expansion of the log-likelihood about  $\eta_0$  yields

$$\begin{aligned} S_2 &= 2 \mathbb{E}_Y [\mathbb{E}_X [- (\hat{\eta} - \eta_0)^\top n \bar{W}_n(\eta_0, X) - \frac{1}{2} (\hat{\eta} - \eta_0)^\top \mathcal{H}(\eta_0, X) (\hat{\eta} - \eta_0) + R_1]] \\ &= \mathbb{E}_Y [2 (\hat{\eta} - \eta_0)^\top \mathbb{E}_X [-n \bar{W}_n(\eta_0, X)] + (\hat{\eta} - \eta_0)^\top \mathbb{E}_X [-\mathcal{H}(\eta_0, X)] (\hat{\eta} - \eta_0) + R_1] \end{aligned}$$

with  $\mathcal{H}$  denoting the Hessian matrix of the log-likelihood and a remainder term  $R_1 \xrightarrow{n \rightarrow \infty} 0$ . Inserting  $\mathbb{E}_X [-\mathcal{H}(\eta_0, X)] = n J_{\text{full}}$  and the asymptotic mean of  $\bar{W}_n$  from Lemma 2.1, we obtain (2.13).

*Proof of (2.14):*

Based on the limiting distribution of the MLE given in (2.8), the expectation  $\mathbb{E}_Y [\sqrt{n}(\hat{\eta} - \eta_0)]$  in  $S_{21}$  can be approximated by

$$\mathbb{E}[Z \cdot \mathbb{1}_{\{Z_3 > 0\}} + TZ \cdot \mathbb{1}_{\{Z_3 \leq 0\}}] = \mathbb{E}[Z | Z_3 > 0] \left[ 1 - \Phi\left(-\frac{\delta}{\kappa}\right) \right] + \mathbb{E}[TZ] \Phi\left(-\frac{\delta}{\kappa}\right).$$

Here,  $\mathbb{E}[Z \mid Z_3 > 0]$  is the mean vector of a truncated trivariate normal distribution. Using the results of Gupta and Tracy (1978) on the moments of such distributions, it can be shown that

$$\mathbb{E}[Z \mid Z_3 > 0] = \kappa \begin{pmatrix} -J_{00}^{-1}J_{01} \\ 1 \end{pmatrix} \frac{\phi(-\delta/\kappa)}{[1 - \Phi(-\delta/\kappa)]} + \begin{pmatrix} 0 \\ \delta \end{pmatrix}.$$

Direct calculations then give (2.14).

*Proof of (2.15)*

According to (2.8) in combination with the continuous mapping theorem, the quadratic form  $\sqrt{n}(\hat{\eta} - \eta_0)^\top J_{\text{full}} \sqrt{n}(\hat{\eta} - \eta_0)$  of the MLE in (2.13) converges in distribution to

$$Z^\top J_{\text{full}} Z \mathbb{1}_{\{Z_3 > 0\}} + (TZ)^\top J_{\text{full}} TZ \mathbb{1}_{\{Z_3 \leq 0\}}.$$

Thus, the mean  $S_{22}$  of the quadratic form is asymptotically equal to

$$S_{22} \approx \mathbb{E}[(Z|Z_3 > 0)^\top J_{\text{full}}(Z|Z_3 > 0)] \left[ 1 - \Phi\left(-\frac{\delta}{\kappa}\right) \right] + \mathbb{E}[(TZ)^\top J_{\text{full}}(TZ)] \Phi\left(-\frac{\delta}{\kappa}\right),$$

which is a weighted sum of means of quadratic forms in the random vectors  $(Z|Z_3 > 0)$  and  $TZ$ , respectively. For any  $k$ -dimensional random vector  $X$  with  $\mathbb{E}[X] = \mu$ ,  $\text{Cov}[X] = \Sigma$  and a constant, symmetric  $(k \times k)$ -matrix  $A$ ,  $\mathbb{E}[X^\top AX] = \text{tr}(A\Sigma) + \mu^\top A\mu$ , where  $\text{tr}(\cdot)$  denotes the trace of a matrix. To use this result here, the covariance matrix of the truncated  $(Z|Z_3 > 0)$  is required. It can be obtained from the results of Gupta and Tracy (1978) as  $J_{\text{full}}^{-1} + G$ , where  $G$  equals

$$\left\{ \delta\kappa \frac{\phi(-\delta/\kappa)}{1 - \Phi(-\delta/\kappa)} - \kappa^2 \left[ \frac{\phi(-\delta/\kappa)}{1 - \Phi(-\delta/\kappa)} \right]^2 \right\} \begin{pmatrix} J_{00}^{-1}J_{01}J_{10}J_{00}^{-1} & -J_{00}^{-1}J_{01} \\ -J_{10}J_{00}^{-1} & 1 \end{pmatrix}.$$

We omit the remaining straightforward computations, which result in (2.15).

*Proof of the asymptotic equivalence of  $S_1$  and  $S_{22}$ :*

Another Taylor expansion of the log-likelihood about  $\hat{\eta}$  yields

$$\ell_n(\eta_0, Y) = \ell_n(\hat{\eta}, Y) + (\eta_0 - \hat{\eta})^\top n\bar{W}_n(\hat{\eta}, Y) + \frac{1}{2}(\eta_0 - \hat{\eta})^\top \mathcal{H}(\hat{\eta}, Y)(\eta_0 - \hat{\eta}) + R_2,$$

with a remainder term  $R_2 \xrightarrow{n \rightarrow \infty} 0$ . Thus,  $S_1$  can be approximated by

$$\mathbb{E}_Y [2(\hat{\eta} - \eta_0)^\top n\bar{W}_n(\hat{\eta}, Y)] + \mathbb{E}_Y [(\hat{\eta} - \eta_0)^\top [-\mathcal{H}(\hat{\eta}, Y)](\hat{\eta} - \eta_0) + R_2]. \quad (2.17)$$

To evaluate the first summand, we study  $\bar{W}_n(\hat{\eta}, Y)$  for the two cases  $\hat{\sigma}^2 > 0$  and  $\hat{\sigma}^2 = 0$ . In the first case, the MLE is an inner point of the parameter space, and, given the definition of the MLE, we have  $\bar{W}_n(\hat{\eta}, Y) = 0$ , such that in this case  $(\hat{\eta} - \eta_0)^\top n\bar{W}_n(\hat{\eta}, Y) = 0$ . In the second case, we know that  $\bar{U}_n(\hat{\eta}, Y) = 0$  and  $\bar{V}_n(\hat{\eta}, Y) \leq 0$ , but as  $(\hat{\gamma} - \gamma_0) = 0$ , we still



have  $(\hat{\eta} - \eta_0)^\top n\bar{W}_n(\hat{\eta}, Y) = 0$ . Consequently, the first term in (2.17) vanishes. With regard to the second summand, we argue that  $-\mathcal{H}(\hat{\eta}, Y)$  approximates reasonably well  $nJ_{\text{full}}$  (see also Hjort and Claeskens, 2003). Thus,  $S_1$  is asymptotically equal to

$$\mathbb{E}_Y [\sqrt{n}(\hat{\eta} - \eta_0)^\top J_{\text{full}} \sqrt{n}(\hat{\eta} - \eta_0)],$$

which is equivalent to  $S_{22}$  in (2.13).

## References

- Abbring, J. H. and G. J. van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika* 94(1), 87–99.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.
- Gavrilova, N. S. and L. A. Gavrilov (2015). Biodemography of old-age mortality in humans and rodents. *The Journals of Gerontology: Series A* 70(1), 1–9.
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* 115, 513–583.
- Gupta, A. K. and D. S. Tracy (1978). Hermite polynomials and truncated trivariate normal distributions. *Journal of Statistical Computation and Simulation* 7, 269–286.
- Hjort, N. (1994). The exact amount of  $t$ -ness that the normal model can tolerate. *Journal of the American Statistical Association* 89, 665–675.
- Hjort, N. L. and G. Claeskens (2003). Frequentist model average estimators. *Journal of the American Statistical Association* 98, 879–899.
- Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer texts in statistics. New York: Springer.
- Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 82(398), 605–610.
- Thatcher, A. R., V. Kannisto, and J. W. Vaupel (1998). *The force of mortality at ages 80 to 120*. Odense Monographs on Population Aging 5. Odense University Press.
- Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3), 439–454.
- Wienke, A. (2010). *Frailty Models in Survival Analysis*. Biostatistics Series. Chapman & Hall/CRC.

# 3

## Information measures and design issues in the study of mortality deceleration: Findings for the gamma-Gompertz model

### Abstract

Mortality deceleration, or the slowing down of death rates at old ages, has been repeatedly investigated, but empirical studies of this phenomenon have produced mixed results. The scarcity of observations at the oldest ages complicates the statistical assessment of mortality deceleration, even in the parsimonious parametric framework of the gamma-Gompertz model considered here. The need for thorough verification of the ages at death can further limit the available data. As logistical constraints may only allow to validate survivors beyond a certain (high) age, samples may be restricted to a certain age range. If we can quantify the effects of the sample size and the age range on the assessment of mortality deceleration, we can make recommendations for study design. For that purpose, we propose applying the concept of the Fisher information and ideas from the theory of optimal design. We compute the Fisher information matrix in the gamma-Gompertz model, and derive information measures for comparing the performance of different study designs. We then discuss interpretations of these measures. The special case in which the

---

This chapter has been published as: M. Böhnstedt, J. Gampe, and H. Putter (2021). Information measures and design issues in the study of mortality deceleration: findings for the gamma-Gompertz model. *Lifetime Data Analysis* 27, 333-356.

frailty variance takes the value of zero and lies on the boundary of the parameter space is given particular attention. The changes in information related to varying sample sizes or age ranges are investigated for specific scenarios. The Fisher information also allows us to study the power of a likelihood ratio test to detect mortality deceleration depending on the study design. We illustrate these methods with a study of mortality among late 19<sup>th</sup>-century French-Canadian birth cohorts.

### 3.1 Introduction

Accurately describing, understanding, and, finally, projecting the trajectory of human mortality over age is crucial for assessing the future of human longevity, but it is also important in actuarial sciences, population forecasting and health care planning. The scientific modeling of human mortality over age has a long tradition. Around two centuries ago, Benjamin Gompertz published his finding that the death rates of humans increase exponentially from mid-life ages onwards (Gompertz, 1825), and this regularity has since been confirmed time and time again in many populations, epochs, and circumstances. Recently, however, improved and more accurate vital registration has revealed that the increase in death rates slows down at higher ages (see, for example, Thatcher et al., 1998; Thatcher, 1999). This decrease in the increase of death rates at older ages is termed *mortality deceleration*.

An explanation for this initially perplexing observation was provided early on by Beard (1959) via the so-called heterogeneity hypothesis. If the individuals in a birth cohort are subjected to non-identical mortality risks, then those with higher risks tend to die earlier, resulting in an increasingly selected group of survivors with lower mortality risks. Hence, even if the individual hazards increase exponentially, the population hazard will increase more slowly (Vaupel et al., 1979).

Although this explanation is plausible, empirical investigations have repeatedly produced mixed results (Bebbington et al., 2014). While some studies have found evidence of a downward deviation from the exponential hazard at the oldest ages (Feehan, 2018), others have suggested that exponential growth continues even through advanced ages (Gavrilov and Gavrilova, 2019).

The empirical study of mortality deceleration is complicated by several issues. It is a phenomenon that manifests in the tail of the lifespan distribution where observations necessarily become sparse, even for sizable cohorts. Whether we are able to detect mortality deceleration will depend on the actual strength of the effect and the size of the sample.

The Gompertz model originated as an actuarial device, but its ability to capture the age-trajectory of adult mortality in a multitude of circumstances prompted numerous attempts to find underlying mechanisms that would produce exponentially increasing hazards. Most attempts come from reliability theory (Gavrilov and Gavrilova, 2001) and the biology of aging (see Kirkwood, 2015, and references therein). Whether and which of the mechanisms will eventually apply is still an open question, however, the repeatedly

confirmed exponential increase of mortality over much of the adult lifespan established the Gompertz model in demography, biology, and epidemiology.

When examining mortality deceleration, we have to decide over what age range death rates should be analyzed in order to uncover potential deviations from a Gompertz hazard. On the one hand, using a rather wide age range – that is, starting from relatively young ages – may run the risk that the observations at younger ages dominate the analysis, and thus mask the deceleration that is based on relatively fewer observations at older ages. This line of thought suggests that observations of higher ages at death might be more informative about a potential deceleration than observations of younger ages at death. On the other hand, using a wider age range might enable us to detect deviations from the exponential increase of the hazard early. Moreover, using a wider age range yields a larger sample size, and can increase the precision of the parameter estimates, particularly of the parameters describing the exponential increase. This might enable us to detect more easily deviations from it at higher ages. How the trade-off between these two opposing effects would play out is not clear.

Another important aspect in all studies involving old-age mortality is data quality. In particular, age misreporting is known to induce a downward bias of mortality at advanced ages (Preston et al., 1999). Therefore, scientific age validation is indispensable in studies involving individuals of very high ages (Jeune and Vaupel, 1999). In practice, performing such individual checks is costly and time-consuming, and logistics can limit the number of cases that can be verified. In the application presented in Section 3.6, the ages at death could be validated for individual members of French-Canadian birth cohorts (born 1880-1896) who survived to age 90 or older. Since extending the age range by another, say, five years, to ages 85 and above, would imply a drastic increase in the number of cases to be validated, a practically relevant question is how much the extra effort would expand the information about mortality deceleration in the resulting larger data set.

All of the considerations discussed above are questions related to optimal design. While the theory of optimal design is applied in various research fields (see Berger and Wong, 2009, and the references therein), applications are less numerous in the area of survival analysis. Hwang and Brookmeyer (2003) attempted to find the optimal spacing between consecutive waves of a panel study. Becker et al. (1989) and Konstantinou et al. (2015) discussed optimal covariate settings in proportional hazards models, and McGree and Eccleston (2010) investigated the design aspects of covariates and sample size in accelerated failure-time models. Here, we will study the effects of the sample size and the age range covered by a data set on the assessment of mortality deceleration; specifically, on the downward deviation from a Gompertz hazard.

The most commonly used approaches for describing individually heterogeneous death risks are proportional hazards frailty models (Vaupel et al., 1979; Duchateau and Janssen, 2008; Balan and Putter, 2020). In this chapter, we focus on one specific model from this class, the gamma-Gompertz model. The individuals share an exponentially increasing Gompertz baseline hazard, but a multiplicative gamma distributed random effect (the frailty) introduces heterogeneity of the individual mortality risks. The amount of hetero-

geneity is determined by the frailty variance. If the frailty variance is zero the population hazard will follow the exponential Gompertz trajectory, while a positive frailty variance implies that the population hazard decelerates at older ages. Consequently, the statistical assessment of mortality deceleration in the gamma-Gompertz model is reduced to inference about the frailty variance. In particular, the likelihood ratio test for a zero frailty variance is a commonly used approach to assess this phenomenon. However, zero is a boundary point of the parameter space for the variance parameter which violates the usual regularity assumptions. Consequently, standard asymptotic results are not directly applicable.

In this chapter, we propose using the concepts of the Fisher information and of optimal design to address issues that arise in planning and evaluating studies that assess mortality deceleration in the setting of the gamma-Gompertz model. Within the likelihood framework, the Fisher information measures the amount of information about the model parameters that is contained in the data (Lehmann, 1999). Therefore, the Fisher information can serve as a basis for identifying optimal designs that maximize the information about the model parameters.

The chapter is organized as follows. Section 3.2 lays out the framework for our study by formally introducing the gamma-Gompertz model, as well as the general concepts of the Fisher information and of optimal designs. In Section 3.3, we present the Fisher information and a specific information measure for the gamma-Gompertz model, and relate them to the power of the likelihood ratio test to detect mortality deceleration. In Section 3.4, we discuss in detail the design issues that arise in studies of mortality deceleration. In Section 3.5, we assess the effects of different design choices on the information measure, and on the power of the test for specific scenarios. In Section 3.6, we apply the proposed concepts and methods to a French-Canadian mortality data set. In Section 3.7, we conclude with a discussion of our findings.

## 3.2 Framework: Gamma-Gompertz model, Fisher information & study design

### 3.2.1 Gamma-Gompertz model

We consider a continuous random variable  $X$  that describes adult lifespans (above some young adult age, such as 30). Its distribution is determined by the hazard function

$$h(x) = \lim_{\Delta x \searrow 0} \frac{P(x < X \leq x + \Delta x \mid X > x)}{\Delta x}.$$

The heterogeneity hypothesis can be formalized in frailty proportional hazards models of the form  $h(x \mid Z = z) = z \cdot h_0(x)$ . The unobserved heterogeneity of the individuals is modeled via the positive random effect  $Z$  that affects a common baseline hazard  $h_0(x)$  in a multiplicative way. Individuals with higher values  $z$  have a higher risk at any age  $x$ , as specified by the conditional hazard  $h(x \mid Z = z)$ ; thus,  $Z$  is called the frailty.

A popular choice for the distribution of frailties is the gamma distribution. It leads to closed-form expressions for marginal survival and hazard functions. Furthermore, for the gamma distribution the frailty among survivors at any age  $x > 0$  again is gamma distributed, only with different parameters (Vaupel et al., 1979; Hougaard, 1984; Economou and Caroni, 2008). Moreover, Abbring and van den Berg (2007) showed that even if the frailty at  $x = 0$  is not gamma distributed the frailty among survivors converges with increasing  $x$  to a gamma distribution for many proportional hazards frailty models.

As the name suggests, in the gamma-Gompertz model the baseline hazard has an exponentially increasing Gompertz form,  $h_0(x) = ae^{bx}$ . Here the parameter  $a > 0$  represents the initial level of mortality for  $x = 0$  and  $b > 0$  is the rate of aging. The frailty is gamma distributed, with a mean of one and a variance of  $\sigma^2$ . The heterogeneity in frailty, and, hence, in mortality risks, is measured by the variance parameter  $\sigma^2$ . In a heterogeneous population with  $\sigma^2 > 0$ , there is a tendency of individuals with higher frailty values to die at younger ages, such that the population of survivors to higher ages consists mainly of individuals with lower mortality risks. Therefore, the marginal hazard,

$$h(x) = \frac{ae^{bx}}{1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)}, \quad (3.1)$$

shows a downward deviation from the exponential increase at higher ages. In a homogeneous population with  $\sigma^2 = 0$ , there is no such selection effect, and the marginal hazard is again of the Gompertz form,  $h(x) = ae^{bx}$ . Thus, the presence or the absence of mortality deceleration is determined by the parameter  $\sigma^2$  and can, for instance, be assessed by a likelihood ratio test for  $H_0: \sigma^2 = 0$  against  $H_1: \sigma^2 > 0$ .

While the parameter  $\sigma^2$  describes the heterogeneity in frailty and mortality risks at the starting age of the model – that is, at  $x = 0$  – the increasingly selected population of survivors to higher ages will be less heterogeneous in terms of their frailty and mortality risks. For example, the heterogeneity in mortality risks will be lower in the subset of survivors to ages 90 and above than among the survivors to ages 80 and above. Consequently, the age range covered by a data set will affect the ability to assess the frailty variance, and, hence, mortality deceleration.

It is important to note that the frailty variance  $\sigma^2$  takes a value on the boundary of its parameter space if there is no heterogeneity ( $\sigma^2 = 0$ ). As this violates common regularity assumptions, some standard asymptotic results for likelihood inference might not hold, which will also affect the interpretation of the information measures in the following.

### 3.2.2 The Fisher information

We briefly recap the concept of the Fisher information, and refer to Chapter 7 in Lehmann (1999) for further details. For a random variable  $X$  with density  $f_X(\cdot; \boldsymbol{\theta})$  and parameter vector  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K)^\top$ , the Fisher information matrix  $\mathbf{I}(\boldsymbol{\theta})$  is defined as

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbb{E} \left[ \left( \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_X(X; \boldsymbol{\theta}) \right) \left( \frac{\partial}{\partial \boldsymbol{\theta}} \ln f_X(X; \boldsymbol{\theta}) \right)^\top \right], \quad (3.2)$$

where the expectation  $\mathbb{E}$  is with respect to the distribution of  $X$ . Under mild regularity conditions, expression (3.2) can be rewritten in terms of the second-order partial derivatives of the log-density of  $X$ ,

$$\mathbf{I}(\boldsymbol{\theta}) = -\mathbb{E} \left[ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \ln f_X(X; \boldsymbol{\theta}) \right]. \quad (3.3)$$

The Fisher information  $\mathbf{I}(\boldsymbol{\theta})$  is often interpreted as the amount of information a single observation of  $X$  contains about the model parameters  $\boldsymbol{\theta}$ . The information  $\mathbf{I}_n(\boldsymbol{\theta})$  of an iid sample  $X_1, X_2, \dots, X_n$  of size  $n$  from the distribution of  $X$  is then  $n$ -times as large,  $\mathbf{I}_n(\boldsymbol{\theta}) = n\mathbf{I}(\boldsymbol{\theta})$ . In many cases, the Fisher information (3.3) cannot be computed directly because it depends on the true unknown parameter value, or because the expectation is not analytically tractable. Thus, we often use the observed Fisher information matrix, which for an iid sample of size  $n$  is given by the negative second-order partial derivatives of the log-likelihood, evaluated at the maximum likelihood estimate (MLE)  $\hat{\boldsymbol{\theta}}_n$ ,

$$\mathcal{J}(\hat{\boldsymbol{\theta}}_n) = -\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \sum_{i=1}^n \ln f_X(X_i; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}. \quad (3.4)$$

The interpretation of  $\mathbf{I}(\boldsymbol{\theta})$  as a measure of information is based on two different arguments. Analytically, the partial derivatives  $\frac{\partial}{\partial \boldsymbol{\theta}} \ln f_X(x; \boldsymbol{\theta}) = \frac{\frac{\partial}{\partial \boldsymbol{\theta}} f_X(x; \boldsymbol{\theta})}{f_X(x; \boldsymbol{\theta})}$  in (3.2) describe the relative change of the density  $f_X(\cdot; \boldsymbol{\theta})$  with respect to  $\boldsymbol{\theta}$  at the point  $x$ . If this change is large for one  $\boldsymbol{\theta}_0$ , this parameter value can be better identified from a range of possible values  $\boldsymbol{\theta}$ . Similarly, the second-order partial derivatives  $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \ln f_X(X; \boldsymbol{\theta})$  in (3.3) describe the curvature of the log-density  $\ln f_X(\cdot; \boldsymbol{\theta})$  with respect to  $\boldsymbol{\theta}$ , and, thus, the curvature of the contributions to the log-likelihood function. A sample for which the log-likelihood shows a clearer peak at some  $\boldsymbol{\theta}_0$ , and for which this value is, therefore, more clearly distinguished from other values  $\boldsymbol{\theta}$ , is viewed as more informative about the parameter than samples with a flatter log-likelihood.

A second justification for the notion of information rests on the following result for asymptotically normal estimators. If an estimator  $\delta_n$  of  $\theta_k$  satisfies  $\sqrt{n}(\delta_n - \theta_k) \xrightarrow{d} \mathcal{N}(0, v(\boldsymbol{\theta}))$ , then its variance is bounded below by  $[\mathbf{I}(\boldsymbol{\theta})]_{kk}^{-1}$ , which denotes the  $k^{\text{th}}$  diagonal element of the inverse of the information matrix  $\mathbf{I}(\boldsymbol{\theta})$  (see Lehmann and Casella, 1998, p. 462). In particular, the MLE  $\hat{\boldsymbol{\theta}}_n$  attains this lower bound under suitable regularity conditions (cf. Lehmann and Casella, 1998, p. 463),

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, [\mathbf{I}(\boldsymbol{\theta})]^{-1}), \quad (3.5)$$

such that each  $\hat{\theta}_{nk}$  is asymptotically efficient,  $\sqrt{n}(\hat{\theta}_{nk} - \theta_k) \xrightarrow{d} \mathcal{N}(0, [\mathbf{I}(\boldsymbol{\theta})]_{kk}^{-1})$ . In this sense, a sample is more informative if the parameters can be estimated with higher precision.

In summary, following the exposition above, the Fisher information serves as a suitable measure of the information contained in a sample about the unknown parameter in likelihood-based inference.

### 3.2.3 Optimal design

The Fisher information can be instrumental for determining optimal designs. The aim is to find a design that maximizes some scalar function of the information matrix, and that therefore maximizes the information, in a suitably defined way, about all or some particular model parameters.

Different criteria for defining and assessing the optimality of a design have been suggested (see Silvey, 1980, for an early monograph, and Atkinson, 1988, for an early review). If all elements of the parameter vector  $\theta$  are of interest, two popular scalar measures of information are  $D$ - and  $A$ -optimality. A design is called  $D$ -optimal if the design maximizes the determinant of the information matrix,  $\det(\mathbf{I}(\theta))$ . Alternatively, the criterion of  $A$ -optimality refers to the trace of the inverse information matrix  $[\mathbf{I}(\theta)]^{-1}$  and states that a design is optimal if the design attains the maximum possible value for the inverse of this trace, that is, for  $1/\text{tr}([\mathbf{I}(\theta)]^{-1})$ .

Both of these information measures are functions of the eigenvalues of the information matrix. The determinant equals the product and the trace equals the sum of the eigenvalues of a matrix, respectively; and the eigenvalues of  $[\mathbf{I}(\theta)]^{-1}$  are the reciprocals of the eigenvalues of  $\mathbf{I}(\theta)$ . These eigenvalues of the information matrix are related to estimator precision. For an estimator  $\hat{\theta}_n$  that is asymptotically normal with the covariance matrix given by the inverse Fisher information matrix  $[\mathbf{I}(\theta)]^{-1}$  as in (3.5), a confidence region for the parameter vector  $\theta$  takes the form of an ellipsoid. The axes of the ellipsoid are characterized by the eigenvalues and the eigenvectors of the matrix  $[\mathbf{I}(\theta)]^{-1}$ . More precisely, the eigenvectors determine the direction of the axes of the ellipsoid, and the eigenvalues are proportional to the squared lengths of the axes. Therefore, the size of the confidence ellipsoid and the precision of the estimator largely depend on the eigenvalues of  $[\mathbf{I}(\theta)]^{-1}$ . In particular, the volume of the confidence ellipsoid is proportional to the product of the eigenvalues of  $[\mathbf{I}(\theta)]^{-1}$ . As a consequence, maximizing the information in terms of  $D$ -optimality corresponds to minimizing the volume of the confidence ellipsoid.

The criteria of  $D$ - and  $A$ -optimality weigh all dimensions of the problem equally. In contrast, the criterion of  $E$ -optimality seeks to maximize only the smallest eigenvalue of the information matrix. This is equivalent to minimizing the largest eigenvalue of  $[\mathbf{I}(\theta)]^{-1}$ , which measures the uncertainty about the parameters in the direction of the largest axis of the confidence ellipsoid. Because the parameters are estimated with least precision in this direction, an  $E$ -optimal design maximizes the precision in the estimation of the least well-estimated parameter combinations.

If one particular linear combination of the parameters is of specific interest, the criterion of  $D_A$ -optimality is applied. If  $\mathbf{A}\theta$  denotes the linear combination of the parameters, where  $\mathbf{A}$  is a  $p \times K$  matrix of rank  $p < K$ , then, by analogy with  $D$ -optimality, one maximizes the determinant of the inverse of  $\mathbf{A}[\mathbf{I}(\theta)]^{-1}\mathbf{A}^\top$ . The criterion of  $D_A$ -optimality also allows us to focus on only one parameter  $\theta_k$ . For that purpose, a matrix  $\mathbf{A}$  of dimension  $1 \times K$  is defined with entry 1 for the  $k^{\text{th}}$  element, and with zeros otherwise. The information measure then simplifies to  $(\mathbf{A}[\mathbf{I}(\theta)]^{-1}\mathbf{A}^\top)^{-1} = 1/[\mathbf{I}(\theta)]_{kk}^{-1}$ . In a regular setting with an asymptotically normal estimator  $\hat{\theta}_k$  that satisfies (3.5), maximizing the in-



formation measure  $1/[I(\boldsymbol{\theta})]_{kk}^{-1}$  is equivalent to minimizing the asymptotic variance of  $\hat{\theta}_k$ , or, in other words, maximizing its precision.

If the information matrix  $I(\boldsymbol{\theta})$  and the derived information measures depend on the unknown parameter vector  $\boldsymbol{\theta}$ , these designs are said to be only locally optimal designs for the given values of the parameter(s). However, we can still evaluate the information measures over a range of possible parameter values to assess the robustness of the optimality of the design against changes in the parameter values.

### 3.3 Information measures in the gamma-Gompertz model

In this section, we develop the concepts of Sections 3.2.2 and 3.2.3 specifically for the gamma-Gompertz model. After providing details on the computation of the Fisher information matrix, we specify an information measure for  $D_A$ -optimality, and discuss its interpretation. In Section 3.3.4, we show that this measure also plays a role in the calculation of the power of the likelihood ratio test to detect mortality deceleration.

#### 3.3.1 Fisher information in the gamma-Gompertz model

The aim is to derive the Fisher information matrix according to (3.3) specifically for an observation from the gamma-Gompertz model (3.1). For this model, the parameter vector consists of three components: the Gompertz baseline parameters  $a$  and  $b$  and the frailty variance  $\sigma^2$ , so that  $\boldsymbol{\theta} = (a, b, \sigma^2)^\top$ . The density of lifespan  $X$  is given by

$$f_X(x; a, b, \sigma^2) = a e^{bx} \left[ 1 + \sigma^2 \frac{a}{b} (e^{bx} - 1) \right]^{-\left(1 + \frac{1}{\sigma^2}\right)}.$$

(The value  $x = 0$  here marks the age from which the exponentially increasing Gompertz hazard has been established as a good model for human mortality, commonly a mid-adult age such as 30 or 40.)

As is common in the analysis of time-to-event data, the observations are often subject to censoring or truncation. Left truncation occurs in our context if the data are limited to individuals who have survived beyond a certain age  $\check{x}$ , as discussed in Section 3.1. In our case, this left-truncation age is identical for all individuals ( $\check{x} = 90$ ). Censoring occurs if some individuals are still alive at the end of follow-up. In our study, we only analyze birth cohorts who are already extinct – that is, all members have already died – and we will not consider right censoring. (For the calculation of the Fisher information with censoring and truncation for a class of location-scale distributions, see Escobar and Meeker, 1998.)

For left-truncated observations, the Fisher information needs to be calculated for the truncated ( $X \mid X > \check{x}$ ) with density  $f_{X \mid X > \check{x}}(\cdot; \boldsymbol{\theta}) = f_X(\cdot; \boldsymbol{\theta})/S_X(\check{x}; \boldsymbol{\theta})$  on  $(\check{x}, \infty)$ , where  $S_X(x; \boldsymbol{\theta}) = P(X > x; \boldsymbol{\theta})$  denotes the survival function of  $X$ . Consequently, formula (3.3) for the information matrix is adapted as

$$I(\boldsymbol{\theta}) = -\mathbb{E} \left[ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \ln f_{X \mid X > \check{x}}(X; \boldsymbol{\theta}) \mid X > \check{x} \right]$$

$$= - \int \left( \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \ln f_{X|X>\check{x}}(u; \boldsymbol{\theta}) \right) f_{X|X>\check{x}}(u; \boldsymbol{\theta}) du. \quad (3.3')$$

Computing the Fisher information matrix in the gamma-Gompertz model requires the second-order partial derivatives of the log-density of the gamma-Gompertz model  $X$  for complete data, or of  $(X | X > \check{x})$  for left-truncated data with respect to the parameters. The formulas are given in Section 3.8.1. Based on these, we obtain explicit formulas for the observed Fisher information matrix  $\mathcal{J}(\hat{\boldsymbol{\theta}}_n)$ , defined in (3.4), for a given sample with corresponding MLE  $\hat{\boldsymbol{\theta}}_n$ .

In contrast, the exact calculation of the Fisher information matrix  $\mathbf{I}(\boldsymbol{\theta})$  in (3.3') requires taking the (negative) expectations of the second-order partial derivatives. As closed-form expressions for these integrals do not exist, we propose approximating the expectations using numerical integration.

In the absence of an analytical expression for the Fisher information matrix  $\mathbf{I}(\boldsymbol{\theta})$  in the gamma-Gompertz model, there is no closed-form function of  $\mathbf{I}(\boldsymbol{\theta})$  of the parameters  $\boldsymbol{\theta}$ , the sample size  $n$ , and the age at left truncation  $\check{x}$ . However,  $\mathbf{I}(\boldsymbol{\theta})$  can be evaluated over a range of relevant values for these quantities in order to get an impression of how they affect the information matrix. Further computational details are given in Section 3.8.2.

### 3.3.2 $D_A$ -optimality in the gamma-Gompertz model

In the gamma-Gompertz model, the presence or the absence of mortality deceleration is determined by the frailty variance, which also to a large extent controls how strongly the hazard decelerates. Thus, for the assessment of mortality deceleration, our main interest lies in the parameter  $\sigma^2$ , while the Gompertz parameters  $a$  and  $b$  are treated as nuisance. Hence, we will evaluate designs primarily according to the criterion of  $D_A$ -optimality, and define the matrix  $\mathbf{A}$  from Section 3.2.3 as  $\mathbf{A} = (0, 0, 1)$ . The resulting information measure is then  $1/[\mathbf{I}(\boldsymbol{\theta})]_{33}^{-1}$ ; in the following, we will denote  $[\mathbf{I}(\boldsymbol{\theta})]_{33}^{-1}$  as  $\kappa^2$ . It is important to note that  $\kappa^2$  still depends on the true parameter value  $\boldsymbol{\theta}$ , but also on the observation scheme (such as left-truncation age  $\check{x}$ ), although this is suppressed in the notation. A design will be preferred over another if it has a smaller  $\kappa^2$ .

### 3.3.3 Interpretation of information measures in a non-standard setting

As we noted in Section 3.2.2, the use of the Fisher information for study design can be motivated by the result that the asymptotic covariance matrix of the MLEs is given by the inverse Fisher information; see (3.5). This result holds under standard conditions, which are, however, violated in the present framework of the gamma-Gompertz model, because the frailty variance takes a value on the boundary of the parameter space if there is no mortality deceleration ( $\sigma^2 = 0$ ). The asymptotic distribution of the MLE in the gamma-Gompertz model was derived in Böhnstedt and Gampe (2019). For sufficiently large  $\sigma^2 > 0$ , the MLE  $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{\sigma}^2)^\top$  is still asymptotically normal with covariance

matrix  $[\mathbf{I}_n(\boldsymbol{\theta})]^{-1}$  as in (3.5); but for  $\sigma^2 = 0$ , the MLE has a two-component mixture distribution. As a result, minimizing the element  $\kappa^2$  of the inverse Fisher information in order to find an optimal design corresponds to minimizing the asymptotic variance of the parameter estimate  $\hat{\sigma}^2$  only if the true  $\sigma^2 > 0$  is sufficiently large. If  $\sigma^2 = 0$ , the quantity  $n^{-1}\kappa^2$  does not correspond to the variance of  $\hat{\sigma}^2$ .

Nonetheless, the element  $\kappa^2$  can be used for the evaluation of certain design choices – for example, for comparing different alternatives for the age range covered by a sample. In a simulation study (for details about the scenarios see Section 3.5), we found that the relative changes in  $n^{-1}\kappa^2$  were very close to the relative changes in the variance of  $\hat{\sigma}^2$  even if  $\sigma^2 = 0$  (see bottom panels of Figure 3.6). This finding suggests that comparative statements about the amount of information or the variance of  $\hat{\sigma}^2$  for subsets of a sample that cover different age ranges can still be based on ratios of the corresponding  $\kappa^2$ . Unfortunately, this does not apply for comparisons across different scenarios defined by different  $\boldsymbol{\theta}$ .

Consequently, the quantity  $\kappa^{-2}$  should only be related to estimator variance in cases in which this is known to be appropriate. Otherwise, we should stick to the notion of a measure of information; e.g., in the sense of local curvature of the log-likelihood.

### 3.3.4 Power of the likelihood ratio test

A common approach for assessing mortality deceleration in the framework of the gamma-Gompertz model is a likelihood ratio test for  $H_0: \sigma^2 = 0$  against  $H_1: \sigma^2 > 0$ . Under the null hypothesis, the value of the variance parameter lies on the boundary of the parameter space so that the likelihood ratio test statistic is not asymptotically chi-squared distributed with one degree of freedom. Instead, one can adopt the results of Self and Liang (1987) to show that, if  $H_0$  holds, the test statistic asymptotically follows a 50:50 mixture of a chi-squared distribution with one degree of freedom and a point mass at zero. Incorrectly assuming a chi-squared distribution with one degree of freedom for the test statistic implies a larger critical value, hence fewer rejections of  $H_0$ , and ultimately lower power to detect a positive  $\sigma^2 > 0$ .

An explicit formula for the asymptotic power of the likelihood ratio test based on a sample from a gamma-Gompertz model with frailty variance  $\sigma^2$  was derived by Böhnstedt and Gampe (2019). According to their Lemma 6, the power  $\beta_n$  of the likelihood ratio test at level  $\alpha$  and sample size  $n$  can be approximated by

$$\beta_n(\sigma^2) \approx 1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{\sqrt{n}\sigma^2}{\kappa}\right), \quad (3.6)$$

where  $\Phi(\cdot)$  is the standard normal distribution function and  $\kappa$  is the square root of the element of the inverse Fisher information, as defined above. (The proof can be found in the online supplementary material of Böhnstedt and Gampe, 2019.) Thus, through  $\kappa$ , the power of the likelihood ratio test depends on the true parameter  $\boldsymbol{\theta}$ , but also on possible left truncation; that is, on the age range of the data. Based on our computation

of the Fisher information matrix and the resulting  $\kappa^2$ , we can now also determine the power of the likelihood ratio test theoretically, without performing extensive simulation studies. Moreover, with regard to study design, we see from formula (3.6) that designs that minimize  $\kappa^2$  simultaneously maximize the power of the likelihood ratio test to detect mortality deceleration.

### 3.4 Design considerations in assessing mortality deceleration

The precision of parameter estimates and the power of statistical tests to detect mortality deceleration depend on the characteristics of the data set under study. We want to quantify the effects that the size of the sample as well as the age range that it covers have on the information contained in the data about the phenomenon. For that purpose, we denote by  $\mathcal{I}$  a scalar measure of information that is derived from the Fisher information matrix  $\mathbf{I}(\boldsymbol{\theta})$ , such as  $\mathcal{I} = \det(\mathbf{I}(\boldsymbol{\theta}))$  or  $\mathcal{I} = \kappa^{-2}$ .

In the first part of this section, we will discuss how we can assess the effect of the age range that is covered by the data. The age range of a data set is usually restricted because accurate age validation is required, but it is often not feasible to perform the validation for an extensive part of a birth cohort. As mortality deceleration occurs at the tail of the survival distribution, studies that examine this phenomenon focus on the older ages, and, therefore, usually collect information only on survivors beyond a certain age  $x$ . On the one hand, the observation of a death at older ages might be expected to carry more information about mortality deceleration than a death at younger ages. On the other hand, the continuing selection of more robust individuals with lower frailty values leads to a decrease in the variance of frailty among survivors to higher ages (Vaupel et al., 1979; Hougaard, 1984; Economou and Caroni, 2008). Therefore it could become more difficult to assess mortality deceleration for higher left-truncation ages. Moreover, observations of deaths at younger ages can provide indirect information about the parameter  $\sigma^2$ , because they lead to increased precision in the estimation of the Gompertz parameters  $a$  and  $b$ .

To see how these effects trade off, we look at  $\mathcal{I}_{x+}$ , the information measure for an observation left-truncated at age  $x$  for a given  $\boldsymbol{\theta}$ . The pattern of the absolute measure  $\mathcal{I}_{x+}$  across different  $x$  tells us which age range is most informative. In addition, ratios like  $\mathcal{I}_{80+}/\mathcal{I}_{90+}$  quantify the change in information if observations are left-truncated at an earlier age; here, at  $x = 80$ , rather than at a later age, like  $x = 90$ .

The Fisher information matrix  $\mathbf{I}(\boldsymbol{\theta})$  and derived measures such as  $\mathcal{I} = \kappa^{-2}$  correspond to a single observation. If we want to compare the amount of information that is available in a situation in which all survivors to ages  $x = 80$  and above (80+) can be studied to a situation in which only survivors to ages 90+ can be studied, we should also take into account that the 80+ data set will include more individuals than the 90+ data set, because for studies on mortality deceleration, all members of a cohort who survive beyond a certain age will usually be included in the sample. For that purpose, we scale the information measure  $\mathcal{I}_{x+}$  by the probability of obtaining an observation of a death at some age  $x+$ , and define the scaled measure as  $\mathcal{I}_{x+}^{(s)} = \mathcal{I}_{x+} \cdot \mathbf{P}(X > x)$ .

For scenarios with sufficiently large  $\sigma^2$ , the variance of  $\hat{\sigma}^2$  can be approximated by  $n^{-1}\kappa^2$ . Thus, we can also draw conclusions about the precision of the estimate of the frailty variance based on  $\mathcal{I} = \kappa^{-2}$ . In this case, the inverse of the ratio  $\mathcal{I}_{80+}^{(s)}/\mathcal{I}_{90+}^{(s)}$  describes the relative change in the variance of  $\hat{\sigma}^2$  when data on all deaths between ages 80 and 89 could be added to a data set that currently contains information only on all survivors to ages 90+. On the basis of such numbers, practitioners could decide whether it is worthwhile to extend an existing data set to also include information on deaths at earlier ages.

Second, let us turn to some sample size considerations. Mortality studies are generally based on populations or a specific subset thereof, such as all survivors of a birth cohort beyond a certain age. Thus, the sample size is not actively chosen, but simply results from the size of the population under study. Nonetheless, sample size calculations are useful either for judging a priori whether a data set provides enough information to produce meaningful results, or for adequately interpreting the results in comparative studies across different countries. Let us assume that some countries or regions are expected to have similar mortality patterns, and that mortality deceleration has been detected in one of them from an 80+ sample. The information contained in that sample is known to be  $n_{80+}\mathcal{I}_{80+}$ . Then, if for a second country or region only 90+ data are available, we might ask whether these data still contain enough information to detect mortality deceleration. To get an idea of the sample size that is required to draw reliable conclusions under the given mortality pattern  $\theta$ , we could determine the sample size  $n_{90+}$  of the subset of survivors to ages 90 and above, which satisfies  $n_{80+}\mathcal{I}_{80+} = n_{90+}\mathcal{I}_{90+}$ .

Alternatively, sample size considerations could concern the precision of the estimate  $\hat{\sigma}^2$ , which is given by  $n_{x+}^{-1}\kappa_{x+}^2$ , if for the (assumed) mortality regime  $\theta$ , the frailty variance  $\sigma^2$  is sufficiently large to make  $\kappa^2$  the correct variance term (see Section 3.3.3). For a given value of  $\kappa_{x+}^2$ , we can either get an initial idea of the precision of  $\hat{\sigma}^2$  if the size  $n_{x+}$  of the  $x+$  data of the country is known, or we can determine the minimum sample size  $n_{x+}$  that is needed for a desired precision, and see whether potential data sets would fulfill this requirement.

Finally, both the age range of a data set and its sample size affect the power of the likelihood ratio test to detect mortality deceleration. For a given mortality regime  $\theta$ , formula (3.6) allows us to assess what level of power the test will achieve if inference is based on all survivors to ages 90+, or on all survivors to ages 80+.

As all of the above quantities for evaluating the design aspects of the age range and the sample size depend on the true unknown parameter  $\theta$ , we will present some empirical results for specific scenarios in the next section.

### 3.5 Empirical results

In this section, we will study the effects of different designs on the information contained in a data set for some specific scenarios. We assume that  $X$  follows a gamma-Gompertz distribution and describes lifespan after age 60; that is,  $x = 0$  corresponds to age 60.

We choose three scenarios for  $\theta$ , all with the same Gompertz parameters,  $a = 0.015$  and  $b = 0.085$ , but with different values for the frailty variance. Scenario  $S_1$  with  $\sigma^2 = 0.043$  corresponds to the gamma-Gompertz model estimated from the female sample in Section 3.6. For Scenario  $S_2$  with  $\sigma^2 = 0.021$ , the frailty variance is roughly halved, representing a less heterogeneous population for which mortality deceleration is less pronounced. In Scenario  $S_3$  with  $\sigma^2 = 0$ , there is no mortality deceleration. In Section 3.8.3, we present results for additional Scenarios  $S_4$  to  $S_6$  with the same three values for the frailty variance as above, but different values for the Gompertz parameters. In particular, we set  $a = 0.021$  and  $b = 0.082$  equal to the estimates obtained from a gamma-Gompertz fit to the male sample in Section 3.6. With regard to the age range, we assess the information measures for complete observations of  $X$  (that is, ages 60 and above, 60+), as well as for left-truncated observations corresponding to survivors to ages 80 and above (80+), 85+, and 90+. The effect of the sample size will be examined by considering different sizes of the subset of survivors to ages 90+; namely,  $n_{90+} = 10,000$  (small),  $n_{90+} = 20,000$  (medium), or  $n_{90+} = 105,000$  (large). The small and medium sizes are close to the sizes of the male and female samples in Section 3.6. All computations are run in R (R Core Team, 2019), and the numerical integration to calculate the Fisher information  $\mathbf{I}(\theta)$  is performed using function `integrate()`.

In a preliminary analysis, we assessed the performance of our approach of using numerical integration to calculate  $\mathbf{I}(\theta)$ . For that purpose, we generated 1,000 samples for each of the scenarios  $S_1$  to  $S_3$ , in which three initial sample sizes at age 60 were determined to yield the desired  $n_{90+}$  given above. For each sample, we estimated the parameters of the gamma-Gompertz model based on the full sample (60+), and based on the subsets of survivors to ages 80+, 85+, and 90+, by maximizing the log-likelihoods numerically using function `nlm()`. We then calculated the averages of the observed Fisher information matrices evaluated at the MLEs across the 1,000 samples of each fixed setting,  $\bar{\mathcal{J}} = \frac{1}{1000} \sum_{r=1}^{1000} \mathcal{J}(\hat{\theta}_n^{(r)})$ . Finally, these averages were compared to the Fisher information matrices  $\mathbf{I}(\theta)$ , that were scaled by the theoretical size  $n_{\cdot+}$  of a sample from the respective setting. The results are reported in Table 3.3 in Section 3.8.3. As expected, mean relative differences decrease with sample size, width of age range and size of the frailty variance. For ages 60+ and 80+ differences are negligible throughout, and for 85+ surpass 0.02 only in the no-frailty scenario ( $n_{90+} = 10,000$ : 0.03389;  $n_{90+} = 20,000$ : 0.02114). For ages 90+ and smallest sample size  $n_{90+} = 10,000$ , the values are  $S_1$ : 0.05487,  $S_2$ : 0.06592, and  $S_3$ : 0.12065.

### 3.5.1 Effect of the age at left truncation

In the following, we quantify how different restrictions of the age range covered by a data set affect the amount of information that is provided by the data. We mainly focus on the criterion for  $D_A$ -optimality (see Section 3.3.2), that is,  $\mathcal{I} = \kappa^{-2}$ .

The left panel of Figure 3.1 shows the values of  $\mathcal{I}$  and its scaled version  $\mathcal{I}^{(s)}$  for Scenario  $S_1$  when the observations are complete (60+) or left-truncated at higher ages (80+,

85+, or 90+). We find that the amount of information contained in an observation decreases as the age of left truncation increases. This effect is even more pronounced for the scaled measure of information  $\mathcal{I}^{(s)}$ , because the probability of observing deaths decreases at the higher ages. Hence, in terms of  $\kappa^{-2}$ , a situation in which only survivors to ages 90+ can be studied indeed provides less information than a situation in which survivors to ages 80+ can be studied, indicating that mortality deceleration is more difficult to assess for higher ages of left truncation. The right panel of Figure 3.1 displays the ratios  $\mathcal{I}_{x+}/\mathcal{I}_{80+}$  and  $\mathcal{I}_{x+}^{(s)}/\mathcal{I}_{80+}^{(s)}$  for  $x = 80, 85, 90$  in Scenario  $S_1$ . We see that if only data on survivors to ages 90+ are available, more than half of the information is lost compared to the situation in which data on survivors to ages 80+ are available. Taking into account the smaller size of the subset of survivors to ages 90+, the loss even amounts to around 87%. The results for Scenarios  $S_2$  and  $S_3$  are similar (cf. Figure 3.2).

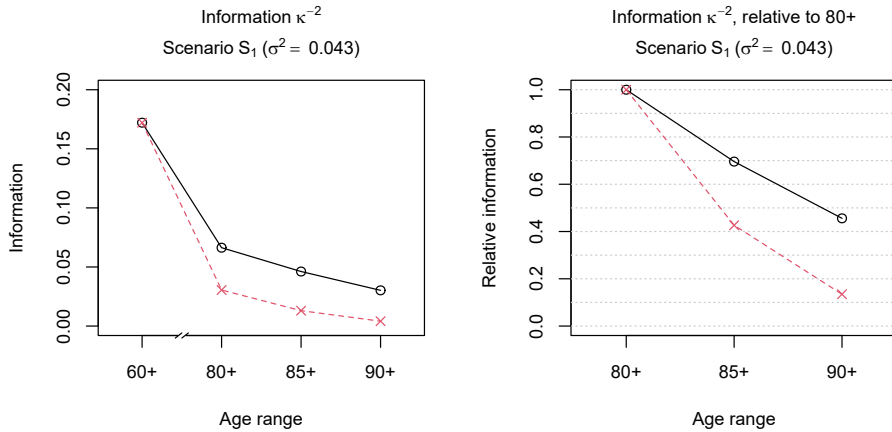


Figure 3.1: Information measure  $\mathcal{I} = \kappa^{-2}$  (black-solid line, circles) and scaled measure  $\mathcal{I}^{(s)}$  (red-dashed line, crosses) under Scenario  $S_1$  depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+). Left: absolute values of (scaled)  $\mathcal{I}$ , right: (scaled) ratios  $\mathcal{I}_{x+}/\mathcal{I}_{80+}$  for  $x = 80, 85, 90$  (connection of the values by lines is only for ease of visual inspection).

We have already briefly discussed in Section 3.3.3 the relationship between the information measure  $\kappa^{-2}$  and the asymptotic variance of the estimator  $\hat{\sigma}^2$ . In settings with sufficiently large  $\sigma^2$ , the variance of  $\hat{\sigma}^2$  is approximately equal to  $\kappa^2$  scaled by the inverse of the sample size. The top-left panel of Figure 3.6 in Section 3.8.3 verifies this for the medium-sized Scenario  $S_1$  with different observation schemes (60+, 80+, 85+, and 90+) by comparing the empirical variance of  $\hat{\sigma}^2$  across the 1,000 replications with the scaled  $\kappa^2$ . In contrast, if  $\sigma^2 = 0$ , the asymptotic variance of  $\hat{\sigma}^2$  is not given by the scaled  $\kappa^2$ , as shown in the bottom-left panel of Figure 3.6 for the medium-sized Scenario  $S_3$ . However, the relative changes in the scaled  $\kappa_{x+}^2$  across different age ranges  $x+$  are in line

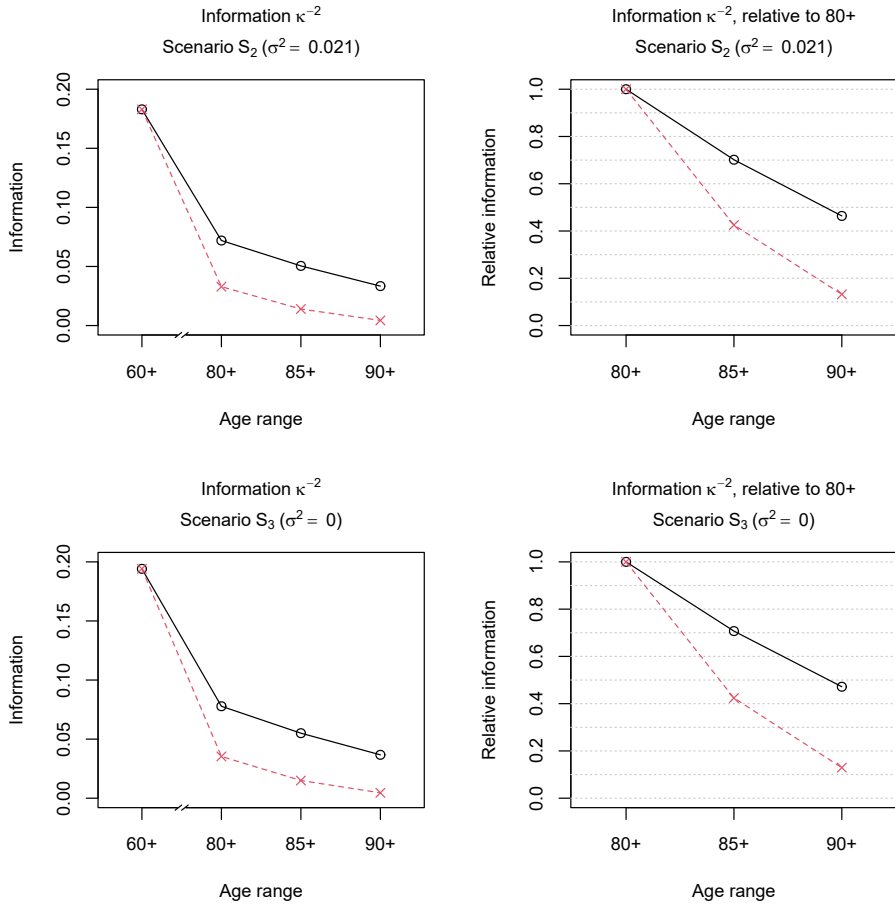


Figure 3.2: Information measure  $\mathcal{I} = \kappa^{-2}$  (black-solid line, circles) and scaled measure  $\mathcal{I}^{(s)}$  (red-dashed line, crosses) under Scenarios  $S_2$  (top) and  $S_3$  (bottom) depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+). Left: absolute values of (scaled)  $\mathcal{I}$ , right: (scaled) ratios  $\mathcal{I}_{x+}/\mathcal{I}_{80+}$  for  $x = 80, 85, 90$ .

with the relative changes in the empirical variances for both Scenario  $S_1$  (top-right panel of Figure 3.6) and Scenario  $S_3$  (bottom-right panel of Figure 3.6). Consequently, ratios  $\mathcal{I}_{x+}^{(s)}/\mathcal{I}_{y+}^{(s)}$  can be readily interpreted in terms of information gain or variance reduction when considering different age ranges  $x+$  and  $y+$ , even if  $\sigma^2 = 0$ . For example, in Scenario  $S_3$ , about 82% of the information in the full sample 60+ is lost if only survivors to ages 80+ can be studied.



While the importance of the variance parameter  $\sigma^2$  for the assessment of mortality deceleration suggested that we should use the criterion of  $D_A$ -optimality, we also looked into the performance of the other information measures introduced in Section 3.2.3. The results for Scenarios  $S_1$  and  $S_3$  are presented in Figures 3.7 and 3.8 in Section 3.8.3. The absolute values of the information measures and their patterns across the different age ranges for  $A$ - and  $E$ -optimality are very close to the ones we observed for  $D_A$ -optimality with  $\mathcal{I} = \kappa^{-2}$ . This can be explained by the fact that the information measure for  $A$ -optimality, that is, the inverse of the sum of the eigenvalues of  $[\mathbf{I}(\boldsymbol{\theta})]^{-1}$ , is dominated by one eigenvalue of relatively large magnitude in the current setting. As this largest eigenvalue of the inverse information matrix is at the same time the target of the criterion for  $E$ -optimality and in this case also closely related to the measure  $\kappa^2$  for  $D_A$ -optimality, the three criteria yield very similar results. In contrast, the criterion of  $D$ -optimality suggests somewhat larger relative losses in information when restricting the age range. This is a consequence of directly accounting for the increased uncertainty in all directions of the parameter space, as the eigenvalues of  $\mathbf{I}(\boldsymbol{\theta})$  are multiplied in the measure  $\det(\mathbf{I}(\boldsymbol{\theta}))$ .

The quantity  $\kappa^{-2}$ , considered for  $D_A$ -optimality, measures the information contained in an observation about the parameter  $\sigma^2$ , and takes into account the correlation between  $\hat{\sigma}^2$  and the estimates of the Gompertz parameters  $\hat{a}$  and  $\hat{b}$ . Alternatively, we could study the information measure  $[\mathbf{I}(\boldsymbol{\theta})]_{33}$  given by

$$[\mathbf{I}(\boldsymbol{\theta})]_{33} = -\mathbb{E} \left[ \frac{\partial^2}{\partial (\sigma^2)^2} \ln f_X(X; \boldsymbol{\theta}) \right].$$

This element of the Fisher information matrix describes the average curvature of the log-density of the gamma-Gompertz model with respect to  $\sigma^2$  for fixed Gompertz parameters  $a$  and  $b$ . Figure 3.3 shows that in terms of this measure, the information increases with the increasing age of left truncation in Scenarios  $S_1$  and  $S_3$ . This supports the idea that observations of later ages at death carry more information about the potential deceleration, as measured by  $\sigma^2$ . However, looking at the scaled measure reveals that, in practice, this effect is compensated for by the decreasing number of survivors to higher ages.

The above findings regarding the effect of the age at left truncation on the different information measures generally hold also for the Scenarios  $S_4$  and  $S_6$  with modified Gompertz parameters (see Figures 3.9 to 3.12 in Section 3.8.3). Increases in the age at left truncation result in considerable information loss according to all the criteria, with the exception of  $\mathcal{I} = [\mathbf{I}(\boldsymbol{\theta})]_{33}$ , for which the information loss is only revealed if the smaller size of the subset of survivors to higher ages is taken into account. Compared to Scenarios  $S_1$  and  $S_3$ , the higher initial level of mortality  $a$  in Scenarios  $S_4$  and  $S_6$  leads to smaller absolute values of the information measures and slightly larger information losses when restricting the age range. This is expected because the higher initial mortality leads to stronger selection effects and thus, a stronger decrease in the variance of frailty among survivors to higher ages, as well as to lower probabilities of surviving to these ages.

Finally, the information measures computed in this subsection refer to single observations, such that the derived conclusions about the effects of the age at left truncation

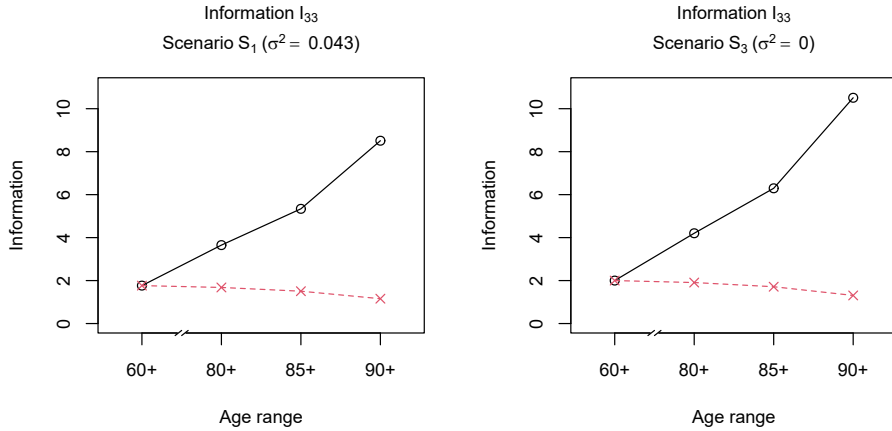


Figure 3.3: Information measure  $\mathcal{I} = [\mathbf{I}(\boldsymbol{\theta})]_{33}$  (black-solid line, circles) and scaled measure  $\mathcal{I}^{(s)}$  (red-dashed line, crosses) depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+) under Scenarios  $S_1$  (left) and  $S_3$  (right).

should be valid for samples of any reasonable size under the different scenarios. Further aspects regarding the size of data sets under study are discussed in the following.

### 3.5.2 Sample size considerations

Let us suppose that old-age mortality is studied in countries that experience the same or very similar mortality patterns. It is clear that the different population sizes affect the assessment of mortality deceleration in the different countries. If the data for the different countries also cover different age ranges, then we want to know, for example, what size  $n_{90+}$  a subset of survivors to ages 90+ would have to be to carry as much information as a subset of survivors to ages 80+ from a different country; that is,  $n_{80+}\mathcal{I}_{80+} = n_{90+}\mathcal{I}_{90+}$ . In Scenarios  $S_1$  to  $S_3$ , the ratios  $\mathcal{I}_{80+}/\mathcal{I}_{90+}$  for  $\mathcal{I} = \kappa^{-2}$  are 2.194, 2.156, and 2.120, respectively. Thus, a sample of survivors to ages 90+ needs to be more than twice as large as a sample of survivors to ages 80+ in order to provide the same amount of information. Moreover, the ratios are increasing in the level of heterogeneity  $\sigma^2$ . For Scenarios  $S_4$  to  $S_6$ , we obtain the slightly larger ratios 2.322, 2.278, and 2.237, reflecting the stronger selection effects in these settings.

If the underlying  $\sigma^2$  is sufficiently large, we can use the measure  $\kappa^2$  to draw conclusions about the precision of  $\hat{\sigma}^2$  for specific sample sizes. In Scenario  $S_1$ , if the sample of survivors to ages 90+ consists of about  $n_{90+} = 20,000$  individuals, this would yield a precision of the estimated frailty variance of about  $\text{var}(\hat{\sigma}^2) \approx n_{90+}^{-1} \kappa_{90+}^2 \sigma^2 = 0.00165$ .

### 3.5.3 Power of the likelihood ratio test

We now evaluate how the age range and the sample size of a data set affect the performance of the likelihood ratio test for assessing mortality deceleration in the gamma-Gompertz model. The power of the test to detect a positive  $\sigma^2$  in Scenarios  $S_1$  and  $S_2$  based on different subsets and sample sizes is calculated from (3.6) at a level of  $\alpha = 0.05$ , and is reported in Table 3.1. As expected, we find that the power of the test increases if the sample size increases, if the age at left truncation decreases, and if the frailty variance is larger. In both scenarios, the power to detect mortality deceleration decreases by more than two-thirds if inference is based only on survivors to ages 90+, rather than on all survivors to ages 80+ for samples of medium size.

Table 3.1: Power  $\beta$  of the likelihood ratio test, performed at the 5% level, according to formula (3.6), under Scenarios  $S_1$  ( $\sigma^2 = 0.043$ ) and  $S_2$  ( $\sigma^2 = 0.021$ ) for three sample size settings (s – small, m – medium, l – large) and varying age range.

Scen.	$n$	Survivors to ages							
		60+		80+		85+		90+	
		$n_{60+}$	$\beta_{60+}$	$n_{80+}$	$\beta_{80+}$	$n_{85+}$	$\beta_{85+}$	$n_{90+}$	$\beta_{90+}$
$S_1$	s	73,558	0.999	33,841	0.653	20,740	0.377	10,000	0.185
	m	147,116	1.000	67,681	0.892	41,480	0.593	20,000	0.278
	l	772,361	1.000	355,327	1.000	217,771	0.996	105,000	0.782
$S_2$	s	76,853	0.801	35,123	0.278	21,290	0.169	10,000	0.104
	m	153,706	0.970	70,245	0.440	42,581	0.251	20,000	0.135
	l	806,956	1.000	368,788	0.962	223,548	0.721	105,000	0.344

The performance of the likelihood ratio test under the Scenarios  $S_4$  and  $S_5$ , with modified Gompertz parameters, is documented in Table 3.4 in Section 3.8.3 and leads to the same conclusions as above. In addition, we see that in Scenario  $S_4$  the test has lower power to detect the positive  $\sigma^2$  based on the survivors to ages 85+ or 90+ than in the corresponding Scenario  $S_1$ . This is due to the higher initial level of mortality  $a$  in  $S_4$  which reduces the heterogeneity in the mortality risks at the higher ages. The power of the test based on survivors to ages 60+ and 80+ in Scenarios  $S_4$  and  $S_5$  is not directly comparable to the power in the corresponding Scenarios  $S_1$  and  $S_2$ , because of the different sizes of the subsets 60+ and 80+ under the two settings for the Gompertz parameters.

Apart from calculating power values for given parameter configurations, formula (3.6) provides a tool for determining what age range a data set should cover to ensure that the likelihood ratio test will achieve a certain level of power. From Table 3.1, we see that for the medium-sized Scenario  $S_1$ , the likelihood ratio test will detect mortality deceleration in about 89.2% of cases based on the sample of survivors to ages 80+. As for any left-truncation age  $x$ , the size of the sample  $x+$  can be calculated based on the given (or estimated) gamma-Gompertz parameters and a known subset size (e.g., here,  $n_{90+} = 20,000$ ), we can determine the left-truncation age  $x$  such that the power is increased to

95%. For the medium-sized Scenario  $S_1$ , we need to include all survivors to ages 78 and above for the test to reach a power of 95%.

The above power calculations are based on relatively large sample sizes as motivated by real cohort data. We have seen that the assessment of mortality deceleration can be demanding even based on data sets of such size especially in case of restricted age ranges. Smaller sample sizes will lower the power of the test to detect a positive frailty variance also for the subsets of survivors to ages 60+ or 80+. A larger underlying frailty variance leads to stronger deceleration in the hazard rate and will generally be favorable for detecting the phenomenon. However, smaller sample sizes and thus increased uncertainty about the parameters can counteract this effect, as can be seen from formula (3.6). While the formula provides only a large-sample approximation to the power of the likelihood ratio test, the sample sizes in human mortality studies are expected to be generally large enough for the approximation to yield valid results.

### 3.6 A study on old-age mortality among French-Canadians

In this section, we apply the proposed methods for evaluating study design and deriving design recommendations to a study on old-age mortality among Catholic French-Canadians born at the end of the 19<sup>th</sup> century.

The data set contains information on 20,917 females and 10,878 males who were born in the Province of Quebec between 1880 and 1896, and who died in Quebec at ages 90 and above between 1970 and 2009. To validate the individual exact survival times, birth registration documents and death certificates from Quebec's parish register archives were linked. Further details on the data and the validation procedure can be found in Ouellette and Bourbeau (2014) and Ouellette (2016), who studied earlier versions of this data set that covered only the centenarians, that is, survivors to ages 100 and above.

The analysis of the French-Canadian mortality data based on the gamma-Gompertz model is conducted separately for the female and the male sample. The starting age of the model is assumed to be 60, and the likelihood is adapted for the left truncation at age 90. We obtain estimates of the frailty variance of  $\hat{\sigma}^2 = 0.043$  for the females and  $\hat{\sigma}^2 = 0.037$  for the males. The likelihood ratio test for  $H_0: \sigma^2 = 0$  leads to  $p$ -values of 0.121 for the females and 0.283 for the males, indicating that the data do not provide much evidence against the null hypothesis of no mortality deceleration. These findings are in contrast to those for the fitted hazards and the empirical death rates, which are displayed in Figure 3.4, and suggest a deceleration, at least for the females. Indeed, it turns out that the likelihood ratio test has relatively low power to detect mortality deceleration in the given settings. According to formula (3.6) with the estimated values of the parameters, the power of the likelihood ratio test at the 5% level based on a 90+ sample of the given size is 28.7% in the female setting and 14.2% in the male setting, respectively. Therefore, we want to investigate how a further extension of the data set that would include deaths at earlier ages – say, between ages 85 and 89, or between 80 and 89 – could impact the assessment of mortality deceleration.

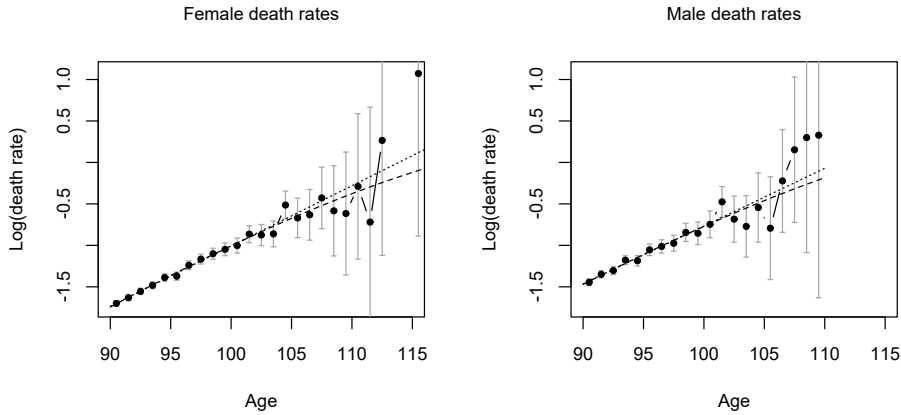


Figure 3.4: Death rates of French-Canadian females (left) and males (right): empirical death rates (solid line, circles) with 95%-confidence intervals (gray), gamma-Gompertz fit (dashed) and Gompertz fit (dotted).

First, we examine the behavior of the information measure  $\mathcal{I} = \kappa^{-2}$  for observations from a gamma-Gompertz model with parameter values equal to the estimates obtained from the female or the male sample, respectively. Figure 3.5 depicts the ratios  $\mathcal{I}_{x+}/\mathcal{I}_{90+}$  for left truncation at the ages  $x = 80, 85, 90$ , as well as the ratios of the scaled measure  $\mathcal{I}^{(s)}$ . In both the female and the male setting, an observation left-truncated at age 80 would be more than twice as informative as an observation left-truncated at age 90. Taking into account the increasing number of observations when younger ages at death are included by looking at  $\mathcal{I}^{(s)}$ , observations left-truncated at age 85 are already more than twice as informative as observations left-truncated at age 90. In other words, the extended data set that includes all survivors to ages 85 and above would contain more than twice as much information as the current data set of survivors to ages 90 and above. Indeed, compared to the 90+ sample, the female 85+ sample would provide about three times as much information, and the male 85+ sample would provide about four times as much information.

Second, Table 3.2 summarizes the effects of expanding the age range of the current data set on the power of the likelihood ratio test which is performed at a level of 5%. The calculations are again based on formula (3.6), with the estimates obtained from the female and the male samples, respectively, inserted for the parameter values. The sizes of the expanded 85+ and 80+ data sets are computed from the known size of the population of survivors to ages 90+ and the fitted gamma-Gompertz model. For both the female and the male data, we find that expanding the data set to deaths between ages 85 and 89 would more than double the power of the likelihood ratio test. For the female setting, the power

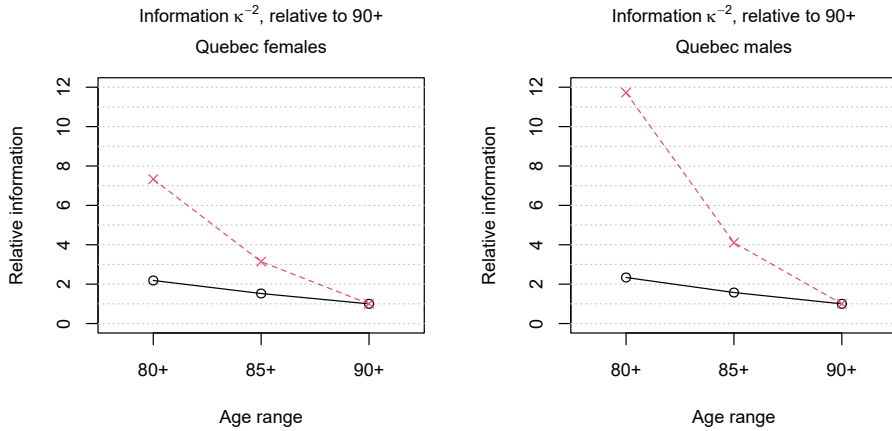


Figure 3.5: Ratios  $\mathcal{I}_{x+}/\mathcal{I}_{90+}$  (black-solid line, circles) or  $\mathcal{I}_{x+}^{(s)}/\mathcal{I}_{90+}^{(s)}$  (red-dashed line, crosses) for the information measure  $\mathcal{I} = \kappa^{-2}$  depending on the age of left truncation  $x = 80, 85, 90$ , under the parameter settings estimated from the samples of French-Canadian females (left) and males (right).

of the test based on the 85+ sample is around 60.9%. In line with the considerations in Section 3.5.3, the age of left truncation that is required in this setting to achieve the desired power level of 80% is found to be age 82.

Table 3.2: Power  $\beta$  of the likelihood ratio test, performed at the 5% level, according to formula (3.6), under the parameter settings estimated from the French-Canadian data for different age ranges  $x+$  and resulting sample sizes  $n_{x+}$ .

Scenario	Survivors to ages					
	80+		85+		90+	
	$n_{80+}$	$\beta_{80+}$	$n_{85+}$	$\beta_{85+}$	$n_{90+}$	$\beta_{90+}$
Females	70,085	0.901	43,126	0.609	20,917	0.287
Males	54,577	0.627	28,369	0.316	10,878	0.142

In conclusion, our results show that the failure of the likelihood ratio test to reject the hypothesis of no mortality deceleration in the female 90+ sample can be explained to some extent by the low power of the test in the specific setting. Both the proposed information measure  $\kappa^{-2}$  and the power calculations for the likelihood ratio test demonstrate that an expansion of the existing data set on French-Canadian mortality could greatly improve the assessment of mortality deceleration in this population. In practice, these potential improvements have to be weighed against the costs of collecting – and validating – the additional data.

### 3.7 Discussion

We have investigated the use of Fisher information-based criteria for planning and evaluating studies that assess mortality deceleration in the framework of the gamma-Gompertz model. Our aim was to derive recommendations for settings in which the parameters of interest could be reliably estimated, and a deceleration in the death rates could be detected with a high probability. As validation of the ages at death is often required in old-age mortality studies, these settings are characterized by the age range covered by the data and the sample size.

The essential component of the proposed methods is the computation of the Fisher information matrix for potentially left-truncated observations from a gamma-Gompertz model. Due to a lack of closed-form expressions, the information matrix is obtained using numerical integration of analytically determined second-order partial derivatives of the log-density. Different criteria for evaluating study designs can be derived from the Fisher information. Given the importance of the frailty variance parameter in assessing mortality deceleration in the gamma-Gompertz model, we focus primarily on a criterion of  $D_A$ -optimality, whereby the Gompertz parameters are treated as nuisance. The resulting measure of information is the reciprocal of the element of the inverse Fisher information that corresponds to  $\sigma^2$ . It allows us to quantify the effects of the sample size and the age range covered by a data set on the amount of information that this data set contains about  $\sigma^2$ . Based on the computation of the Fisher information matrix, we are also able to calculate the power of the likelihood ratio test to detect mortality deceleration in specific scenarios. As a result, recommendations can be given about what age range a data set needs to cover for the likelihood ratio test to achieve a certain power. In the illustration with a study on French-Canadian mortality, the information measures and the power calculations clearly demonstrate that the assessment of mortality deceleration could be greatly improved if the current data set, which includes only survivors to ages 90 and above, was extended to also include deaths at the earlier ages 85 to 89.

Here, we only consider changes in the age at left truncation (the age range covered by the data) that apply to all survivors in a cohort, which is the most common setting in demographic studies. We could, however, extend these considerations to more complex design questions for which actually random samples could be drawn from the observed survivors to particular ages. In such situations, the quantification of information in particular observations would be even more crucial for attaining an optimal design. However, addressing such questions is beyond the scope of the current research.

The present work has some limitations. As our focus is on the assessment of mortality deceleration – the deviation from the log-linear hazard trajectory of the Gompertz model at high ages –, we have stayed within the framework of the gamma-Gompertz model here. However, as the concepts of the Fisher information and of optimal design are defined for any parametric model, the proposed methods should be applicable in a wider context.

When the age range can be extended to appreciably lower ages the exponential increase of senescent mortality may no longer hold and questions of model choice for the

baseline hazard arise. Established model selection techniques can help here (see Burnham and Anderson, 2002). In such cases most problems discussed in this chapter would be obsolete though. It should be noted, however, that also commonly used model choice criteria, such as Akaike's Information Criterion (AIC), are affected by non-standard conditions induced by a boundary parameter (see Böhnstedt and Gampe, 2019).

Although the approach of using numerical integration to compute the Fisher information matrix in the gamma-Gompertz model seems to perform well, it prevents us from deriving general analytical formulas that describe the effects of the sample size and the age range of a data set. Nonetheless, empirical studies for specific parameter settings and design choices as presented here can serve as a basis for formulating recommendations.

In this context, we have to bear in mind that the gamma-Gompertz model provides a non-standard setting, and, hence, that the information measures are not directly related to estimator variability in the boundary case ( $\sigma^2 = 0$ ). We have, however, shown that comparative statements on information gain or variance reduction are still meaningful when designs covering different age ranges are compared.

Finally, as the assessment of the performance of different study designs for the gamma-Gompertz model based on the Fisher information depends on the true underlying parameter values, it is valid only locally. Still, the robustness of the design's performance can be checked by evaluating the information measures using a range of possible parameter values.

## 3.8 Supplementary material

### 3.8.1 Derivatives of gamma-Gompertz log-densities

In the following subsections, we give explicit formulas of the second-order partial derivatives of the log-density of complete or left-truncated observations from a gamma-Gompertz model. These derivatives are the basis for computing the Fisher information matrix  $\mathbf{I}(\boldsymbol{\theta})$  according to formula (3.3) or (3.3'),

$$\mathbf{I}(\boldsymbol{\theta}) = -\mathbb{E} \left[ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \ln f_X(X; \boldsymbol{\theta}) \right]. \quad (3.3)$$

#### Complete data

For the gamma-Gompertz model (3.1), the log-density  $\ln f_X(\cdot; a, b, \sigma^2)$  of complete data  $X$  takes the form

$$\ln f_X(x; a, b, \sigma^2) = \ln a + bx - \left( 1 + \frac{1}{\sigma^2} \right) \ln \left[ 1 + \sigma^2 \frac{a}{b} (e^{bx} - 1) \right].$$

Its partial derivatives with respect to the parameters are calculated as

$$\frac{\partial \ln f_X}{\partial a} = \frac{1}{a} - \frac{(\sigma^2 + 1)}{b} \cdot \frac{e^{bx} - 1}{1 + \sigma^2 \frac{a}{b} (e^{bx} - 1)},$$



$$\frac{\partial \ln f_X}{\partial b} = x - \frac{a(\sigma^2 + 1)}{b^2} \cdot \frac{bx e^{bx} - (e^{bx} - 1)}{1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)}, \quad \text{and}$$

$$\frac{\partial \ln f_X}{\partial \sigma^2} = \frac{1}{\sigma^4} \ln \left[ 1 + \sigma^2 \frac{a}{b} (e^{bx} - 1) \right] - \left( 1 + \frac{1}{\sigma^2} \right) \frac{a}{b} \cdot \frac{e^{bx} - 1}{1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)}.$$

The second-order partial derivatives equal

$$\begin{aligned} \frac{\partial^2 \ln f_X}{\partial a^2} &= -\frac{1}{a^2} + \frac{\sigma^2(\sigma^2 + 1)}{b^2} \cdot \frac{(e^{bx} - 1)^2}{[1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)]^2} \\ \frac{\partial^2 \ln f_X}{\partial a \partial b} &= \frac{(\sigma^2 + 1)}{b^2} \cdot \frac{(e^{bx} - 1) - bx e^{bx}}{[1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)]^2} \\ \frac{\partial^2 \ln f_X}{\partial a \partial \sigma^2} &= \frac{1}{b^2} \cdot \frac{a(e^{bx} - 1)^2 - b(e^{bx} - 1)}{[1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)]^2} \\ \frac{\partial^2 \ln f_X}{\partial b^2} &= \frac{a(\sigma^2 + 1)}{b} \cdot \frac{\frac{2}{b} x e^{bx} - \frac{2}{b^2}(e^{bx} - 1) - \sigma^2 \frac{a}{b^3}(e^{bx} - 1)^2 + (\sigma^2 \frac{a}{b} - 1)x^2 e^{bx}}{[1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)]^2} \\ \frac{\partial^2 \ln f_X}{\partial b \partial \sigma^2} &= \frac{a}{b} \cdot \frac{[\frac{1}{b}(e^{bx} - 1) - x e^{bx}][1 - \frac{a}{b}(e^{bx} - 1)]}{[1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)]^2} \\ \frac{\partial^2 \ln f_X}{\partial (\sigma^2)^2} &= -\frac{2}{\sigma^6} \ln \left[ 1 + \sigma^2 \frac{a}{b} (e^{bx} - 1) \right] + \frac{2}{\sigma^4} \frac{a}{b} \cdot \frac{e^{bx} - 1}{1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)} \\ &\quad + \left( 1 + \frac{1}{\sigma^2} \right) \frac{a^2}{b^2} \cdot \frac{(e^{bx} - 1)^2}{[1 + \sigma^2 \frac{a}{b}(e^{bx} - 1)]^2}. \end{aligned} \quad (3.7)$$

In the limit  $\sigma^2 \rightarrow 0$ , we obtain

$$\begin{aligned} \frac{\partial^2 \ln f_X}{\partial a^2} &= -\frac{1}{a^2} \\ \frac{\partial^2 \ln f_X}{\partial a \partial b} &= \frac{1}{b^2}(e^{bx} - 1) - \frac{1}{b} x e^{bx} \\ \frac{\partial^2 \ln f_X}{\partial a \partial \sigma^2} &= \frac{a}{b^2}(e^{bx} - 1)^2 - \frac{1}{b}(e^{bx} - 1) \\ \frac{\partial^2 \ln f_X}{\partial b^2} &= \frac{2a}{b^2} x e^{bx} - \frac{2a}{b^3}(e^{bx} - 1) - \frac{a}{b} x^2 e^{bx} \\ \frac{\partial^2 \ln f_X}{\partial b \partial \sigma^2} &= \frac{a}{b} \left[ \frac{1}{b}(e^{bx} - 1) - x e^{bx} \right] \left[ 1 - \frac{a}{b}(e^{bx} - 1) \right] \\ \frac{\partial^2 \ln f_X}{\partial (\sigma^2)^2} &= -\frac{2a^3}{3b^3}(e^{bx} - 1)^3 + \frac{a^2}{b^2}(e^{bx} - 1)^2, \end{aligned}$$

where we have applied the rule of L'Hôpital for the last equation.

### Left-truncated data

For the gamma-Gompertz model (3.1), the log-density  $\ln f_{X|X>y}(\cdot; a, b, \sigma^2)$  for data  $(X | X > y)$ , left-truncated at age  $y$ , takes the form

$$\begin{aligned} \ln f_{X|X>y}(x; a, b, \sigma^2) &= \ln f_X(x; a, b, \sigma^2) - \ln S_X(y; a, b, \sigma^2) \\ &= \ln f_X(x; a, b, \sigma^2) + \frac{1}{\sigma^2} \ln \left[ 1 + \sigma^2 \frac{a}{b} (e^{by} - 1) \right], \end{aligned}$$

for  $x > y$ . The partial derivatives of the first summand have already been presented in the preceding subsection. Thus, we focus on the second summand here, which we denote as  $g(y; a, b, \sigma^2)$ . The partial derivatives of  $g$  with respect to the parameters are computed as

$$\begin{aligned} \frac{\partial g}{\partial a} &= \frac{1}{b} \cdot \frac{e^{by} - 1}{1 + \sigma^2 \frac{a}{b} (e^{by} - 1)}, \\ \frac{\partial g}{\partial b} &= \frac{a}{b^2} \cdot \frac{bye^{by} - (e^{by} - 1)}{1 + \sigma^2 \frac{a}{b} (e^{by} - 1)}, \quad \text{and} \\ \frac{\partial g}{\partial \sigma^2} &= -\frac{1}{\sigma^4} \ln \left[ 1 + \sigma^2 \frac{a}{b} (e^{by} - 1) \right] + \frac{a}{b\sigma^2} \cdot \frac{(e^{by} - 1)}{1 + \sigma^2 \frac{a}{b} (e^{by} - 1)}. \end{aligned}$$

The second-order partial derivatives read

$$\begin{aligned} \frac{\partial^2 g}{\partial a^2} &= -\frac{\sigma^2}{b^2} \cdot \frac{(e^{by} - 1)^2}{\left[ 1 + \sigma^2 \frac{a}{b} (e^{by} - 1) \right]^2} \\ \frac{\partial^2 g}{\partial a \partial b} &= \frac{1}{b^2} \cdot \frac{1 - e^{by} + bye^{by}}{\left[ 1 + \sigma^2 \frac{a}{b} (e^{by} - 1) \right]^2} \\ \frac{\partial^2 g}{\partial a \partial \sigma^2} &= -\frac{a}{b^2} \cdot \frac{(e^{by} - 1)^2}{\left[ 1 + \sigma^2 \frac{a}{b} (e^{by} - 1) \right]^2} \\ \frac{\partial^2 g}{\partial b^2} &= \frac{a}{b} \cdot \frac{\frac{2}{b^2} (e^{by} - 1) + \sigma^2 \frac{a}{b^3} (e^{by} - 1)^2 - \frac{2}{b} ye^{by} + (1 - \sigma^2 \frac{a}{b}) y^2 e^{by}}{\left[ 1 + \sigma^2 \frac{a}{b} (e^{by} - 1) \right]^2} \\ \frac{\partial^2 g}{\partial b \partial \sigma^2} &= -\frac{a^2}{b^3} \cdot \frac{bye^{by} (e^{by} - 1) - (e^{by} - 1)^2}{\left[ 1 + \sigma^2 \frac{a}{b} (e^{by} - 1) \right]^2} \\ \frac{\partial^2 g}{\partial (\sigma^2)^2} &= \frac{2}{\sigma^6} \ln \left[ 1 + \sigma^2 \frac{a}{b} (e^{by} - 1) \right] - \frac{2}{\sigma^4} \cdot \frac{\frac{a}{b} (e^{by} - 1)}{1 + \sigma^2 \frac{a}{b} (e^{by} - 1)} \\ &\quad - \frac{1}{\sigma^2} \left[ \frac{\frac{a}{b} (e^{by} - 1)}{1 + \sigma^2 \frac{a}{b} (e^{by} - 1)} \right]^2. \end{aligned} \quad (3.8)$$

In the limit  $\sigma^2 \rightarrow 0$ , we have

$$\frac{\partial^2 g}{\partial a^2} = 0$$

$$\begin{aligned}\frac{\partial^2 g}{\partial a \partial b} &= \frac{1}{b^2} \cdot [1 - e^{by} + bye^{by}] \\ \frac{\partial^2 g}{\partial a \partial \sigma^2} &= -\frac{a}{b^2} (e^{by} - 1)^2 \\ \frac{\partial^2 g}{\partial b^2} &= \frac{a}{b} \left[ \frac{2}{b^2} (e^{by} - 1) - \frac{2}{b} ye^{by} + y^2 e^{by} \right] \\ \frac{\partial^2 g}{\partial b \partial \sigma^2} &= -\frac{a^2}{b^3} [bye^{by}(e^{by} - 1) - (e^{by} - 1)^2] \\ \frac{\partial^2 g}{\partial (\sigma^2)^2} &= \frac{2a^3}{3b^3} (e^{by} - 1)^3,\end{aligned}$$

by again applying the rule of L'Hôpital for the last equation.

### 3.8.2 Computational details on the calculation of the observed Fisher information matrix

The calculation of the observed Fisher information matrix  $\mathcal{J}(\hat{\theta}_n)$  in the gamma-Gompertz model is based on the negative second-order partial derivatives of the log-likelihood and the maximum likelihood estimate (MLE)  $\hat{\theta}_n$  of the parameter vector  $\theta = (a, b, \sigma^2)^\top$ .

The MLE can be determined by numerical optimization of the log-likelihood using function `nlm()` in R. Optimization over the log-scale of the parameters ensures non-negativity of the parameter estimates. The numerical stability of the estimation problem for values of  $\sigma^2$  close to zero can be improved by providing also the analytic gradient of the log-likelihood to the optimization routine as well as by using Taylor expansions of the log-likelihood and the gradient if the current value of  $\sigma^2$  is smaller than  $10^{-5}$ . In addition, a number of different starting values for the parameter  $\sigma^2$  should be considered.

Although we have derived explicit formulas for the partial derivatives of the log-density of the gamma-Gompertz model, it turns out that the expressions for the second-order partial derivatives with respect to  $\sigma^2$ , given in (3.7) and (3.8), are not numerically stable if  $\sigma^2$  approaches zero. Therefore, when calculating  $\mathcal{J}(\hat{\theta}_n)$ , we approximate the term  $\ln \left[ 1 + \sigma^2 \frac{a}{b} (e^{bx} - 1) \right]$  in expressions (3.7) and (3.8) by a Taylor expansion if  $\hat{\sigma}^2 < 10^{-5}$ .

### 3.8.3 Additional figures and tables for empirical studies

In this section, we present additional figures and tables displaying some results of our empirical studies in Section 3.5.

- Table 3.3 reports on the performance of the numerical integration approach for computing the Fisher information  $I(\theta)$ .
- The relation between the information measure  $\kappa^{-2}$  and the variance of  $\hat{\sigma}^2$ , as discussed in Section 3.3.3, is illustrated in Figure 3.6.

- The information measures corresponding to the criteria of  $D$ -,  $A$ -, and  $E$ -optimality (see Section 3.2.3) are examined in Figures 3.7 and 3.8 for Scenarios  $S_1$  and  $S_3$ , respectively.
- On pages 55ff., we study the various information measures and the performance of the likelihood ratio test for scenarios with different values for the Gompertz parameters. More precisely, for Scenarios  $S_4$  to  $S_6$ , we set  $a = 0.021$  and  $b = 0.082$ , while the values for the frailty variance are the same as in the previous scenarios, that is,  $\sigma^2 = 0.043$  in Scenario  $S_4$ ,  $\sigma^2 = 0.021$  in Scenario  $S_5$ , and  $\sigma^2 = 0$  in Scenario  $S_6$ . Figure 3.9 depicts the patterns of the criterion of  $D_A$ -optimality across different age ranges for Scenarios  $S_4$  and  $S_6$ , while in Figures 3.10 and 3.11 the criteria of  $D$ -,  $A$ -, and  $E$ -optimality are presented. Figure 3.12 displays the criterion  $[\mathbf{I}(\boldsymbol{\theta})]_{33}$  for Scenarios  $S_4$  and  $S_6$ . Finally, the power of the likelihood ratio test to detect a positive  $\sigma^2$  in Scenarios  $S_4$  ( $\sigma^2 = 0.043$ ) and  $S_5$  ( $\sigma^2 = 0.021$ ) based on different age ranges and sample sizes at a level of  $\alpha = 0.05$  was calculated based on formula (3.6). The results are presented in Table 3.4.

Table 3.3: Mean relative difference between the Fisher information  $\mathbf{I}_n(\boldsymbol{\theta})$  and the average  $\bar{\mathcal{J}}$  of observed Fisher information matrices across 1,000 replications of Scenarios  $S_1$ ,  $S_2$ , and  $S_3$  for different sample sizes and age ranges.

Scenario	$n_{90+}$	Survivors to ages			
		60+	80+	85+	90+
$S_1: \sigma^2 = 0.043$	10,000	0.00027	0.00344	0.01116	0.05487
	20,000	0.00019	0.00114	0.00512	0.02553
	105,000	0.00009	0.00025	0.00136	0.00579
$S_2: \sigma^2 = 0.021$	10,000	0.00053	0.00356	0.01382	0.06592
	20,000	0.00023	0.00134	0.00614	0.03601
	105,000	0.00009	0.00033	0.00119	0.00591
$S_3: \sigma^2 = 0$	10,000	0.00816	0.01540	0.03389	0.12065
	20,000	0.00611	0.01073	0.02114	0.07251
	105,000	0.00249	0.00504	0.00997	0.02829

**Relation between information measures and estimator precision**

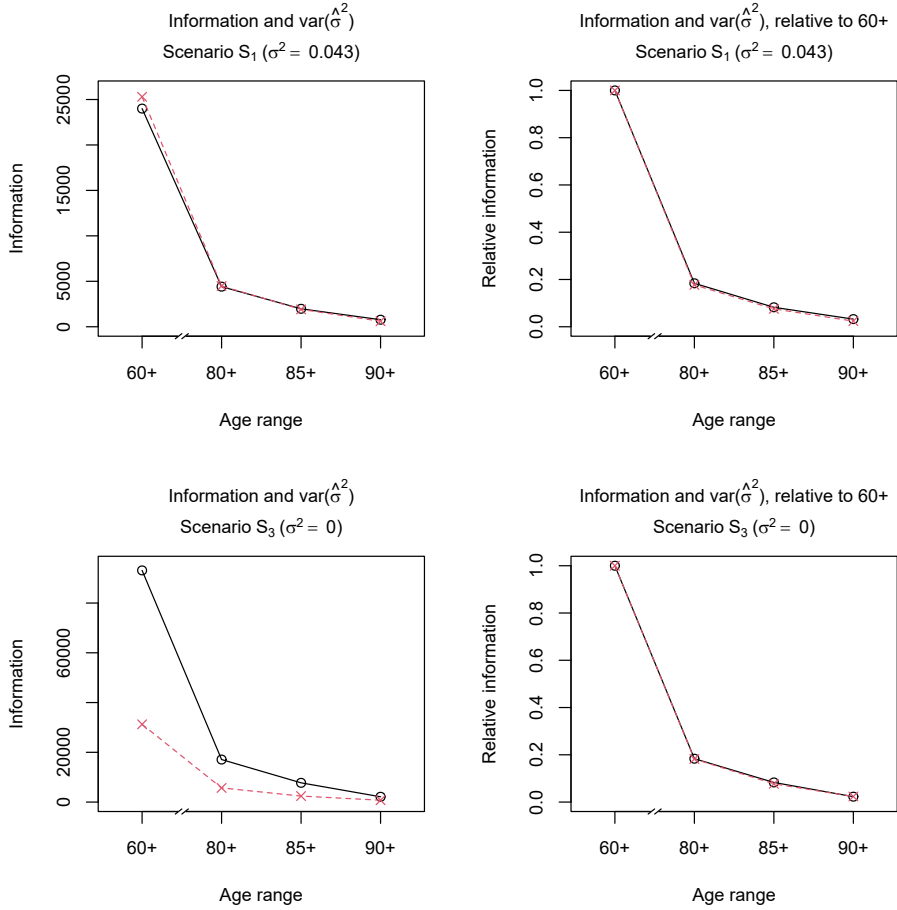


Figure 3.6: Information measure  $n\kappa^{-2}$  (red-dashed line, crosses) and inverse of the empirical variance of  $\hat{\sigma}^{-2}$  (black-solid line, circles) based on 1,000 samples from a gamma-Gompertz model under the medium-sized Scenarios  $S_1$  (top) and  $S_3$  (bottom) depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+). Left: absolute values, right: relative to the value for the 60+ setting.

## Alternative information measures

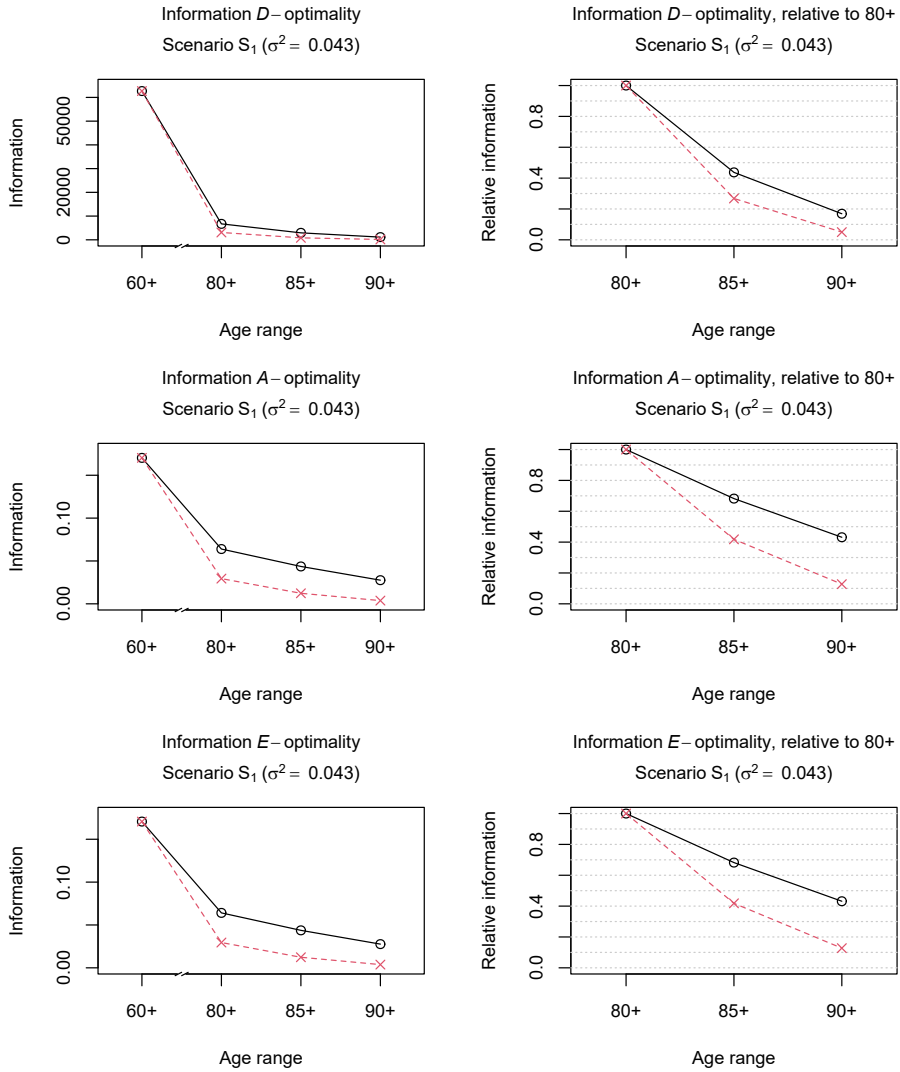


Figure 3.7: Information measures  $\mathcal{I}$  (black-solid line, circles) and scaled measures  $\mathcal{I}^{(s)}$  (red-dashed line, crosses) under Scenario  $S_1$  depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+). Top:  $\mathcal{I} = \det(\mathbf{I}(\theta))$  for  $D$ -optimality, middle:  $\mathcal{I} = 1/\text{tr}([\mathbf{I}(\theta)]^{-1})$  for  $A$ -optimality, bottom:  $\mathcal{I}$  as the minimum eigenvalue of  $\mathbf{I}(\theta)$  for  $E$ -optimality. Left: absolute values of (scaled)  $\mathcal{I}$ , right: (scaled) ratios  $\mathcal{I}_{x+}/\mathcal{I}_{80+}$  for  $x = 80, 85, 90$ .

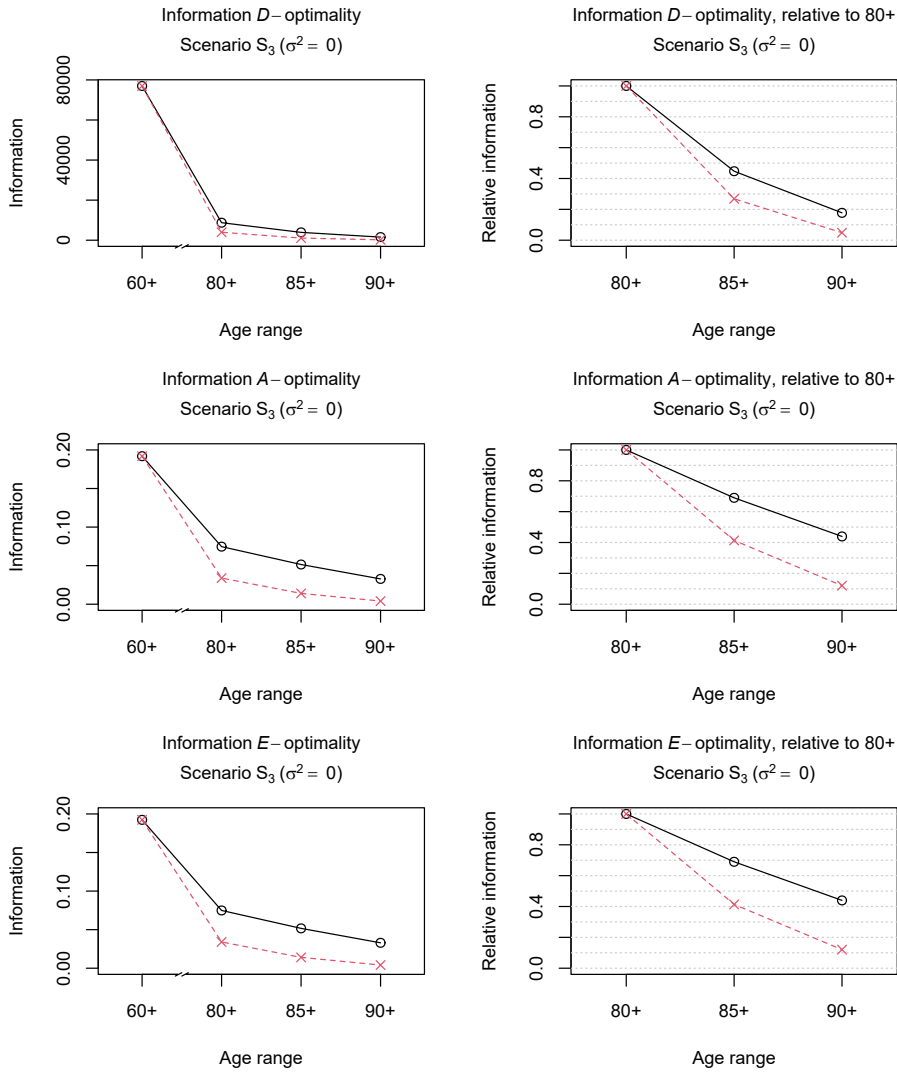


Figure 3.8: Information measures  $\mathcal{I}$  (black-solid line, circles) and scaled measures  $\mathcal{I}^{(s)}$  (red-dashed line, crosses) under Scenario  $S_3$  depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+). Top:  $\mathcal{I} = \det(\mathbf{I}(\theta))$  for  $D$ -optimality, middle:  $\mathcal{I} = 1/\text{tr}([\mathbf{I}(\theta)]^{-1})$  for  $A$ -optimality, bottom:  $\mathcal{I}$  as the minimum eigenvalue of  $\mathbf{I}(\theta)$  for  $E$ -optimality. Left: absolute values of (scaled)  $\mathcal{I}$ , right: (scaled) ratios  $\mathcal{I}_{x+}/\mathcal{I}_{80+}$  for  $x = 80, 85, 90$ .

### Information measures and power of the likelihood ratio test for scenarios with different values of the Gompertz parameters

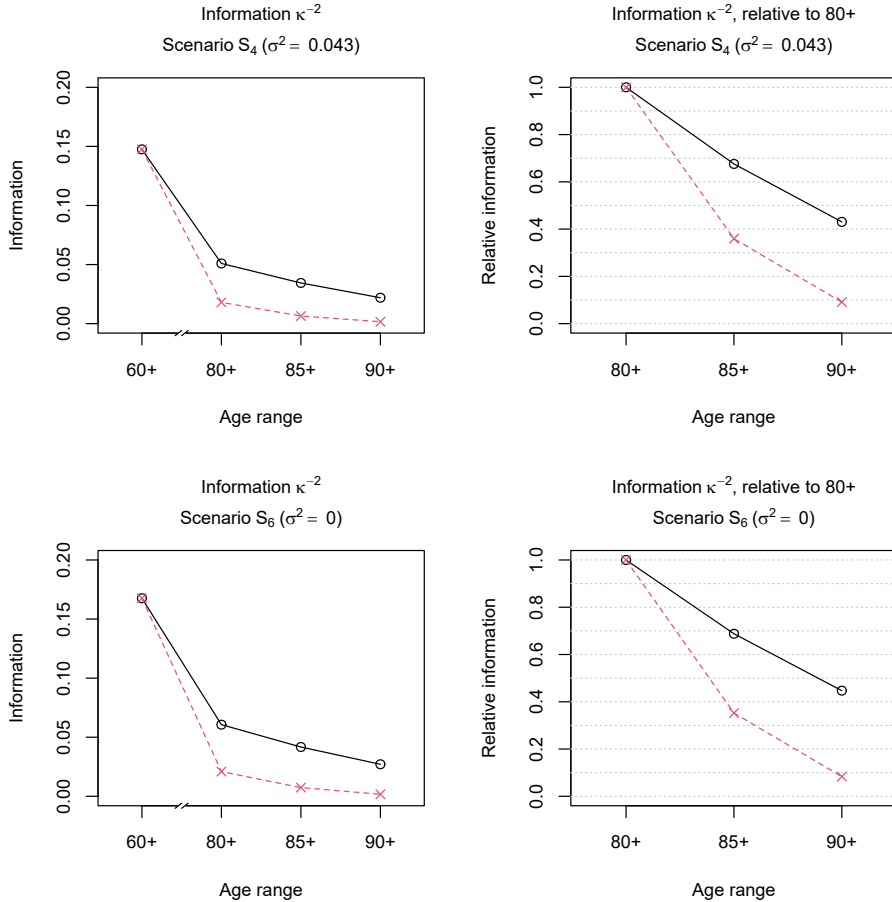


Figure 3.9: Information measure  $\mathcal{I} = \kappa^{-2}$  (black-solid line, circles) and scaled measure  $\mathcal{I}^{(s)}$  (red-dashed line, crosses) under Scenarios  $S_4$  (top) and  $S_6$  (bottom) depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+). Left: absolute values of (scaled)  $\mathcal{I}$ , right: (scaled) ratios  $\mathcal{I}_{x+}/\mathcal{I}_{80+}$  for  $x = 80, 85, 90$ .



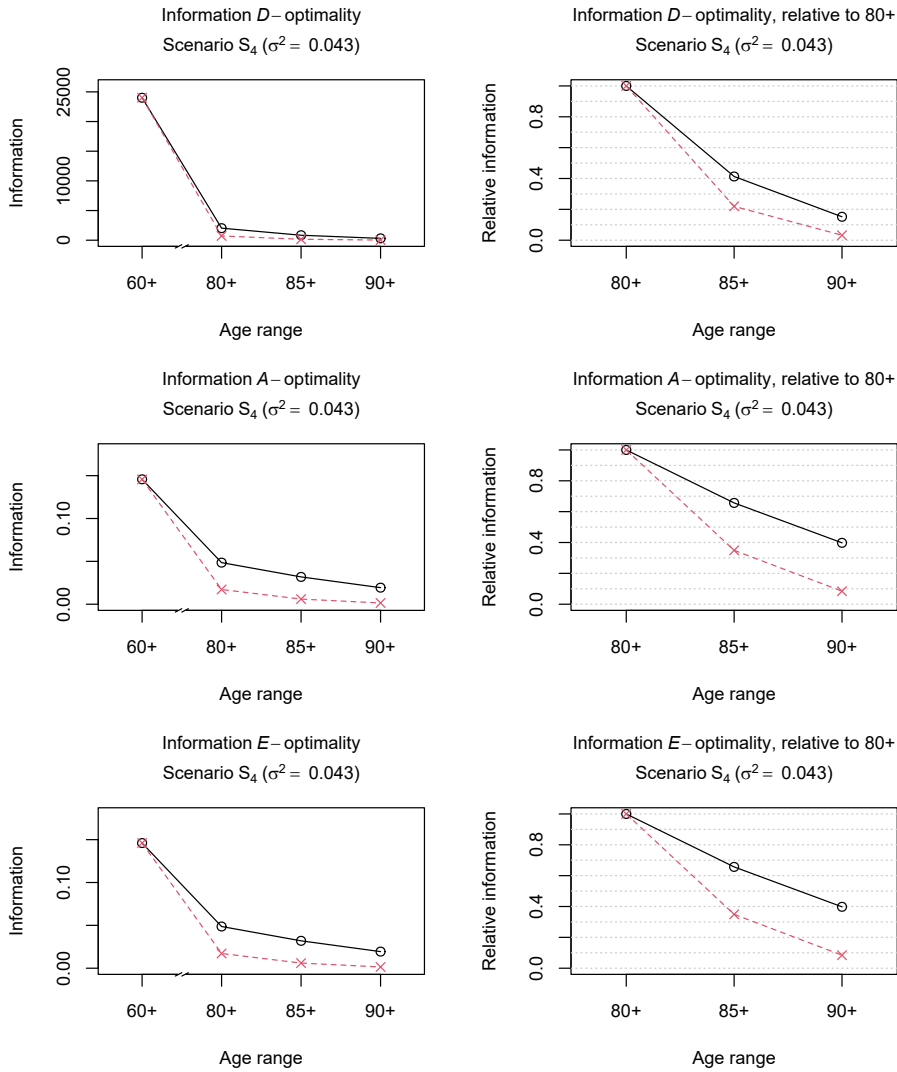


Figure 3.10: Information measures  $\mathcal{I}$  (black-solid line, circles) and scaled measures  $\mathcal{I}^{(s)}$  (red-dashed line, crosses) under Scenario  $S_4$  depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+). Top:  $\mathcal{I} = \det(\mathbf{I}(\theta))$  for D-optimality, middle:  $\mathcal{I} = 1/\text{tr}([\mathbf{I}(\theta)]^{-1})$  for A-optimality, bottom:  $\mathcal{I}$  as the minimum eigenvalue of  $\mathbf{I}(\theta)$  for E-optimality. Left: absolute values of (scaled)  $\mathcal{I}$ , right: (scaled) ratios  $\mathcal{I}_{x+}/\mathcal{I}_{80+}$  for  $x = 80, 85, 90$ .

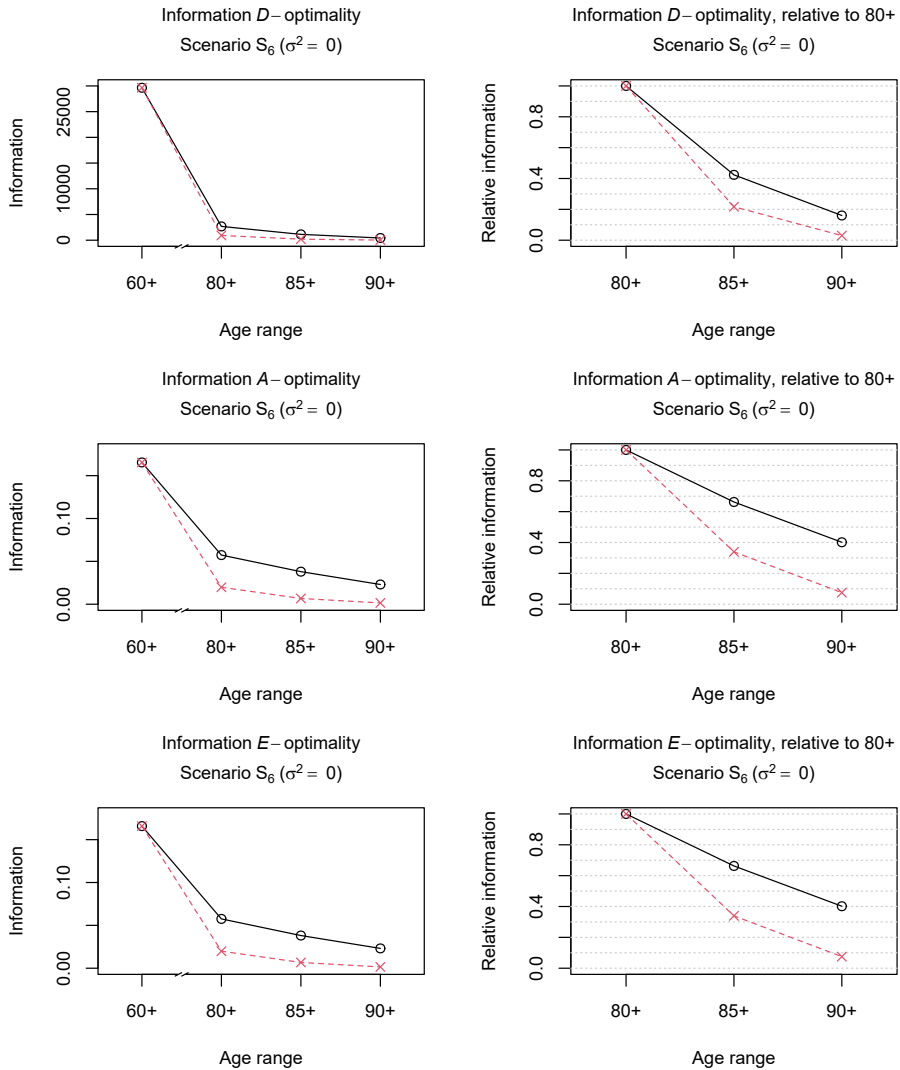


Figure 3.11: Information measures  $\mathcal{I}$  (black-solid line, circles) and scaled measures  $\mathcal{I}^{(s)}$  (red-dashed line, crosses) under Scenario  $S_6$  depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+). Top:  $\mathcal{I} = \det(\mathbf{I}(\theta))$  for  $D$ -optimality, middle:  $\mathcal{I} = 1/\text{tr}([\mathbf{I}(\theta)]^{-1})$  for  $A$ -optimality, bottom:  $\mathcal{I}$  as the minimum eigenvalue of  $\mathbf{I}(\theta)$  for  $E$ -optimality. Left: absolute values of (scaled)  $\mathcal{I}$ , right: (scaled) ratios  $\mathcal{I}_{x+}/\mathcal{I}_{80+}$  for  $x = 80, 85, 90$ .

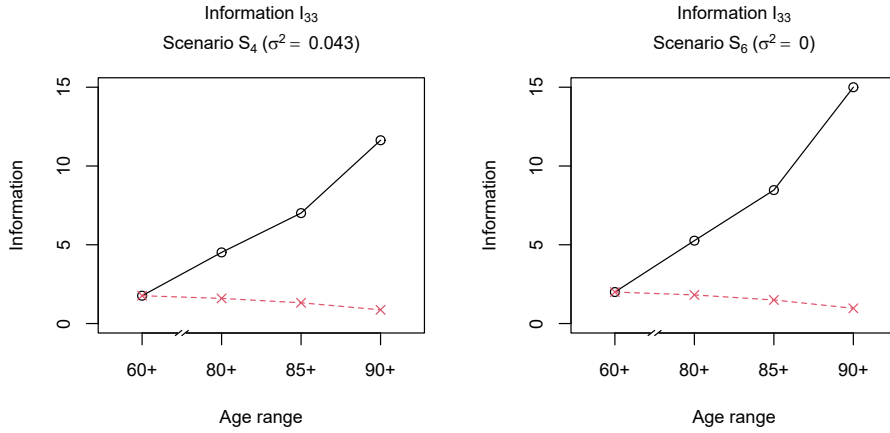


Figure 3.12: Information measure  $\mathcal{I} = [I(\theta)]_{33}$  (black-solid line, circles) and scaled measure  $\mathcal{I}^{(s)}$  (red-dashed line, crosses) depending on the age range of the data (left to right: 60+, 80+, 85+, or 90+) under Scenarios  $S_4$  (left) and  $S_6$  (right).

Table 3.4: Power  $\beta$  of the likelihood ratio test, performed at the 5% level, according to formula (3.6), under Scenarios  $S_4$  ( $\sigma^2 = 0.043$ ) and  $S_5$  ( $\sigma^2 = 0.021$ ) for three sample size settings (s – small, m – medium, l – large) and varying age range.

		Survivors to ages							
		60+		80+		85+		90+	
Scen.	$n$	$n_{60+}$	$\beta_{60+}$	$n_{80+}$	$\beta_{80+}$	$n_{85+}$	$\beta_{85+}$	$n_{90+}$	$\beta_{90+}$
$S_4$	s	133,506	1.000	47,165	0.678	25,090	0.352	10,000	0.157
	m	267,012	1.000	94,329	0.909	50,179	0.557	20,000	0.228
	l	1,401,813	1.000	495,229	1.000	263,441	0.993	105,000	0.662
$S_5$	s	143,746	0.935	50,181	0.296	26,196	0.163	10,000	0.094
	m	287,493	0.998	100,362	0.470	52,392	0.239	20,000	0.119
	l	1,509,337	1.000	526,901	0.975	275,058	0.692	105,000	0.281

## Acknowledgements

The authors thank Nadine Ouellette and Robert Bourbeau for creating and providing the data set on French-Canadian mortality.

## References

- Abbring, J. H. and G. J. van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika* 94(1), 87–99.
- Atkinson, A. C. (1988). Recent developments in the methods of optimum and related experimental designs. *International Statistical Review* 56(2), 99–115.
- Balan, T. A. and H. Putter (2020). A tutorial on frailty models. *Statistical Methods in Medical Research* 29, 3424–3454.
- Beard, R. E. (1959). Note on some mathematical mortality models. In G. E. W. Wolstenholme and M. O'Connor (Eds.), *The Lifespan of Animals*, Ciba Foundation Colloquia on Ageing, pp. 302–311. Boston: Little, Brown.
- Bebbington, M., R. Green, C.-D. Lai, and R. Zitikis (2014). Beyond the Gompertz law: exploring the late-life mortality deceleration phenomenon. *Scandinavian Actuarial Journal* 2014(3), 189–207.
- Becker, N., B. McDonald, and C. Khoo (1989). Optimal designs for fitting a proportional hazards regression model to data subject to censoring. *Australian Journal of Statistics* 31(3), 449–468.
- Berger, M. P. F. and W. K. Wong (2009). *An Introduction to Optimal Designs for Social and Biomedical Research*. Statistics in Practice. John Wiley & Sons.
- Böhnstedt, M. and J. Gampe (2019). Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model. *Statistics and Probability Letters* 150, 68–73.
- Burnham, K. P. and D. R. Anderson (2002). *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*. Springer.
- Duchateau, L. and P. Janssen (2008). *The Frailty Model*. Springer.
- Economou, P. and C. Caroni (2008). Closure properties and diagnostic plots for the frailty distribution in proportional hazards models. In F. Vonta, M. Nikulin, N. Limnios, and C. Huber-Carol (Eds.), *Statistical Models and Methods for Biomedical and Technical Systems*, Statistics for Industry and Technology, pp. 43–54. Birkhäuser Boston.
- Escobar, L. A. and W. Q. Meeker (1998). Fisher information matrices with censoring, truncation, and explanatory variables. *Statistica Sinica* 8, 221–237.
- Feehan, D. M. (2018). Separating the signal from the noise: Evidence for deceleration in old-age death rates. *Demography* 55, 2025–2044.
- Gavrilov, L. A. and N. S. Gavrilova (2001). The reliability theory of aging and longevity. *Journal of Theoretical Biology* 213, 527 – 545.

- Gavrilov, L. A. and N. S. Gavrilova (2019). New trend in old-age mortality: Gompertzialization of mortality trajectory. *Gerontology* 65, 451–457.
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* 115, 513–583.
- Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* 71, 75–83.
- Hwang, W.-T. and R. Brookmeyer (2003). Design of panel studies for disease progression with multiple stages. *Lifetime Data Analysis* 9, 261–274.
- Jeune, B. and J. W. Vaupel (Eds.) (1999). *Validation of exceptional longevity*. Odense Monographs on Population Aging 6. Odense University Press.
- Kirkwood, T. B. L. (2015). Deciphering death: a commentary on Gompertz (1825) ‘On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies’. *Philosophical Transactions of the Royal Society of London B* 370. doi: 10.1098/rstb.2014.0379.
- Konstantinou, M., S. Biedermann, and A. C. Kimber (2015). Optimal designs for full and partial likelihood information – With application to survival models. *Journal of Statistical Planning and Inference* 165, 27–37.
- Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer texts in statistics. New York: Springer.
- Lehmann, E. L. and G. Casella (1998). *Theory of Point Estimation* (2 ed.). Springer texts in statistics. New York: Springer.
- McGree, J. M. and J. A. Eccleston (2010). Investigating design for survival models. *Metrika* 72, 295–311.
- Ouellette, N. (2016). La forme de la courbe de mortalité des centenaires canadiens-français. *Gérontologie et société* 38(151), 41–53.
- Ouellette, N. and R. Bourbeau (2014). Measurement of mortality among centenarians in Canada. In *Living to 100 Monograph*. Society of Actuaries. 17 pages.
- Preston, S. H., I. T. Elo, and Q. Stewart (1999). Effects of age misreporting on mortality estimates at older ages. *Population Studies* 53(2), 165–177.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

- 
- Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 82(398), 605–610.
- Silvey, S. D. (1980). *Optimal Design*. Monographs on Applied Probability and Statistics. London: Chapman and Hall.
- Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society. Series A* 162(1), 5–43.
- Thatcher, A. R., V. Kannisto, and J. W. Vaupel (1998). *The force of mortality at ages 80 to 120*. Odense Monographs on Population Aging 5. Odense University Press.
- Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3), 439–454.



# 4

## Shifting attention to old age: Detecting mortality deceleration using focused model selection

### **Abstract**

The decrease in the increase in death rates at old ages is a phenomenon that has repeatedly been discussed in demographic research. While mortality deceleration can be explained as an effect of selection in heterogeneous populations, this phenomenon can be difficult to assess statistically because it relates to the tail of the lifespan distribution. By using a focused information criterion (FIC) for model selection, we can directly target model performance at those advanced ages at death. We analyze this question in the framework of the gamma-Gompertz model that is reduced to the competing Gompertz model without mortality deceleration if the variance parameter lies on the boundary of the parameter space. We develop a new version of the FIC for this non-standard condition. In a simulation study, the new FIC is shown to outperform other methods in detecting mortality deceleration. We apply the approach to mortality data for extinct French-Canadian birth cohorts, and we extend the method to include additional covariate information.

---

This chapter has been submitted for publication as: M. Böhnstedt, H. Putter, N. Ouellette, G. Claeskens, and J. Gampe. Shifting attention to old age: Detecting mortality deceleration using focused model selection.



## 4.1 Introduction

Almost two centuries ago, Benjamin Gompertz noted that death rates of humans increase exponentially with age from mid-life onwards (Gompertz, 1825). Makeham (1860) supplemented Gompertz's mortality model by a constant, and thus age-independent, component to provide a better fit at younger ages. By the 1930s, Perks (1932) had become aware that the Gompertz-Makeham hazard overestimated actual death rates at advanced ages, and suggested replacing the exponential part with a logistic function for graduation of mortality. The logistic hazard follows the Gompertz trajectory for the lower ages, but gradually deviates – that is, increases less quickly – at advanced ages. This slowdown in mortality rates late in life is now commonly known as mortality deceleration.

The study of death rates at high ages gained increased attention when, starting in the 1970s, progress against mortality became noticeable also among the elderly. This continuing improvement of death rates at old ages (Rau et al., 2008) is the primary reason for longevity increases in high income countries. Thus, accurately describing the trajectory of mortality at advanced ages is of great interest to actuaries, demographers, and aging researchers, as it has important implications for pension funds, life insurance, social support systems, and public health planning. Understanding and projecting the numbers of the oldest-old are essential for aging populations, and hinge on having proper estimates of mortality late in life.

As vital statistics improved and more detailed information became available for individuals who survive to very old ages, the early findings of mortality deceleration were replicated for more recent data (Horiuchi and Wilmoth, 1998; Thatcher et al., 1998; Richards, 2008; Feehan, 2018). However, the phenomenon of mortality deceleration has also been contested, with some scholars arguing that there is a continued exponential increase in mortality with age (Gavrilov and Gavrilova, 2011, 2019), and that the apparent slowdown in death rates with advancing age is primarily attributable to data errors (Newman, 2018). It is indeed the case that exaggeration in the reporting of age and the failure to remove deceased individuals from registers (due to unreported deaths) can result in an overestimation of the number of long-lived individuals, which will bias death rates downwards at the most advanced ages (Preston et al., 1999). Thus, for individuals who die at very old ages, a thorough scientific validation of the reported age at death is mandatory (Jeune and Vaupel, 1999).

Such age validation procedures involve linking individual birth (or baptism) records with death certificates, which is usually a tedious and time-consuming task. Thus, the amount of available data is often limited. Furthermore, mortality deceleration occurs in the tail of the survival distribution, where data are unavoidably scarce. For these reasons, the statistical assessment of this phenomenon is challenging, and standard methods may fail to identify deviations from the Gompertz hazard for the very old. The questions of whether and, if so, how statistical inference can be improved to tackle these challenges are addressed in this chapter.

To illustrate how rapidly the number of observations declines with age, Figure 4.1 shows for French-Canadians born between 1880 and 1896 the empirical death rates by age at ages 90 and above, as well as the number of deaths at each age. We can see that 75% of all deaths in this population had already occurred by age 96 for women and by age 95 for men. While we have sizable samples for this population (about 20,000 women and about 10,000 men), the information about (potential) mortality deceleration has to be extracted from relatively scant data in the tail.

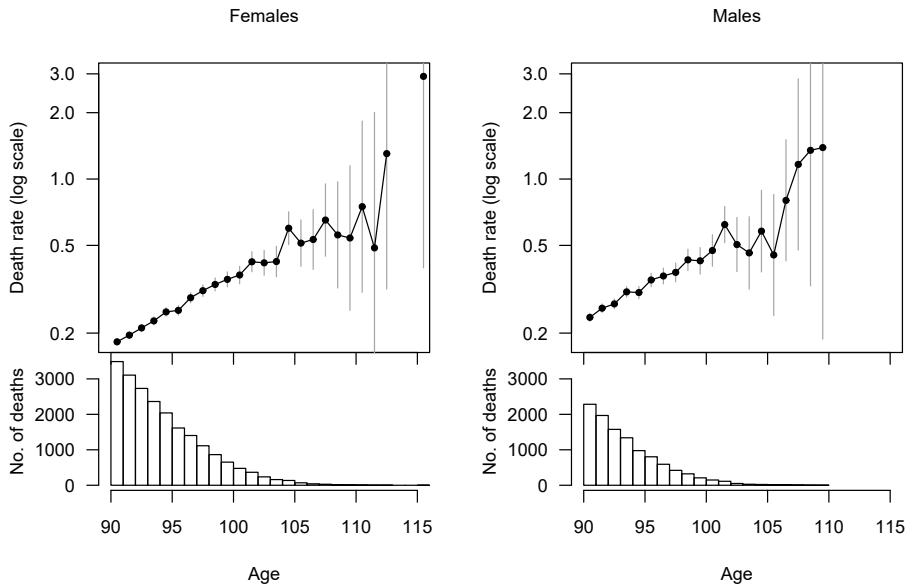


Figure 4.1: Top: Death rates (on log scale) with 95% confidence intervals for French-Canadian females (left) and males (right). Bottom: Frequency distribution of ages at death.

Much of present-day survival analysis is nonparametric, and thus avoids parametric distributional assumptions about the times to event. For the specific questions addressed here, a parametric framework proves both adequate and instrumental. The exponential increase of mortality over much of the adult lifespan has been established empirically for numerous populations, in different periods, and under varied social circumstances. Research results that dismiss mortality deceleration (Gavrilov and Gavrilova, 2011, 2019) support the use of a Gompertz hazard also at advanced ages. The possibility of treating this problem in a parsimonious parametric framework is attractive given the limited amount of data from which information on deceleration can be drawn.

In this chapter, we will discuss the statistical assessment of mortality deceleration in the framework of the gamma-Gompertz model, which is a particular parametrization of

the logistic hazard introduced by Perks (1932). This model belongs to the class of proportional hazards frailty models, which represent the standard approach for formalizing individually heterogeneous hazards of death (Vaupel et al., 1979; Wienke, 2011). As was already noted by Beard (1959), in such a model, frail individuals with higher mortality levels tend to die at younger ages, while more robust individuals with lower death risks tend to survive to higher ages, which leads to a deceleration of the average hazard with age.

In the gamma-Gompertz model, an exponentially increasing (Gompertz) baseline hazard is multiplied by a gamma distributed random effect (the frailty). If the variance  $\sigma^2$  of the gamma frailty takes a positive value, then the population hazard shows a downward deviation from the Gompertz hazard at advanced ages. If the variance parameter takes the value of zero, the population hazard is exponentially increasing. Thus, answering the question of whether mortality does or does not decelerate at advanced ages corresponds to selecting the gamma-Gompertz model or the Gompertz model. However, the single additional parameter  $\sigma^2$  of the gamma-Gompertz model lies on the boundary of the parameter space if the true model is the Gompertz model. As this boundary constraint on the parameter violates the usual regularity assumptions, the inference and the model selection have to be adapted to this non-standard condition.

Traditional approaches to inference in this context either employ a likelihood ratio test for  $\sigma^2 = 0$ , which has low power to detect actual deceleration, or model selection via Akaike's information criterion (AIC, Akaike, 1974). While information criteria, like the AIC, select a 'best' model regardless of the specific estimand that is of interest, we propose using a focused information criterion (FIC, Claeskens and Hjort, 2003) that selects the model that performs 'best' for a specific parameter of interest, called the focus parameter. Applying the FIC is particularly appealing in our context, as it will allow us to choose a focus parameter that is directly affected by the presence or the absence of mortality deceleration; for example, the hazard at some advanced age.

Technically, the FIC is constructed as an unbiased estimator of the limiting risk of an estimator of the focus parameter, and the candidate model with the smallest FIC value is selected. While the standard version of the FIC aims to minimize the mean squared error (MSE) of the estimator of the focus parameter, the criterion has been generalized to other risk measures, such as  $L_p$ -risks (Claeskens et al., 2006). Still, all of these model selection criteria have been developed based on general likelihood theory under the standard regularity assumptions, which are violated in our setting. Therefore, we will derive versions of the FIC that allow us to choose between two models in which the additional parameter may lie on the boundary of the parameter space.

The chapter is structured as follows. In Section 4.2, we summarize models for late-life mortality, and discuss the specific parametrization of the gamma-Gompertz model. We also present traditional methods for detecting mortality deceleration in this framework. Then, in Section 4.3, we propose the FIC as a new approach for assessing this phenomenon. We introduce the method for a single sample, and, in Section 4.4, we study the performance of the FIC in a simulation study, and compare it with the performance

of an AIC that is adjusted to the presence of the boundary constraint. In Section 4.5, we apply the new model selection criteria to the French-Canadian mortality data presented in Figure 4.1. We describe the data source and the age validation procedure, and apply the method to the samples of females and males separately. Additionally, we incorporate effects for different birth cohorts. We conclude with a discussion in Section 4.6.

## 4.2 Mortality at advanced ages

### 4.2.1 Hazard models

To model mortality at advanced ages, we consider the continuous random variable  $Y$ , which describes adult lifespans from mid-life onwards. Its distribution can be characterized by the hazard function, the instantaneous death rate at age  $y$  given survival up to  $y$ ,

$$h(y) = \lim_{\Delta y \searrow 0} \frac{P(y < Y \leq y + \Delta y \mid Y > y)}{\Delta y}.$$

The value  $y = 0$  corresponds to the age from which we start modeling, typically age 50 or 60. The Gompertz distribution is characterized by an exponentially increasing trajectory  $h(y) = ae^{by}$ , with two positive parameters  $a$  and  $b$ . Makeham (1860) extended the hazard by an age-independent constant  $c \geq 0$  to  $h(y) = ae^{by} + c$  to achieve a better fit at younger ages.

The replacement of the exponential component by a logistic term was suggested by Perks (1932) in order to allow for a slower-than-exponential increase in death rates at the highest ages, while keeping the proven Gompertz shape before that point. In its general form, the logistic hazard is

$$h(y) = \frac{Ae^{by}}{1 + Be^{by}} \quad \text{with } A > 0, B \geq 0. \quad (4.1)$$

If  $B = 0$ , then the hazard (4.1) reduces to the Gompertz model with  $A = a$ . The logistic hazard can be viewed as a simple device to capture potential mortality deceleration by one additional parameter. However, Beard (1959) already noted that a hazard of the form (4.1) arises when individuals are submitted to Gompertz hazards with individually varying parameters  $a$ . If the distribution of  $a$  follows a gamma distribution, then an average hazard of logistic shape results.

In contemporary statistical terminology, Beard's finding would be called a proportional hazards frailty model of the form  $h(y|Z = z) = z \cdot h_0(y)$  (Duchateau and Janssen, 2008; Wienke, 2011). Here, a positive random effect  $Z$  (called frailty) acts multiplicatively on a common baseline hazard  $h_0(y)$ , such that  $h(y|Z = z)$  denotes the conditional hazard of an individual at age  $y$ , given that his or her frailty is  $Z = z$ . The random effect  $Z$  assigns heterogeneous mortality risks to individuals in a cohort who, apart from those risks, share a common 'law' of mortality (the baseline). As a result of selection, frail individuals with higher mortality levels tend to die at younger ages, while the more

robust individuals with lower death risks tend to survive to higher ages. Consequently, the population hazard, averaged over the survivors at each age, deviates from the baseline shape; and the larger the initial heterogeneity, the stronger the deviation is.

The frailty  $Z$  is often assumed to follow a gamma distribution with mean one and variance  $\sigma^2$  (Vaupel et al., 1979). The choice of the gamma distribution, which may be deemed ad hoc, is not only mathematically convenient, but is also theoretically justified. Abbring and van den Berg (2007) proved that the distribution of the heterogeneity among survivors, once selection took effect, converges to a gamma distribution for a large class of proportional hazards frailty models, even if the frailty is not gamma distributed from the outset.

The so-called gamma-Gompertz model is obtained if the gamma frailty is multiplied to an exponentially increasing Gompertz baseline hazard,  $h_0(y) = ae^{by}$ . The resulting marginal hazard is

$$h(y) = \frac{ae^{by}}{1 + \sigma^2 \frac{a}{b} (e^{by} - 1)}. \quad (4.2)$$

If  $\sigma^2 > 0$ , there is heterogeneity in the risk of death, and the selection of more robust individuals will take place. If  $\sigma^2 = 0$ , there is no heterogeneity, and the marginal hazard is exponentially increasing, such that  $h(y) = ae^{by}$ . Hence, in the framework of the gamma-Gompertz model, the statistical assessment of mortality deceleration is reduced to inference on the parameter  $\sigma^2$ .

The gamma-Gompertz hazard (4.2) is a straightforward reparametrization of the logistic hazard (4.1), with  $A = a/(1 - \sigma^2 \frac{a}{b})$  and  $B = \frac{\sigma^2 a}{b}/(1 - \sigma^2 \frac{a}{b})$ ; and in both versions the Gompertz hazard is retained if the third parameter equals zero, that is,  $\sigma^2 = 0$  or  $B = 0$ , respectively. The gamma-Gompertz formulation explicitly signalizes the role of individual heterogeneity and its impact on and the progression of selection among survivors.

It is important to note that the parameter  $\sigma^2$  measures population heterogeneity at the starting age of the model (corresponding to  $y = 0$ ). Due to the continuing selection of robust individuals, the variance of frailty among the survivors decreases with age. Thus, the higher the age at which we start our observation, the lower the heterogeneity in mortality is among the individuals in the sample. The age at the beginning of the observation will, therefore, have an impact on the resulting inference.

The inference in the gamma-Gompertz model involves the frailty variance  $\sigma^2$ , which is a parameter that lies on the boundary of its parameter space in the absence of mortality deceleration ( $\sigma^2 = 0$ ). This violates the standard assumptions that underlie the asymptotic properties of the likelihood-based inference, which, in turn, affects the traditional approaches for assessing mortality deceleration that are presented in the following section.

## 4.2.2 Traditional approaches to inference

Two methods are commonly used for assessing mortality deceleration in the framework of the gamma-Gompertz model: a likelihood ratio test for a zero frailty variance, and

model selection between the gamma-Gompertz model and the Gompertz model based on the AIC.

The likelihood ratio test for homogeneity in the gamma-Gompertz model, where  $H_0 : \sigma^2 = 0$  and  $H_1 : \sigma^2 > 0$ , is non-standard in that, under the null hypothesis, the parameter  $\sigma^2$  lies on the boundary of the parameter space. Consequently, the asymptotic distribution of the likelihood ratio test statistic under  $H_0$  is no longer a chi-squared distribution with one degree of freedom. However, using the results of Self and Liang (1987), it can be shown that under the null hypothesis, the likelihood ratio test statistic asymptotically follows a 50:50 mixture of a point mass at zero and a chi-squared distribution with one degree of freedom,  $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ . Tests based on the wrong assumption of a  $\chi_1^2$ -distribution of the test statistic occasionally appear in studies of mortality deceleration (Pletcher, 1999). Ignoring the issue of the boundary parameter and using the incorrect distribution of the test statistic lowers the power to (correctly) decide in favor of the gamma-Gompertz model. But even when the test statistic is correctly assumed to follow a  $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ -distribution, the likelihood ratio test has low power to detect mortality deceleration in the gamma-Gompertz model. This is especially likely to be the case when the inference has to be based on age-restricted samples, such as a sample of individuals who survived beyond age 90 (see Section 4.7.2 for an illustration).

A popular alternative approach for assessing mortality deceleration is model selection based on the AIC (Richards, 2008; Gavrilova and Gavrilov, 2015). The AIC targets an unbiased estimate of the Akaike information; that is, of the expected relative Kullback-Leibler distance between the true data-generating mechanism and the best parametric approximation. Under standard conditions, the AIC is therefore defined as  $-2\ell + 2k$ , where the log-likelihood  $\ell$ , evaluated at the maximum likelihood estimate, is penalized by the number  $k$  of parameters in the model. This common definition has, however, been found to be biased under the non-standard conditions of the gamma-Gompertz model (Böhnstedt and Gampe, 2019). Thus, the standard version of the AIC is not a valid tool for model selection in the setting of the gamma-Gompertz model. In Section 4.3.4, we will present a modified version of the AIC that is adjusted to the presence of a boundary parameter.

### 4.3 Focused information criterion for mortality deceleration

The preceding considerations indicate that neither a testing strategy, particularly if it is low-powered, nor an all-purpose model selection criterion will adequately assess the occurrence of mortality deceleration. Focused information criteria (FIC) have been introduced to address problems of this kind, and we propose selecting the model based on a new version of the FIC that takes the boundary constraint on the frailty variance into account.

### 4.3.1 Rationale for FIC

Statistical analyses are performed for particular purposes, and acknowledging the specific purpose when choosing the statistical model is the key concept of a FIC. In the following exposition, we use the terminology and notation of Claeskens and Hjort (2003).

Observations  $y_i, i = 1, \dots, n$  (here: ages at death) are assumed to be generated by a parametric density  $f(y)$ . The parameters of the model are split into a  $d$ -vector  $\theta$ , which characterizes the narrow model, and an additional  $q$ -vector  $\gamma$  for the extended model. The narrow model is obtained for one particular value  $\gamma_0$ , which is fixed and known. In the current application, the density of the gamma-Gompertz model (4.2) is

$$f(y) = ae^{by} \left\{ 1 + \sigma^2 \frac{a}{b} (e^{by} - 1) \right\}^{-\left(1 + \frac{1}{\sigma^2}\right)}.$$

The parameter  $\theta = (a, b)^\top$  is the Gompertz part of the model, so  $d = 2$ . The single additional parameter is  $\gamma = \sigma^2$  with  $\gamma_0 = 0$ , so  $q = 1$ .

The original FIC is derived in a framework of local misspecification (Hjort and Claeskens, 2003), where a sample of size  $n$  is assumed to be generated from a density

$$f_{\text{true}}(y) = f(y, \theta_0, \gamma_0 + \delta/\sqrt{n}), \quad (4.3)$$

with the parameter vector  $\gamma = \gamma_0 + \delta/\sqrt{n}$  perturbed in the direction of  $\delta$ . Selection is between the null model, where  $\gamma$  is fixed at the known value  $\gamma_0$ ; the full model, including both  $\theta$  and  $\gamma$ ; and, if  $q > 1$ , any model including  $\theta$ , but only a subset of the components of  $\gamma$  and the remaining fixed at the respective values in  $\gamma_0$ . For the current setting, selection is only between the null model with  $\sigma^2 = 0$ , that is, the Gompertz model; and the full model including  $\sigma^2$ , that is, the gamma-Gompertz model. Due to the boundary constraint on the frailty variance,  $\delta = \sqrt{n}\sigma^2$  is subject to the a priori restriction  $\delta \geq 0$ . Therefore, we will restrict the framework in the following to the choice of including or not including a single parameter with a boundary constraint; that is,  $q = 1$  and  $\gamma \geq \gamma_0$ .

The focus is the parameter of interest, which depends on the underlying density (4.3) via  $\theta$  and  $\gamma$ . The focus is commonly denoted by  $\mu$ , and we define  $\mu_{\text{true}} = \mu(\theta, \gamma_0 + \delta/\sqrt{n})$ . Based on the maximum likelihood estimators  $\hat{\theta}_{\text{null}}$  in the null model and  $(\hat{\theta}_{\text{full}}, \hat{\gamma})$  in the full model, the focus parameter is estimated as  $\hat{\mu}_{\text{null}} = \mu(\hat{\theta}_{\text{null}}, \gamma_0)$  or  $\hat{\mu}_{\text{full}} = \mu(\hat{\theta}_{\text{full}}, \hat{\gamma})$ . For each model  $M, M \in \{\text{null}, \text{full}\}$ , the estimator  $\hat{\mu}_M$  converges in distribution,  $\sqrt{n}(\hat{\mu}_M - \mu_{\text{true}}) \xrightarrow{d} \Lambda_M$ .

The FIC selects the model that performs ‘best’ for the focus parameter  $\mu$ . If it is based on the general  $L_p$ -loss, the FIC aims to estimate without bias the limiting  $L_p$ -risk of  $\hat{\mu}_M$ ; that is,  $r_p(M) = \mathbb{E}[|\Lambda_M|^p]$ . The model for which this limiting risk is smaller is selected by the criterion. Of particular interest is a FIC based on the MSE ( $p = 2$ , as for the original version, Claeskens and Hjort, 2003), constructed as an estimator of  $\mathbb{E}[\Lambda_M^2]$ ; and a FIC based on the mean absolute error (MAE,  $p = 1$ ), constructed as an estimator of  $\mathbb{E}[|\Lambda_M|]$ .

### 4.3.2 FIC with a parameter on the boundary of the parameter space

Under standard regularity conditions, when general likelihood theory applies, the asymptotic normality of the maximum likelihood estimator implies that the  $\Lambda_M$  are normally distributed (Claeskens and Hjort, 2003). In the non-standard setting considered here,  $\Lambda_{\text{full}}$  is not normally distributed because the maximum likelihood estimator  $(\hat{\theta}_{\text{full}}, \hat{\gamma})$  converges in distribution to a mixture with two components (Böhnhstedt and Gampe, 2019). The limiting distribution depends on the information matrix  $J_{\text{full}}$  of the full model evaluated at the null model  $(\theta_0, \gamma_0)$ . We denote by  $J_{00}, J_{01}, J_{10}$ , and  $J_{11}$ , the four blocks of  $J_{\text{full}}$  corresponding to the components  $\theta$  and  $\gamma$  of the parameter vector; and by  $\kappa^2$  the element of the inverse information matrix  $J_{\text{full}}^{-1}$ , which corresponds to  $\gamma$ . Then, the following convergence in distribution holds for the estimator of the frailty variance

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \xrightarrow{d} \max(0, D) \quad \text{with } D \sim \mathcal{N}(\delta, \kappa^2).$$

For the limiting distribution of the estimator of the focus parameter, it can be shown that

$$\begin{aligned} \sqrt{n}(\hat{\mu}_{\text{null}} - \mu_{\text{true}}) &\xrightarrow{d} \Lambda_{\text{null}} = \Lambda_0 + \omega\delta \quad \text{and} \\ \sqrt{n}(\hat{\mu}_{\text{full}} - \mu_{\text{true}}) &\xrightarrow{d} \Lambda_{\text{full}} = \begin{cases} \Lambda_0 + \omega(\delta - D) & \text{if } D > 0 \\ \Lambda_0 + \omega\delta & \text{if } D \leq 0 \end{cases}, \end{aligned} \quad (4.4)$$

where  $\Lambda_0 \sim \mathcal{N}(0, \tau_0^2)$  is independent of  $D$ ,  $\tau_0^2 = \left(\frac{\partial \mu}{\partial \theta}\right)^\top J_{00}^{-1} \frac{\partial \mu}{\partial \theta}$  and  $\omega = J_{10} J_{00}^{-1} \frac{\partial \mu}{\partial \theta} - \frac{\partial \mu}{\partial \gamma}$  (cf. Section 10.2 in Claeskens and Hjort, 2008).

To define a FIC, we need to derive  $\mathbb{E}[|\Lambda|]$  or  $\mathbb{E}[\Lambda^2]$  from (4.4), depending on whether we intend to base the criterion on the limiting  $L_1$ - or  $L_2$ -risk of the estimator  $\hat{\mu}$ .

As in the original version of the FIC, the limiting MSE of  $\hat{\mu}$  is considered first. However, as we will demonstrate in the following, the FIC based on the  $L_2$ -risk has some drawbacks in the current setting, which makes the  $L_1$ -risk an attractive alternative.

From equation (4.4) we can determine  $\mathbb{E}[\Lambda^2]$  for the null and the full model:

$$\begin{aligned} \mathbb{E}[\Lambda_{\text{null}}^2] &= \tau_0^2 + \omega^2 \delta^2 \quad \text{and} \\ \mathbb{E}[\Lambda_{\text{full}}^2] &= \tau_0^2 + \omega^2 \left\{ \delta^2 \Phi\left(-\frac{\delta}{\kappa}\right) - \kappa \delta \phi\left(\frac{\delta}{\kappa}\right) + \kappa^2 \Phi\left(\frac{\delta}{\kappa}\right) \right\}, \end{aligned} \quad (4.5)$$

where  $\Phi(\cdot)$  and  $\phi(\cdot)$  denote the cdf and the pdf of the standard normal distribution, respectively. The  $\text{FIC}_{\text{MSE}}$  would be constructed as an unbiased estimator of the MSEs in (4.5), and the model with the smaller FIC value would be selected.

As has already been pointed out by Claeskens and Hjort (2008, Section 5.3), in the case of a single additional parameter  $\gamma$ , the so-called tolerance radius does not depend on the focus  $\mu$ . This radius signifies the deviation  $\delta$  for which the MSE of the null model estimator is smaller than that of the full model estimator; that is,  $\mathbb{E}[\Lambda_{\text{null}}^2] \leq \mathbb{E}[\Lambda_{\text{full}}^2]$ . From (4.5), we see that the two risks are the same for  $\omega = 0$ , and that if  $\omega \neq 0$  the



tolerance radius encompasses all  $\delta$  with  $\delta < 0.8399\kappa$ . We can still define a pre-test strategy for assessing mortality deceleration, which is based on the quantity  $\hat{\delta}/\hat{\kappa}$ , where  $\hat{\delta} = \sqrt{n}(\hat{\gamma} - \gamma_0) = \sqrt{n}\hat{\sigma}^2$  and  $\hat{\kappa}$  is derived from the observed Fisher information. If  $\hat{\delta}/\hat{\kappa} \leq 0.8399$ , the estimator  $\hat{\mu}_{\text{null}}$  based on the Gompertz model is used; whereas if  $\hat{\delta}/\hat{\kappa} > 0.8399$ , the estimator  $\hat{\mu}_{\text{full}}$  based on the gamma-Gompertz model is used. We note here that  $\hat{\delta}$  is not an unbiased estimator of  $\delta$ , with the bias depending in a complex way on  $\delta$  and  $\kappa$ . In appraising this pre-test-based model choice, we can see that for large samples, the local power of this strategy is approximately the same as the power of a likelihood ratio test for  $H_0: \sigma^2 = 0$  at the 20% level (cf. Section 4.7.3).

Although strategies based on the limiting  $L_2$ -risks of the estimator  $\hat{\mu}$  are common, the derived pre-test strategy has drawbacks. On the one hand, the performance of this strategy does not depend on the chosen focus parameter; while on the other, the equal penalty for squared bias and variance of the estimators in the  $L_2$ -risk might not be suitable for choosing whether to include a heterogeneity parameter.

Consequently, using risk measures other than the  $L_2$ -risk can be more appropriate, as was already suggested in Claeskens et al. (2006). Formulas for the general limiting  $L_p$ -risk of  $\hat{\mu}_M$  were derived there under regularity conditions where  $\Lambda_M$  follows a normal distribution for each of the models. In our non-standard setting, the limiting distribution of the full model estimator in (4.4) is not normal, but we can still derive the limiting  $L_1$ -risk of the estimators  $\hat{\mu}_{\text{null}}$  and  $\hat{\mu}_{\text{full}}$  as follows (see Section 4.7.4 for details):

$$\begin{aligned} \mathbb{E}[|\Lambda_{\text{null}}|] &= 2\tau_0\phi\left(\frac{\omega\delta}{\tau_0}\right) + 2\omega\delta\left\{\Phi\left(\frac{\omega\delta}{\tau_0}\right) - \frac{1}{2}\right\} \quad \text{and} \\ \mathbb{E}[|\Lambda_{\text{full}}|] &= \left[2\tau_0\phi\left(\frac{\omega\delta}{\tau_0}\right) + 2\omega\delta\left\{\Phi\left(\frac{\omega\delta}{\tau_0}\right) - \frac{1}{2}\right\}\right]\left\{1 - \Phi\left(\frac{\delta}{\kappa}\right)\right\} \\ &\quad + \sqrt{\tau_0^2 + \omega^2\kappa^2} \cdot \sqrt{\frac{2}{\pi}}\Phi\left(\frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2\kappa^2}}{\tau_0}\right) - \omega\kappa\phi\left(\frac{\delta}{\kappa}\right) \cdot 2\left\{\Phi\left(\frac{\omega\delta}{\tau_0}\right) - \frac{1}{2}\right\}. \end{aligned} \quad (4.6)$$

Thus, we define the  $\text{FIC}_{\text{MAE}}$  of the null model and the full model as the estimators

$$\begin{aligned} \text{FIC}_{\text{MAE}}(\text{null}) &= 2\hat{\tau}_0\phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) + 2\hat{\omega}\hat{\delta}\left\{\Phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) - \frac{1}{2}\right\} \quad \text{and} \\ \text{FIC}_{\text{MAE}}(\text{full}) &= \left[2\hat{\tau}_0\phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) + 2\hat{\omega}\hat{\delta}\left\{\Phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) - \frac{1}{2}\right\}\right]\left\{1 - \Phi\left(\frac{\hat{\delta}}{\hat{\kappa}}\right)\right\} \\ &\quad + \sqrt{\hat{\tau}_0^2 + \hat{\omega}^2\hat{\kappa}^2} \cdot \sqrt{\frac{2}{\pi}}\Phi\left(\frac{\hat{\delta}}{\hat{\kappa}} \cdot \frac{\sqrt{\hat{\tau}_0^2 + \hat{\omega}^2\hat{\kappa}^2}}{\hat{\tau}_0}\right) - \hat{\omega}\hat{\kappa}\phi\left(\frac{\hat{\delta}}{\hat{\kappa}}\right) \cdot 2\left\{\Phi\left(\frac{\hat{\omega}\hat{\delta}}{\hat{\tau}_0}\right) - \frac{1}{2}\right\}, \end{aligned}$$

respectively. Based on this new model selection criterion  $\text{FIC}_{\text{MAE}}$ , the full model is chosen if the estimated MAE of its estimator of the focus parameter  $\mu$  is smaller than the MAE for the null model estimator. In contrast to the MSE, the tolerance radius determined by the MAE of  $\hat{\mu}_M$  does depend on the focus parameter via  $\omega$  and  $\tau_0$ .

### 4.3.3 Choice of the focus parameter

The central concept and virtue of the FIC approach is that it allows us to consolidate a scientific question in a focus parameter, and to customize the model selection to the specific focus. In the context of mortality deceleration, two focus parameters suggest themselves. The first parameter is the frailty variance, since it determines whether mortality deceleration is present, so  $\mu = \sigma^2$ . The second focus parameter targets the deceleration of the hazard function, measured by the second derivative of the log-hazard at some (high) age  $y$  so that  $\mu = [\ln h(y)]''$ .

For  $\mu = \sigma^2$  the expressions in (4.6) take the form

$$\mathbb{E}[|\Lambda_{\text{null}}|] = \delta \quad \text{and} \quad \mathbb{E}[|\Lambda_{\text{full}}|] = \kappa \sqrt{\frac{2}{\pi}} - \kappa \phi\left(-\frac{\delta}{\kappa}\right) + \delta \Phi\left(-\frac{\delta}{\kappa}\right).$$

Consequently, model choice based on the  $\text{FIC}_{\text{MAE}}$  results in the gamma-Gompertz model if  $\hat{\delta}/\hat{\kappa} > 0.6399$ . If we view this as a pre-test strategy, then it has asymptotically the same local power as the likelihood ratio test for  $H_0: \sigma^2 = 0$  at a level of 26%.

If we choose  $\mu = [\ln h(y)]''$  the choice of the age  $y$  should be such that it marks an age in the tail of the distribution where deceleration occurs, but which still lies within the range of observed lifespans.

While the above choices of the focus parameter are natural and allow for immediate interpretations, we could also select as the focus any function that characterizes the distribution of lifespans, such as the survival function or the log-hazard. The effects of different focus parameters on the model selection will be briefly illustrated in the simulation study in Section 4.4, and recommendations will be given in Section 4.6.

### 4.3.4 A modified AIC for the gamma-Gompertz model

As we mentioned in Section 4.2.2, the standard AIC is biased as an estimator of the Akaike information in the presence of a boundary parameter, and should therefore not be used for assessing mortality deceleration. However, Böhstedt and Gampe (2019) explicitly derived the bias of the standard AIC for the gamma-Gompertz model (4.2) under the local misspecification framework (4.3) as  $2\Phi(-\delta/\kappa)$ . This bias depends via  $\delta = \sqrt{n}\sigma^2$  on the unknown value of the frailty variance, and it cannot be estimated without bias if the true variance is small. Thus, the bias cannot be removed completely, but it can be reduced if we correct the standard AIC using the estimator  $2\Phi(-\hat{\delta}/\hat{\kappa})$  of the bias term. Hence, we define a modified version of the AIC for the gamma-Gompertz model as

$$\text{AIC}^* = -2\ell + 2 \cdot 3 - 2\Phi\left(-\frac{\hat{\delta}}{\hat{\kappa}}\right). \quad (4.7)$$

The performance of this modified  $\text{AIC}^*$  for detecting mortality deceleration is studied in the next section.

## 4.4 Simulation study

To examine the performance of the proposed  $\text{FIC}_{\text{MAE}}$  in assessing mortality deceleration, we conducted a simulation study. In addition to considering different choices for the focus, the study compares the behavior of the  $\text{FIC}_{\text{MAE}}$  with that of the pre-test based on  $L_2$ -risks, and with that of the  $\text{AIC}^*$  defined in (4.7).

The following factors will affect the performance of the different strategies: the size of the true frailty variance  $\sigma^2$ ; the sample size  $n$ ; and the starting age used when observing lifespans, with a younger starting age being more favorable for detecting actual mortality deceleration.

For the frailty variance (at  $y = 0$ ), three different scenarios were considered:  $\sigma^2 = 0.0625$  ( $S_1$ ) and  $\sigma^2 = 0.03$  ( $S_2$ ) with Gompertz parameters  $a = 0.013$ ,  $b = 0.092$ . Scenario  $S_3$  is a pure Gompertz model with  $a = 0.0198$ ,  $b = 0.0726$  (and  $\sigma^2 = 0$ ). Scenario  $S_1$  was chosen to resemble the mortality pattern in the data on French-Canadian females that are analyzed in the following section. Scenario  $S_2$  mimics a population with the same Gompertz baseline hazard, but lower heterogeneity in the risk of death. By comparing Scenarios  $S_1$  and  $S_2$ , we will therefore be able to single out the effect of the size of the frailty variance on the performance of the selection strategies. Note that the frailty variances were set to levels comparable to those estimated from the French-Canadian data. Lastly, Scenario  $S_3$  was inspired by a fit of the Gompertz model to the female data.

To cover the latter two aspects, survival times were generated from the gamma-Gompertz model (4.2), with  $y = 0$  corresponding to age 60. However, the model selection was based only on subsets of individuals reaching certain ages. Motivated by the French-Canadian data, we considered individuals who survived to ages 90 or higher (90+). Additional comparisons based on the larger subsets of individuals who survived to ages 85+ and 80+ are presented in Section 4.7.6.

For each scenario  $S_1$  to  $S_3$ , three different initial sample sizes (at age 60) were chosen, such that the size of the 90+ subset approximately equals  $n_{90+} = 10,000$  (small),  $n_{90+} = 20,000$  (medium) or  $n_{90+} = 105,000$  (large). The sample sizes may look unusually large, but they cover a realistic range of population-based data. Recall that the French-Canadian data presented in Figure 4.1 contain information on about 20,000 women and 10,000 men.

For each 90+ sample, the log-likelihoods for the Gompertz model and the gamma-Gompertz model were maximized using function `n.lm()` in R (R Core Team, 2018); further computational details are given in Section 4.7.1. Then, the best model is selected based on the  $\text{FIC}_{\text{MAE}}$  for different focus parameters, the MSE pre-test of  $\delta < 0.8399\kappa$ , and the  $\text{AIC}^*$ . We ran 1,000 replications for each setting.

The left panel of Figure 4.2 compares the performance of the three selection approaches in Scenario  $S_1$  ( $\sigma^2 = 0.0625$ ) across the various sample sizes. The  $\text{FIC}_{\text{MAE}}$  with focus parameter  $\mu = [\ln h(100)]''$  clearly outperforms the other two methods, as it detects mortality deceleration more often. The proportion of correct decisions in favor of the gamma-Gompertz model increases with the sample size for all three methods, and is close to one for the setting with a large sample size. However, for the setting with a small

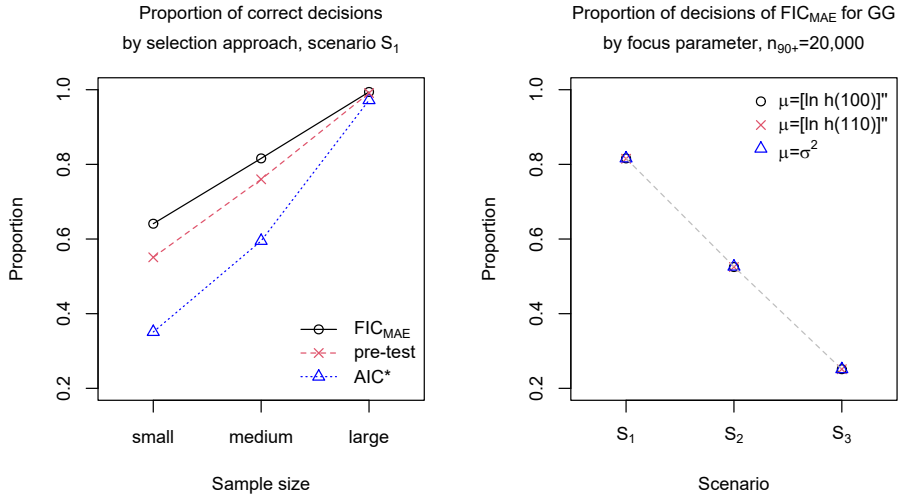


Figure 4.2: Proportion of decisions in favor of the gamma-Gompertz model. Left: Scenario  $S_1$  for sample sizes  $n_{90+} = 10,000$ ,  $n_{90+} = 20,000$ , and  $n_{90+} = 105,000$  (left to right) based on  $FIC_{MAE}$  with  $\mu = [\ln h(100)]''$  (black-solid-circle), pre-test (red-dashed-cross), and  $AIC^*$  (blue-dotted-triangle). Right: Scenarios  $S_1$ ,  $S_2$ , and  $S_3$  (left to right) all with  $n_{90+} = 20,000$  based on  $FIC_{MAE}$  with  $\mu = [\ln h(100)]''$  (black circle),  $\mu = [\ln h(110)]''$  (red cross), and  $\mu = \sigma^2$  (blue triangle).

(medium) sample size, the proportion of correct decisions based on the  $FIC_{MAE}$  is 82.6% (37.1%) higher than that based on the  $AIC^*$ .

The right panel of Figure 4.2 illustrates the performance of the  $FIC_{MAE}$  depending on the magnitude of the frailty variance, and on the choice of the focus parameter in the medium sample size setting. We display the results for the focus parameters  $\mu = \sigma^2$ ,  $\mu = [\ln h(100)]''$ , and  $\mu = [\ln h(110)]''$ . The ability of the method to detect deviations from the Gompertz hazard naturally decreases when the frailty variance decreases. For Scenario  $S_2$ , in which the frailty variance is about half as large as it is in Scenario  $S_1$ , the proportion of correct decisions is about 35% smaller than it is for  $S_1$ . If the true model is the Gompertz model ( $S_3$ ), then the proportion of decisions in favor of the gamma-Gompertz model is about 25% for the medium sample size. As the  $FIC_{MAE}$  performs equally well for all three focus parameters, the age  $y$  at which  $\mu = [\ln h(y)]''$  is evaluated does not seem to matter. It also turns out that the focus parameters  $\mu = \sigma^2$  and  $\mu = [\ln h(y)]''$  perform better than, for instance,  $\mu = \ln h(y)$  or  $\mu = S(y)$ ; as is shown in Section 4.7.5. Although the focus age  $y$  did not affect the results in the simulation study, other aspects may render one choice more reasonable than another. In the medium-sized Scenario  $S_1$ , in which around 20,000 individuals reach age 90, more than a thousand will, on average, also reach age 100, but fewer than 10 will reach age 110. Consequently, a

focus age of  $y = 100$  will probably produce more reliable results than a focus age of  $y = 110$ .

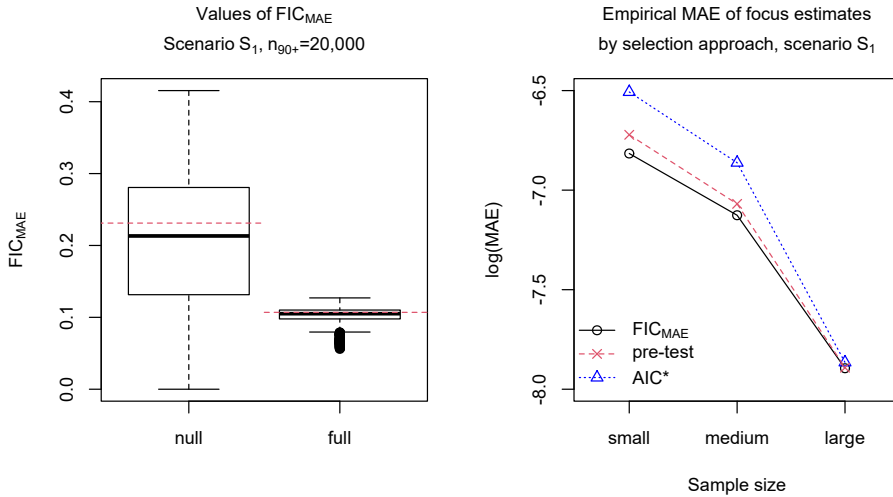


Figure 4.3: Left: Box plots of  $FIC_{MAE}$  values with  $\mu = [\ln h(100)]''$  for the null and the full model in Scenario  $S_1$  with  $n_{90+} = 20,000$  and empirical MAE of focus estimates  $\hat{\mu}$  (red-dashed). Right: Empirical MAE of selected  $\hat{\mu}$  for  $\mu = [\ln h(100)]''$  in Scenario  $S_1$  for sample sizes  $n_{90+} = 10,000$ ,  $n_{90+} = 20,000$ , and  $n_{90+} = 105,000$  (left to right) based on  $FIC_{MAE}$  (black-solid-circle), pre-test (red-dashed-cross), and  $AIC^*$  (blue-dotted-triangle).

The concept of the  $FIC_{MAE}$  as an estimator of the limiting MAE of  $\hat{\mu}$  is illustrated in the left panel of Figure 4.3, which shows a box plot of the  $FIC_{MAE}$  values with  $\mu = [\ln h(100)]''$  for 1,000 replications of the medium-sized Scenario  $S_1$ . We see that for both the null and the full model, the empirical MAEs of the estimators  $\hat{\mu}_{null}$  and  $\hat{\mu}_{full}$  are close to the average of the respective FIC scores. As a consequence, the empirical MAE of the selected estimators in the 1,000 replications – that is,  $\hat{\mu}_{full}$  for those replications, where  $FIC_{MAE}(full) < FIC_{MAE}(null)$ , and  $\hat{\mu}_{null}$  otherwise – should be smaller than it is for other selection criteria. The right panel of Figure 4.3 verifies for Scenario  $S_1$ , that the estimator  $\hat{\mu}$  of  $\mu = [\ln h(100)]''$  has the smallest empirical MAE when the model selection is based on the  $FIC_{MAE}$  with  $\mu = [\ln h(100)]''$ , rather than on the pre-test or the  $AIC^*$ .

Overall, the findings of the simulation study support the claim that the proposed  $FIC_{MAE}$  is a suitable tool for detecting mortality deceleration in the framework of the gamma-Gompertz model, which outperforms the competing approaches of the pre-test and  $AIC^*$ .

## 4.5 Mortality of French-Canadians at high ages

### 4.5.1 Data and age validation

To demonstrate the performance of the proposed FIC, we analyze data on French-Canadian cohorts born between 1880 and 1896. The data were illustrated in Figure 4.1.

When French missionaries settled in Quebec in the 17<sup>th</sup> century, they followed the Catholic tradition of registering all baptisms, marriages, and burials in parish registers. Starting in 1679, two copies of the registrations were made, with one being kept at the parish, while the other was sent to the government body responsible for civil registration. This practice was continued until the end of the 20<sup>th</sup> century. Thus, this data collection is an invaluable resource for historical demography, among other disciplines (Desjardins, 1998). The data set that we analyze here covers virtually all Catholic French-Canadians (20,917 females and 10,878 males) who were born in the Province of Quebec during the 1880-1896 period, and who died at age 90 or older in Quebec between 1970 and 2009. These 1880-1896 birth cohorts were fully extinct by the end of 2009.

To confirm a reported age at death (date of death minus date of birth), the date of birth reported on the death certificate has to be verified using the baptism record in the corresponding parish register, and is often additionally verified using historic census returns. Although the historic French-Canadian parish registers contain information on all baptisms and are available on microfilm, the identification of individual entries in the registers continues to be time-consuming. The appropriate parish has to be identified using the information on the province and the names of the parents, and the individual register entry has to be looked up (for details, see Bourbeau and Desjardins, 2002).

The core of the data set was compiled by Beaudry-Godin (2010), who individually validated ages at death using the parish registers of all Catholic French-Canadians who died as centenarians – i.e., at age 100 or older – between 1970 and 2004. The data were further extended and validated to cover deaths until 2009 (Ouellette and Bourbeau, 2014; Ouellette, 2016), and the deaths of individuals aged 90-99 were added.

### 4.5.2 Comparison of information criteria

We fit the gamma-Gompertz model and the Gompertz model to the female and the male data separately via maximum likelihood. For that purpose, we set age 60 as the starting age of the models, and take into account left truncation at age 90. Then, we choose between the gamma-Gompertz model and the Gompertz model based on the AIC\*, pre-test, and FIC<sub>MAE</sub> for the focus parameters  $\mu = \sigma^2$  and  $\mu = [\ln h(100)]''$ .

Figure 4.4 shows the fit of the gamma-Gompertz model and the Gompertz model, respectively, to the empirical death rates (single years of age) for the French-Canadian cohorts. The estimated frailty variance in the gamma-Gompertz model is  $\hat{\sigma}^2 = 0.043$  for the female population and  $\hat{\sigma}^2 = 0.037$  for the male population. A likelihood ratio test for  $H_0: \sigma^2 = 0$  results in a  $p$ -value of 0.121 for females and 0.283 for males, such that the hypothesis of no mortality deceleration would not be rejected at the usual levels of

significance. Table 4.1 also shows that based on the modified AIC\*, the Gompertz model is selected, if only just, for females and for males. By contrast, based on the pre-test and the  $FIC_{MAE}$ , the gamma-Gompertz model is selected for females, and the Gompertz model is selected for males. Hence, it appears that unlike other methods, the  $FIC_{MAE}$  detects mortality deceleration in the female sample. Figure 4.4 also supports this finding of a deceleration in the female mortality rates.

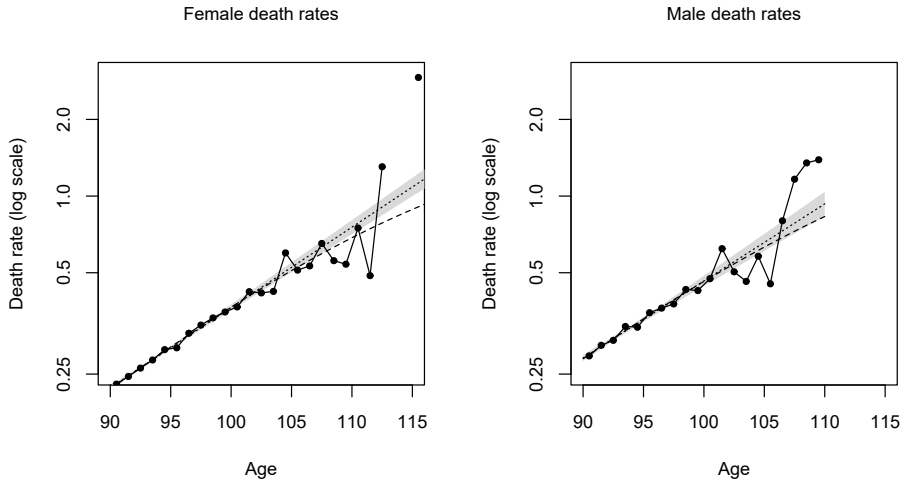


Figure 4.4: Death rates (on log scale) of French-Canadian females (left) and males (right): empirical death rates (solid-circle), gamma-Gompertz fit (dashed), Gompertz fit (dotted), and 95%-confidence band for the Gompertz log-hazard (gray).

Table 4.1: Values of different model selection criteria for the gamma-Gompertz model (GG) and the Gompertz model using data on French-Canadians.

	Females		Males	
	GG	Gompertz	GG	Gompertz
AIC*	101390.3	<b>101390.0</b>	48364.10	<b>48363.01</b>
$FIC_{MAE} : \mu = \sigma^2$	<b>4.065</b>	6.200	4.316	<b>3.890</b>
$FIC_{MAE} : \mu = [\ln h(100)]''$	<b>0.098</b>	0.149	0.120	<b>0.108</b>

### 4.5.3 Including effects of year of birth

The analysis so far has split the data by gender, but combined all birth cohorts in one model. This was done primarily in order to retain sufficiently large samples, which, as the simulations showed, is an important issue in the assessment of mortality deceleration, even if the samples can be considered sizable based on common statistical standards. But it is obvious that such a pooling of birth cohorts can increase the heterogeneity in mortality risks, and may thus yield a larger frailty variance, which, in turn, affects the assessment of mortality deceleration. This concern is particularly relevant if mortality patterns have improved over time at advanced ages (Vaupel et al., 2021). Including information on the year of birth has been shown to be important in different contexts (Richards et al., 2006).

We have, therefore, chosen to include the effect of year of birth in the analysis by allowing the parameters of the gamma-Gompertz model to vary by cohort, and have adapted our model selection approach accordingly. While the inclusion of cohort effects implies a more flexible approach to modeling, it also reduces the parsimony of the model, which is critical in this application. In the following, we will focus on the female sample in undertaking this considerably more complex analysis. The male sample is only half of the size of the female sample, and does not allow for a reliable statistical analysis in the extended setting introduced below.

The data cover 17 single-year birth cohorts made up of between about 700 and 2000 individuals. These cohort sizes, in conjunction with the restricted age range of 90+, are insufficient for an analysis by single years of birth. For this reason, we group the data into multi-year birth cohorts of roughly equal sizes by combining individuals with the following birth years: 1880-1884, 1885-1888, 1889-1891, 1892-1894, and 1895-1896. In the remainder of this section, we will refer to these five multi-year birth cohorts, which are, on average, made up of 4183.4 individuals.

As a gamma-Gompertz model in which all parameters are cohort-specific would still have five times as many parameters to be estimated as the plain gamma-Gompertz model considered in Section 4.5.2, we seek to identify a model that is more parsimonious, but is still appropriately flexible. To do so, we consult data for the respective cohorts between ages 60 and 109 from the Canadian Human Mortality Database (CHMD, 2020). The CHMD data are collected at the level of the Canadian provinces and territories, including the Province of Quebec. These data can be used to identify potential trends across cohorts in the Gompertz part of the gamma-Gompertz model, which is largely determined at the mid-adult ages. Although the data for Quebec in the CHMD cover a wider population, French-Canadians constitute the majority in the age groups considered here, so that these data can serve for this purpose. (A comparison of the two data sources has already been provided in Ouellette, 2016.) More precisely, we use the CHMD data (number of deaths and total exposure times at each age) for Quebec to estimate Gompertz models with a starting age of 60, separately for each of the five cohorts (maximizing a Poisson likelihood for death counts; see, for instance, Keiding, 1990). The parameter estimates suggest that there is a linear trend in the Gompertz parameter  $a$  across cohorts, while the Gompertz parameter  $b$  stays roughly constant for the cohorts under study (see Figure 4.9).



This gamma-Gompertz model, in which there is a linear trend in the parameter  $a$  and a single parameter  $b$ , but with each cohort having its own frailty variance, is estimated and analyzed in the following. This model is a restricted version of the gamma-Gompertz model in which all parameters vary by cohort. Nonetheless, it still provides an adequate description of the mortality patterns in the female French-Canadian cohorts studied here. More formally, we modify the gamma-Gompertz model with hazard (4.2) in the following way: The parameters for cohort  $c$ ,  $c = 1, \dots, C$ , are given by

$$a_c = a_0 + a_1 \cdot z_{1c}, \quad b_c \equiv b_0, \quad \text{and} \quad \sigma_c^2 = \sigma_0^2 \exp(\mathbf{z}_2^\top \boldsymbol{\zeta}), \quad (4.8)$$

where  $a_0$  and  $\sigma_0^2$  are the parameters of a reference cohort  $c^*$  and  $\boldsymbol{\zeta} \in \mathbb{R}^{C-1}$  describes cohort-specific deviations from the frailty variance. The parameter  $b_0 > 0$  applies to all cohorts, whereas the linear trend in the initial level of mortality  $a_c$  is governed by the slope  $a_1 \in \mathbb{R}$  (but such that  $a_c > 0$ ) with  $z_{1c} = (c - c^*)$  and  $\mathbf{z}_1 = (z_{11}, \dots, z_{1C})^\top$ . The vector  $\mathbf{z}_2$  of length  $(C - 1)$  consists of cohort dummies.

We fit model (4.8) to the individual survival times of the female cohorts in the French-Canadian data via maximum likelihood. The middle cohort is used as reference and, as before, age 60 is set as the starting age of the model, and left truncation at age 90 is taken into account. The parameter estimates and the corresponding standard errors are summarized in Table 4.2. The negative estimate of  $a_1$  indicates a decrease in the initial level of mortality for the later cohorts, and, hence, that mortality has indeed improved. The estimated frailty variances are comparatively high for the first three cohorts, but are decreasing for the last two cohorts, reaching a value close to zero for the most recent cohort. These results are in line with the impression conveyed by the empirical death rates of the five cohorts displayed in Figure 4.5.

Table 4.2: Parameter estimates (with standard errors) for gamma-Gompertz model (4.8) with cohort effects for French-Canadian females.

$\hat{a}_0$	$\hat{a}_1$	$\hat{b}_0$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	$\hat{\sigma}_4^2$	$\hat{\sigma}_5^2$
0.0122	-0.0006	0.0932	0.1089	0.0989	0.0757	0.0471	0.0089
(0.0035)	(0.0002)	(0.0113)	(0.0379)	(0.0399)	(0.0401)	(0.0406)	(0.0413)

Performing focused model selection is considerably more challenging in this extended setting with cohort effects, even in the still relatively parsimonious gamma-Gompertz model specified in (4.8). In the framework of local misspecification, as introduced in Section 4.3.1, the full model is now defined as the gamma-Gompertz model (4.8) with cohort effects. The null model – that is, the simplest candidate model – should still be the plain Gompertz model without cohort effects. Thus, the parameter vector of the full model is split into the Gompertz part  $\boldsymbol{\theta} = (a_0, b_0)^\top$ , so still with  $d = 2$ , and the additional  $\boldsymbol{\gamma} = (\sigma_0^2, a_1, \boldsymbol{\zeta}^\top)^\top$  with  $\boldsymbol{\gamma}_0 = (0, 0, \mathbf{0}^\top)^\top$ , so  $q = (C + 1)$ .

Thus, selection is no longer only between the full model and the null model (as was the case in Section 4.5.2), as we can now choose from an extended list of candidate

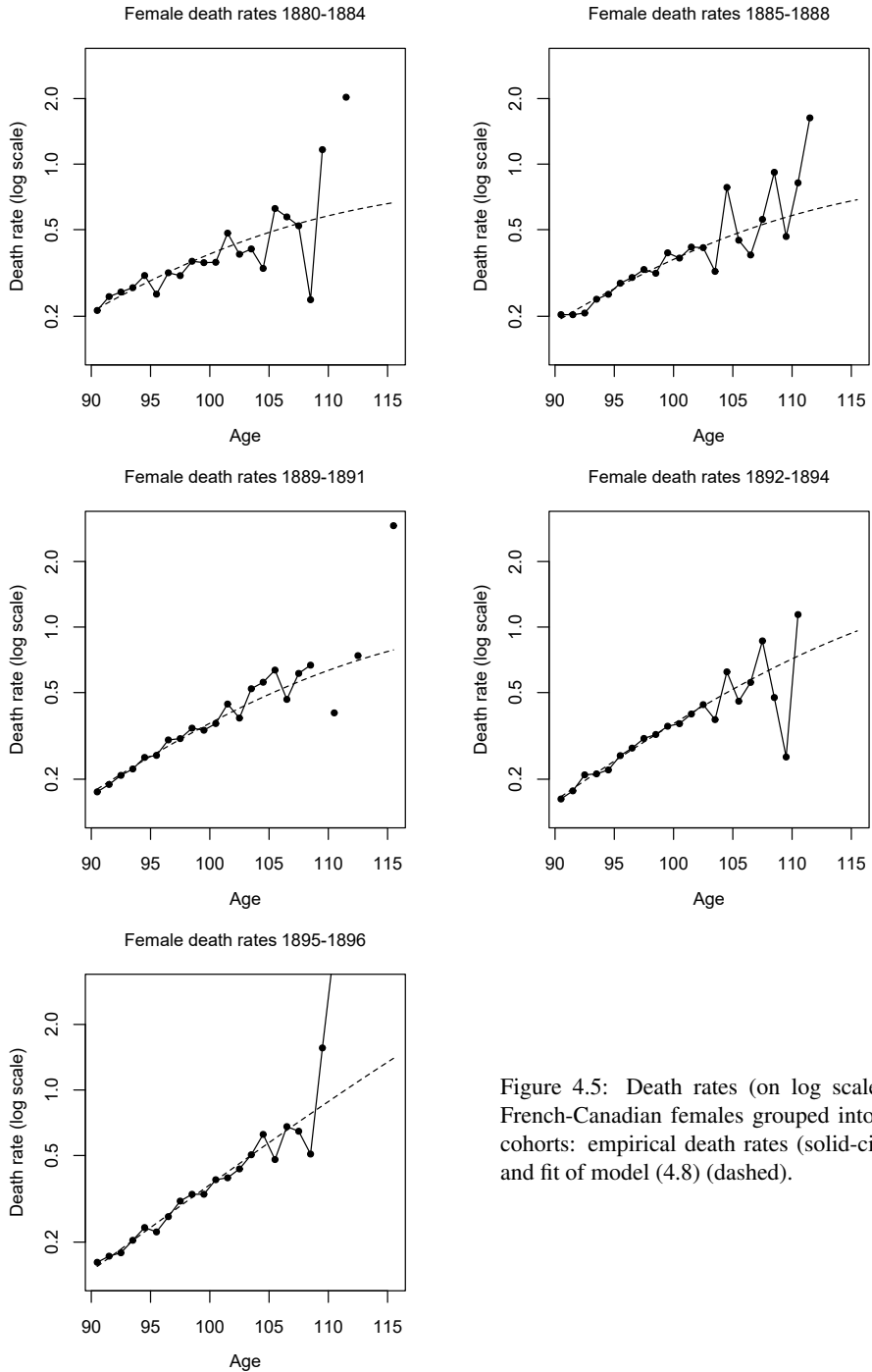


Figure 4.5: Death rates (on log scale) of French-Canadian females grouped into five cohorts: empirical death rates (solid-circle) and fit of model (4.8) (dashed).

models. On this list of candidate models are also the plain gamma-Gompertz model without cohort effects, as well as the Gompertz and gamma-Gompertz models in which only the parameter  $a$  varies linearly across cohorts, while the other parameters are constant. The relevant candidate models are shown in Table 4.3, where a model is denoted by  $M$ ,  $M \subseteq \{1, \dots, q\}$ , if it includes all parameters  $\gamma_j$  with  $j \in M$ . Mortality deceleration is generally present in those models  $M$  that include the parameter  $\sigma_0^2$ . But we have to be careful when drawing conclusions about this phenomenon from the models with cohort-specific frailty variances  $\sigma_c^2$ . If the frailty variances of some of the cohorts are relatively large, while those of other cohorts are close to zero, it may be assumed that mortality deceleration is present for the former, but not for the latter cohorts. The distinct advantage of focused model selection is that we can choose cohort-specific focus parameters, such as the frailty variance  $\mu = \sigma_c^2$  of cohort  $c$ , which then allow us to assess mortality deceleration separately for each cohort.

Table 4.3: Overview of candidate models in the framework of model (4.8).

Model	$M$	Parameters			No. of param.
		$a_c$	$b_c$	$\sigma_c^2$	
Gompertz, plain (null)	$M = \emptyset$	$a_0$	$b_0$	0	2
Gompertz with linear trend in $a$	$M = \{2\}$	$a_0 + a_1 z_{1c}$	$b_0$	0	3
gamma-Gompertz (GG), plain	$M = \{1\}$	$a_0$	$b_0$	$\sigma_0^2$	3
GG with linear trend in $a$	$M = \{1, 2\}$	$a_0 + a_1 z_{1c}$	$b_0$	$\sigma_0^2$	4
GG with cohort-specific $\sigma^2$	$M = \{1, 3, \dots, q\}$	$a_0$	$b_0$	$\sigma_c^2$	$2 + C$
GG as in (4.8) (full)	$M = \{1, \dots, q\}$	$a_0 + a_1 z_{1c}$	$b_0$	$\sigma_c^2$	$3 + C$

The current selection framework also differs from the previous setting in Section 4.3 from a more technical point of view. Previously, we had to choose whether to include a single parameter with a boundary constraint ( $q = 1$  and  $\gamma \geq \gamma_0$ ). Here, we have to choose whether to include none, some, or all of the components of the parameter vector  $\boldsymbol{\gamma}$  where only the first component is subject to a boundary constraint ( $q > 1$  and  $\gamma_1 \geq \gamma_{01}$ ). Consequently, the results from Section 4.3.2 do not apply in the current setting, and new formulas for the  $\text{FIC}_{\text{MAE}}(M)$  for each of the candidate models  $M$  have to be derived. We refer the interested reader to Section 4.7.8 for these derivations of the limiting  $L_1$ -risks of the focus estimators.

To better evaluate the approach of focused model selection in the setting with cohort effects, we seek to compare model selection based on the new  $\text{FIC}_{\text{MAE}}$  with that based on the AIC. However, due to the presence of the boundary parameter  $\sigma_0^2$  in some of the candidate models, the standard AIC is again found to be biased. As in the case of the modified AIC\* for the plain gamma-Gompertz model presented in Section 4.3.4, we derived a modified AIC\* with a bias correction term for the models listed in Table 4.3 (see Section 4.7.9 for a sketch of the derivations).

Finally, we apply the two model selection strategies to the five female cohorts in the French-Canadian data. We choose between the candidate models of Table 4.3 based on the  $AIC^*$ , and based on the  $FIC_{MAE}$  with cohort-specific focus parameters  $\mu = \sigma_c^2$  and  $\mu = [\ln h_c(100)]''$ , where  $h_c(y)$  is the gamma-Gompertz hazard (4.2) with parameters  $a_c$ ,  $b_c$ , and  $\sigma_c^2$  according to (4.8). The selected models are presented in Table 4.4.

Table 4.4: Selected models for data on French-Canadian females.

Multi-year birth cohort $c$	$FIC_{MAE}$		$AIC^*$
	$\mu = \sigma_c^2$	$\mu = [\ln h_c(100)]''$	
1880-1884	gamma-Gompertz	full	} full
1885-1888	gamma-Gompertz	full	
1889-1891	gamma-Gompertz	gamma-Gompertz	
1892-1894	Gompertz	Gompertz	
1895-1896	Gompertz	Gompertz, linear trend in $a$	

While the  $AIC^*$  selects a single best model irrespective of whether the interest is in one particular cohort, the cohort-specific focus parameters for the  $FIC_{MAE}$  make a clear differentiation, and may select different models for different cohorts of interest. More specifically, based on the  $AIC^*$  the full gamma-Gompertz model (4.8) with a linear trend in  $a$ , constant  $b$ , and cohort-specific frailty variances  $\sigma_c^2$  is selected. By contrast, based on the  $FIC_{MAE}$ , models with mortality deceleration are selected only for the first three cohorts, whereas models without mortality deceleration are selected for the two most recent cohorts. These selection results seem plausible given the parameter estimates reported in Table 4.2 and the mortality rates displayed in Figure 4.5. (The reader might notice that for the 1892-1894 cohort, the Gompertz model without mortality deceleration is selected based on the  $FIC_{MAE}$ , although the estimated frailty variance  $\hat{\sigma}_4^2$  is larger than that for the pooled cohorts for which the gamma-Gompertz model was selected. Here, we have to keep in mind the smaller sample size of the cohort, which affects the performance of the  $FIC_{MAE}$ .)

Furthermore, the selection results nicely illustrate that based on the  $FIC_{MAE}$ , different models might be favored for the same cohort if different focus parameters are chosen. If the interest lies in the cohort's frailty variance, the plain gamma-Gompertz model is selected for the first two cohorts; but if the interest lies in the second derivative of the cohort's log-hazard, the full model is selected for these two cohorts. This is reasonable because the estimated frailty variances of the first two cohorts are close ( $\hat{\sigma}_1^2 = 0.1089$  and  $\hat{\sigma}_2^2 = 0.0989$ ), which suggests that cohort-specific deviations from the frailty variance would not markedly improve the estimator performance regarding the frailty variances  $\sigma_1^2$  and  $\sigma_2^2$ . In contrast, all gamma-Gompertz parameters are involved in the estimation of the log-hazard, such that an estimator based on the full model with cohort-specific frailty variances and a trend in  $a$ , which consequently takes into account the higher initial mortality  $a_c$  in the first two cohorts, would, on average, perform better.

## 4.6 Discussion

Motivated by the issue of how mortality deceleration can be assessed at high ages, we have extended the FIC, as introduced by Claeskens and Hjort (2003), to a non-standard setting in which we are choosing between two models that differ by one parameter that takes a value on the boundary of the parameter space if the smaller model is the true model. We considered two versions of the FIC that aim to minimize the limiting MAE or MSE of the estimator of the focus, respectively. When targeting the MAE, we obtained the new model selection criterion  $FIC_{MAE}$ . When targeting the MSE, the model selection does not depend on the chosen focus, but a pre-test strategy was defined. In addition, we presented the new  $AIC^*$ , which reduces the bias of the original AIC that occurs when the selection concerns a parameter that lies on the boundary of the parameter space in the narrow model.

The proposed model selection criteria provide new tools for the assessment of mortality deceleration in the framework of the gamma-Gompertz model. While traditional approaches either have low power to detect mortality deceleration or are not valid in the presence of boundary-constrained parameters, the methods developed here are adapted to the non-standard setting. An advantage of the  $FIC_{MAE}$  is that, by choosing an appropriate focus parameter, it can be targeted directly at the quantities that reveal mortality deceleration. We recommend using as the focus parameter the frailty variance or the second derivative of the log-hazard at some advanced age. Both potential choices readily translate into the presence or the absence of mortality deceleration, as the focus parameter takes a value of zero if there is no deceleration.

The results of our simulation studies indicate that the  $FIC_{MAE}$ , especially with the recommended choices of the focus parameter, outperforms the competing approaches of the pre-test and the  $AIC^*$  in detecting mortality deceleration. This observation was made for different magnitudes of the frailty variance, and with different sample sizes. We found that the  $FIC_{MAE}$  performs substantially better than the  $AIC^*$ , particularly for small samples.

Mortality deceleration addresses properties of the tail of the lifespan distribution, and inference about the tail behavior requires sufficiently large samples. Moreover, it is equally important that the models are sufficiently parsimonious so that the question of interest can be isolated in a few parameters that inference can center on. While the size of our data set on French-Canadian Catholics born at the end of the 19<sup>th</sup> century is relatively large, the male sample contains only half as many observations as the female sample. In contrast to the other methods, the  $FIC_{MAE}$  detected mortality deceleration in the sample of females. However, all of the methods point towards a Gompertz model, and hence towards no mortality deceleration for the French-Canadian males.

As we demonstrated, focused model selection allows us to include additional information, such as differences across birth cohorts. At the same time, however, the number of candidate models and the demands for data tend to increase. The analysis of the female sample with cohort effects pinpointed the differences between traditional ( $AIC^*$ )

and focused criteria. The within-cohort heterogeneity was estimated to decline, and the gamma-Gompertz model was selected only for the three earlier cohorts.

Data on mortality at advanced ages are often limited due to the complexity of age validation procedures or the availability of the historic documents required for verifying alleged ages at death. As the ability to detect features that are contained in the distribution tail depends on having sufficient information in the data, it is worth investigating how our sampling and validation efforts can be optimized. This includes reconsidering both the required sample sizes and the age range over which data are collected. Here, we were limited to ages at death above 90, which is a rather high starting age compared to the starting ages in other studies. Thus, if further data collection efforts are planned, it may be more advantageous to extend the age range to lower ages. Böhnstedt et al. (2021) have offered some tools to address such questions.

While the set-up in this article was restricted to individual-level data, many studies on aging rely on aggregated data in which death counts and exposure times are available for given age-intervals. However, an extension of the approach to aggregated data is straightforward if we keep the assumption of the parametric model. Consequently, the new tools for the assessment of mortality deceleration presented here will be applicable to a variety of data sets collected for different human and non-human populations. For data on humans, the application of the Gompertz hazard is well-studied and well-established, both across time and across populations. For data on non-human species, we might want to consider relaxing the assumption of a parametric model for the hazard. More research is needed to understand how more flexible hazard shapes can be incorporated by, for example, using splines and penalized likelihood.

Although our development of the  $FIC_{MAE}$  was motivated by the specific problem of assessing mortality deceleration, the method could be used in a range of other contexts in which there is a need to choose between parametric models that differ only by one parameter with a boundary constraint, such as when assessing heterogeneity in other proportional hazards frailty models, or when choosing between a Poisson model and an over-dispersed negative binomial model. Linear mixed models are another model class where some parameters, in that case variance components, are restricted to be non-negative and where a focused search is useful (Cunen et al., 2020).

## 4.7 Supplementary material

### 4.7.1 Computational issues

The maximum likelihood estimates of the parameters of the gamma-Gompertz model are determined via numerical optimization in R (R Core Team, 2018). Non-negativity of the parameter estimates is achieved by maximizing the log-likelihood over the log-transformed parameters. Nevertheless, values of the frailty variance  $\sigma^2$  that are close to zero cause numerical difficulties. Here, we briefly describe the steps that we took to increase the numerical stability of the estimation problem.

We maximize the log-likelihood via the R-function `n.lm()`, where we can also supply the analytic gradient of the objective function. In addition, Taylor expansions of the log-likelihood and its gradient are used if the current value of  $\sigma^2$  is smaller than  $10^{-5}$ . The numerically identified maximum  $\hat{\sigma}^2$  might still depend on the starting value that was provided to the optimization routine. We therefore recommend running the optimization with a number of different starting values for the frailty variance and choosing the fit with the largest value of the log-likelihood as the final estimate.

For calculating the  $FIC_{MAE}$  values, we need estimates not only of the model parameters, but also of the information matrix  $J_{full}$ . For that purpose, we analytically derive the matrix  $H(a, b, \sigma^2)$  of second-order derivatives of the log-likelihood for the gamma-Gompertz model, and again use a Taylor expansion if  $\sigma^2 < 10^{-5}$ .  $J_{full}$  is then estimated as  $-n^{-1}H(\hat{a}, \hat{b}, \hat{\sigma}^2)$ ; and  $\hat{\kappa}^2$  is the bottom right element of its inverse.

### 4.7.2 Power of the likelihood ratio test

A likelihood ratio test (LRT) for homogeneity in the gamma-Gompertz model, where  $H_0 : \sigma^2 = 0$  and  $H_1 : \sigma^2 > 0$ , may have low power to detect mortality deceleration. To illustrate this property, we summarize the results for two of the scenarios that were described in Section 4.4. In particular, we show the extent to which a smaller underlying frailty variance or a smaller sample size can decrease the power of the test, which is performed at a significance level of 5%. We also compare the power of the LRT in a situation in which only individuals who survived beyond age 90 can be studied to a situation in which observations for individuals who survived beyond age 80 or 85 are available.

Figure 4.6 illustrates how strongly the power of the test is affected by the three features. The left panel displays the results for Scenario  $S_1$  (frailty variance  $\sigma^2 = 0.0625$ ), while the right panel shows the results for Scenario  $S_2$  in which the frailty variance was roughly halved ( $\sigma^2 = 0.03$ ). Within each panel, we can see the loss in power that occurs if only individuals who survived beyond age 90 (90+) can be studied, instead of individuals who survived beyond age 80 (80+) or 85 (85+). For example, in the medium-sized Scenario  $S_1$ , the power of the LRT decreases by more than 45% if the test is based on the 90+ subset instead of on the 85+ subset.

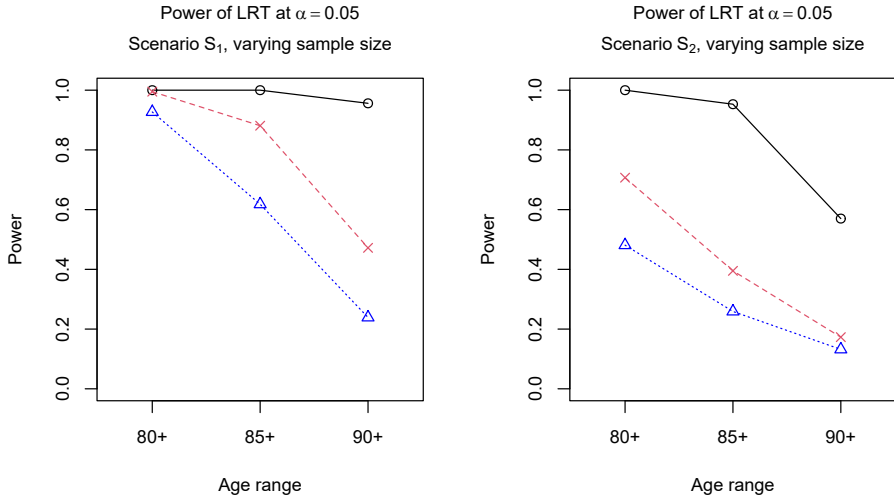


Figure 4.6: Power of the LRT at the 5% level to detect mortality deceleration in the gamma-Gompertz model depending on the age range of the data (left to right: 80+, 85+, or 90+). The depicted scenarios are  $S_1$  (left) and  $S_2$  (right) with the sample sizes  $n_{90+} = 10,000$  (blue-dotted-triangle),  $n_{90+} = 20,000$  (red-dashed-cross), and  $n_{90+} = 105,000$  (black-solid-circle).

### 4.7.3 Local power of the LRT and the pre-test

The local power of the LRT for homogeneity in the gamma-Gompertz model is derived in Böhnstedt and Gampe (2019). Under the sequence of local alternatives (4.3), the power of the LRT for  $H_0: \sigma^2 = 0$  at level  $\alpha$  based on a gamma-Gompertz sample of size  $n$  can be approximated by

$$1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{\delta}{\kappa}\right) = 1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{\sqrt{n}\sigma^2}{\kappa}\right). \quad (4.9)$$

The pre-test derived in Section 4.3.2 selects the gamma-Gompertz model if  $\hat{\delta}/\hat{\kappa} > 0.8399$ . Due to  $\hat{\delta}/\hat{\kappa} \xrightarrow{d} \max(0, D/\kappa)$ , we have  $P[\hat{\delta}/\hat{\kappa} \leq z] \approx \Phi(z - \delta/\kappa) \mathbb{1}_{\{z \geq 0\}}$ . As a consequence, the power of the pre-test with critical region  $\hat{\delta}/\hat{\kappa} > 0.8399$  is determined as

$$P\left[\frac{\hat{\delta}}{\hat{\kappa}} > 0.8399 \mid \text{fixed } \delta\right] \approx 1 - \Phi\left(0.8399 - \frac{\delta}{\kappa}\right). \quad (4.10)$$

Comparing (4.10) and (4.9), we find that for large samples, the pre-test has approximately the same power as the LRT for  $H_0: \sigma^2 = 0$  at level  $\tilde{\alpha}$  satisfying  $\Phi^{-1}(1 - \tilde{\alpha}) = 0.8399$ , which is  $\tilde{\alpha} = 1 - \Phi(0.8399) \approx 0.2005$ .



#### 4.7.4 Derivation of $\text{FIC}_{\text{MAE}}$ with a single additional parameter on the boundary of the parameter space

The  $\text{FIC}_{\text{MAE}}$  of a model with focus estimator  $\hat{\mu}$ , where  $\sqrt{n}(\hat{\mu} - \mu_{\text{true}}) \xrightarrow{d} \Lambda$ , is derived as an estimate of  $\mathbb{E}[|\Lambda|]$ . For the null model,  $\hat{\mu}_{\text{null}}$  converges to a normal distribution,  $\Lambda_{\text{null}} = (\Lambda_0 + \omega\delta) \sim \mathcal{N}(\omega\delta, \tau_0^2)$ . Therefore,  $\mathbb{E}[|\Lambda_{\text{null}}|]$  is calculated as the expected value of the folded normal random variable  $|\Lambda_0 + \omega\delta|$ ; that is,

$$\mathbb{E}[|\Lambda_{\text{null}}|] = \mathbb{E}[|\Lambda_0 + \omega\delta|] = 2\tau_0\phi\left(\frac{\omega\delta}{\tau_0}\right) + 2\omega\delta\left\{\Phi\left(\frac{\omega\delta}{\tau_0}\right) - \frac{1}{2}\right\}. \quad (4.11)$$

For the full model, we have  $\Lambda_{\text{full}} = \Lambda_0 - \omega(D - \delta)\mathbb{1}_{\{D>0\}} + \omega\delta\mathbb{1}_{\{D\leq 0\}}$ , with  $D \sim \mathcal{N}(\delta, \kappa^2)$  independent of  $\Lambda_0$ , such that

$$\begin{aligned} \mathbb{E}[|\Lambda_{\text{full}}|] &= \mathbb{E}[|\Lambda_{\text{full}}| \mid D \leq 0]\text{P}[D \leq 0] + \mathbb{E}[|\Lambda_{\text{full}}| \mid D > 0]\text{P}[D > 0] \\ &= \mathbb{E}[|\Lambda_0 + \omega\delta|]\Phi\left(-\frac{\delta}{\kappa}\right) + \mathbb{E}[|\Lambda_0 - \omega(D - \delta)| \mid D > 0]\Phi\left(\frac{\delta}{\kappa}\right). \end{aligned} \quad (4.12)$$

The first expectation is the same as (4.11). For the computation of the second expectation, we define the normally distributed random vector  $\mathbf{X} = (\Lambda_0, D)^\top$  and its affine transformation  $\mathbf{Y} = (\Lambda_0 - \omega(D - \delta), D)^\top$ , which is also normally distributed, with mean vectors  $\boldsymbol{\mu}_X = \boldsymbol{\mu}_Y = (0, \delta)^\top$  and covariance matrices

$$\text{Cov}[\mathbf{X}] = \begin{pmatrix} \tau_0^2 & 0 \\ 0 & \kappa^2 \end{pmatrix} \quad \text{and} \quad \text{Cov}[\mathbf{Y}] = \begin{pmatrix} \tau_0^2 + \omega^2\kappa^2 & -\omega\kappa^2 \\ -\omega\kappa^2 & \kappa^2 \end{pmatrix}.$$

Then,  $\mathbb{E}[|\Lambda_0 - \omega(D - \delta)| \mid D > 0]$  can be rewritten as

$$\begin{aligned} \mathbb{E}[|Y_1| \mid Y_2 > 0] &= \mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0] \frac{\text{P}[Y_1 > 0, Y_2 > 0]}{\text{P}[Y_2 > 0]} \\ &\quad + \mathbb{E}[-Y_1 \mid -Y_1 \geq 0, Y_2 > 0] \frac{\text{P}[Y_1 \leq 0, Y_2 > 0]}{\text{P}[Y_2 > 0]}. \end{aligned} \quad (4.13)$$

The expected values of one component of a bivariate truncated normal distribution are more easily found for bivariate normal distributions with zero mean vectors, unit variances, and possible correlations. Transforming  $\mathbf{Y}$  into such a normally distributed random vector  $\mathbf{Z} = ((\tau_0^2 + \omega^2\kappa^2)^{-1/2}Y_1, (Y_2 - \delta)/\kappa)^\top$  with covariances  $-\omega\kappa/\sqrt{\tau_0^2 + \omega^2\kappa^2}$ , and noting that

$$\mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0] = \sqrt{\tau_0^2 + \omega^2\kappa^2} \mathbb{E}\left[Z_1 \mid Z_1 > 0, Z_2 > -\frac{\delta}{\kappa}\right],$$

we can apply the results of Tallis (1961) to obtain  $\mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0]$  in (4.13) as

$$\frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\mathbb{P}[Y_1 > 0, Y_2 > 0]} \left\{ \frac{1}{\sqrt{2\pi}} \Phi \left( \frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\tau_0} \right) - \frac{\omega \kappa}{\sqrt{\tau_0^2 + \omega^2 \kappa^2}} \phi \left( \frac{\delta}{\kappa} \right) \Phi \left( \frac{\omega \delta}{\tau_0} \right) \right\}.$$

Analogously,  $\mathbb{E}[-Y_1 \mid -Y_1 \geq 0, Y_2 > 0]$  in (4.13) is computed as

$$\frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\mathbb{P}[Y_1 \leq 0, Y_2 > 0]} \left\{ \frac{1}{\sqrt{2\pi}} \Phi \left( \frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\tau_0} \right) + \frac{\omega \kappa}{\sqrt{\tau_0^2 + \omega^2 \kappa^2}} \phi \left( \frac{\delta}{\kappa} \right) \Phi \left( -\frac{\omega \delta}{\tau_0} \right) \right\}.$$

Combining these two results, we find that  $\mathbb{E}[|\Lambda_0 - \omega(D - \delta)| \mid D > 0]$  in (4.12) is equal to

$$\frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\Phi \left( \frac{\delta}{\kappa} \right)} \cdot \sqrt{\frac{2}{\pi}} \Phi \left( \frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\tau_0} \right) - \frac{\omega \kappa}{\Phi \left( \frac{\delta}{\kappa} \right)} \phi \left( \frac{\delta}{\kappa} \right) \cdot 2 \left\{ \Phi \left( \frac{\omega \delta}{\tau_0} \right) - \frac{1}{2} \right\}. \quad (4.14)$$

Inserting (4.11) and (4.14) into (4.12) yields the postulated result

$$\begin{aligned} \mathbb{E}[|\Lambda_{\text{full}}|] &= \left[ 2\tau_0 \phi \left( \frac{\omega \delta}{\tau_0} \right) + 2\omega \delta \left\{ \Phi \left( \frac{\omega \delta}{\tau_0} \right) - \frac{1}{2} \right\} \right] \left\{ 1 - \Phi \left( \frac{\delta}{\kappa} \right) \right\} \\ &\quad + \sqrt{\tau_0^2 + \omega^2 \kappa^2} \cdot \sqrt{\frac{2}{\pi}} \Phi \left( \frac{\delta}{\kappa} \cdot \frac{\sqrt{\tau_0^2 + \omega^2 \kappa^2}}{\tau_0} \right) - \omega \kappa \phi \left( \frac{\delta}{\kappa} \right) \cdot 2 \left\{ \Phi \left( \frac{\omega \delta}{\tau_0} \right) - \frac{1}{2} \right\}. \end{aligned}$$

### 4.7.5 Other focus parameters

In our simulation studies, we also assessed the performance of the  $FIC_{MAE}$  for several other focus parameters, such as quantiles of the survival distribution or the log-hazard and the survival function at different advanced ages. Overall, the frailty variance,  $\mu = \sigma^2$ , and the second derivative of the log-hazard,  $\mu = [\ln h(y)]''$ , yielded the best results. Figure 4.7 illustrates the proportion of decisions in favor of the gamma-Gompertz model in several settings, when the focus is placed on the second derivative of the log-hazard at age 100, the log-hazard at age 100 or 110, or the survival function at age 100. While the choice of  $\mu = [\ln h(100)]''$  again results in the highest proportion of decisions in favor of the gamma-Gompertz model, the choice of  $\mu = S(100)$  performs almost as well. When the focus is put on the log-hazard, the age at which the function is evaluated apparently makes a difference, in that  $\mu = \ln h(110)$  leads to a better performance of the  $FIC_{MAE}$  than  $\mu = \ln h(100)$ . However, survival beyond age 110 is relatively rare in some of our simulated settings, and we should be careful when putting the focus on ages for which there are too few data points.

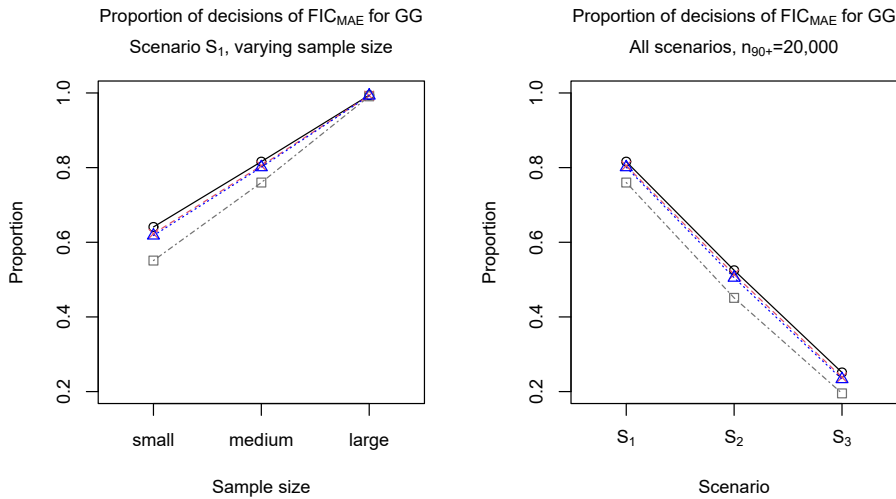


Figure 4.7: Proportion of decisions in favor of the gamma-Gompertz model based on  $FIC_{MAE}$  with  $\mu = [\ln h(100)]''$  (black-solid-circle),  $\mu = \ln h(100)$  (gray-dot-dashed-square),  $\mu = \ln h(110)$  (red-dashed-cross), and  $\mu = S(100)$  (blue-dotted-triangle). Left: Decisions in Scenario  $S_1$  for sample sizes  $n_{90+} = 10,000$ ,  $n_{90+} = 20,000$ , and  $n_{90+} = 105,000$  (left to right). Right: Decisions in Scenarios  $S_1$ ,  $S_2$ , and  $S_3$  (left to right) all with  $n_{90+} = 20,000$ .

### 4.7.6 Impact of the age range on the performance of the $FIC_{MAE}$

So far, we have, motivated by the data application, considered only samples of individuals who survived beyond age 90. However, the amount of heterogeneity in mortality risk within the population decreases with age due to selection. Therefore, it is of interest to study the performance of the  $FIC_{MAE}$  according to the age range of the sample. Figure 4.8 depicts the proportion of correct decisions in favor of the gamma-Gompertz model based on the  $FIC_{MAE}$  with  $\mu = [\ln h(100)]''$  in different settings when the sample consisted of all individuals who had reached at least age 80, 85, or 90. We see that, in general, the probability of detecting mortality deceleration increases if the sample covers a wider age range. For Scenario  $S_1$  with the target sample size of  $n_{90+} = 10,000$ , the proportion of correct decisions increases by more than a third if we observe all individuals who had reached at least age 85 instead of only those individuals who had reached at least age 90. Both the larger sample size of the 85+ subset and the greater amount of heterogeneity in the mortality risk of this subset played a part in this result.

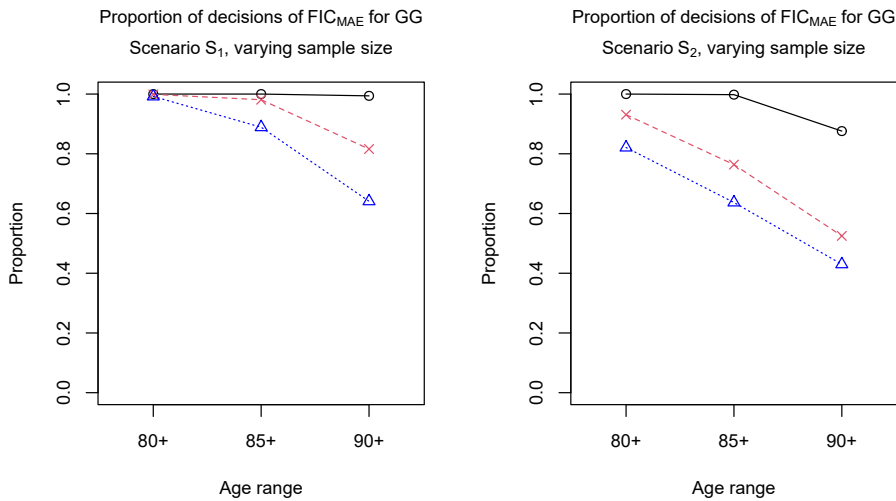


Figure 4.8: Proportion of correct decisions in favor of the gamma-Gompertz model based on the  $FIC_{MAE}$  with  $\mu = [\ln h(100)]''$  depending on the age range of the data (left to right: 80+, 85+, or 90+). The depicted scenarios are  $S_1$  (left) and  $S_2$  (right) with sample sizes  $n_{90+} = 10,000$  (blue-dotted-triangle),  $n_{90+} = 20,000$  (red-dashed-cross), and  $n_{90+} = 105,000$  (black-solid-circle).

### 4.7.7 Preliminary analysis of CHMD data

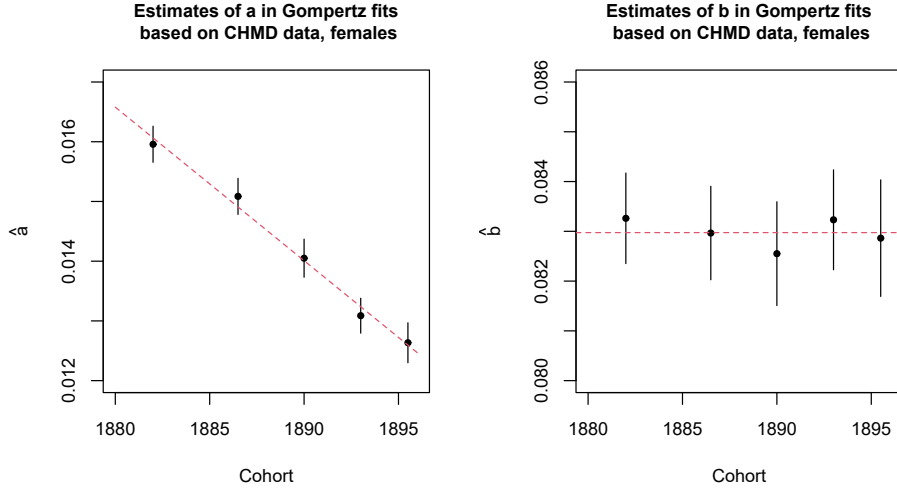


Figure 4.9: Estimates of Gompertz parameters  $a$  (left) and  $b$  (right) with 95% confidence intervals based on data for Quebec females between ages 60 and 109 from the CHMD. Fitted linear and constant trend, respectively, is marked by red, dashed line.

### 4.7.8 Derivation of $\text{FIC}_{\text{MAE}}$ for the model with cohort effects

We derive formulas for the  $\text{FIC}_{\text{MAE}}$  under the framework of local misspecification (4.3) for the setting that the parameter vector  $\boldsymbol{\gamma}$  has dimension  $q > 1$ , but only its first component is subject to a boundary constraint,  $\gamma_1 \geq \gamma_{01}$ .

A candidate model  $M$  includes the components  $\gamma_j$  for  $j \in M$ , while the remaining components, that is,  $\gamma_j$  for  $j \in M^c = \{1, \dots, q\} \setminus M$ , are fixed at the respective values in  $\boldsymbol{\gamma}_0$ . Thus, the focus estimator for model  $M$  is  $\hat{\boldsymbol{\mu}}_M = \boldsymbol{\mu}(\hat{\boldsymbol{\theta}}_M, \hat{\boldsymbol{\gamma}}_M, \boldsymbol{\gamma}_{0, M^c})$  with the maximum likelihood estimator  $(\hat{\boldsymbol{\theta}}_M^\top, \hat{\boldsymbol{\gamma}}_M^\top)^\top$  for model  $M$ . The  $\text{FIC}_{\text{MAE}}(M)$  of a model  $M$  is, as before, determined as an estimate of  $\mathbb{E}[|\Lambda_M|]$ , where  $\sqrt{n}(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu}_{\text{true}}) \xrightarrow{d} \Lambda_M$ . The limiting results will again involve  $\Lambda_0$  and  $\boldsymbol{\omega}$ , defined in Section 4.3.2, where  $\boldsymbol{\omega}$  is now a vector of length  $q$ . Similarly, we now consider normal random vectors

$$\boldsymbol{D} \sim \mathcal{N}_q(\boldsymbol{\delta}, \boldsymbol{Q}) \text{ and } \boldsymbol{E} = \boldsymbol{Q}^{-1}\boldsymbol{D} \sim \mathcal{N}_q(\boldsymbol{Q}^{-1}\boldsymbol{\delta}, \boldsymbol{Q}^{-1}),$$

independent of  $\Lambda_0$ , and the  $q \times q$  lower-right submatrix  $\boldsymbol{Q}$  of the inverse information matrix  $J_{\text{full}}^{-1}$ , which corresponds to  $\boldsymbol{\gamma}$ .

The limiting distribution of focus estimators  $\hat{\mu}_M$  under the framework (4.3) for the case that some or all of the components of  $\boldsymbol{\gamma}$  are subject to a boundary constraint is given in Theorem 10.2 of Claeskens and Hjort (2008) as

$$\sqrt{n}(\hat{\mu}_M - \mu_{\text{true}}) \xrightarrow{d} \Lambda_M = \Lambda_0 + \boldsymbol{\omega}^\top (\boldsymbol{\delta} - \pi_M^\top \hat{\boldsymbol{t}}_M),$$

where  $\hat{\boldsymbol{t}}_M$  is the random maximizer of  $\mathbf{E}_M^\top \boldsymbol{t} - \frac{1}{2} \boldsymbol{t}^\top \mathbf{Q}_M^{-1} \boldsymbol{t}$  over all  $\boldsymbol{t} \in \Omega_M$ . Here, the set  $\Omega_M \subset \mathbb{R}^{|M|}$  describes the parameter space of  $(\boldsymbol{\gamma}_M - \boldsymbol{\gamma}_{0,M})$  and  $\pi_M$  is a projection matrix of dimension  $|M| \times q$  mapping a vector  $\boldsymbol{v} = (v_1, v_2, \dots, v_q)^\top$  onto the vector  $\boldsymbol{v}_M$  that contains the  $|M|$  components  $v_j$  for  $j \in M$ . In particular, we have  $\mathbf{E}_M = \pi_M \mathbf{E} \sim \mathcal{N}_{|M|}(\pi_M \mathbf{Q}^{-1} \boldsymbol{\delta}, \mathbf{Q}_M^{-1})$  and  $\mathbf{Q}_M = (\pi_M \mathbf{Q}^{-1} \pi_M^\top)^{-1}$ .

In the following, we will consider different types of candidate models  $M$  and determine the limiting risk  $\mathbb{E}[|\Lambda_M|]$  based on the specific  $\Omega_M$  and the resulting  $\hat{\boldsymbol{t}}_M$ . The  $\text{FIC}_{\text{MAE}}$  is obtained by replacing the unknowns  $\boldsymbol{\delta}$ ,  $\boldsymbol{\omega}$ ,  $\tau_0^2$ , and  $\mathbf{Q}$  in the formulas for  $\mathbb{E}[|\Lambda_M|]$  by their estimates.

For some choices of the focus parameter  $\mu$ , the derivations given below need to be adapted. If, for example,  $\mu = \sigma_c^2$ , it follows that  $\frac{\partial \mu}{\partial \boldsymbol{\theta}} = \mathbf{0}$ , such that  $\tau_0^2 = 0$  and the variable  $\Lambda_0$  is deterministic.

### $M = \emptyset$

If  $M = \emptyset$ , the focus estimator  $\hat{\mu}_{\text{null}}$  converges to a normal distribution,  $\Lambda_M = (\Lambda_0 + \boldsymbol{\omega}^\top \boldsymbol{\delta}) \sim \mathcal{N}(\boldsymbol{\omega}^\top \boldsymbol{\delta}, \tau_0^2)$ . Hence,  $\mathbb{E}[|\Lambda_M|]$  is calculated as the expected value of the folded normal random variable  $|\Lambda_0 + \boldsymbol{\omega}^\top \boldsymbol{\delta}|$ , so

$$\mathbb{E}[|\Lambda_\emptyset|] = 2\tau_0 \phi\left(\frac{\boldsymbol{\omega}^\top \boldsymbol{\delta}}{\tau_0}\right) + 2\boldsymbol{\omega}^\top \boldsymbol{\delta} \left[ \Phi\left(\frac{\boldsymbol{\omega}^\top \boldsymbol{\delta}}{\tau_0}\right) - \frac{1}{2} \right]. \quad (4.15)$$

### $M \neq \mathbf{1}$

If  $\mathbf{1} \notin M$ , the parameter vector  $\boldsymbol{\gamma}_M$  is not subject to boundary constraints, such that  $\Omega_M = \mathbb{R}^{|M|}$  and  $\hat{\boldsymbol{t}}_M = \mathbf{Q}_M \mathbf{E}_M$ . The limiting variable  $\Lambda_M = \Lambda_0 + \boldsymbol{\omega}^\top (\boldsymbol{\delta} - \pi_M^\top \mathbf{Q}_M \mathbf{E}_M)$  is a linear transformation of the normal random vector  $(\Lambda_0, (\mathbf{Q}_M \mathbf{E}_M)^\top)^\top$  and is therefore normally distributed with mean  $m_M = \boldsymbol{\omega}^\top (I - V_M \mathbf{Q}^{-1}) \boldsymbol{\delta}$  and variance  $\tau_M^2 = \tau_0^2 + \boldsymbol{\omega}^\top V_M \boldsymbol{\omega}$ , where  $V_M = \pi_M^\top \mathbf{Q}_M \pi_M$ . Thus, the limiting  $L_1$ -risk of  $\Lambda_M$  is again found as the mean of a folded normal distribution,

$$\mathbb{E}[|\Lambda_M|] = 2\tau_M \phi\left(\frac{m_M}{\tau_M}\right) + 2m_M \left[ \Phi\left(\frac{m_M}{\tau_M}\right) - \frac{1}{2} \right].$$

$M = \{1\}$

If  $M = \{1\}$ , that is, the model  $M$  includes only the boundary parameter  $\gamma_1$ , the parameter space is restricted to the non-negative real numbers,  $\Omega_{\{1\}} = \mathbb{R}_0^+$ . Therefore, we have  $\hat{\imath}_M = 0 \cdot \mathbb{1}_{\{Q_M E_M \leq 0\}} + Q_M E_M \cdot \mathbb{1}_{\{Q_M E_M > 0\}}$ . The limiting variable is then given by

$$\Lambda_{\{1\}} = \Lambda_0 + \omega^\top \delta \cdot \mathbb{1}_{\{Q_M E_M \leq 0\}} + \omega^\top (\delta - \pi_M^\top Q_M E_M) \cdot \mathbb{1}_{\{Q_M E_M > 0\}},$$

and the risk  $\mathbb{E}[|\Lambda_{\{1\}}|]$  can be calculated as

$$\begin{aligned} \mathbb{E}[|\Lambda_{\{1\}}| \mid Q_M E_M \leq 0] & \mathbb{P}[Q_M E_M \leq 0] + \mathbb{E}[|\Lambda_{\{1\}}| \mid Q_M E_M > 0] \mathbb{P}[Q_M E_M > 0] \\ & = \mathbb{E}[|\Lambda_0 + \omega^\top \delta|] \mathbb{P}[Q_M E_M \leq 0] \\ & \quad + \mathbb{E}[|\Lambda_0 + \omega^\top (\delta - \pi_M^\top Q_M E_M)| \mid Q_M E_M > 0] \mathbb{P}[Q_M E_M > 0]. \end{aligned}$$

The first expectation was given in (4.15) and the computation of the second expectation follows along the lines of the computation of the second expectation in (4.12), see Section 4.7.4. We consider the normal random vector  $Y = (\Lambda_0 + \omega^\top (\delta - \pi_M^\top Q_M E_M), Q_M E_M)^\top$ , which is a linear transformation of  $X = (\Lambda_0, Q_M E_M)^\top$ , and note that  $\mathbb{E}[|\Lambda_0 + \omega^\top (\delta - \pi_M^\top Q_M E_M)| \mid Q_M E_M > 0]$  equals  $\mathbb{E}[|Y_1| \mid Y_2 > 0]$  and thus,

$$\mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0] \mathbb{P}[Y_1 > 0 \mid Y_2 > 0] + \mathbb{E}[-Y_1 \mid -Y_1 > 0, Y_2 > 0] \mathbb{P}[Y_1 \leq 0 \mid Y_2 > 0].$$

In order to apply the results of Tallis (1961) on the expected values of the components of bivariate truncated normal distributions with zero mean vector and unit variances, we work with the transformed  $Z$  with components  $Z_k = \tau_{Y_k}^{-1}(Y_k - m_{Y_k})$ , where  $m_{Y_k} = \mathbb{E}[Y_k]$  and  $\tau_{Y_k}^2 = \text{var}[Y_k]$ ,  $k = 1, 2$ . This yields

$$\begin{aligned} \mathbb{E}[Y_1 \mid Y_1 > 0, Y_2 > 0] & = \tau_{Y_1} \mathbb{E}[Z_1 \mid Z_1 > u_1, Z_2 > u_2] + m_{Y_1} \\ & = \tau_{Y_1} \frac{\phi(u_1) \Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) + \rho \phi(u_2) \Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right)}{\mathbb{P}[Y_1 > 0, Y_2 > 0]} + m_{Y_1}, \end{aligned}$$

where  $u_k = -m_{Y_k} / \tau_{Y_k}$ ,  $\rho = \rho_Y \tau_{Y_1}^{-1} \tau_{Y_2}^{-1}$ , and  $\rho_Y = \text{cov}(Y_1, Y_2)$ , and, analogously,

$$\mathbb{E}[-Y_1 \mid -Y_1 > 0, Y_2 > 0] = \tau_{Y_1} \frac{\phi(-u_1) \Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) - \rho \phi(u_2) \Phi\left(\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right)}{\mathbb{P}[Y_1 \leq 0, Y_2 > 0]} - m_{Y_1}.$$

Combining these two results and using the symmetry of the normal distribution as well as that  $\mathbb{P}[Y_1 \leq 0 \mid Y_2 > 0] = 1 - \mathbb{P}[Y_1 > 0 \mid Y_2 > 0]$ , we find that  $\mathbb{E}[|\Lambda_0 + \omega^\top (\delta - \pi_M^\top Q_M E_M)| \mid Q_M E_M > 0]$  is equal to

$$\begin{aligned} & \frac{\tau_{Y_1}}{\mathbb{P}[Y_2 > 0]} \left\{ 2\phi(u_1) \Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) + \rho \phi(u_2) \left[ 2\Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right) - 1 \right] \right\} \\ & \quad + m_{Y_1} (2\mathbb{P}[Y_1 > 0 \mid Y_2 > 0] - 1). \end{aligned}$$

Consequently, the limiting risk  $\mathbb{E}[|\Lambda_{\{1\}}|]$  is found to be

$$\begin{aligned} \mathbb{E}[|\Lambda_{\{1\}}|] &= \left\{ 2\tau_0\phi\left(\frac{\omega^\top\delta}{\tau_0}\right) + 2\omega^\top\delta \left[ \Phi\left(\frac{\omega^\top\delta}{\tau_0}\right) - \frac{1}{2} \right] \right\} \mathbb{P}[Y_2 \leq 0] \\ &\quad + 2\tau_{Y_1} \left\{ \phi(u_1)\Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) + \rho\phi(u_2) \left[ \Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right) - \frac{1}{2} \right] \right\} \\ &\quad + 2m_{Y_1} \left( \mathbb{P}[Y_1 > 0 \mid Y_2 > 0] - \frac{1}{2} \right) \mathbb{P}[Y_2 > 0]. \end{aligned}$$

$M \supseteq \{1\}$

In the last case, the model  $M$  includes the boundary parameter  $\gamma_1$ , but also some or all of the remaining components of  $\gamma$ . The maximizer  $\hat{t}_M$  of  $\mathbf{E}_M^\top \mathbf{t} - \frac{1}{2} \mathbf{t}^\top \mathbf{Q}_M^{-1} \mathbf{t}$  has to be determined over the parameter space  $\Omega_M = \mathbb{R}_0^+ \times \mathbb{R}^{|M|-1}$ . If the first component of  $\mathbf{Q}_M \mathbf{E}_M$  is positive, then  $\hat{t}_M = \mathbf{Q}_M \mathbf{E}_M$ ; but if the first component of  $\mathbf{Q}_M \mathbf{E}_M$  is negative or zero, then  $\hat{t}_1$  is set to zero and the remaining components of  $\hat{t}_M$  maximize  $\mathbf{E}_M^\top \mathbf{t} - \frac{1}{2} \mathbf{t}^\top \mathbf{Q}_M^{-1} \mathbf{t}$  for  $t_1 = 0$ . More formally, we have

$$\hat{t}_M = \mathbf{G}_M^\top \mathbf{R}_M \mathbf{G}_M \mathbf{E}_M \cdot \mathbb{1}_{\{F_M \mathbf{Q}_M \mathbf{E}_M \leq 0\}} + \mathbf{Q}_M \mathbf{E}_M \cdot \mathbb{1}_{\{F_M \mathbf{Q}_M \mathbf{E}_M > 0\}},$$

where we defined a  $1 \times |M|$  projection matrix  $F_M$  mapping a vector of length  $|M|$  onto its first component, and a  $(|M| - 1) \times |M|$  projection matrix  $G_M$  mapping a vector of length  $|M|$  onto all but its first component, as well as  $\mathbf{R}_M = (\mathbf{G}_M \mathbf{Q}_M^{-1} \mathbf{G}_M^\top)^{-1}$ . The limiting variable takes the form

$$\begin{aligned} \Lambda_M &= \Lambda_0 + \omega^\top (\delta - \pi_M^\top \mathbf{G}_M^\top \mathbf{R}_M \mathbf{G}_M \mathbf{E}_M) \cdot \mathbb{1}_{\{F_M \mathbf{Q}_M \mathbf{E}_M \leq 0\}} \\ &\quad + \omega^\top (\delta - \pi_M^\top \mathbf{Q}_M \mathbf{E}_M) \cdot \mathbb{1}_{\{F_M \mathbf{Q}_M \mathbf{E}_M > 0\}}, \end{aligned}$$

such that the limiting risk of the focus estimator  $\hat{\mu}_M$  is given by

$$\begin{aligned} \mathbb{E}[|\Lambda_M|] &= \mathbb{E}[|\Lambda_0 + \omega^\top (\delta - \pi_M^\top \mathbf{G}_M^\top \mathbf{R}_M \mathbf{G}_M \mathbf{E}_M)| \mid F_M \mathbf{Q}_M \mathbf{E}_M \leq 0] \mathbb{P}[F_M \mathbf{Q}_M \mathbf{E}_M \leq 0] \\ &\quad + \mathbb{E}[|\Lambda_0 + \omega^\top (\delta - \pi_M^\top \mathbf{Q}_M \mathbf{E}_M)| \mid F_M \mathbf{Q}_M \mathbf{E}_M > 0] \mathbb{P}[F_M \mathbf{Q}_M \mathbf{E}_M > 0]. \end{aligned}$$

For computing the first expectation, we consider the normal random vector

$$\mathbf{X} = \begin{pmatrix} \Lambda_0 \\ \mathbf{E}_M \end{pmatrix} \sim \mathcal{N}_{1+|M|} \left( \begin{pmatrix} 0 \\ \pi_M \mathbf{Q}_M^{-1} \delta \end{pmatrix}, \begin{pmatrix} \tau_0^2 & 0 \\ 0 & \mathbf{Q}_M^{-1} \end{pmatrix} \right),$$

and its linear transformation

$$\mathbf{Y} = \begin{pmatrix} \Lambda_0 + \omega^\top (\delta - \pi_M^\top \mathbf{G}_M^\top \mathbf{R}_M \mathbf{G}_M \mathbf{E}_M) \\ F_M \mathbf{Q}_M \mathbf{E}_M \end{pmatrix} \sim \mathcal{N}_2 \left( \begin{pmatrix} m_{Y_1} \\ m_{Y_2} \end{pmatrix}, \begin{pmatrix} \tau_{Y_1}^2 & \rho_Y \\ \rho_Y & \tau_{Y_2}^2 \end{pmatrix} \right).$$

One can show that  $\rho_Y = -\omega^\top \pi_M^\top \mathbf{G}_M^\top \mathbf{R}_M \mathbf{G}_M \mathbf{F}_M^\top = 0$ . Consequently, the components of  $\mathbf{Y}$  are jointly normally distributed and uncorrelated, and hence, independent. The first



expectation in the expression for  $\mathbb{E}[|\Lambda_M|]$  therefore simplifies to  $\mathbb{E}[|Y_1|]$ , which is the mean of a folded normal random variable, that is,

$$\mathbb{E}[|\Lambda_0 + \omega^\top (\boldsymbol{\delta} - \pi_M^\top G_M^\top R_M G_M \mathbf{E}_M)|] = 2\tau_{Y_1} \phi\left(\frac{m_{Y_1}}{\tau_{Y_1}}\right) + 2m_{Y_1} \left[ \Phi\left(\frac{m_{Y_1}}{\tau_{Y_1}}\right) - \frac{1}{2} \right].$$

For computing the second expectation, we define another linear transformation of  $\mathbf{X}$ , namely,

$$\mathbf{W} = \begin{pmatrix} \Lambda_0 + \omega^\top (\boldsymbol{\delta} - \pi_M^\top Q_M \mathbf{E}_M) \\ F_M Q_M \mathbf{E}_M \end{pmatrix} \sim \mathcal{N}_2 \left( \begin{pmatrix} m_{W_1} \\ m_{W_2} \end{pmatrix}, \begin{pmatrix} \tau_{W_1}^2 & \rho_W \\ \rho_W & \tau_{W_2}^2 \end{pmatrix} \right),$$

such that

$$\mathbb{E}[|\Lambda_0 + \omega^\top (\boldsymbol{\delta} - \pi_M^\top Q_M \mathbf{E}_M)| \mid F_M Q_M \mathbf{E}_M > 0] = \mathbb{E}[|W_1| \mid W_2 > 0].$$

Using once again the law of total expectation, the latter can be calculated as

$$\begin{aligned} \mathbb{E}[W_1 \mid W_1 > 0, W_2 > 0] \mathbb{P}[W_1 > 0 \mid W_2 > 0] \\ + \mathbb{E}[-W_1 \mid -W_1 > 0, W_2 > 0] \mathbb{P}[W_1 \leq 0 \mid W_2 > 0]. \end{aligned}$$

We rewrite these expectations in terms of a centered random vector  $\mathbf{Z}$  with unit variances, where  $Z_k = \tau_{W_k}^{-1} (W_k - m_{W_k})$ ,  $k = 1, 2$ , to apply the results of Tallis (1961),

$$\begin{aligned} \mathbb{E}[W_1 \mid W_1 > 0, W_2 > 0] &= \tau_{W_1} \mathbb{E}[Z_1 \mid Z_1 > u_1, Z_2 > u_2] + m_{W_1} \\ &= \tau_{W_1} \frac{\phi(u_1) \Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) + \rho \phi(u_2) \Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right)}{\mathbb{P}[W_1 > 0, W_2 > 0]} + m_{W_1}, \end{aligned}$$

with  $u_k = -m_{W_k} / \tau_{W_k}$  and  $\rho = \rho_W \tau_{W_1}^{-1} \tau_{W_2}^{-1}$ , and

$$\mathbb{E}[-W_1 \mid -W_1 > 0, W_2 > 0] = \tau_{W_1} \frac{\phi(u_1) \Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) - \rho \phi(u_2) \Phi\left(\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right)}{\mathbb{P}[W_1 \leq 0, W_2 > 0]} - m_{W_1}.$$

Combining these two expressions, we obtain that  $\mathbb{E}[|W_1| \mid W_2 > 0]$  equals

$$\begin{aligned} \frac{\tau_{W_1}}{\mathbb{P}[W_2 > 0]} \left\{ 2\phi(u_1) \Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) + \rho \phi(u_2) \left[ 2\Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right) - 1 \right] \right\} \\ + m_{W_1} (2\mathbb{P}[W_1 > 0 \mid W_2 > 0] - 1). \end{aligned}$$

Finally, the above results lead to

$$\begin{aligned} \mathbb{E}[|\Lambda_M|] &= \left\{ 2\tau_{Y_1} \phi\left(\frac{m_{Y_1}}{\tau_{Y_1}}\right) + 2m_{Y_1} \left[ \Phi\left(\frac{m_{Y_1}}{\tau_{Y_1}}\right) - \frac{1}{2} \right] \right\} \mathbb{P}[W_2 \leq 0] \\ &\quad + 2\tau_{W_1} \left\{ \phi(u_1) \Phi\left(-\frac{u_2 - \rho u_1}{\sqrt{1 - \rho^2}}\right) + \rho \phi(u_2) \left[ \Phi\left(-\frac{u_1 - \rho u_2}{\sqrt{1 - \rho^2}}\right) - \frac{1}{2} \right] \right\} \\ &\quad + 2m_{W_1} \left( \mathbb{P}[W_1 > 0 \mid W_2 > 0] - \frac{1}{2} \right) \mathbb{P}[W_2 > 0]. \end{aligned}$$

### 4.7.9 A modified AIC for the model with cohort effects

In this section, we study the AIC for the (gamma-)Gompertz models with cohort effects which are listed in Table 4.3. We show that the standard AIC is biased as an estimator of the Akaike information for the models that include the boundary parameter  $\sigma_0^2$  and define a modified AIC\* with a bias correction term.

#### The modified AIC

We work within the framework of local misspecification (4.3) for the setting that the parameter vector  $\boldsymbol{\gamma}$  has dimension  $q > 1$ , and only its first component is subject to a boundary constraint,  $\gamma_1 \geq \gamma_{01}$ . The candidate models  $M$  always include the full  $d$ -dimensional parameter vector  $\boldsymbol{\theta}$ , but only those components  $\gamma_j$  of  $\boldsymbol{\gamma}$  for which  $j \in M$ . The standard AIC for such a model  $M$  is defined as

$$\text{AIC}_M = -2\ell_M(\hat{\boldsymbol{\eta}}_M; Y) + 2(d + |M|),$$

where the log-likelihood  $\ell_M$  for a sample  $Y$  under model  $M$  is evaluated at the maximum likelihood estimate under model  $M$ , that is,  $\hat{\boldsymbol{\eta}}_M = (\hat{\boldsymbol{\theta}}_M^\top, \hat{\boldsymbol{\gamma}}_M^\top)^\top$ . Let us denote by  $\pi_M$  a  $|M| \times q$  projection matrix mapping a vector  $\mathbf{v} = (v_1, v_2, \dots, v_q)^\top$  onto the vector  $\mathbf{v}_M$  that contains the  $|M|$  components  $v_j$  for  $j \in M$ , and by  $Q$  the  $q \times q$  lower-right submatrix of the information matrix  $J_{\text{full}}^{-1}$ , which corresponds to  $\boldsymbol{\gamma}$ , and finally  $Q_M = (\pi_M Q^{-1} \pi_M^\top)^{-1}$ . We will find that for models  $M$  that include the boundary parameter  $\gamma_1 = \sigma_0^2$ , the standard AIC has bias  $2\Phi\left(-[Q_M]_{11}^{-1/2} [Q_M \pi_M Q^{-1} \boldsymbol{\delta}]_1\right)$ , where  $[Q_M]_{11}$  is the top-left element of  $Q_M$  and  $[Q_M \pi_M Q^{-1} \boldsymbol{\delta}]_1$  is the first component of  $Q_M \pi_M Q^{-1} \boldsymbol{\delta}$ . We therefore define a modified version of the AIC for the models  $M$ , listed in Table 4.3, as

$$\text{AIC}_M^* = \begin{cases} -2\ell_M(\hat{\boldsymbol{\eta}}_M; Y) + 2(d + |M|), & \text{if } M \not\ni 1, \\ -2\ell_M(\hat{\boldsymbol{\eta}}_M; Y) + 2(d + |M|) - 2\Phi\left(-[\hat{Q}_M]_{11}^{-1/2} [\hat{Q}_M \pi_M \hat{Q}^{-1} \hat{\boldsymbol{\delta}}]_1\right), & \text{if } M \ni 1. \end{cases} \quad (4.16)$$

#### Outline of proof

We give a sketch of the derivations leading to the modified AIC in (4.16) here and provide further details on pp. 101 ff.

The AIC of a model  $M$  is derived as an asymptotically unbiased estimator of the expected relative Kullback-Leibler distance

$$-2 \mathbb{E}_Y [\mathbb{E}_X [\ln f(X; \hat{\eta}_M(Y))]], \quad (4.17)$$

which measures the distance between the true underlying distribution from which  $X$  and  $Y$  are generated and the best parametric approximation  $f(\cdot, \hat{\eta}_M)$  (Akaike, 1974). An unbiased estimator of the Akaike information (4.17) is given by

$$\begin{aligned} & -2 \ell_M(\hat{\eta}_M; Y) + 2 \mathbb{E}_Y [\ell_M(\hat{\eta}_M; Y) - \ell_M(\eta_{0,M}; Y)] \\ & \quad \underbrace{\hspace{10em}}_{=:A_1} \\ & + 2 \mathbb{E}_Y [\mathbb{E}_X [\ell_M(\eta_{0,M}; X) - \ell_M(\hat{\eta}_M; X)]], \quad (4.18) \\ & \quad \underbrace{\hspace{10em}}_{=:A_2} \end{aligned}$$

where the MLE  $\hat{\eta}_M = \hat{\eta}_M(Y)$  is based on the sample  $Y$  of size  $n$ . Using a Taylor expansion of the log-likelihood of model  $M$  evaluated at the MLE  $\hat{\eta}_M$  about the null point  $\eta_{0,M} = (\theta_0^\top, \gamma_{0,M}^\top)^\top$ , the term  $A_2$  is found to be asymptotically equivalent to

$$\begin{aligned} & -2 \mathbb{E}_Y [\underbrace{\sqrt{n}(\hat{\eta}_M - \eta_{0,M})^\top}_{=:A_{21}}] \left( \begin{array}{c} J_{01} \delta \\ \pi_M J_{11} \delta \end{array} \right) + \mathbb{E}_Y [\underbrace{\sqrt{n}(\hat{\eta}_M - \eta_{0,M})^\top J_M \sqrt{n}(\hat{\eta}_M - \eta_{0,M})}_{=:A_{22}}]. \quad (4.19) \end{aligned}$$

Here,  $J_M$  is the information matrix for model  $M$ , with blocks  $J_{00}$ ,  $J_{01,M} = J_{01} \pi_M^\top$ ,  $J_{10,M} = \pi_M J_{10}$ , and  $J_{11,M} = \pi_M J_{11} \pi_M^\top$ . The derivation of result (4.19) involves the limiting distribution of the score vector under model  $M$ . Let us denote by  $\mathbf{U}(y)$  and  $\mathbf{V}(y)$  the score functions with respect to  $\theta$  and  $\gamma$  of the log-likelihood of a single observation  $y$  from  $f_{\text{true}}$  in (4.3), where  $\mathbf{U}$  and  $\mathbf{V}$  are evaluated at the null model  $(\theta_0^\top, \gamma_0^\top)^\top$ . For a sample  $Y$  of size  $n$ , the averaged score vectors are  $\bar{\mathbf{U}}_n = n^{-1} \sum_{i=1}^n \mathbf{U}(Y_i)$  and  $\bar{\mathbf{V}}_n = n^{-1} \sum_{i=1}^n \mathbf{V}(Y_i)$ . According to the multivariate central limit theorem, and under the framework (4.3), the score vector of model  $M$  converges in distribution to a normal random vector (see Hjort and Claeskens, 2003, for a proof in the regular setting that can be seen to carry over to the current setting),

$$\left( \begin{array}{c} \sqrt{n} \bar{\mathbf{U}}_n \\ \sqrt{n} \bar{\mathbf{V}}_n \end{array} \right) \xrightarrow{d} \left( \begin{array}{c} J_{01} \delta \\ \pi_M J_{11} \delta \end{array} \right) + \left( \begin{array}{c} \mathbf{K} \\ \mathbf{N}_M \end{array} \right), \quad \text{with } \left( \begin{array}{c} \mathbf{K} \\ \mathbf{N}_M \end{array} \right) \sim \mathcal{N}_{d+|M|}(\mathbf{0}, J_M). \quad (4.20)$$

Further approximations of the expressions  $A_{21}$  and  $A_{22}$  in (4.19) rely on the limiting distribution of the MLE  $\hat{\eta}_M$  under model  $M$ . As stated in Theorem 10.2 in Claeskens and Hjort (2008), under the framework (4.3) in case some or all components of  $\gamma$  are subject to boundary constraints, the MLE tends to the following limiting variable,

$$\left( \begin{array}{c} \sqrt{n}(\hat{\theta}_M - \theta_0) \\ \sqrt{n}(\hat{\gamma}_M - \gamma_{0,M}) \end{array} \right) \xrightarrow{d} \left( \begin{array}{c} J_{00}^{-1} (J_{01} \delta + \mathbf{K} - J_{01,M} \hat{\mathbf{t}}_M) \\ \hat{\mathbf{t}}_M \end{array} \right), \quad (4.21)$$

where  $\hat{\boldsymbol{t}}_M$  is the random maximizer of  $\boldsymbol{E}_M^\top \boldsymbol{t} - \frac{1}{2} \boldsymbol{t}^\top \boldsymbol{Q}_M^{-1} \boldsymbol{t}$  over all  $\boldsymbol{t} \in \Omega_M$ , and  $\Omega_M$  is the parameter space of  $(\boldsymbol{\gamma}_M - \boldsymbol{\gamma}_{0,M})$ . Based on this, we obtain

$$\begin{aligned} A_{21} &\approx \boldsymbol{\delta}^\top J_{10} J_{00}^{-1} J_{01} \boldsymbol{\delta} + \mathbb{E}[\hat{\boldsymbol{t}}_M]^\top \pi_M \boldsymbol{Q}^{-1} \boldsymbol{\delta}, \\ A_{22} &\approx \boldsymbol{\delta}^\top J_{10} J_{00}^{-1} J_{01} \boldsymbol{\delta} + \mathbb{E}[\mathbf{K}^\top J_{00}^{-1} \mathbf{K}] + \mathbb{E}[\hat{\boldsymbol{t}}_M^\top \boldsymbol{Q}_M^{-1} \hat{\boldsymbol{t}}_M]. \end{aligned} \quad (4.22)$$

As for a  $k$ -dimensional random vector  $\boldsymbol{X}$  with  $\mathbb{E}[\boldsymbol{X}] = \boldsymbol{m}$ ,  $\text{Cov}[\boldsymbol{X}] = \Sigma$ , and a constant  $k \times k$  symmetric matrix  $\boldsymbol{B}$ , it holds that

$$\mathbb{E}[\boldsymbol{X}^\top \boldsymbol{B} \boldsymbol{X}] = \text{tr}(\boldsymbol{B} \Sigma) + \boldsymbol{m}^\top \boldsymbol{B} \boldsymbol{m}, \quad (4.23)$$

where  $\text{tr}(\cdot)$  denotes the trace of a matrix, we can rewrite  $A_{22}$  in (4.22) as

$$A_{22} \approx \boldsymbol{\delta}^\top J_{10} J_{00}^{-1} J_{01} \boldsymbol{\delta} + d + \text{tr}(\boldsymbol{Q}_M^{-1} \mathbb{E}[\hat{\boldsymbol{t}}_M \hat{\boldsymbol{t}}_M^\top]). \quad (4.24)$$

Regarding  $A_1$  in (4.18), it can be shown to be asymptotically equivalent to  $A_{22}$  in (4.19). Combining the above results yields

$$-2\ell_M(\hat{\boldsymbol{\eta}}_M; Y) + 2d - 2\mathbb{E}[\hat{\boldsymbol{t}}_M]^\top \pi_M \boldsymbol{Q}^{-1} \boldsymbol{\delta} + 2\text{tr}(\boldsymbol{Q}_M^{-1} \mathbb{E}[\hat{\boldsymbol{t}}_M \hat{\boldsymbol{t}}_M^\top]) \quad (4.25)$$

as an asymptotically unbiased estimator of the Akaike information (4.17) for model  $M$ . Similar to the derivations of the  $\text{FIC}_{\text{MAE}}$  in Section 4.7.8, we need to consider the different types of candidate models  $M$  with the specific  $\Omega_M$  and resulting  $\hat{\boldsymbol{t}}_M$  to determine more specific forms of (4.25).

### $M = \emptyset$

If  $M = \emptyset$ , then (4.25) simplifies to the standard AIC,

$$\text{AIC}_\emptyset^* = -2\ell_{\text{null}}(\hat{\boldsymbol{\eta}}_{\text{null}}; Y) + 2d.$$

### $M \not\equiv \mathbf{1}$

If  $\mathbf{1} \notin M$ , the parameter vector  $\boldsymbol{\gamma}_M$  is not subject to boundary constraints, such that  $\Omega_M = \mathbb{R}^{|M|}$  and  $\hat{\boldsymbol{t}}_M = \boldsymbol{Q}_M \boldsymbol{E}_M \sim \mathcal{N}_{|M|}(\boldsymbol{Q}_M \pi_M \boldsymbol{Q}^{-1} \boldsymbol{\delta}, \boldsymbol{Q}_M)$ . The term  $\text{tr}(\boldsymbol{Q}_M^{-1} \mathbb{E}[\hat{\boldsymbol{t}}_M \hat{\boldsymbol{t}}_M^\top])$  is then calculated as

$$\text{tr}(\boldsymbol{Q}_M^{-1} \text{Cov}[\hat{\boldsymbol{t}}_M]) + \text{tr}(\boldsymbol{Q}_M^{-1} \mathbb{E}[\hat{\boldsymbol{t}}_M] \mathbb{E}[\hat{\boldsymbol{t}}_M]^\top) = |M| + \boldsymbol{\delta}^\top \boldsymbol{Q}^{-1} \pi_M^\top \boldsymbol{Q}_M \pi_M \boldsymbol{Q}^{-1} \boldsymbol{\delta},$$

because the trace is a linear mapping and invariant under cyclic permutations. Hence, the asymptotically unbiased estimator (4.25) in this case again takes the form of the standard AIC,

$$\text{AIC}_M^* = -2\ell_M(\hat{\boldsymbol{\eta}}_M; Y) + 2(d + |M|).$$

$M = \{1\}$

If the model  $M$  includes only the boundary parameter  $\gamma_1$ , then

$$\hat{\boldsymbol{\tau}}_M = 0 \cdot \mathbb{1}_{\{Q_M E_M \leq 0\}} + Q_M E_M \cdot \mathbb{1}_{\{Q_M E_M > 0\}},$$

with the indicator function  $\mathbb{1}_{\{\cdot\}}$ , a scalar  $Q_M$ , and  $Q_M E_M \sim \mathcal{N}_1(Q_M \pi_M Q^{-1} \boldsymbol{\delta}, Q_M)$ . Consequently, the moments of  $\hat{\boldsymbol{\tau}}_M$  can be computed as

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{\tau}}_M] &= 0 \cdot \mathbb{P}[Q_M E_M \leq 0] + \mathbb{E}[Q_M E_M \mid Q_M E_M > 0] \cdot \mathbb{P}[Q_M E_M > 0], \\ \mathbb{E}[\hat{\boldsymbol{\tau}}_M^2] &= \mathbb{E}[(Q_M E_M)^2 \mid Q_M E_M > 0] \cdot \mathbb{P}[Q_M E_M > 0]. \end{aligned}$$

Applying the formulas for the moments of the truncated normal distribution, we find that

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{\tau}}_M] &= Q_M^{1/2} u [1 - \Phi(-u)] + Q_M^{1/2} \phi(-u), \\ \mathbb{E}[\hat{\boldsymbol{\tau}}_M^2] &= \left\{ Q_M \left[ 1 - \frac{u\phi(-u)}{1 - \Phi(-u)} - \left( \frac{\phi(-u)}{1 - \Phi(-u)} \right)^2 \right] + \left[ Q_M^{1/2} u + Q_M^{1/2} \frac{\phi(-u)}{1 - \Phi(-u)} \right]^2 \right\} \\ &\quad \cdot [1 - \Phi(-u)], \end{aligned}$$

where  $u = \mathbb{E}[Q_M E_M] / \sqrt{\text{var}[Q_M E_M]} = Q_M^{1/2} \pi_M Q^{-1} \boldsymbol{\delta}$ .

After inserting these expressions into formula (4.25), some straightforward calculations and replacing the unknowns by their estimates lead to

$$\text{AIC}_{\{1\}}^* = -2\ell_{\{1\}}(\hat{\boldsymbol{\eta}}_{\{1\}}; Y) + 2(d+1) - 2\Phi(-\hat{Q}_{\{1\}}^{1/2} \pi_{\{1\}} \hat{Q}^{-1} \hat{\boldsymbol{\delta}}).$$

$M \not\subseteq \{1\}$

If the model  $M$  includes the boundary parameter  $\gamma_1$  and some or all of the remaining components of  $\boldsymbol{\gamma}$ , then

$$\hat{\boldsymbol{\tau}}_M = G_M^\top R_M G_M E_M \cdot \mathbb{1}_{\{F_M Q_M E_M \leq 0\}} + Q_M E_M \cdot \mathbb{1}_{\{F_M Q_M E_M > 0\}},$$

where  $F_M$  is a  $1 \times |M|$  projection matrix mapping a vector of length  $|M|$  onto its first component,  $G_M$  is a  $(|M| - 1) \times |M|$  projection matrix mapping such a vector onto all but its first component, and  $R_M = (G_M Q_M^{-1} G_M^\top)^{-1}$ . The moments of  $\hat{\boldsymbol{\tau}}_M$  can again be determined based on the law of total expectation, that is,

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{\tau}}_M] &= \mathbb{E}[G_M^\top R_M G_M E_M \mid F_M Q_M E_M \leq 0] \cdot \mathbb{P}[F_M Q_M E_M \leq 0] \\ &\quad + \mathbb{E}[Q_M E_M \mid F_M Q_M E_M > 0] \cdot \mathbb{P}[F_M Q_M E_M > 0]. \end{aligned}$$

We first note that  $G_M^\top R_M G_M E_M$  and  $F_M Q_M E_M$  are independent (cf. p. 95), such that the condition  $F_M Q_M E_M \leq 0$  in the first expectation can be dropped. Moreover, from  $F_M Q_M E_M \sim \mathcal{N}_1(F_M Q_M \pi_M Q^{-1} \boldsymbol{\delta}, F_M Q_M F_M^\top)$ , we have  $\mathbb{P}[F_M Q_M E_M \leq 0] = \Phi(-u_M)$

with  $u_M = (F_M Q_M F_M^\top)^{-1/2} F_M Q_M \pi_M Q^{-1} \delta = [Q_M]_{11}^{-1/2} [Q_M \pi_M Q^{-1} \delta]_1$ . Finally, the moments of  $Q_M E_M$  conditional on  $F_M Q_M E_M > 0$  are found by applying the results of Gupta and Tracy (1978) on truncated trivariate normal distributions. In the end, we arrive at

$$\begin{aligned} \mathbb{E}[\hat{\mathbf{t}}_M] &= G_M^\top R_M G_M \pi_M Q^{-1} \delta \cdot \Phi(-u_M) \\ &\quad + Q_M \pi_M Q^{-1} \delta \cdot [1 - \Phi(-u_M)] + (F_M Q_M F_M^\top)^{-1/2} Q_M F_M^\top \phi(-u_M), \\ \mathbb{E}[\hat{\mathbf{t}}_M \hat{\mathbf{t}}_M^\top] &= G_M^\top R_M G_M \Phi(-u_M) + G_M^\top R_M G_M \pi_M Q^{-1} \delta \delta^\top Q^{-1} \pi_M^\top G_M^\top R_M G_M \Phi(-u_M) \\ &\quad + Q_M [1 - \Phi(-u_M)] - \frac{Q_M F_M^\top F_M Q_M}{F_M Q_M F_M^\top} u_M \phi(-u_M) \\ &\quad + 2Q_M \pi_M Q^{-1} \delta \frac{F_M Q_M}{(F_M Q_M F_M^\top)^{1/2}} \phi(-u_M) \\ &\quad + Q_M \pi_M Q^{-1} \delta \delta^\top Q^{-1} \pi_M^\top Q_M [1 - \Phi(-u_M)]. \end{aligned} \quad (4.26)$$

Inserting these expressions in (4.25), using some algebra, and replacing the unknowns by their estimates results in

$$\text{AIC}_M^* = -2\ell_M(\hat{\boldsymbol{\eta}}_M; Y) + 2(d + |M|) - 2\Phi\left(- (F_M \hat{Q}_M F_M^\top)^{-1/2} F_M \hat{Q}_M \pi_M \hat{Q}_M^{-1} \hat{\boldsymbol{\delta}}\right).$$

### Additional proofs and derivations

#### Proof of (4.19)

A Taylor expansion of the log-likelihood of model  $M$  at the MLE  $\hat{\boldsymbol{\eta}}_M$  about  $\boldsymbol{\eta}_{0,M}$  yields

$$\begin{aligned} A_2 &= 2 \mathbb{E}_Y [\mathbb{E}_X [-(\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})^\top n \bar{\mathbf{W}}_{n,M}(\boldsymbol{\eta}_{0,M}; X) \\ &\quad - \frac{1}{2} (\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})^\top \mathcal{H}_M(\boldsymbol{\eta}_{0,M}; X) (\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M}) + R_2]] \\ &\approx \mathbb{E}_Y [2\sqrt{n} (\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})^\top \mathbb{E}_X [-\sqrt{n} \bar{\mathbf{W}}_{n,M}(\boldsymbol{\eta}_{0,M}; X) \\ &\quad + (\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})^\top \mathbb{E}_X [-\mathcal{H}_M(\boldsymbol{\eta}_{0,M}; X)] (\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})], \end{aligned}$$

where  $\bar{\mathbf{W}}_{n,M} = (\bar{\mathbf{U}}_n^\top, \bar{\mathbf{V}}_{n,M}^\top)^\top$ ,  $\mathcal{H}_M$  is the Hessian matrix of the log-likelihood of model  $M$ , and  $R_2$  is a remainder term with  $R_2 \xrightarrow{n \rightarrow \infty} 0$ . Using the limiting distribution of the score (4.20) and that  $\mathbb{E}_X [-\mathcal{H}_M(\boldsymbol{\eta}_{0,M}; X)] = nJ_M$ , we obtain (4.19).

#### Proof of (4.22)

Using the limiting distribution of the MLE given in (4.21),  $A_{21}$  can be rewritten as,

$$\begin{aligned} A_{21} &\approx \mathbb{E} [(J_{01} \delta + \mathbf{K} - J_{01,M} \hat{\mathbf{t}}_M)^\top J_{00}^{-1}] J_{01} \delta + \mathbb{E} [\hat{\mathbf{t}}_M]^\top \pi_M J_{11} \delta \\ &= \delta^\top J_{10} J_{00}^{-1} J_{01} \delta + \mathbb{E} [\mathbf{K}]^\top J_{00}^{-1} J_{01} \delta + \mathbb{E} [\hat{\mathbf{t}}_M]^\top \pi_M (J_{11} - J_{10} J_{00}^{-1} J_{01}) \delta. \end{aligned}$$

Because  $\mathbb{E}[\mathbf{K}] = \mathbf{0}$  and noting that  $\mathbf{Q} = (J_{11} - J_{10}J_{00}^{-1}J_{01})^{-1}$ , this is the first part of (4.22). Similarly, we can approximate  $A_{22}$ ,

$$\begin{aligned}
A_{22} &\approx \mathbb{E} \left[ (J_{01}\boldsymbol{\delta} + \mathbf{K} - J_{01,M}\hat{\boldsymbol{\imath}}_M)^\top J_{00}^{-1}J_{00}J_{00}^{-1}(J_{01}\boldsymbol{\delta} + \mathbf{K} - J_{01,M}\hat{\boldsymbol{\imath}}_M) \right. \\
&\quad \left. + 2(J_{01}\boldsymbol{\delta} + \mathbf{K} - J_{01,M}\hat{\boldsymbol{\imath}}_M)^\top J_{00}^{-1}J_{01,M}\hat{\boldsymbol{\imath}}_M + \hat{\boldsymbol{\imath}}_M^\top J_{11,M}\hat{\boldsymbol{\imath}}_M \right] \\
&= \mathbb{E} \left[ (J_{01}\boldsymbol{\delta} + \mathbf{K})^\top J_{00}^{-1}(J_{01}\boldsymbol{\delta} + \mathbf{K}) \right] - 2\mathbb{E} \left[ (J_{01}\boldsymbol{\delta} + \mathbf{K})^\top J_{00}^{-1}J_{01,M}\hat{\boldsymbol{\imath}}_M \right] \\
&\quad + \mathbb{E}[\hat{\boldsymbol{\imath}}_M^\top J_{10,M}J_{00}^{-1}J_{01,M}\hat{\boldsymbol{\imath}}_M] + 2\mathbb{E} \left[ (J_{01}\boldsymbol{\delta} + \mathbf{K})^\top J_{00}^{-1}J_{01,M}\hat{\boldsymbol{\imath}}_M \right] \\
&\quad - 2\mathbb{E}[\hat{\boldsymbol{\imath}}_M^\top J_{10,M}J_{00}^{-1}J_{01,M}\hat{\boldsymbol{\imath}}_M] + \mathbb{E}[\hat{\boldsymbol{\imath}}_M^\top J_{11,M}\hat{\boldsymbol{\imath}}_M] \\
&= \boldsymbol{\delta}^\top J_{10}J_{00}^{-1}J_{01}\boldsymbol{\delta} + 2\boldsymbol{\delta}^\top J_{10}J_{00}^{-1}\mathbb{E}[\mathbf{K}] + \mathbb{E}[\mathbf{K}^\top J_{00}^{-1}\mathbf{K}] \\
&\quad + \mathbb{E}[\hat{\boldsymbol{\imath}}_M^\top (J_{11,M} - J_{10,M}J_{00}^{-1}J_{01,M})\hat{\boldsymbol{\imath}}_M],
\end{aligned}$$

which with  $\mathbb{E}[\mathbf{K}] = \mathbf{0}$  and  $\mathbf{Q}_M = (J_{11,M} - J_{10,M}J_{00}^{-1}J_{01,M})^{-1}$  is the second part of (4.22).

#### Proof of (4.24)

According to (4.23), we have

$$\begin{aligned}
\mathbb{E}[\mathbf{K}^\top J_{00}^{-1}\mathbf{K}] &= \text{tr}(J_{00}^{-1}J_{00}) + \mathbb{E}[\mathbf{K}]^\top J_{00}^{-1}\mathbb{E}[\mathbf{K}] = d, \\
\mathbb{E}[\hat{\boldsymbol{\imath}}_M^\top \mathbf{Q}_M^{-1}\hat{\boldsymbol{\imath}}_M] &= \text{tr}(\mathbf{Q}_M^{-1}\text{Cov}[\hat{\boldsymbol{\imath}}_M]) + \mathbb{E}[\hat{\boldsymbol{\imath}}_M]^\top \mathbf{Q}_M^{-1}\mathbb{E}[\hat{\boldsymbol{\imath}}_M] \\
&= \text{tr}(\mathbf{Q}_M^{-1}\mathbb{E}[\hat{\boldsymbol{\imath}}_M\hat{\boldsymbol{\imath}}_M^\top]) - \text{tr}(\mathbf{Q}_M^{-1}\mathbb{E}[\hat{\boldsymbol{\imath}}_M]\mathbb{E}[\hat{\boldsymbol{\imath}}_M]^\top) + \mathbb{E}[\hat{\boldsymbol{\imath}}_M]^\top \mathbf{Q}_M^{-1}\mathbb{E}[\hat{\boldsymbol{\imath}}_M],
\end{aligned}$$

where the last two expressions cancel because the trace is invariant under cyclic permutations and the last term is scalar.

#### Proof of the asymptotic equivalence of $A_1$ and $A_{22}$ in (4.19)

To show that  $A_1$  in (4.18) is asymptotically equivalent to  $A_{22}$  in (4.19), we consider the following Taylor expansion of the log-likelihood of model  $M$  at the null point about the MLE,

$$\begin{aligned}
\ell_M(\boldsymbol{\eta}_{0,M}; Y) &= \ell_M(\hat{\boldsymbol{\eta}}_M; Y) + (\boldsymbol{\eta}_{0,M} - \hat{\boldsymbol{\eta}}_M)^\top n\bar{\mathbf{W}}_{n,M}(\hat{\boldsymbol{\eta}}_M; Y) \\
&\quad + \frac{1}{2}(\boldsymbol{\eta}_{0,M} - \hat{\boldsymbol{\eta}}_M)^\top \mathcal{H}_M(\hat{\boldsymbol{\eta}}_M; Y)(\boldsymbol{\eta}_{0,M} - \hat{\boldsymbol{\eta}}_M) + R_1,
\end{aligned}$$

with a remainder term  $R_1 \xrightarrow{n \rightarrow \infty} 0$ . Then, we see that

$$\begin{aligned}
A_1 &= 2\mathbb{E}_Y[\ell_M(\hat{\boldsymbol{\eta}}_M; Y) - \ell_M(\boldsymbol{\eta}_{0,M}; Y)] \\
&= 2\mathbb{E}_Y[(\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})^\top n\bar{\mathbf{W}}_{n,M}(\hat{\boldsymbol{\eta}}_M; Y)] \\
&\quad + \mathbb{E}_Y[(\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M})^\top [-\mathcal{H}_M(\hat{\boldsymbol{\eta}}_M; Y)](\hat{\boldsymbol{\eta}}_M - \boldsymbol{\eta}_{0,M}) + R_1].
\end{aligned}$$

The term involving the score vector is equal to zero. If the MLE  $\hat{\eta}_M$  is an inner point of the parameter space, then by definition  $\bar{W}_{n,M}(\hat{\eta}_M; Y) = \mathbf{0}$ . If some components of the MLE, say  $\hat{\gamma}_j$  for  $j \in B \subset M$ , lie on the boundary of the parameter space, then  $\bar{U}_n(\hat{\eta}_M; Y) = \mathbf{0}$  and  $\bar{V}_{n,B^c}(\hat{\eta}_M; Y) = \mathbf{0}$ , while  $\bar{V}_{n,B}(\hat{\eta}_M; Y) \neq \mathbf{0}$ , but because  $(\hat{\gamma}_B - \gamma_{0,B}) = \mathbf{0}$ , we still have  $(\hat{\eta}_M - \eta_{0,M})^\top n \bar{W}_{n,M}(\hat{\eta}_M; Y) = 0$ .

Regarding the term involving the Hessian matrix, one can argue that  $-\mathcal{H}_M(\hat{\eta}_M; Y)$  reasonably well approximates  $nJ_M$  (see also Claeskens and Hjort, 2008, Section 6.5), which proves the assertion.

### Proof of (4.26)

The first lines in the expressions for  $\mathbb{E}[\hat{t}_M]$  and  $\mathbb{E}[\hat{t}_M \hat{t}_M^\top]$ , respectively, in (4.26) follow from  $G_M^\top R_M G_M \mathbf{E}_M \sim \mathcal{N}_{|M|}(G_M^\top R_M G_M \pi_M Q^{-1} \delta, G_M^\top R_M G_M)$ .

The remaining lines correspond to the product of  $\mathbb{P}[F_M Q_M \mathbf{E}_M > 0]$  and  $\mathbb{E}[Q_M \mathbf{E}_M | F_M Q_M \mathbf{E}_M > 0]$  or  $\mathbb{E}[(Q_M \mathbf{E}_M)(Q_M \mathbf{E}_M)^\top | F_M Q_M \mathbf{E}_M > 0]$ , respectively. The conditional expectations can be computed componentwise using the results of Gupta and Tracy (1978). For that purpose, we define random vectors  $X = ([Q_M \mathbf{E}_M]_1, [Q_M \mathbf{E}_M]_k, [Q_M \mathbf{E}_M]_l)^\top$  made up of the first,  $k^{\text{th}}$ , and  $l^{\text{th}}$  component of  $Q_M \mathbf{E}_M$ , for  $k, l \in M$ . Then, the  $X$  have a trivariate normal distribution with mean vector given by the corresponding components of  $Q_M \pi_M Q^{-1} \delta$  and covariance matrix equal to the corresponding submatrix of  $Q_M$ . The transformed vectors  $Y = \text{diag}([Q_M]_{11}^{-1/2}, [Q_M]_{kk}^{-1/2}, [Q_M]_{ll}^{-1/2})(X - \mathbb{E}[X])$  have a trivariate normal distribution with zero mean vector, unit variances, and correlations  $\rho_{st} = [Q_M]_{st} [Q_M]_{ss}^{-1/2} [Q_M]_{tt}^{-1/2}$ , for  $s, t \in \{1, k, l\}$ . The  $k^{\text{th}}$  component of  $\mathbb{E}[Q_M \mathbf{E}_M | F_M Q_M \mathbf{E}_M > 0]$  can then be expressed as

$$\mathbb{E}[X_2 | X_1 > 0] = [Q_M]_{kk}^{1/2} \mathbb{E}[Y_2 | Y_1 > -u_M] + [Q_M \pi_M Q^{-1} \delta]_k, \quad k \in M,$$

where still  $u_M = [Q_M]_{11}^{-1/2} [Q_M \pi_M Q^{-1} \delta]_1$ . According to the formulas in Section 4 of Gupta and Tracy (1978),

$$\mathbb{E}[Y_2 | Y_1 > -u_M] = \rho_{1k} \frac{\phi(-u_M)}{1 - \Phi(-u_M)},$$

such that

$$\mathbb{E}[Q_M \mathbf{E}_M | F_M Q_M \mathbf{E}_M > 0] = Q_M \pi_M Q^{-1} \delta + (F_M Q_M F_M^\top)^{-1/2} Q_M F_M^\top \frac{\phi(-u_M)}{1 - \Phi(-u_M)}.$$

Similarly, the components  $\mathbb{E}[[Q_M \mathbf{E}_M]_k [Q_M \mathbf{E}_M]_l | [Q_M \mathbf{E}_M]_1 > 0]$  of the conditional expectation of  $(Q_M \mathbf{E}_M)(Q_M \mathbf{E}_M)^\top$  correspond to  $\mathbb{E}[X_2 X_3 | X_1 > 0]$ , and hence, equal

$$\begin{aligned} & [Q_M]_{kk}^{1/2} [Q_M]_{ll}^{1/2} \mathbb{E}[Y_2 Y_3 | Y_1 > -u_M] + [Q_M]_{kk}^{1/2} \mathbb{E}[Y_2 | Y_1 > -u_M] \cdot [Q_M \pi_M Q^{-1} \delta]_l \\ & + [Q_M]_{ll}^{1/2} \mathbb{E}[Y_3 | Y_1 > -u_M] \cdot [Q_M \pi_M Q^{-1} \delta]_k + [Q_M \pi_M Q^{-1} \delta]_k \cdot [Q_M \pi_M Q^{-1} \delta]_l. \end{aligned}$$



Consulting again Section 4 in Gupta and Tracy (1978), to find

$$\mathbb{E}[Y_2 Y_3 \mid Y_1 > -u_M] = \rho_{kl} - u_M \rho_{1k} \rho_{1l} \frac{\phi(-u_M)}{1 - \Phi(-u_M)},$$

this leads to  $\mathbb{E}[(Q_M \mathbf{E}_M)(Q_M \mathbf{E}_M)^\top \mid F_M Q_M \mathbf{E}_M > 0]$  being equal to

$$Q_M - \frac{Q_M F_M^\top F_M Q_M}{F_M Q_M F_M^\top} \cdot u_M \frac{\phi(-u_M)}{1 - \Phi(-u_M)} \\ + 2Q_M \pi_M Q^{-1} \delta \frac{F_M Q_M}{(F_M Q_M F_M^\top)^{1/2}} \cdot \frac{\phi(-u_M)}{1 - \Phi(-u_M)} + Q_M \pi_M Q^{-1} \delta \delta^\top Q^{-1} \pi_M^\top Q_M.$$

## References

- Abbring, J. H. and G. J. van den Berg (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika* 94(1), 87–99.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723.
- Beard, R. E. (1959). Note on some mathematical mortality models. In G. E. W. Wolstenholme and M. O’Connor (Eds.), *The Lifespan of Animals*, Ciba Foundation Colloquia on Ageing, pp. 302–311. Boston: Little, Brown.
- Beaudry-Godin, M. (2010). *La démographie des centenaires québécois: validation des âges au décès, mesure de la mortalité et composante familiale de la longévité*. Ph. D. thesis, Université de Montréal.
- Böhnhstedt, M. and J. Gampe (2019). Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model. *Statistics and Probability Letters* 150, 68–73.
- Böhnhstedt, M., J. Gampe, and H. Putter (2021). Information measures and design issues in the study of mortality deceleration: findings for the gamma-Gompertz model. *Lifetime Data Analysis* 27, 333–356.
- Bourbeau, R. and B. Desjardins (2002). Dealing with problems in data quality for the measurement of mortality at advanced ages in Canada. *North American Actuarial Journal* 6(3), 1–13.
- Canadian Human Mortality Database (2020). Department of Demography, Université de Montréal (Canada). Available at [www.bdlc.umontreal.ca/CHMD/](http://www.bdlc.umontreal.ca/CHMD/) (data downloaded on March 30, 2020).
- Claeskens, G., C. Croux, and J. Van Kerckhoven (2006). Variable selection for logistic regression using a prediction-focused information criterion. *Biometrics* 62(4), 972–979.

- Claeskens, G. and N. L. Hjort (2003). The focused information criterion. *Journal of the American Statistical Association* 98, 900–916.
- Claeskens, G. and N. L. Hjort (2008). *Model Selection and Model Averaging*. New York: Cambridge University Press.
- Cunen, C., L. Walløe, and N. L. Hjort (2020). Focused model selection for linear mixed models with an application to whale ecology. *Annals of Applied Statistics* 14, 872–904.
- Desjardins, B. (1998). Le registre de la population du Québec ancien. *Annales de démographie historique* 2, 215–226.
- Duchateau, L. and P. Janssen (2008). *The Frailty Model*. New York: Springer.
- Feehan, D. M. (2018). Separating the signal from the noise: Evidence for deceleration in old-age death rates. *Demography* 55, 2025–2044.
- Gavrilov, L. A. and N. S. Gavrilova (2011). Mortality measurement at advanced ages: A study of the Social Security Administration Death Master File. *North American Actuarial Journal* 15(3), 432–447.
- Gavrilov, L. A. and N. S. Gavrilova (2019). New trend in old-age mortality: Gompertzialization of mortality trajectory. *Gerontology* 65, 451–457.
- Gavrilova, N. S. and L. A. Gavrilov (2015). Biodemography of old-age mortality in humans and rodents. *Journals of Gerontology, Series A* 70(1), 1–9.
- Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* 115, 513–583.
- Gupta, A. K. and D. S. Tracy (1978). Hermite polynomials and truncated trivariate normal distributions. *Journal of Statistical Computation and Simulation* 7, 269–286.
- Hjort, N. L. and G. Claeskens (2003). Frequentist model average estimators. *Journal of the American Statistical Association* 98, 879–899.
- Horiuchi, S. and J. R. Wilmoth (1998). Deceleration in the age pattern of mortality at older ages. *Demography* 35(4), 391–412.
- Jeune, B. and J. W. Vaupel (Eds.) (1999). *Validation of exceptional longevity*. Odense Monographs on Population Aging 6. Odense University Press.
- Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philosophical Transactions: Physical Sciences and Engineering* 332(1627), 487–509.
- Makeham, W. M. (1860). On the law of mortality and the construction of annuity tables. *The Assurance Magazine and Journal of the Institute of Actuaries* 8, 301–310.

- Newman, S. J. (2018). Errors as a primary cause of late-life mortality deceleration and plateaus. *PLOS Biology* 16(12), 1–12.
- Ouellette, N. (2016). La forme de la courbe de mortalité des centenaires canadiens-français. *Gérontologie et société* 38(151), 41–53.
- Ouellette, N. and R. Bourbeau (2014). Measurement of mortality among centenarians in Canada. In *Living to 100 Monograph*, pp. 17 pages. Society of Actuaries.
- Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries* 61(1), 12–57.
- Pletcher, S. D. (1999). Model fitting and hypothesis testing for age-specific mortality data. *Journal of Evolutionary Biology* 12, 430–439.
- Preston, S. H., I. T. Elo, and Q. Stewart (1999). Effects of age misreporting on mortality estimates at older ages. *Population Studies* 53, 165–177.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rau, R., E. Soroko, D. Jasilionis, and J. W. Vaupel (2008). Continued reductions in mortality at advanced ages. *Population and Development Review* 34(4), 747–768.
- Richards, S. J. (2008). Applying survival models to pensioner mortality data. *British Actuarial Journal* 14(2), 257–303.
- Richards, S. J., J. G. Kirkby, and I. D. Currie (2006). The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal* 12(1), 5–61.
- Self, S. G. and K.-Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 82(398), 605–610.
- Tallis, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society. Series B* 23(1), 223–229.
- Thatcher, A. R., V. Kannisto, and J. W. Vaupel (1998). *The force of mortality at ages 80 to 120*. Odense Monographs on Population Aging 5. Odense University Press.
- Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16(3), 439–454.
- Vaupel, J. W., F. Villavicencio, and M.-P. Bergeron-Boucher (2021). Demographic perspectives on the rise of longevity. *PNAS* 118(9). doi: 10.1073/pnas.2019536118.
- Wienke, A. (2011). *Frailty Models in Survival Analysis*. Biostatistics Series. Chapman & Hall/CRC.

## **Part II**

# **Joint modeling of recurrent events and death**



# 5

## Joint modeling of interval counts of recurrent events and death

### Abstract

When a recurrent event process is ended by death, this may imply dependent censoring if the two processes are associated. Such dependent censoring would have to be modeled to obtain a valid inference. Moreover, the dependence between the recurrence process and the terminal event may be the primary topic of interest. Joint frailty models for recurrent events and death, which include a separate dependence parameter, have been proposed for exactly observed recurrence times. However, in many situations, only the number of events experienced during consecutive time intervals are available. We propose a method for estimating a joint frailty model based on such interval counts and observed or independently censored terminal events. The baseline rates of the two processes are modeled by piecewise constant functions, and Gaussian quadrature is used to approximate the marginal likelihood. Covariates can be included in a proportional rates setting. The observation intervals for the recurrent event counts can differ between individuals. Furthermore, we adapt a score test for the association between recurrent events and death to the setting in which only individual interval counts are observed. We study the performance of both approaches via simulation studies, and exemplify the methodology in

---

This chapter has been published as: M. Böhnstedt, H. Putter, A. Daňko, M.J. Daňko, and J. Gampe (2021). Joint modeling of interval counts of recurrent events and death. *Biometrical Journal* 63, 323–340.

a biodemographic study of the dependence between budding rates and mortality in the species *Eleutheria dichotoma*.

## 5.1 Introduction

Studies of recurrent events, in which an individual can experience the same type of event repeatedly over time, are common in various fields of applications (Cook and Lawless, 2007). Examples range from medical studies of the recurrence of adverse symptoms, such as epileptic seizures, asthma attacks, or tumor relapse; to investigations of repeated insurance claims; to biodemographic studies of fertility (recurrent reproductive events) in particular animal species.

In some cases, the exact occurrence times can be observed, but often only the numbers of events that were experienced in specific time intervals are available. Such interval counts of recurrent events may result if, for example, patients only report the number of adverse events that they experience between two hospital visits, or the number of offspring produced by an animal is collected on a monthly basis only. In the latter example, the observation intervals would be the same for all individuals, while in the former example, we would expect the intervals between the two visits to vary from patient to patient.

The recurrent event process is often terminated by another event – most commonly by death – which usually cannot be assumed to be independent of the recurrent event process. Consequently, the terminal event introduces dependent censoring of the recurrence process, and this has to be taken into account to render a valid inference. Therefore, the two processes, the recurrent event process and the terminal event, have to be modeled jointly.

In many medical applications, the dependence of the two processes will be positive; that is, a higher recurrence rate of the (adverse) symptoms will be accompanied by a higher hazard of death. In other contexts, however, the direction of the association between the recurrence process and the terminal event is not clear at the outset, and will be a matter of interest in itself.

Our motivating example examines fertility and mortality in *Eleutheria dichotoma*, a marine hydrozoan for which the association between fertility and mortality has not previously been studied in detail. Reproduction and survival in *E. dichotoma* were investigated in a laboratory experiment for several months (Daňko et al., 2020). Individual survival times and the number of offspring that were produced by each individual within successive intervals of several days were recorded. The intervals resulted from the laboratory procedures and varied across individuals. These data were used to estimate the patterns of fertility and mortality over age, but the dependence between the two processes is also of biological interest.

It has been suggested that in some species, there is a trade-off between reproduction and survival. This idea is based on the assumption that individuals that produce a higher number of offspring are able to devote fewer resources to maintenance, and will, therefore, tend to die earlier. The claim that there is a cost of reproduction effect is in contrast to the

hypothesis that individuals who are stronger will be able to both produce more offspring and survive longer. Therefore, in addition to modeling the shape of age-specific fertility and mortality, the aim of the analysis is to find out which of the two explanations the data on *E. dichotoma* support. Thus, we will model fertility and mortality jointly to assess how the two processes are related in this species.

Several approaches to jointly modeling recurrent events and death have been proposed. We focus here on the joint frailty model introduced by Liu et al. (2004), because it allows for positive and negative associations between the recurrences and death. The dependence between the two processes is modeled by a shared individual random effect that acts on both the rate of recurrence and the hazard of death, possibly in different directions. In other frailty models, frailty has the same effect on the recurrence rate and the hazard of death. Thus, these models are restricted to a positive association (see Huang and Wang (2004) in the setting with exact recurrence times, or Lancaster and Intrator (1998) in the setting with interval counts). As marginal models for recurrent events in the presence of death leave the dependence between the two processes unspecified, they are not suitable for our purposes (see Cook and Lawless (1997) and Ghosh and Lin (2003) in the setting with exact recurrence times, or Zhao et al. (2013) in the setting with interval counts). Sinha and Maiti (2004) proposed a model similar to that of Liu et al. (2004), which is based on interval counts, but assumes that observation intervals are the same for all individuals, and that the termination time is discrete.

Estimation of the joint frailty model introduced by Liu et al. (2004) has so far only been based on observed recurrence times. For this setting, several methods of estimation have been developed: Liu et al. (2004) used a Monte Carlo EM algorithm, whereas Liu and Huang (2008) and Rondeau et al. (2007) applied Gaussian quadrature to the marginal likelihood. Moreover, a test for the association between recurrent events and a terminal event in the joint frailty model was derived by Balan et al. (2016), which was also based on observed recurrence times. It builds on concepts that are similar to the test proposed by Huang et al. (2004) for the association between two event processes in clustered survival data.

In this chapter, we propose methods for making inferences in the joint frailty model when only individual interval counts of the recurrent events are observed, and these observation intervals can vary between individuals. We will adapt the method of Liu and Huang (2008) for the estimation of the joint frailty model, and we will adjust the score test developed by Balan et al. (2016) to the setting in which only interval counts are available.

The chapter is structured as follows. In Section 5.2, we describe the joint frailty model for recurrent events and death, as well as the setting of individual interval counts. In Section 5.3, we present our approach of using Gaussian quadrature to estimate the joint frailty model based on interval counts, and adapt the score test for association in the joint frailty model. In Section 5.4, we assess the performance of the estimation method and the test in simulation studies. In Section 5.5, we apply the proposed methods to the data on *E. dichotoma*, followed by a discussion in Section 5.6.



## 5.2 Joint frailty model and interval counts

Before presenting the estimation and test procedure in the next section, we will introduce in the following the joint frailty model, which allows us to model the dependence of the recurrent event process and the terminal event. We then derive the likelihood function for data that only contain interval counts of the recurrent events.

We consider a sample of  $m$  independent individuals denoted by  $i, i = 1, \dots, m$ . Each individual  $i$  is observed from time  $t_0 = 0$  until the end of its follow-up  $X_i$ . The time  $X_i$  is either a censoring time  $C_i$ , which is independent of the recurrent event process and the terminal event, such as end of study; or it is the time  $D_i$  of the terminal event, whichever comes first:  $X_i = \min(C_i, D_i)$ . For simplicity, we assume that the terminal event is death, and denote by  $\delta_i = \mathbb{1}\{D_i \leq C_i\}$  the death indicator, where  $\mathbb{1}\{\cdot\}$  is the indicator function.  $Y_i(t) = \mathbb{1}\{t \leq X_i\}, t \geq 0$ , is the at-risk indicator at time  $t$ . We define two additional counting processes  $N_i^{D^*}(t) = \mathbb{1}\{D_i \leq t\}$  and  $N_i^D(t) = \mathbb{1}\{X_i \leq t, \delta_i = 1\}$ , where  $N_i^{D^*}(t)$  refers to the actual (but potentially unobserved) terminal event, whereas  $N_i^D(t)$  is the counting process of an observed terminal event, respectively.

For the recurrent event process, we denote with  $N_i^{R^*}(t)$  the number of events of individual  $i$  in the interval  $[0, t]$ . However, we only observe  $N_i^R(t) = N_i^{R^*}(\min(t, X_i))$ . The increments of the recurrence process over small intervals  $[t, t + dt)$  are  $dN_i^{R^*}(t) = N_i^{R^*}((t + dt)^-) - N_i^{R^*}(t^-)$ . Here,  $t^-$  denotes the left-hand limit.

Additional observed characteristics of individual  $i$  are collected in the covariate vector  $\mathbf{z}_i$ , whereas unobserved characteristics are summarized in a frailty value  $u_i$ . The  $u_i$  are realizations of a positive random variable  $U$ , independent across individuals. The observed data on individual  $i$  up to time  $t$  are collected in  $O_i(t) = \{Y_i(s), N_i^R(s), N_i^D(s), 0 \leq s \leq t; \mathbf{z}_i\}$ .

As in Liu et al. (2004), the recurrence process is characterized by the intensity  $Y_i(t)\lambda_i(t)$ , for which we assume

$$\begin{aligned} \text{P}(dN_i^R(t) = 1 \mid \mathcal{F}_{t^-}, D_i \geq t) &= Y_i(t)\lambda_i(t)dt \quad \text{with} \\ \lambda_i(t)dt &= d\Lambda_i(t) = \text{P}(dN_i^{R^*}(t) = 1 \mid \mathbf{z}_i, u_i, D_i \geq t), \end{aligned} \quad (5.1)$$

where  $\mathcal{F}_t = \sigma\{O_i(s), 0 \leq s \leq t, u_i; i = 1, \dots, m\}$ .

Analogously, for the terminal event

$$\begin{aligned} \text{P}(dN_i^D(t) = 1 \mid \mathcal{F}_{t^-}) &= Y_i(t)h_i(t)dt \quad \text{with} \\ h_i(t)dt &= dH_i(t) = \text{P}(dN_i^{D^*}(t) = 1 \mid \mathbf{z}_i, u_i, D_i \geq t). \end{aligned}$$

The joint frailty model for recurrent events and death, as proposed by Liu et al. (2004), is then specified as

$$\begin{aligned} \lambda_i(t) &= u_i e^{\boldsymbol{\beta}^\top \mathbf{z}_i} \lambda_0(t), \\ h_i(t) &= u_i^\gamma e^{\boldsymbol{\alpha}^\top \mathbf{z}_i} h_0(t), \end{aligned} \quad (5.2)$$

with baseline rates  $\lambda_0(t)$  and  $h_0(t)$  that are common to all individuals.

The shared frailty  $u$  enters both the recurrent event rate and the hazard rate of the terminal event, thereby introducing both dependence between the recurrences within one individual, as well as the association between the recurrences and the terminal event. The parameter  $\gamma$  determines the direction and the strength of the association between the two processes. If  $\gamma > 0$ , a higher rate of recurrence implies a higher mortality risk; if  $\gamma < 0$ , a higher rate of recurrence implies a lower mortality risk. If  $\gamma = 0$ , the rate of recurrence does not affect the mortality risk.

The frailties  $u_i$  are often assumed to follow a gamma distribution with mean one and variance  $\theta$ . In the following, we will more generally assume that the  $u_i$  stem from a distribution with density  $g_\theta(u)$  with parameter  $\theta$ .

The covariates  $z_i$  enter in (5.2) in a proportional rates or a proportional hazards formulation, but can have different effects  $\alpha$  and  $\beta$  on the terminal event and the recurrence process, respectively.

In the simplest setting, the exact times of event occurrence are observed. In many cases, however, only the number of events that occurred in a sequence of consecutive time intervals is available. More precisely, we observe individual interval counts  $n_{ij}$  as realizations of  $N_{ij} = N_i^R(t_{ij}) - N_i^R(t_{ij-1})$ . The  $N_{ij}$  give the number of recurrent events experienced by individual  $i$  in the interval  $I_{ij} = (t_{ij-1}, t_{ij}]$ ,  $j = 1, \dots, J_i$ . The  $t_{ij}$  correspond to the observation times of individual  $i$ , for instance, times of hospital visits in medical studies or, as in our example, generated by the lab logistics. Thus, both the positions of the  $I_{ij}$  and the total number of intervals  $J_i$  can vary across individuals. The follow-up times  $X_i$  are still exactly observed so that  $t_{iJ_i} = X_i$ .

As the frailties  $u_i$  are unobservable, the inference is based on the marginal likelihood that is obtained by integrating the conditional likelihood given the frailties  $u_i$  over the frailty distribution  $g_\theta(u)$ . The conditional likelihood of the joint frailty model (5.2) based on exactly observed recurrence times was developed in Liu et al. (2004). For the current setting, the likelihood factor for the recurrence times (formula (7) in Liu et al., 2004) has to be replaced by the contribution of the interval counts of the recurrent events. From (5.1) and (5.2), it follows that  $(N_{ij} \mid z_i, u_i, D_i \geq t_{ij})$  has a Poisson distribution with mean  $u_i \mu_{ij}$  where  $\mu_{ij} = \int_{I_{ij}} e^{\beta^\top z_i} \lambda_0(s) ds$ . Therefore, the likelihood contribution  $L_i^{(c)}(u_i)$  of individual  $i$  conditional on its frailty value  $u_i$  is given by

$$L_i^{(c)}(u_i) = \prod_{j=1}^{J_i} \frac{\exp(-u_i \mu_{ij}) (u_i \mu_{ij})^{n_{ij}}}{n_{ij}!} [u_i^\gamma e^{\alpha^\top z_i} h_0(x_i)]^{\delta_i} \exp\left\{-\int_0^{x_i} u_i^\gamma e^{\alpha^\top z_i} h_0(s) ds\right\}.$$

This leads to the marginal likelihood contributions

$$L_i = \int_0^\infty L_i^{(c)}(u) g_\theta(u) du. \quad (5.3)$$

In general, this integral does not have a closed-form expression.

## 5.3 Methods

In the first part of this section, we elaborate the estimation procedure of the joint frailty model based on interval counts with different individual observation intervals. Then, in Section 5.3.2, we demonstrate how the score test for dependence between the two processes can be adapted to the case of interval-count data.

### 5.3.1 Estimation of the joint frailty model based on interval counts

For observed recurrence times, Liu and Huang (2008) suggested using Gaussian quadrature to approximate the marginal likelihood of the joint frailty model. The approximated likelihood can then be maximized directly, since the integral is replaced by a weighted sum of function values. More specifically, Liu and Huang (2008) applied Gauss-Hermite quadrature, which for a function  $f(x)$  uses the approximation

$$\int_{-\infty}^{\infty} f(x) e^{-x^2} dx \approx \sum_{q=1}^Q w_q f(x_q).$$

The quadrature points  $x_q$  are the roots of the  $Q^{\text{th}}$ -order Hermite polynomial, and  $w_q$  are the corresponding weights. This approach is applicable to marginal likelihoods that are integrated over normal random effects, for which

$$\int_{-\infty}^{\infty} L^{(c)}(v) \phi(v) dv \approx \sum_{q=1}^Q \tilde{w}_q L^{(c)}(\tilde{x}_q) \phi(\tilde{x}_q), \quad (5.4)$$

with the standard normal density  $\phi(\cdot)$ , and modified quadrature points  $\tilde{x}_q = \sqrt{2}x_q$  together with weights  $\tilde{w}_q = \sqrt{2}w_q e^{x_q^2}$ . For non-normal random effects, Liu and Huang (2008) used the probability integral transformation proposed by Nelson et al. (2006). If the random effect density is  $g_\theta(u)$  with corresponding distribution function  $G_\theta(u)$ , then the integral over the density  $g_\theta(u)$  is transformed into an integral over standard normal random effects. This is achieved by noting that  $a = \Phi^{-1}(G_\theta(u))$  follows a standard normal distribution if the  $G_\theta(u)$ , which follow a standard uniform, are transformed by the inverse of the standard normal distribution function  $\Phi(\cdot)$ .

We apply this quadrature approach to the marginal likelihood of the joint frailty model based on interval counts of recurrent events; see equation (5.3). Substituting  $u = G_\theta^{-1}(\Phi(a))$  in the marginal likelihood contributions, we obtain

$$L_i = \int_0^\infty L_i^{(c)}(u) g_\theta(u) du = \int_{-\infty}^\infty L_i^{(c)}(G_\theta^{-1}(\Phi(a))) \phi(a) da.$$

These  $L_i$  can directly be approximated using Gauss-Hermite quadrature as

$$L_i \approx \sum_{q=1}^Q L_i^{(c)}(G_\theta^{-1}(\Phi(\tilde{x}_q))) \phi(\tilde{x}_q) \tilde{w}_q,$$

with  $\tilde{x}_q$  and  $\tilde{w}_q$  as defined in (5.4). The approximate marginal likelihood of the joint frailty model is then given by

$$\prod_{i=1}^m \sum_{q=1}^Q L_i^{(c)}(G_\theta^{-1}(\Phi(\tilde{x}_q))) \phi(\tilde{x}_q) \tilde{w}_q. \quad (5.5)$$

To actually maximize the approximate likelihood (5.5), we have to specify the baseline rates  $\lambda_0(t)$  and  $h_0(t)$ , as well as the frailty distribution  $g_\theta(u)$ . Similar to Liu and Huang (2008), we model the baseline rates as piecewise constant functions

$$\lambda_0(t) = \sum_{k=1}^{K^R} \lambda_{0k} \mathbb{1}\{t \in I_k^R\} \quad \text{and} \quad h_0(t) = \sum_{k=1}^{K^D} h_{0k} \mathbb{1}\{t \in I_k^D\}. \quad (5.6)$$

This choice is particularly suitable if no prior knowledge of the shapes of the two rates  $\lambda_0(t)$  and  $h_0(t)$  is available. The specifications of the intervals  $I_k^R = (t_{k-1}^R, t_k^R]$ ,  $k = 1, \dots, K^R$ , and  $I_k^D = (t_{k-1}^D, t_k^D]$ ,  $k = 1, \dots, K^D$ , can differ between the recurrent event process and the death process in terms of both their lengths  $\Delta_k^R = t_k^R - t_{k-1}^R$  and  $\Delta_k^D = t_k^D - t_{k-1}^D$  and their total numbers  $K^R$  and  $K^D$ . (The intervals for the piecewise constant baseline rates should not, however, be confused with the intervals in which the numbers of recurrent events  $n_{ij}$  are observed; see Section 5.2.)

The baseline rate  $\lambda_0(t)$  of the recurrence process enters the likelihood (5.5) through the conditional means  $u_i \mu_{ij}$  of the interval counts  $N_{ij}$  given the frailty value  $u_i$ . Under the piecewise constant rate model,  $\mu_{ij}$  is computed as

$$\mu_{ij} = \int_{I_{ij}} e^{\beta^\top z_i} \lambda_0(s) ds = e^{\beta^\top z_i} \sum_{k=1}^{K^R} \lambda_{0k} \max\{0, \min(t_k^R, t_{ij}) - \max(t_{k-1}^R, t_{ij-1})\}.$$

Piecewise constant baseline rates offer more flexibility than parametric models, such as the Weibull model, while at the same time remaining more tractable than purely nonparametric models. Previous studies have suggested that a moderate number of intervals – i.e., between 8 and 10 intervals – yields satisfactory estimation results (Liu and Huang, 2008; Lawless and Zhan, 1998).

The performance of the piecewise constant model is usually improved when the interval cut-points  $t_k$  are based on quantiles of the recurrence and the survival times, respectively. In the current setting in which only interval counts of recurrent events are observed, the exact recurrence times are unknown. If, however, the individual observation intervals  $I_{ij}$  are relatively short compared with the total follow-up, we can approximate quantiles by creating a set from the observation times  $t_{ij}$ , with each repeated  $n_{ij}$  times,  $j = 1, \dots, J_i$ ,  $i = 1, \dots, m$ ; and then determining the cut-points  $t_k^R$  as quantiles of this set of times.

Parameter estimation in the joint frailty model is then done by maximizing the approximate marginal log-likelihood; that is, the logarithm of (5.5). The standard errors for the parameter estimates can be obtained from the inverse of the negative Hessian of the approximate marginal log-likelihood. Further computational details can be found in Section 5.7.1.

### 5.3.2 Score test for the association between recurrences and death

The joint frailty model for recurrent events and death is rather complex, with estimation procedures that are more involved than those needed for the fitting of two separate models, one for the recurrent event process and one for the survival process. Investigating whether the two processes are associated – and, consequently, whether joint modeling is required – is a useful first step. Moreover, the question of whether there is an association – and, if so, whether it is positive or negative – can be a stand-alone issue, not necessarily followed by fitting a joint model.

Balan et al. (2016) proposed a correlation score test for the association between recurrences and terminal events in the joint frailty model for settings with observed recurrence times. Their test can be performed by fitting separate models for the recurrence times and the survival data, and, thus, without fitting the joint frailty model. In addition, the sign of the test statistic is an indicator of the direction of the association. In the following, we show that this score test can be adapted to the setting with interval counts of recurrent events, provided the survival times are exactly observed.

A test for association in the joint frailty model (5.2) corresponds to a test of  $H_0: \gamma = 0$  against  $H_1: \gamma \neq 0$ . All other parameters, including those for the baseline rate models, are treated as nuisance parameters, and denoted by  $\boldsymbol{\eta}$ . The score test of Balan et al. (2016) is based on the score function for  $\gamma$  under the null hypothesis; that is,

$$U_\gamma(0, \boldsymbol{\eta}) = \frac{\partial}{\partial \gamma} \ell(\gamma, \boldsymbol{\eta}) \Big|_{\gamma=0},$$

where  $\ell$  is the marginal log-likelihood of the joint frailty model. The authors showed that the score, evaluated at the maximum likelihood estimate under  $H_0$ , that is,  $(0, \hat{\boldsymbol{\eta}}_0)$ , is proportional to the covariance of the estimated martingale residuals of the terminal event and the ‘posterior’ estimates of the log-frailties for the recurrent events given the observed data. More formally, defining

$$K_i(u, t) = u^{N_i^R(t-) + \gamma N_i^D(t-)} \exp \left\{ -u e^{\boldsymbol{\beta}^\top \mathbf{z}_i} \Lambda_0(t) - u^\gamma e^{\boldsymbol{\alpha}^\top \mathbf{z}_i} H_0(t) \right\}, \quad (5.7)$$

with the cumulative baseline rates  $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$  and  $H_0(t) = \int_0^t h_0(s) ds$ , Balan et al. (2016) derived that

$$\begin{aligned} U_\gamma(0, \hat{\boldsymbol{\eta}}_0) &= \sum_{i=1}^m \left[ N_i^D(x_i) - e^{\hat{\boldsymbol{\alpha}}^\top \mathbf{z}_i} \widehat{H}_0(x_i) \right] \frac{\int_0^\infty \ln(u) \widehat{K}_i(u, x_i) g_\theta(u) du}{\int_0^\infty \widehat{K}_i(u, x_i) g_\theta(u) du} \\ &= \sum_{i=1}^m \widehat{M}_i^D \cdot \widehat{\ln}(u_i). \end{aligned} \quad (5.8)$$

The  $\widehat{M}_i^D$  are estimates of the martingale residuals of the terminal event model  $M_i^D = N_i^D(x_i) - \int_0^{x_i} e^{\boldsymbol{\alpha}^\top \mathbf{z}_i} h_0(s) ds$ . The  $\widehat{\ln}(u_i)$  are the posterior estimates of the log-frailty values

given the observed data from the recurrent event process:  $\widehat{\ln(u_i)} = E[\ln U_i | O_i(x_i)]$ . For gamma distributed frailties  $u_i$  with mean one and variance  $\theta$ , one obtains

$$\widehat{\ln(u_i)} = \psi\left(\frac{1}{\hat{\theta}} + N_i^R(x_i)\right) - \ln\left(\frac{1}{\hat{\theta}} + e^{\hat{\beta}^\top z_i} \widehat{\Lambda}_0(x_i)\right), \quad (5.9)$$

where  $\psi(\cdot)$  is the digamma function.

Because of the zero-mean constraint of the  $\widehat{M}_i^D$ , the second line of (5.8) is proportional to the correlation  $r = \text{cor}(\widehat{M}^D, \widehat{\ln(u)})$  between the martingale residuals and the estimated log-frailties. Consequently, Balan et al. (2016) based the correlation score test on the test statistic

$$t = r \sqrt{\frac{m-2}{1-r^2}}, \quad (5.10)$$

which, under the null hypothesis, asymptotically follows a  $t$ -distribution with  $m-2$  degrees of freedom.

It turns out that in the setting in which only interval counts of the recurrent events are available, but exact survival times (or censoring times) are observed, equation (5.8) still holds. We show this result in Section 5.7.2. Therefore, the test statistic  $t$  in (5.10) is still valid. Furthermore, for gamma distributed frailties, the estimates  $\widehat{\ln(u_i)}$  can still be determined using formula (5.9). The latter formula involves estimates  $\hat{\theta}$  of the frailty variance, the covariate effect  $\hat{\beta}$  on the recurrence rate, and the cumulative baseline rate  $\widehat{\Lambda}_0$ ; all determined under  $H_0$ . These can be obtained by fitting a mixed Poisson model to the interval counts of the recurrent events (see, for instance, Lawless and Zhan, 1998, who assume a piecewise constant baseline rate function  $\lambda_0$ ). We generally recommend to estimate the shared frailty model for the interval counts using a flexible specification such as (5.6) for the baseline rate. A simple parametric model like, for example, the Weibull model, although appealing due to its parsimony, always bears the risk of misspecification and consequently misleading test results. The estimates of the martingale residuals  $\widehat{M}_i^D$  can be derived from a Cox proportional hazards model fitted to the survival data  $\{X_i, \delta_i, z_i; i = 1, \dots, m\}$ .

## 5.4 Simulation study

### 5.4.1 Performance of the estimation method

To evaluate the performance of the proposed method for estimating the parameters of the joint frailty model based on interval counts of recurrent events and survival times, we conducted a simulation study.

Several different aspects will affect the estimation results, both on the part of the model specification but also on the part of the observable data. The latter include the number and lengths of the intervals for the recurrent event counts and whether they are the same for all individuals in the sample or not. The amount of independent censoring

will also have an impact. Among the former aspects, the size of the frailty variance and the sign of the dependence parameter are expected to influence the estimation, but also the number of intervals used in the piecewise constant specification of the baseline rates (and consequently the total number of parameters to be estimated) will matter.

We generated data for  $m = 200$  individuals from the joint frailty model (5.2). For the baseline rates  $\lambda_0(t)$  and  $h_0(t)$ , we chose the form of Weibull hazards,  $(a/b)(t/b)^{a-1}$ , with shape parameter  $a$  equal to 1.5 or 3 and scale parameter  $b$  equal to  $1/3$  and  $1.35$ , respectively. A single binary covariate that takes values 0 or 1 with probability 0.5 was included, and had the same effect on the hazard of death and the rate of recurrence ( $\alpha = \beta = 1$ ). Frailties were simulated from a gamma distribution with mean one and different variances  $\theta \in \{0.25, 0.5, 0.75\}$ . We considered both cases of positive ( $\gamma = 1$ ) and negative association ( $\gamma = -1$ ) between the recurrence process and the terminal event. Additionally, we looked at one setting with a relatively small value for the frailty variance,  $\theta = 0.05$ , that was inspired by the results of the data set on *Eleutheria dichotoma*, see Section 5.5. In this setting, two values for the dependence parameter  $|\gamma| = 1$  or  $|\gamma| = 5$  were studied.

Independent censoring was considered in two versions: either an end-of-study censoring at time  $t = 2$  for all individuals still alive then or individual censoring times which occurred uniformly over the total follow-up window  $[0, 2]$ .

Regarding the observation times  $t_{ij}$ , which determine the intervals during which the recurrent events are counted, we examined two scenarios. In Scenario I the observation times were the same for all individuals. We set  $t_{ij} = 0.2j$ ,  $j = 0, \dots, 10$ , such that we observed individual interval counts in up to 10 intervals of equal length 0.2. We also studied one scenario with  $t_{ij} = 0.1j$ ,  $j = 0, \dots, 20$ , leading to up to 20 individual interval counts. In Scenario II the observation times  $t_{ij}$  varied across individuals. This scenario mimicked a study in which participants have scheduled visit times, but actually could be early or late for their visits. This was implemented as follows: the scheduled times were  $t_j^0 = 0.2j$ ,  $j = 0, \dots, 10$ , but the actual observation times for each individual were obtained by adding a random noise  $\varepsilon_{ij}$  to the scheduled times so that  $t_{ij} = t_j^0 + \varepsilon_{ij}$ . The  $\varepsilon_{ij}$  were drawn from a uniform distribution on  $[-0.1, 0.1]$  (except for  $\varepsilon_{i0} = 0$  and  $\varepsilon_{i10}$  as uniform on  $[-0.1, 0]$ ) and hence the  $t_{ij}$  varied across individuals.

For the estimation of the joint frailty model, we assumed that the frailties were gamma distributed, and that the baseline rates were specified as piecewise constant functions on 10 intervals of equal length,  $t_k^R = t_k^D = 0.2k$ . Thus, under Scenario I and  $t_{ij} = 0.2j$  the intervals for the rate pieces coincided with the observation intervals for the counts. We also studied the impact of an increase of the number of intervals to 20. Approximation of the likelihood was based on  $Q = 30$  quadrature points. All computations were run in R (R Core Team, 2019), see Section 5.7.1 for further details. In all settings we ran 200 replications.

The results of the simulation study under Scenario I with frailty variance  $\theta = 0.5$  are shown in Figures 5.1 and 5.2, both for positive and negative dependence  $|\gamma| = 1$  and for both independent censoring schemes (A: end-of-study censoring at  $t = 2$ , and B: uniform censoring times in  $[0, 2]$ ).

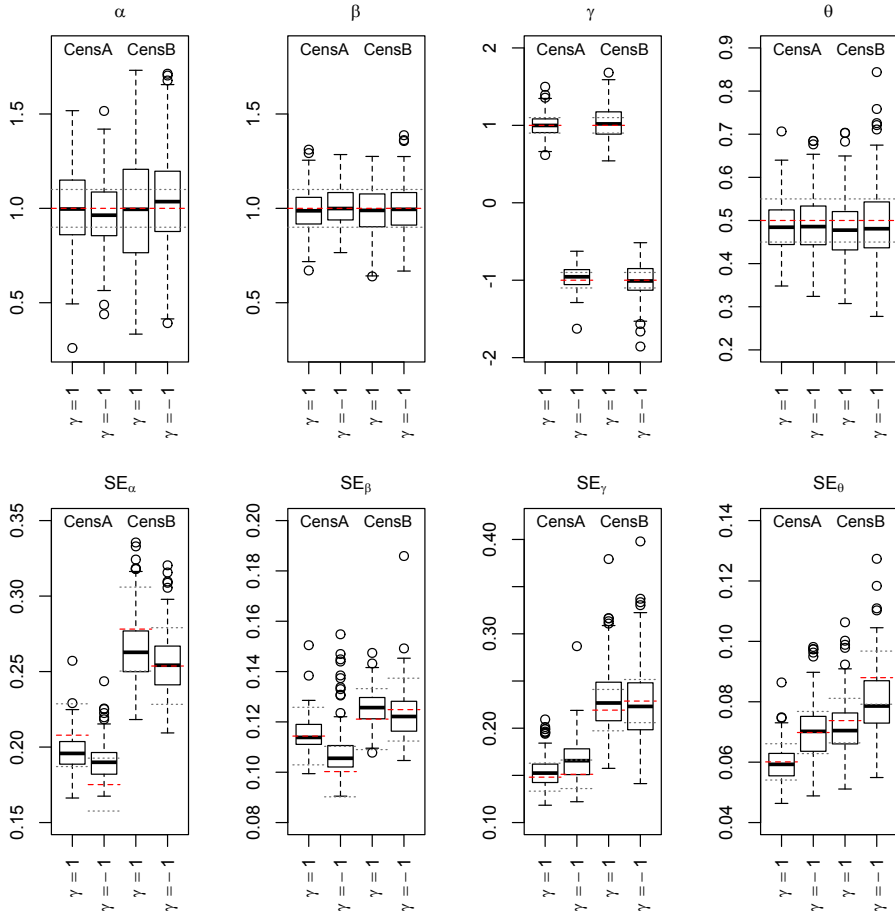


Figure 5.1: Box plots of the parameter estimates (top) and estimated standard errors (bottom) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ) under two schemes of independent censoring (CensA: at end of study  $C = 2$ , CensB: uniform on  $[0, 2]$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true parameter value (top) or empirical standard deviation (bottom); gray, dotted lines mark 10% deviations from respective value.



From the box plots of the estimates of the covariate effects  $\alpha$  and  $\beta$ , the dependence parameter  $\gamma$ , and the frailty variance  $\theta$  in Figure 5.1, we can see that the method performs satisfactorily. Censoring scheme A is considerably milder than scheme B (for an illustration see Table 5.4 in Section 5.8.1), and the corresponding loss in information is reflected in a moderately increased variability in the estimates and, as could be expected, also in the estimated standard errors (Figure 5.1, bottom). For the frailty variance  $\theta$ , variability in estimates and estimated standard errors is higher for negative dependence ( $\gamma = -1$ ) than for positive dependence. This is also true for the dependence parameter  $\gamma$  itself. The estimated standard errors for the covariate effects  $\alpha$  and  $\beta$ , the dependence parameter  $\gamma$ , and the frailty variance  $\theta$  are compared with the empirical standard deviations of the respective estimates across the replications in the bottom panels of Figure 5.1. The general magnitude of the standard errors is well captured.

Figure 5.2 displays the estimates of the cumulative baseline rates for positive dependence for both censoring schemes A and B. The averages of the estimates are very close to the true underlying rates. The stronger loss of observations in scheme B leads to markedly increased variability of the estimates toward the end of the follow-up window, which is an obvious consequence of the decreasing number of individuals under study over time. This low number of observations results in part from individuals experiencing a terminal event. Additionally, even fewer individuals (in this setting, less than 5%) are observed in the interval  $[1.8, 2)$ , which corresponds to the last piece of the baseline rates, due to independent censoring under scheme B. In contrast, censoring scheme A is benign.

For the other settings of the simulation study we restricted our attention to the independent censoring scheme B. With its uniformly distributed censoring times it produces a relatively challenging loss of observations. So the results we report here are conservative in the sense that they hold despite this appreciable amount of censoring. Section 5.8.1 illustrates the results of the other simulation settings analogous to Figures 5.1 and 5.2. Here we summarize the main results of the various scenarios.

If we change the frailty variance to  $\theta = 0.25$  or  $\theta = 0.75$ , we note that also in these settings the biases of the parameter estimates, if present at all, are small. In general, varying the true underlying frailty variance mainly affects the variability of the estimates. The estimates of the covariate effects and the frailty variance show increased variability for larger frailty variance. In contrast, the estimates of the dependence parameter show less variability for larger frailty variance, because the increased heterogeneity among the individual patterns makes it easier to assess the dependence between the recurrence process and survival.

If the true frailty variance is close to zero ( $\theta = 0.05$ ) the estimates of the dependence parameter are still only modestly biased but the variability of the  $\hat{\gamma}$  increases strongly. This had to be expected, since the identification of the dependence hinges on the variation in the individual frailty. If the true value of  $|\gamma| = 1$ , then the estimated values may turn out with the wrong sign of the dependence parameter, particularly if the true dependence is negative. Thus, weak dependence is difficult to identify in case of low variability in

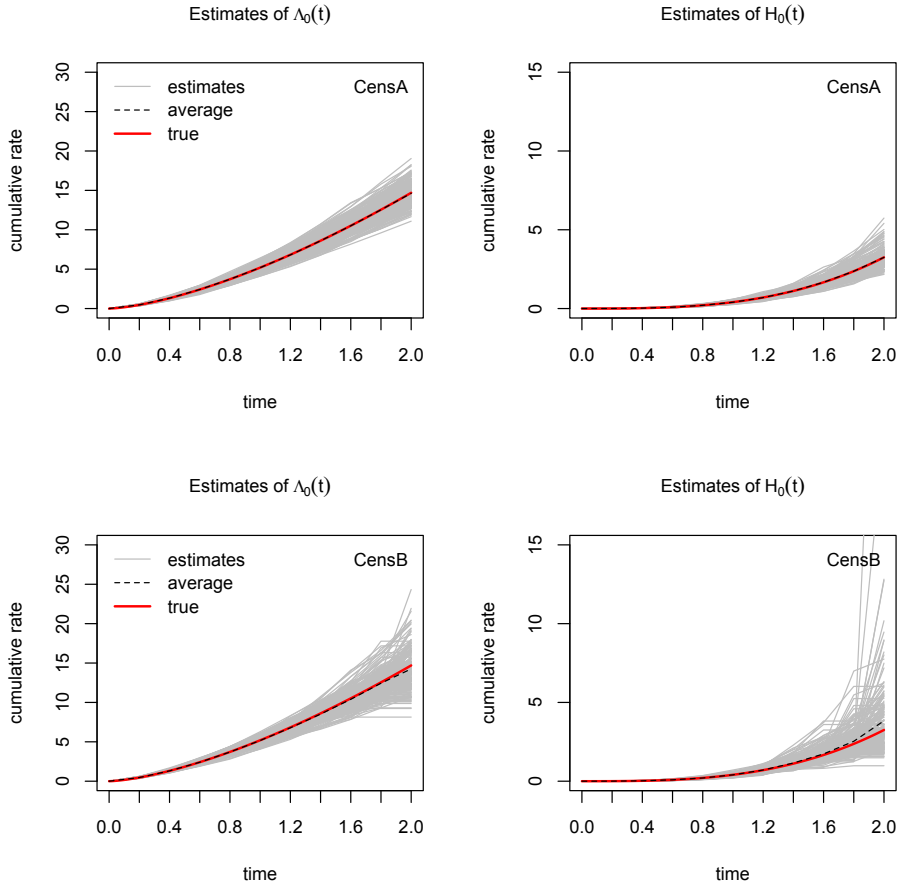


Figure 5.2: Estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) based on samples of size  $m = 200$  which were generated from a joint frailty model with positive dependence for two schemes of independent censoring. Top: end-of-study censoring at  $t = 2$  (CensA), bottom: uniform censoring on  $[0, 2]$  (CensB). Red, solid line gives the true cumulative baseline rate; black, dashed line is the mean of the 200 estimates.

frailty. However, if the dependence is strong,  $|\gamma| = 5$ , then, despite the variability of the estimates, the sign of the dependence is correctly estimated.

If we modify the width of the observation intervals from 0.2 to 0.1 we first note the following: if we keep the specification of the baseline rate of recurrence as piecewise constant on the 10 intervals with  $t_k^R = 0.2k$  as before, then the additional information provided by the finer observation intervals would not be used by the estimation method

(see Section 5.7.3 for a proof). Therefore, we also specify the baseline rates as piecewise constant functions on 20 intervals, that is  $t_k^R = t_k^D = 0.1k$ . Apart from increased variability in the rate estimates near the end of the follow-up window, there is little change to the estimation results with width 0.2.

If we allow the observation intervals to vary across individuals (Scenario II), then we find that the method performs equally well as for fixed observation intervals, with the exception of the estimates for the last baseline rate pieces. In this simulation, we used fixed intervals for the baseline rates though, to unify the presentation of results. In applications we would recommend to choose the cut-points  $t_k^R$  and  $t_k^D$  based on (approximate) quantiles of the event times, as described earlier, to increase precision of the baseline rate estimates.

If we replace the baseline rates by a parametric Weibull specification (which here is correct), then there is only very modest change in the other parameter estimates of the model as well as their standard errors, so the main advantage is in the less variable estimation of the two baseline rates. This is, however, counterbalanced by the risk of model misspecification so we advise the use of piecewise constant rates unless one has a good understanding of the underlying processes that justifies a parametric choice.

Overall, the proposed method for estimating the joint frailty model based on interval counts yielded reliable results in this simulation study and different scenarios behaved in the way we had anticipated beforehand.

## 5.4.2 Performance of the score test

In a second set of simulations, we investigated the performance of the score test in the setting with interval counts of recurrent events.

For this purpose, we again generated data for  $m = 200$  individuals from the joint frailty model (5.2). The covariate effects, values for the variance of the gamma frailties, and baseline rates were the same as in Section 5.4.1. Here, however, we considered not only different directions but also different strengths of the association between the recurrence process and the terminal event; namely,  $\gamma \in \{-1, -0.5, 0, 0.5, 1\}$ . Counts were again observed in 10 intervals of equal length 0.2 (Scenario I) or of varying length (Scenario II). We studied two schemes of independent censoring, scheme A with censoring time  $C = 2$ , and scheme B with uniformly distributed censoring times on  $[0, 2]$ . We ran 1000 replications for each setting to determine the size or power of the test.

The score test involves fitting separate models to the recurrence data and the survival data. First, we fitted a shared gamma frailty model to the individual interval counts of the recurrent events, assuming that  $(N_{ij} \mid u_i)$  follows a Poisson distribution with mean  $u_i \tilde{\mu}_{ij}$ , where  $\tilde{\mu}_{ij} = \int_{I_{ij}} e^{\beta^T z_i} \tilde{\lambda}_0(s) ds$ . The baseline recurrence rate  $\tilde{\lambda}_0(\cdot)$  was modeled as piecewise constant, as in (5.6), with pieces defined by the cut-points  $t_k^R = 0.2k$ ,  $k = 0, \dots, 10$ . The estimates  $\ln(\widehat{u_i})$  were then determined according to formula (5.9). Second, we estimated a Cox proportional hazards model from the survival data  $\{X_i, \delta_i, z_i; i = 1, \dots, m\}$  using function `coxph()` from package `survival` (Therneau and Grambsch,

2000) to obtain the martingale residuals  $\widehat{M}_i^D$  of the terminal event. Finally, we calculated the test statistic based on the correlation between the  $\widehat{M}_i^D$  and the  $\overline{\ln(u_i)}$ .

Table 5.1 reports the size and the power of the score test, performed at a level of 5%, depending on the true underlying dependence parameter  $\gamma$  and frailty variance  $\theta$ . The proportion of a type I error, that is, of falsely rejecting the hypothesis of no dependence ( $\gamma = 0$ ), was affected most strongly by the censoring scheme. If  $C = 2$ , which implies a modest proportion of independently censored cases (see Table 5.6 in Section 5.8.2), then the level of the test is met or exceeded only slightly. For strong independent censoring (scheme B), which implies that roughly half of the observations are censored and quite some of them early during the follow-up, then this loss of information increases the proportion of a type I error, particularly for the frailty variance  $\theta = 0.75$ . This is a noteworthy result, and hence the test should be regarded with caution, if there is strong (and early) censoring in the data.

Table 5.1: Power and size of the score test, performed at the 5% level, in the joint frailty model with varying dependence parameter  $\gamma \in \{-1, -0.5, 0, 0.5, 1\}$ , frailty variance  $\theta \in \{0.25, 0.5, 0.75\}$ , and independent censoring  $C \sim \mathcal{U}[0, 2]$  or  $C = 2$ , across 1000 replications each.

censoring	$\theta$	Dependence $\gamma$				
		-1	-0.5	0	0.5	1
$C \sim \mathcal{U}[0, 2]$	0.25	0.960	0.586	0.066	0.514	0.932
	0.5	1.000	0.905	0.062	0.872	1.000
	0.75	1.000	0.984	0.091	0.974	1.000
$C = 2$	0.25	0.999	0.788	0.057	0.799	1.000
	0.5	1.000	0.990	0.039	0.992	1.000
	0.75	1.000	1.000	0.040	1.000	1.000

Regarding the power of the score test, as expected, we found that the power increased with the strength of the dependence, that is,  $|\gamma|$ , and with the magnitude of the frailty variance. In the settings with the larger frailty variances  $\theta = 0.5$  and  $\theta = 0.75$  and stronger dependence  $|\gamma| = 1$ , the score test detected the association in all cases. We note that for the samples for which the association was detected by the score test, the direction of the dependence was always identified correctly. An extension of the simulation settings suggested that the score test can detect associations even for small values of the frailty variance as long as the association is sufficiently strong, that is,  $|\gamma|$  is sufficiently large (see Table 5.7 in Section 5.8.2).

Furthermore, we assessed the performance of the score test for Scenario II in which the observation intervals vary across individuals. Results for the setting with 10 scheduled visit times and 10 pieces for the baseline rate are displayed in Table 5.10 of Section 5.8.2. The proportion of false rejections of the hypothesis of no dependence are again a bit higher than the nominal level due to the high percentage of censoring. The power of the

test is comparable to the values obtained in Scenario I, although the power in the settings with negative dependence is a bit lower here.

Additional results on the performance of the score test for modifications of Scenario I and II are given in Section 5.8.2.

In conclusion, the results of the simulation study provide further evidence that the score test is a powerful method for assessing the association between recurrent events and the terminal event, also in a setting in which only interval counts of the recurrent events are observed.

## 5.5 Fertility and mortality in *Eleutheria dichotoma*

To illustrate the proposed methods, we use data from a biodemographic study on the fertility and mortality of *Eleutheria dichotoma* (Daňko et al., 2020), which was briefly introduced in Section 5.1. *E. dichotoma* is a marine organism that passes through several life-cycle stages: i.e., planula larva, polyp, and medusa stages. The colonial polyps (hydroids) asexually produce medusae. The medusae can then reproduce both sexually (by producing larvae) and asexually (by producing medusa buds).

In this study, asexually budded medusae were collected directly from the hydroid colony and reared for three generations – each of which was, in turn, obtained through the asexual budding of medusae. In our analysis, we focus on the age trajectory of the budding rate (asexual reproduction) and the mortality of one of the medusa generations, as well as on the association between the two processes.

Age  $t_0 = 0$  of a medusa corresponds to the point in time when it detaches from its hydroid colony or from its ancestor medusa. The medusae were followed individually, and were observed until death or censoring. This occurred when either the study ended, a laboratory accident (e.g., water evaporation) led to the loss of the medusa, or the medusa was absorbed by a large bud of the same individual. The animals were checked for newly released larvae and buds roughly three times per week. The resulting observation times differed across individuals, with interval lengths between 1 and 11 days.

Salinity is an important factor that affects the physiological responses of species like *E. dichotoma*, both at the level of the hydroid colony that produced the medusae under study, and at the level of the medusae themselves. Four combinations of salinity levels were studied here (low(hydroid)–low(medusa), medium–medium, low–medium, and medium–low); for more details, see Daňko et al. (2020).

The data set contains  $m = 141$  individuals, with the following group sizes in the four experimental conditions: 36 low–low, 40 medium–medium, 32 low–medium, and 33 medium–low. The individuals produced between 0 and 27 buds over their life course, with a mean of 8.99. Follow-up times varied between 9 and 217 days, with a median of 98 days; and 34 individuals (24%) were censored. Figure 5.3 exemplarily shows the data for 14 individuals.

To assess whether the recurrent budding process and survival are associated in *E. dichotoma*, we first conducted the score test that was presented in Section 5.3.2. For fitting

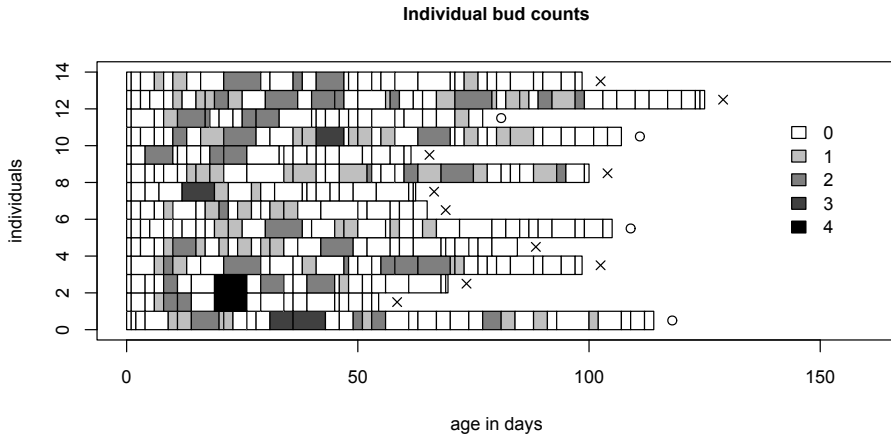


Figure 5.3: Interval counts of budding and survival (dead: cross, censored: circle) for 14 medusae *E. dichotoma*.

the shared frailty model to the individual bud counts, we assumed that the frailty variable was gamma distributed. Moreover, for the piecewise constant baseline budding rate, we defined cut-points  $t_k^R$  at 0, 16, 22, 27, 32, 40, 47, 63, 85, 111, and 220 days. These cut-points were obtained as (approximative) deciles of the recurrent event times, as described at the end of Section 5.3.1. We included two binary covariates for the salinity levels at the polyp stage and the medusa stage, respectively.

The martingale residuals of the terminal event were obtained from a Cox proportional hazards model fitted to the survival data, including the two covariates on the salinity levels.

The parameter estimates of the two separately fitted models, the Cox model for the survival data, and the shared frailty model for the budding rate based on interval counts, are shown in Table 5.2.

The correlation between the martingale residuals from the Cox regression and the estimated log-frailties from the budding model is  $r = -0.434$ , yielding a test statistic of  $t = -5.676$  with a  $p$ -value of  $7.729 \cdot 10^{-8}$ . Thus, the result of the score test clearly suggests that reproduction and mortality are negatively associated – i.e., that a higher budding rate is associated with lower mortality – which implies that a joint model should be used to analyze these data on *E. dichotoma*.

Therefore, we estimated a joint frailty model for reproduction and mortality for these data, including the two covariates on salinity. Frailties were again assumed to stem from a gamma distribution. For the baseline budding rate, the same specification was used as in the separate model (see above). For the baseline hazard of death, we used a piecewise

Table 5.2: Parameter estimates (with standard errors) for different models fitted to the *E. dichotoma* data set.

	Joint frailty model	Separate models	
		Shared frailty model	Cox PH model
Mortality			
Polyp (low salinity)	-0.493 (0.299)	-	-0.376 (0.201)
Medusa (low salinity)	2.201 (0.432)	-	1.392 (0.218)
Budding			
Polyp (low salinity)	0.052 (0.069)	0.057 (0.066)	-
Medusa (low salinity)	0.574 (0.071)	0.600 (0.068)	-
Association			
Dependence $\gamma$	-4.941 (1.557)	-	-
Frailty variance $\theta$	0.051 (0.018)	0.036 (0.017)	-

constant function with cut-points  $t_k^D$  at 0, 60, 67, 72, 85, 98, 103, 114, 132, 153, and 220 days; again, the cut-points were taken from approximate deciles of the survival times. For the Gaussian quadrature in the marginal likelihood (see equation (5.5)),  $Q = 30$  quadrature points were used.

The parameter estimates of the joint model are also displayed in Table 5.2. Interestingly, the dependence parameter  $\gamma$  was estimated as  $-4.941$  along with a frailty variance of  $0.051$ . The negative value of  $\hat{\gamma}$  indicated that individuals with higher rates of asexual reproduction tended to have a lower mortality risk than individuals with lower rates of asexual reproduction. This finding that higher rates of reproduction were accompanied by longer survival stands in contrast to the idea of a trade-off between reproduction and survival. Regarding the salinity levels, we found that the salinity experienced by the polyps did not have a noticeable effect on the reproduction or survival of the medusae. In contrast, individuals who were exposed to low salinity at the medusa stage were found to have both higher fertility and mortality rates than those exposed to medium salinity. Finally, Figure 5.4 shows the estimates of the age-specific budding and death rates. The budding rate peaked at about 20 days of age before gradually declining to a non-zero level at later ages. The death rates increased markedly after about 60 days of age.

To conclude, while the results presented here are in agreement with the findings of Daňko et al. (2020), they also provide additional insight into the association between the asexual reproduction and survival of *E. dichotoma*. In addition to the biological implications of the dependence itself, due to the association between the two processes, the recurrent event process is censored by the terminal event (death) non-independently, which warrants the joint modeling of the two processes.

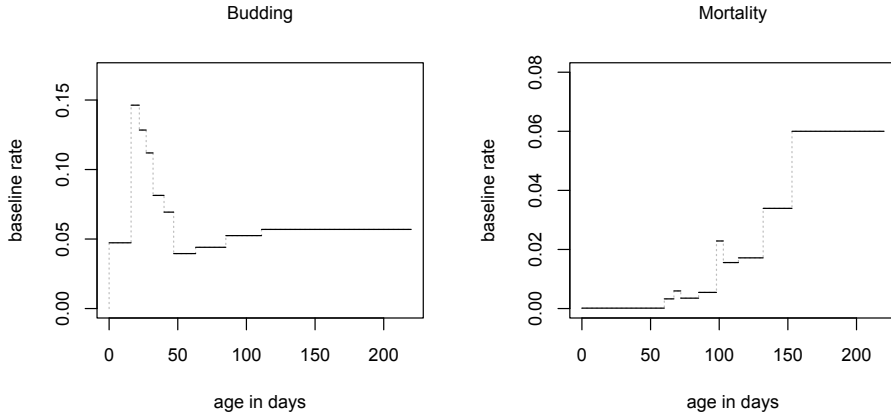


Figure 5.4: Estimated baseline rates of budding (left) and death (right) for the *E. dichotoma* data set.

## 5.6 Discussion

We have presented a method for estimating the joint frailty model for recurrent events and death in situations in which only individual interval counts of the recurrence process are observed. When modeling the baseline rates as piecewise constant, the marginal likelihood can be approximated using Gaussian quadrature, and can then be maximized directly. In addition, we have shown that the score test for the association between recurrences and death (Balan et al., 2016) is also applicable in the setting with interval counts. The test is based on the correlation between the martingale residuals of the terminal event and the estimates of the log-frailties, which are obtained by separately fitting a Cox proportional hazards model to the survival data and a shared frailty model to the interval counts of recurrent events. Our simulation studies demonstrated that both the estimation method and the score test perform well.

We also found that when applying the proposed methods to data on fertility and mortality in *E. dichotoma*, the rate of asexual reproduction and the mortality risk are negatively associated. While this finding is interesting from a biological point of view, it also demonstrates the necessity of allowing positive and negative dependence in a model such as the relatively complex joint frailty model (5.2). Moreover, the *E. dichotoma* example illustrates the advantages of using the piecewise constant rate model. On the one hand, the shape of the budding rate, which is characterized by a sharp peak at younger ages and a gradual leveling off to a non-zero level at older ages, is hard to capture using a simple parametric model. On the other hand, the data structure of the interval counts,



in which the observation times vary across individuals, makes it difficult to construct a purely nonparametric estimate of the baseline rate.

Implementing the estimation method in the joint frailty model involves making several choices, such as decisions regarding the distribution of the frailties, the number of quadrature points, and the number of pieces included in the baseline rate models. For the frailty distribution, we assumed that frailties follow a gamma distribution in both the simulation study and the application. The use of a gamma distribution is a popular choice for the distribution of frailties, and, with respect to the score test, it has the benefit of yielding closed-form expressions for the estimated log-frailties. However, the quadrature approach is equally able to accommodate the log-normal distribution or any frailty distribution that has a closed-form inverse distribution function. Yet, the performance of the quadrature approach relies on the quality of the approximation of the marginal likelihood, which, in turn, depends on the number of quadrature points. Liu and Huang (2008) suggested using  $Q = 30$  quadrature points for gamma frailty models, and in our experience, this number yields reliable results. In practice, we recommend fitting the model for several increasing values of  $Q$  until the estimates stabilize. For the piecewise constant rate models, using a moderate number of up to 10 intervals for the rates seems to produce good results. Using a larger number of intervals for the baseline rates generates a larger number of parameters to be estimated, which, in turn, increases computational costs, and might affect the numeric stability of the method.

One of the limitations of the approach presented here is that, in the joint frailty model, the association between recurrences and death is modeled via a dependence parameter acting on a shared frailty. If, however, the individuals are not sufficiently heterogeneous – that is, if the frailty variance is not sufficiently large – the dependence parameter is not meaningful, and the association cannot be assessed.

Another restriction is imposed by the piecewise constant rate models. While these models can capture a variety of different shapes of the baseline rates, using more flexible rates – and, in particular, smooth rates – might be desirable in some applications. Further work is needed on how to incorporate smooth rate models with automatic smoothing parameter selection.

Finally, the observation times in our application are fixed by the experimental set-up, even though in real-world applications, particularly in medicine, the observation times may depend on the recurrence process. For instance, patients may visit the doctor more often when they are in worse condition. This is taken into account in a model developed by Zhao et al. (2013), which considers interval counts of recurrent events in the presence of death and a dependent observation process. However, the authors adopted a marginal approach that left the association between the recurrences and death unspecified.

## 5.7 Appendix

### 5.7.1 Computational details

We implemented our estimation approach in R (R Core Team, 2019), using function `gauss.quad()` from package `statmod` (Smyth, 1998) to determine the quadrature points and weights, and function `nlm()` for numerical optimization of the approximate marginal log-likelihood.

The Hessian of the log-likelihood can be obtained directly from the output of the function `nlm()`. However, in some cases, computation of the Hessian using function `hessian()` from package `numDeriv` (Gilbert and Varadhan, 2019) yields more stable results.

To ensure that the Hessian of the approximate log-likelihood with piecewise constant baseline rates is invertible, it can be necessary to fit the joint frailty model with small, fixed ridge penalties on the logarithm of the baseline rates. The standard errors are then calculated based on the Hessian of the penalized log-likelihood.

### 5.7.2 Derivation of the score test

In this section, we show that the score  $U_\gamma(\gamma, \boldsymbol{\eta})$  of the joint frailty model (5.2) has the same form independently of whether exact recurrence times or only interval counts of the recurrent events are available, as long as the recurrence process is observed up to exactly known follow-up times.

Let us start by rewriting the individual contributions (5.3) to the marginal likelihood for interval counts

$$\begin{aligned}
 L_i &= \int_0^\infty \prod_{j=1}^{J_i} \left[ \frac{1}{n_{ij}!} \exp \left\{ - \int_{I_{ij}} u e^{\beta^\top z_i} \lambda_0(s) ds \right\} \left( \int_{I_{ij}} u e^{\beta^\top z_i} \lambda_0(s) ds \right)^{n_{ij}} \right] \\
 &\quad \cdot [u^\gamma e^{\alpha^\top z_i} h_0(x_i)]^{\delta_i} \exp \left\{ - \int_0^{x_i} u^\gamma e^{\alpha^\top z_i} h_0(s) ds \right\} g_\theta(u) du \\
 &= \prod_{j=1}^{J_i} \left[ \frac{1}{n_{ij}!} \left( \int_{I_{ij}} e^{\beta^\top z_i} \lambda_0(s) ds \right)^{n_{ij}} \right] [e^{\alpha^\top z_i} h_0(x_i)]^{\delta_i} \cdot \int_0^\infty \left[ \prod_{j=1}^{J_i} u^{n_{ij}} \right] \\
 &\quad \cdot u^\gamma \delta_i \exp \left\{ - u \sum_{j=1}^{J_i} \int_{I_{ij}} e^{\beta^\top z_i} \lambda_0(s) ds \right\} \exp \left\{ - \int_0^{x_i} u^\gamma e^{\alpha^\top z_i} h_0(s) ds \right\} g_\theta(u) du \\
 &= \prod_{j=1}^{J_i} \left[ \frac{1}{n_{ij}!} \mu_{ij}^{n_{ij}} \right] [e^{\alpha^\top z_i} h_0(x_i)]^{\delta_i} \\
 &\quad \cdot \int_0^\infty u^{n_i + \gamma \delta_i} \exp \left\{ - u e^{\beta^\top z_i} \Lambda_0(x_i) \right\} \exp \left\{ - u^\gamma e^{\alpha^\top z_i} H_0(x_i) \right\} g_\theta(u) du
 \end{aligned}$$

$$= \prod_{j=1}^{J_i} \left[ \frac{1}{n_{ij}!} \mu_{ij}^{n_{ij}} \right] [e^{\alpha^\top z_i} h_0(x_i)]^{\delta_i} \cdot \int_0^\infty K_i(u, x_i) g_\theta(u) du,$$

with  $n_i = \sum_{j=1}^{J_i} n_{ij}$  and  $K_i(u, t)$  as defined in (5.7).

The marginal log-likelihood  $\ell(\gamma, \boldsymbol{\eta}) = \sum_{i=1}^m \ln L_i$  takes the form

$$\sum_{i=1}^m \sum_{j=1}^{J_i} [-\ln(n_{ij}!) + n_{ij} \ln(\mu_{ij})] + \delta_i [\alpha^\top z_i + \ln(h_0(x_i))] + \ln \int_0^\infty K_i(u, x_i) g_\theta(u) du,$$

and the score with respect to  $\gamma$  reads

$$U_\gamma(\gamma, \boldsymbol{\eta}) = \sum_{i=1}^m \frac{\int_0^\infty \left[ \frac{\partial}{\partial \gamma} K_i(u, x_i) \right] g_\theta(u) du}{\int_0^\infty K_i(u, x_i) g_\theta(u) du}.$$

Since

$$\begin{aligned} \frac{\partial}{\partial \gamma} K_i(u, t) &= u^{N_i^R(t-) + \gamma N_i^D(t-)} \cdot N_i^D(t-) \cdot \ln(u) \cdot \exp \left\{ -u e^{\beta^\top z_i} \Lambda_0(t) - u^\gamma e^{\alpha^\top z_i} H_0(t) \right\} \\ &\quad - u^{N_i^R(t-) + \gamma N_i^D(t-)} \cdot \exp \left\{ -u e^{\beta^\top z_i} \Lambda_0(t) - u^\gamma e^{\alpha^\top z_i} H_0(t) \right\} \\ &\quad \cdot u^\gamma \cdot e^{\alpha^\top z_i} H_0(t) \cdot \ln(u) \\ &= K_i(u, t) \left[ N_i^D(t-) \cdot \ln(u) - u^\gamma \cdot e^{\alpha^\top z_i} H_0(t) \cdot \ln(u) \right], \end{aligned}$$

we have

$$U_\gamma(\gamma, \boldsymbol{\eta}) = \sum_{i=1}^m \frac{\int_0^\infty [N_i^D(x_i) - u^\gamma \cdot e^{\alpha^\top z_i} H_0(x_i)] \ln(u) K_i(u, x_i) g_\theta(u) du}{\int_0^\infty K_i(u, x_i) g_\theta(u) du}.$$

If we now evaluate the score at  $(\gamma, \boldsymbol{\eta}) = (0, \hat{\boldsymbol{\eta}}_0)$ , we obtain

$$U_\gamma(0, \hat{\boldsymbol{\eta}}_0) = \sum_{i=1}^m \frac{\int_0^\infty [N_i^D(x_i) - e^{\hat{\alpha}^\top z_i} \hat{H}_0(x_i)] \ln(u) \hat{K}_i(u, x_i) g_\theta(u) du}{\int_0^\infty \hat{K}_i(u, x_i) g_\theta(u) du},$$

which is the first line of (5.8).

### 5.7.3 Likelihood with fixed observation times

In this section, we study the likelihood of the joint frailty model (5.2) based on individual interval counts of recurrent events in one particular setting. Specifically, the observation times are assumed to be the same for all individuals and the baseline rate of recurrence  $\lambda_0(t)$  is modeled as a piecewise constant function. We will show that if the

observation intervals are finer than the intervals for the rate pieces, the score depends on the individual's interval counts only through the sums of these counts over each baseline rate piece.

For that purpose, suppose the observation times  $t_{ij}$  are given by  $t_{ij} = t_j$  for  $j = 0, \dots, J_i - 1$ , and the last observation time  $t_{iJ_i} = X_i$  is equal to the follow-up time  $X_i$ . The interval counts  $n_{ij}$ , observed over the intervals  $I_{ij} = (t_{ij-1}, t_{ij}]$ , enter the likelihood contribution  $L_i^{(c)}(u_i)$  of individual  $i$  given its frailty value  $u_i$  (see (5.3)) only through the factor

$$\prod_{j=1}^{J_i} \frac{1}{n_{ij}!} \exp(-u_i \mu_{ij}) (u_i \mu_{ij})^{n_{ij}} = \left( \prod_{j=1}^{J_i} \frac{1}{n_{ij}!} \right) \cdot u_i^{n_i} \cdot \exp\left(-u_i \sum_{j=1}^{J_i} \mu_{ij}\right) \prod_{j=1}^{J_i} \mu_{ij}^{n_{ij}} \quad (5.11)$$

with  $n_i = \sum_{j=1}^{J_i} n_{ij}$  and  $\mu_{ij} = \int_{I_{ij}} e^{\beta^\top z_i} \lambda_0(s) ds$ .

For a baseline rate of recurrence specified as a piecewise constant function on intervals  $I_k^R = (t_{k-1}^R, t_k^R]$  of length  $\Delta_k^R = (t_k^R - t_{k-1}^R)$ ,  $k = 1, \dots, K^R$ , as in (5.6), we have

$$\mu_{ij} = e^{\beta^\top z_i} \sum_{k=1}^{K^R} \lambda_{0k} \int_0^\infty \mathbb{1}\{s \in I_{ij} \cap I_k^R\} ds = e^{\beta^\top z_i} \sum_{k=1}^{K^R} \lambda_{0k} |I_{ij} \cap I_k^R|, \quad (5.12)$$

where  $|I|$  denotes the length of the interval  $I$ . Now let  $K_i$  be the index  $k$  of the interval  $I_k^R$  for which  $X_i \in I_{K_i}^R$ . We can write the sum in the third term in (5.11) as

$$\begin{aligned} \sum_{j=1}^{J_i} \mu_{ij} &= e^{\beta^\top z_i} \sum_{j=1}^{J_i} \sum_{k=1}^{K^R} \lambda_{0k} |I_{ij} \cap I_k^R| = e^{\beta^\top z_i} \sum_{k=1}^{K^R} \lambda_{0k} \sum_{j=1}^{J_i} |I_{ij} \cap I_k^R| \\ &= e^{\beta^\top z_i} \sum_{k=1}^{K^R} \lambda_{0k} \left| \bigcup_{j=1}^{J_i} I_{ij} \cap I_k^R \right| = e^{\beta^\top z_i} \left\{ \sum_{k=1}^{K_i-1} \lambda_{0k} \Delta_k^R + \lambda_{0K_i} (X_i - t_{K_i-1}^R) \right\}, \quad (5.13) \end{aligned}$$

which depends only on the individual's follow-up time  $X_i$ , but not on the other observation times  $t_{ij}$ ,  $j = 0, \dots, J_i - 1$ .

Using (5.12), the last term in (5.11) equals

$$\prod_{j=1}^{J_i} \mu_{ij}^{n_{ij}} = (e^{\beta^\top z_i})^{n_i} \prod_{j=1}^{J_i} \left( \sum_{k=1}^{K^R} \lambda_{0k} |I_{ij} \cap I_k^R| \right)^{n_{ij}}.$$

Suppose now we choose the intervals  $I_k^R$  for the baseline rate of recurrence to be rougher than the observation intervals, but such that the  $t_k^R$  are a subset of the  $t_j$ . Then, each  $I_{ij}$ ,  $j = 1, \dots, J_i$ , is a subset of exactly one  $I_k^R$ , thus,

$$\prod_{j=1}^{J_i} \left( \sum_{k=1}^{K^R} \lambda_{0k} |I_{ij} \cap I_k^R| \right)^{n_{ij}} = \prod_{j=1}^{J_i} \prod_{k=1}^{K^R} \left( \lambda_{0k} |I_{ij} \cap I_k^R| \right)^{n_{ij} \cdot \mathbb{1}\{I_{ij} \subset I_k^R\}}$$

$$\begin{aligned}
&= \prod_{k=1}^{K^R} \prod_{j=1}^{J_i} \lambda_{0k}^{n_{ij} \cdot \mathbb{1}\{I_{ij} \subset I_k^R\}} |I_{ij} \cap I_k^R|^{n_{ij} \cdot \mathbb{1}\{I_{ij} \subset I_k^R\}} \\
&= \left( \prod_{k=1}^{K^R} \lambda_{0k}^{\sum_{j=1}^{J_i} n_{ij} \cdot \mathbb{1}\{I_{ij} \subset I_k^R\}} \right) \left( \prod_{k=1}^{K^R} \prod_{j=1}^{J_i} (|I_{ij} \cap I_k^R|)^{n_{ij} \cdot \mathbb{1}\{I_{ij} \subset I_k^R\}} \right). \quad (5.14)
\end{aligned}$$

Note that  $\tilde{n}_{ik} = \sum_{j=1}^{J_i} n_{ij} \cdot \mathbb{1}\{I_{ij} \subset I_k^R\}$  gives the number of recurrent events which individual  $i$  experiences over the time interval  $[0, X_i] \cap I_k^R$ . Hence, the first factor of expression (5.14) depends on the counts  $n_{ij}$  which were observed over the smaller intervals  $I_{ij}$  only via the counts  $\tilde{n}_{ik}$  corresponding to the larger intervals  $I_k^R$ .

Inserting (5.13) and (5.14) into (5.11) yields

$$\begin{aligned}
&\left( \prod_{j=1}^{J_i} \frac{1}{n_{ij}!} \right) \cdot u_i^{n_i} \cdot \exp \left\{ -u_i e^{\beta^\top z_i} \left( \sum_{k=1}^{K_i-1} \lambda_{0k} \Delta_k^R + \lambda_{0K_i} (X_i - t_{K_i-1}^R) \right) \right\} \\
&\cdot (e^{\beta^\top z_i})^{n_i} \prod_{k=1}^{K^R} \lambda_{0k}^{\tilde{n}_{ik}} \prod_{k=1}^{K^R} \prod_{j=1}^{J_i} (|I_{ij} \cap I_k^R|)^{n_{ij} \cdot \mathbb{1}\{I_{ij} \subset I_k^R\}} \\
&\propto (u_i e^{\beta^\top z_i})^{n_i} \exp \left\{ -u_i e^{\beta^\top z_i} \left( \sum_{k=1}^{K_i-1} \lambda_{0k} \Delta_k^R + \lambda_{0K_i} (X_i - t_{K_i-1}^R) \right) \right\} \left( \prod_{k=1}^{K^R} \lambda_{0k}^{\tilde{n}_{ik}} \right).
\end{aligned}$$

Consequently, if the observation intervals are finer than the rate intervals the likelihood is proportional to an expression that depends only on recurrent event interval counts corresponding to the larger rate intervals.

## 5.8 Supplementary material: Extended simulation study

In this section, we present additional figures and tables illustrating the results of our simulation study. Subsection 5.8.1 focuses on the performance of the method for estimating the joint frailty model based on individual interval counts of recurrent events. Subsection 5.8.2 deals with the performance of the score test for the association between the recurrent events and death.

### 5.8.1 Performance of the estimation method

The basic set-up for the simulation study was described in Section 5.4.1. Table 5.3 gives an overview of the scenarios covered in this supplementary material. All samples include  $m = 200$  individuals. Data were generated from a joint frailty model with a binary covariate having effects  $\alpha = \beta = 1$  on the hazard of death and the rate of recurrence. Independent censoring occurs uniformly over the total follow-up window  $[0, 2]$ .

In Section 5.4.1, we focused on scenarios in which the values of the frailty variance  $\theta$  and the dependence parameter  $\gamma$  are fixed at  $\theta = 0.5$  and  $\gamma = \pm 1$ . Here, we also study the effect of varying the frailty variance by including scenarios with  $\theta \in \{0.25, 0.75\}$ . Moreover, we present settings with  $\theta = 0.05$  which are motivated by the estimates obtained in the application (see Section 5.5).

Regarding the observation times  $t_{ij}$ , that define the intervals for which the recurrent event counts are observed, these times can be fixed and the same for all individuals (Scenario I) or varying across individuals (Scenario II), as explained in Section 5.4.1. In Scenarios I.G and II.D, the widths of these observation intervals are changed as well.

The baseline rates are generally specified as piecewise constant (pwc) functions on  $K^R = K^D = K^*$  intervals of equal length, where we also assess the impact of increasing  $K^*$  from 10 to 20 (Scenarios I.F and I.G). In contrast, in Scenario I.E, a simple parametric model is used for the baseline rates by assuming (in this case correctly) a Weibull model.

Concerning the independent censoring, in Section 5.4.1, we have compared uniform censoring over the total follow-up window with the alternative of end-of-study censoring at  $t = 2$ . Table 5.4 reports on the resulting proportions of censoring under the different mechanisms across 200 replications.

Table 5.3: Simulation settings.

Scenario	$ \gamma $	$\theta$	obs. times $t_{ij}$	rates	$K^*$	Figure
Varying the frailty variance in Scenario I						
I.A	1	0.25	$0.2j$	pwc	10	5.5
I.B	1	0.75	$0.2j$	pwc	10	5.6
I.C	1	0.05	$0.2j$	pwc	10	5.7
I.D	5	0.05	$0.2j$	pwc	10	5.8
Varying the specification of the baseline rates in Scenario I						
I.E	1	0.5	$0.2j$	Weibull	-	5.9
I.F	1	0.5	$0.2j$	pwc	20	5.10
Varying the width of the observation intervals in Scenario I						
I.G	1	0.5	$0.1j$	pwc	20	5.11
Varying the frailty variance in Scenario II						
II.A	1	0.25	$0.2j + \varepsilon_{ij}$	pwc	10	5.12
II.B	1	0.5	$0.2j + \varepsilon_{ij}$	pwc	10	5.13
II.C	1	0.75	$0.2j + \varepsilon_{ij}$	pwc	10	5.14
Varying the width of the observation intervals in Scenario II						
II.D	1	0.5	$0.1j + \varepsilon_{ij}$	pwc	10	5.15

The  $\varepsilon_{ij}$  are drawn from a uniform distribution (see Section 5.4.1).

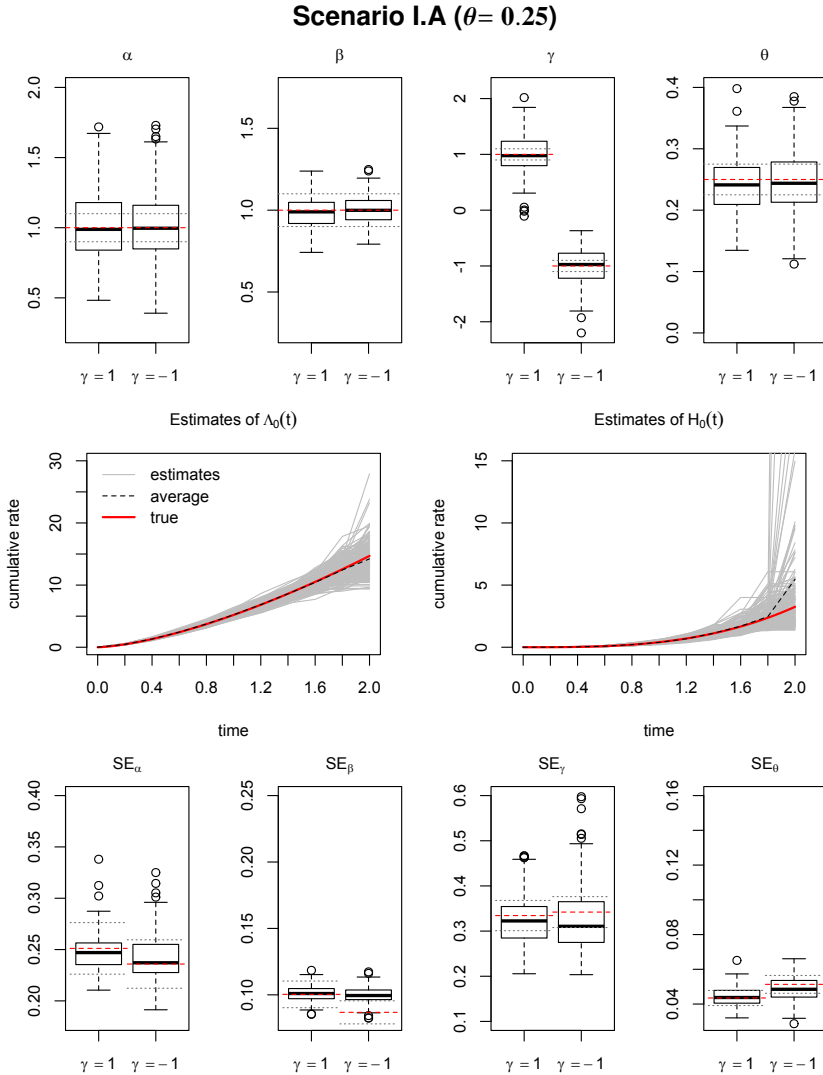


Figure 5.5: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .

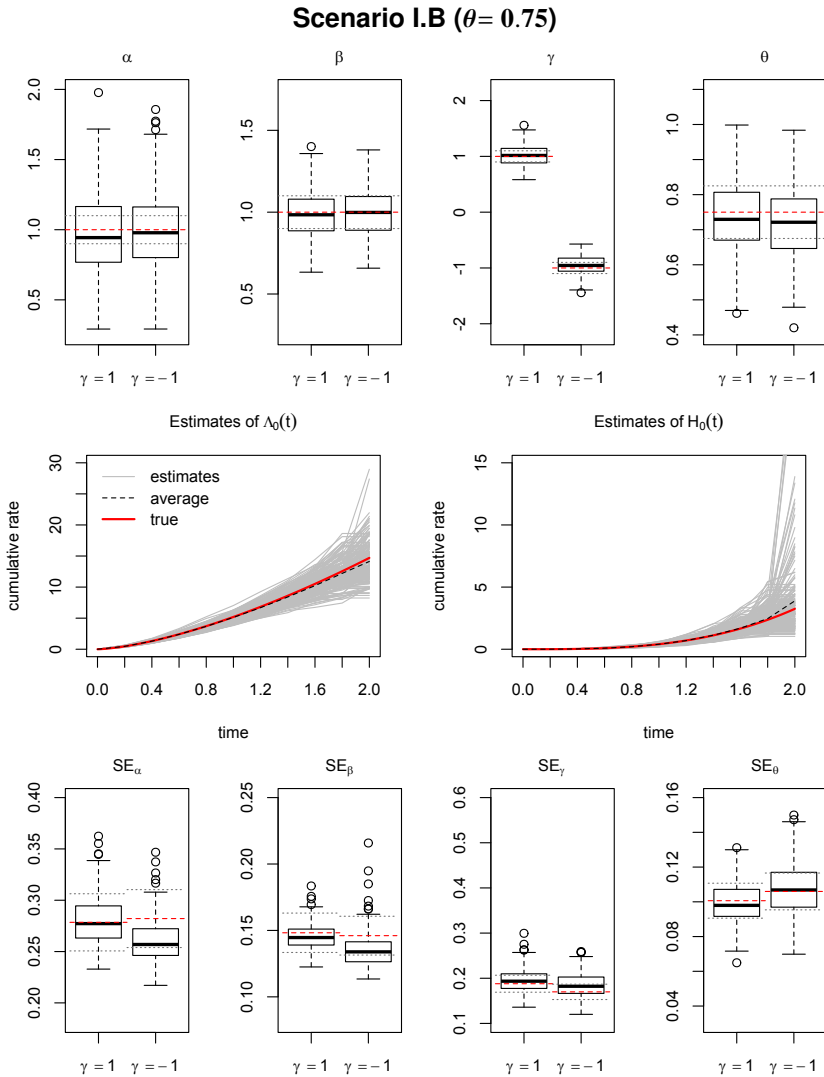


Figure 5.6: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .



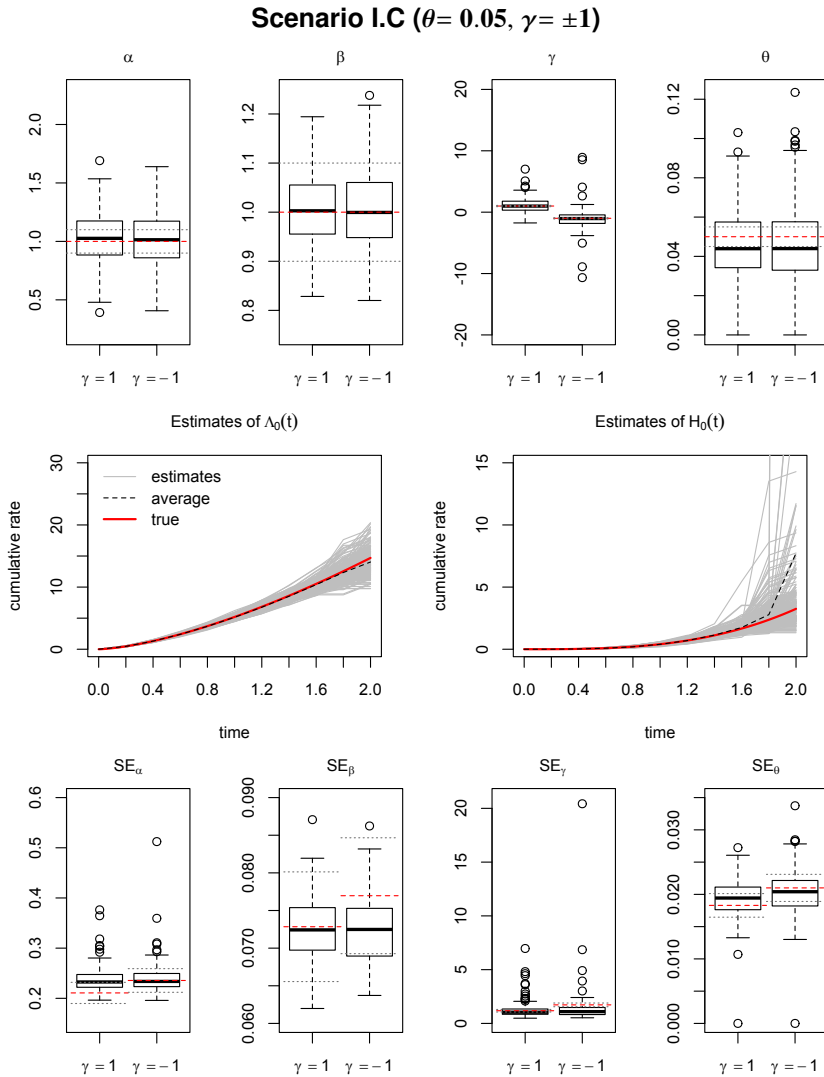


Figure 5.7: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .

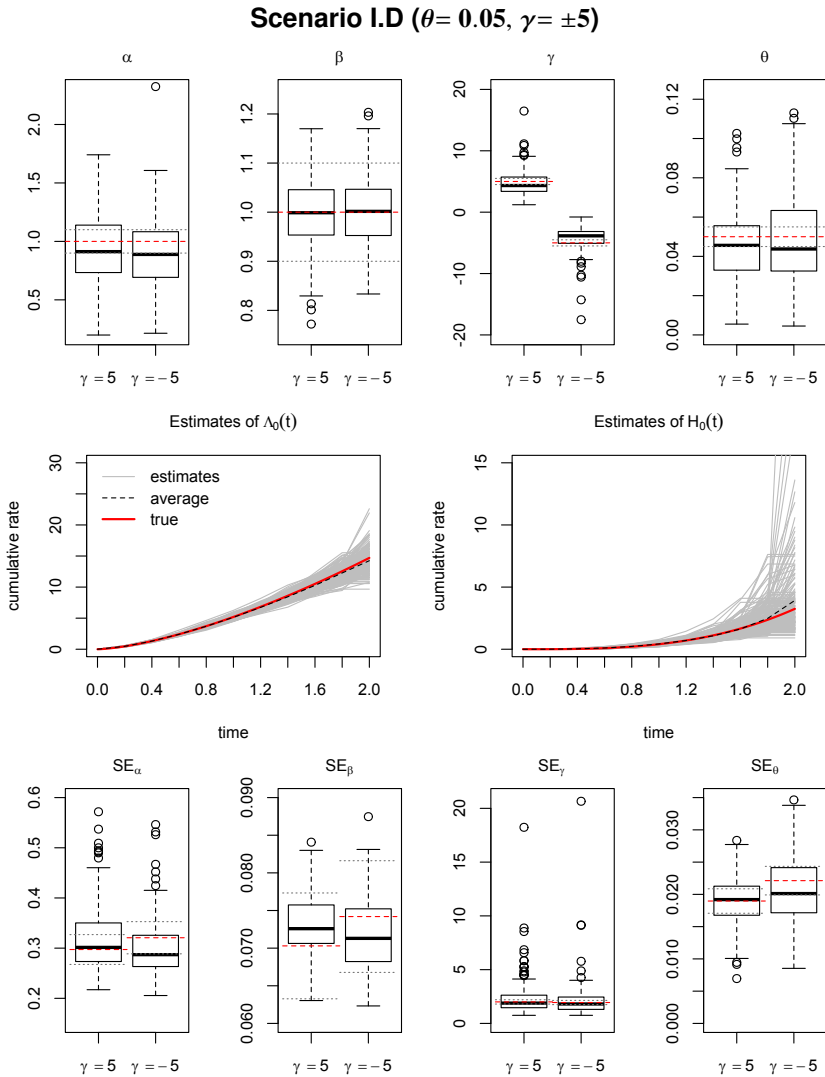


Figure 5.8: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 5$ ) and negative dependence ( $\gamma = -5$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 5$ .

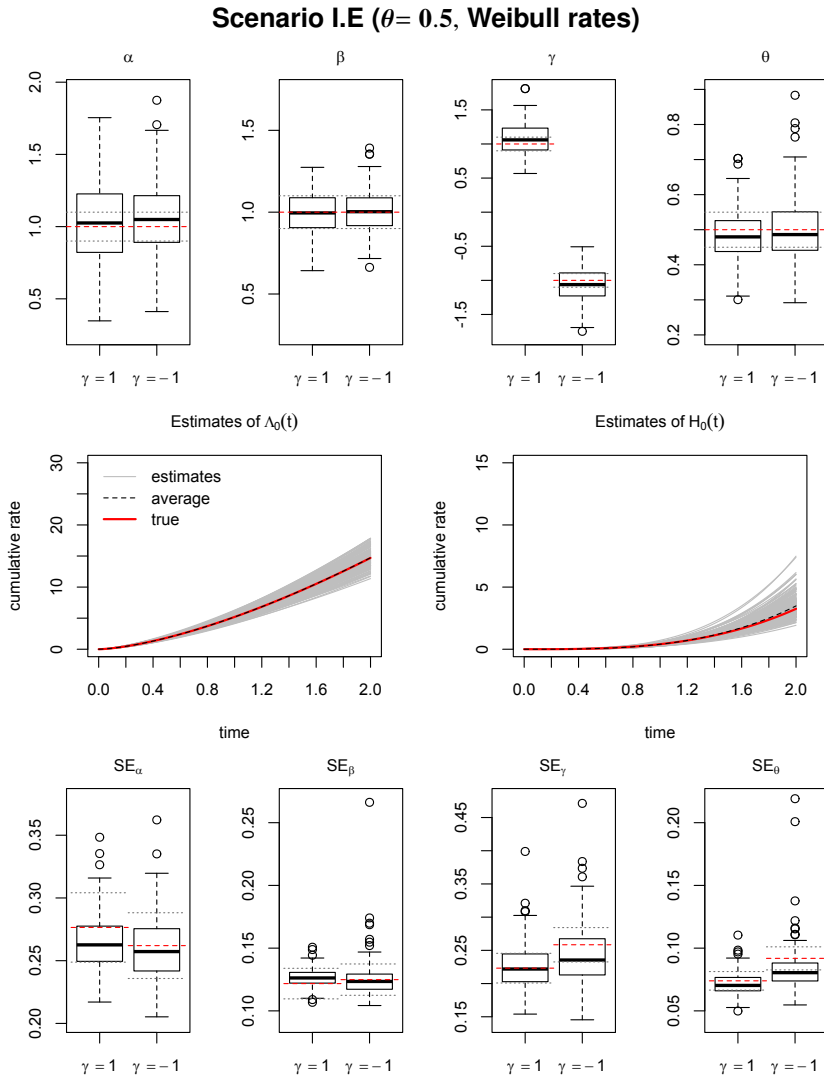


Figure 5.9: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .

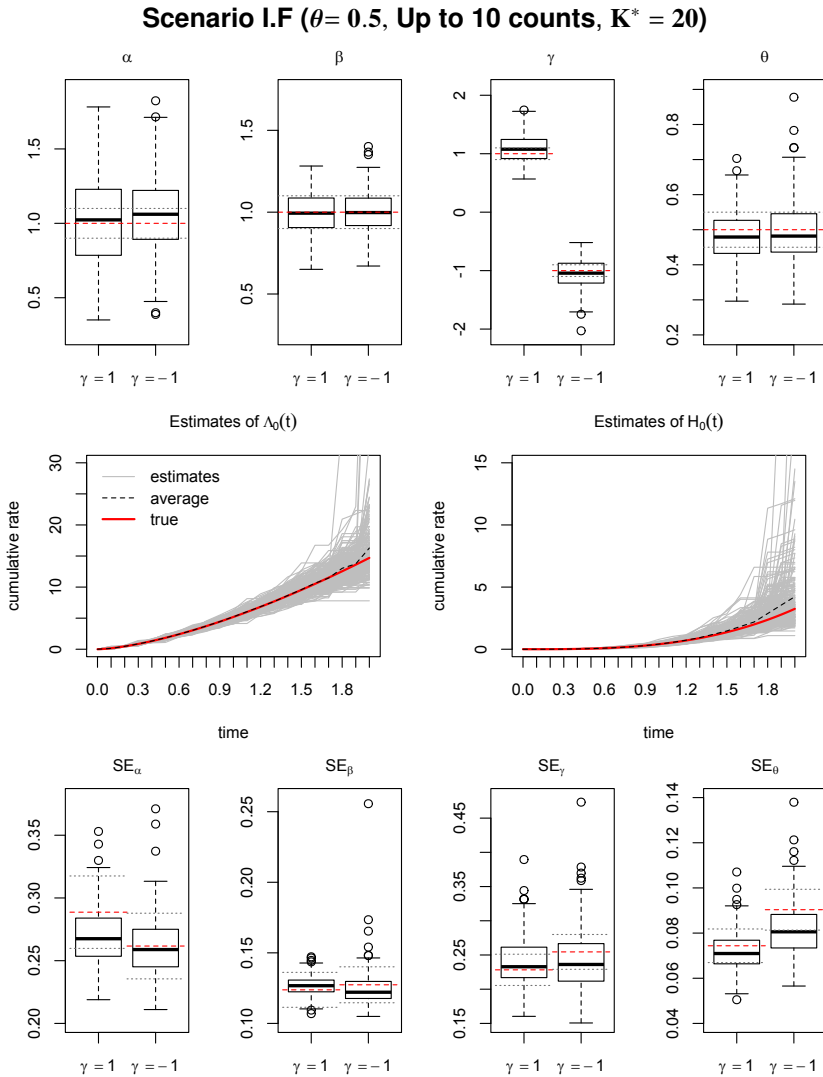


Figure 5.10: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .

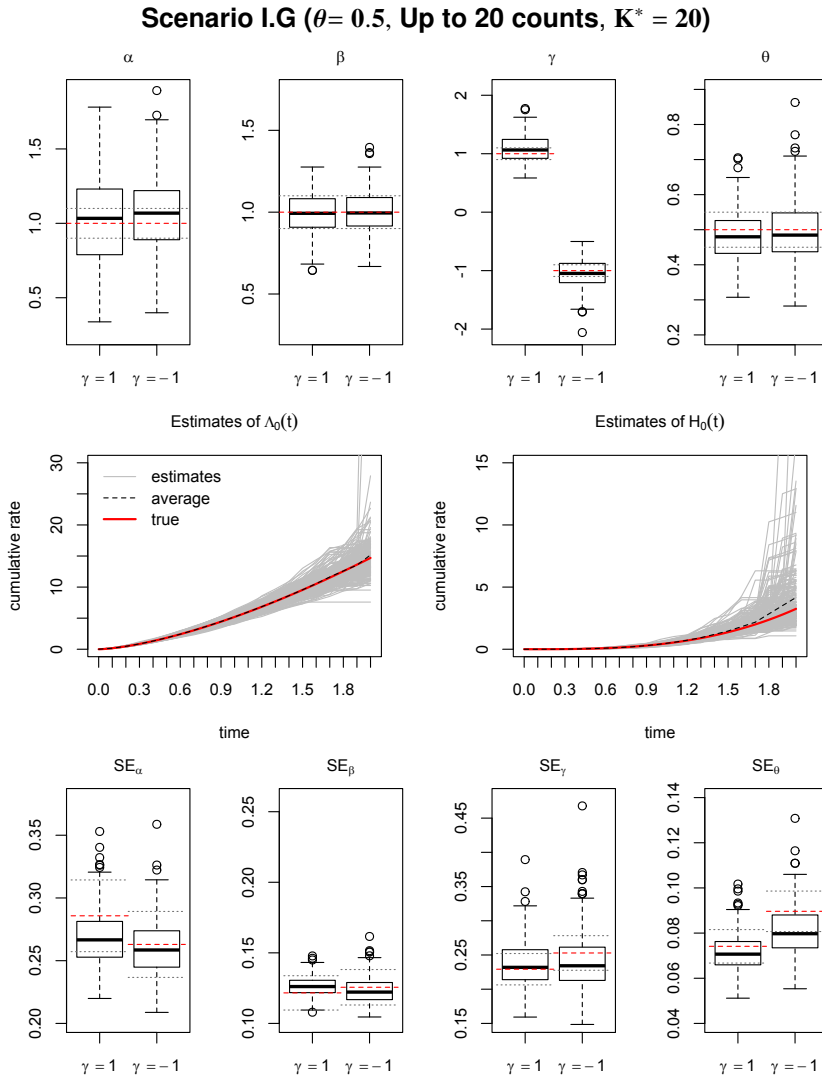


Figure 5.11: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .

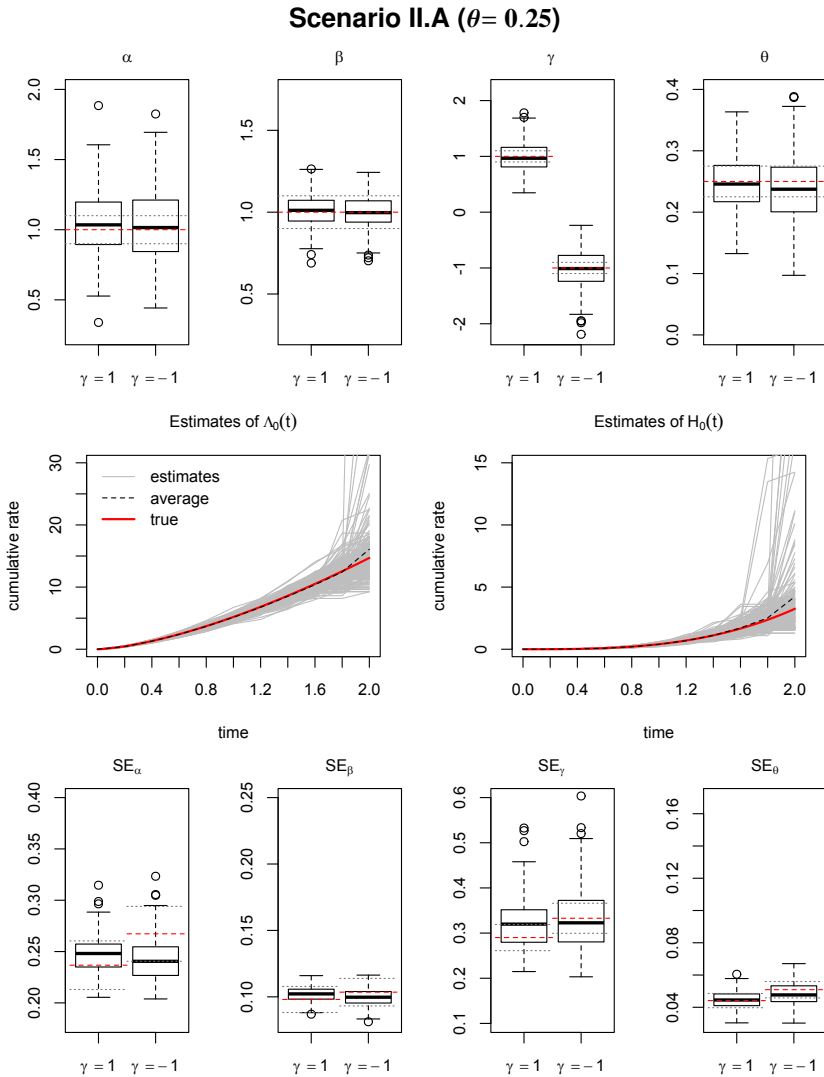


Figure 5.12: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .

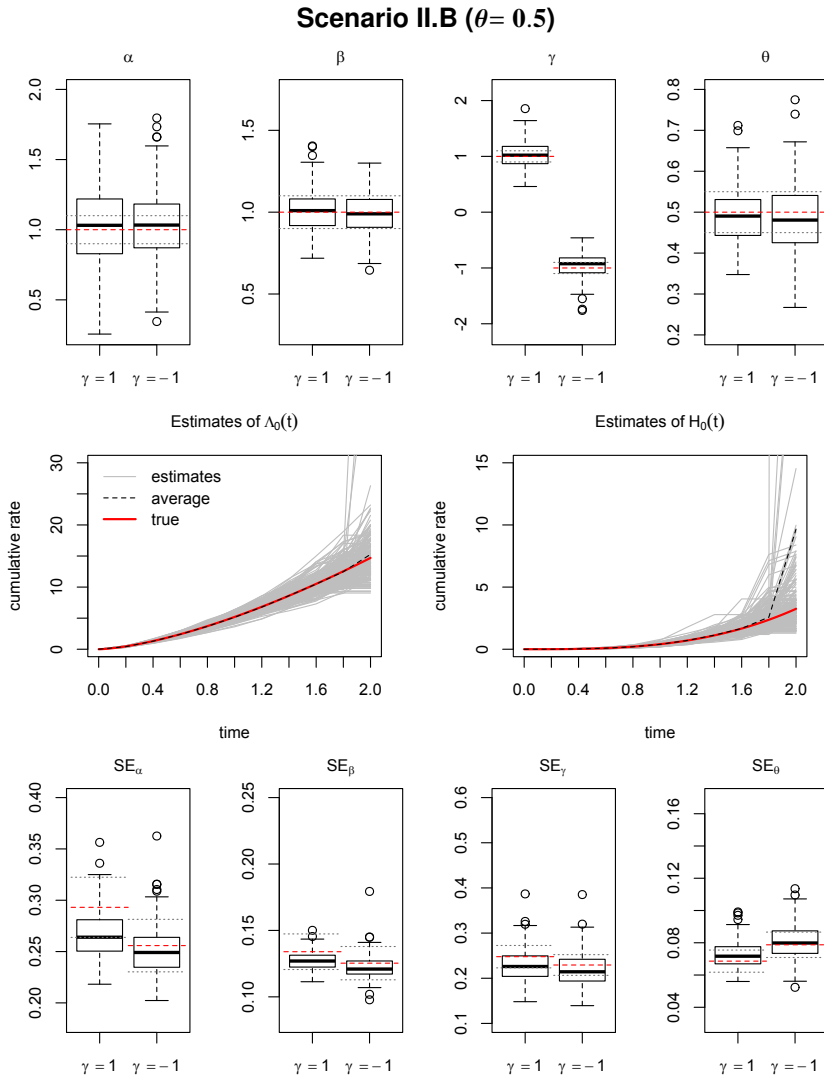


Figure 5.13: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .

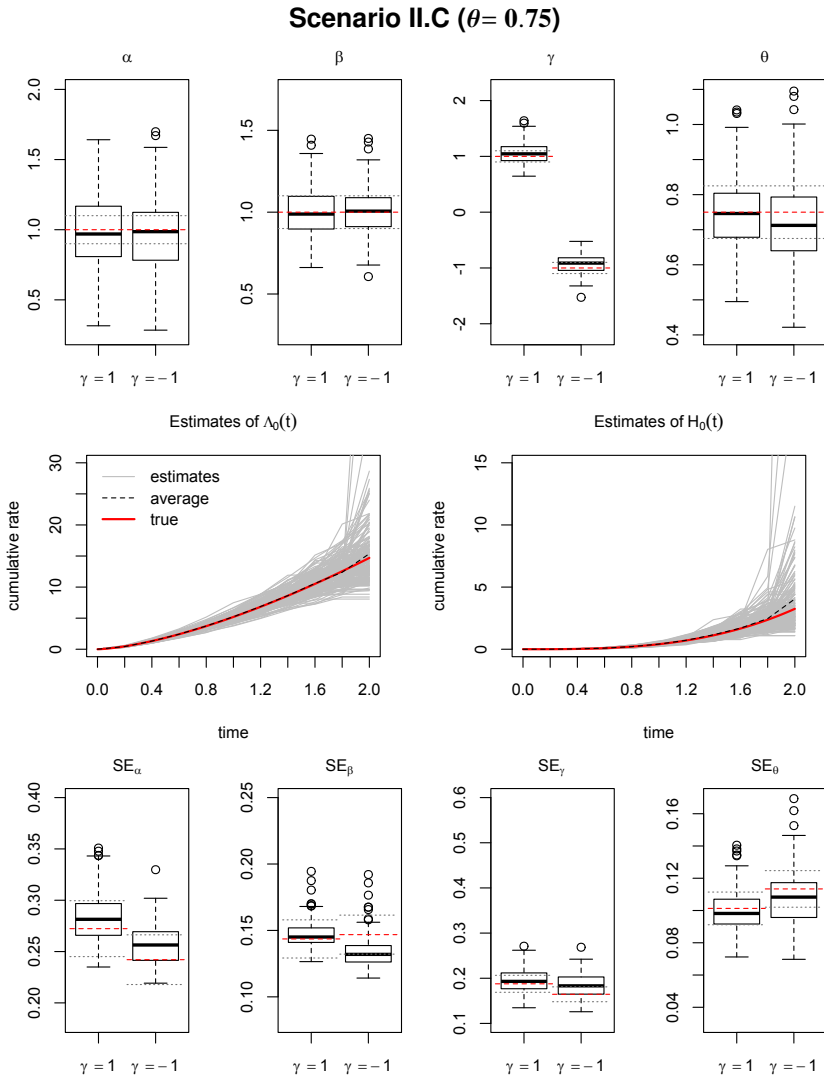


Figure 5.14: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .



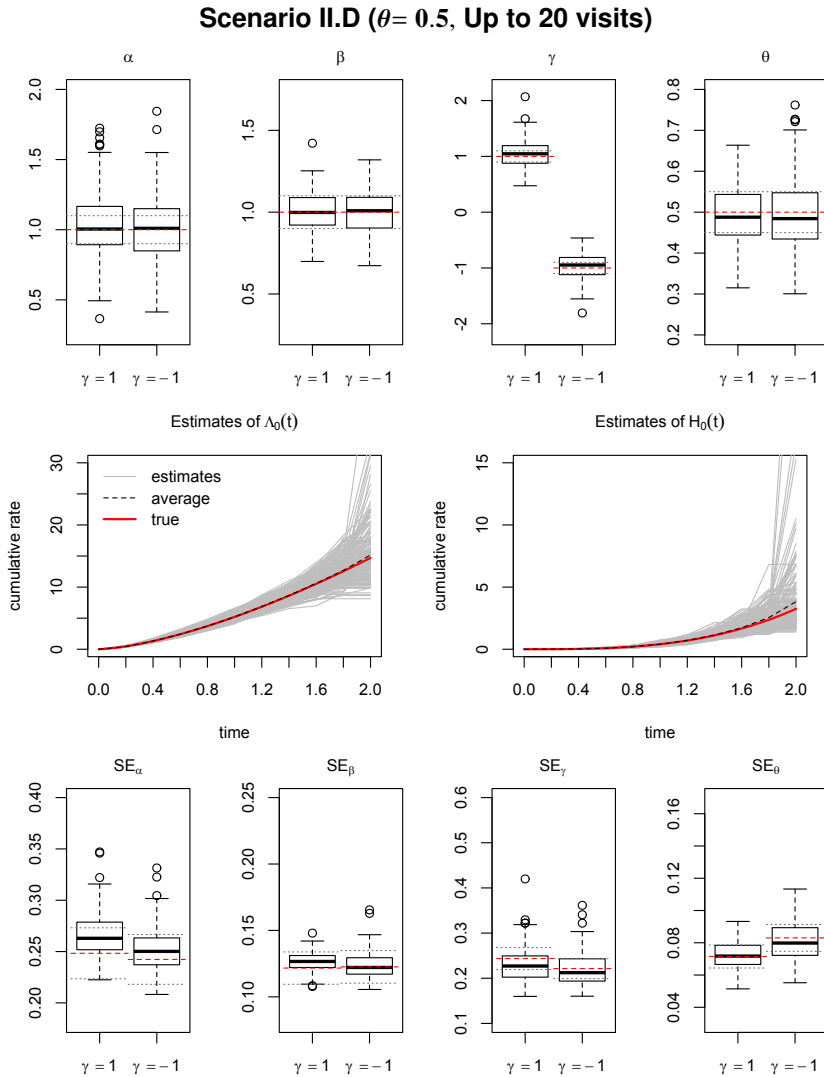


Figure 5.15: Box plots of the parameter estimates (top row) and estimated standard errors (bottom row) in the joint frailty model for positive ( $\gamma = 1$ ) and negative dependence ( $\gamma = -1$ ). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ), based on 200 samples of size  $m = 200$ . Red, dashed line marks true value (top row) or empirical standard deviation (bottom row); gray, dotted lines mark 10% deviations from respective value. Middle row: estimates (gray, solid) of the cumulative rate of recurrence (left) and of death (right) in the joint frailty model with  $\gamma = 1$ .

Table 5.4: Proportion of censoring under different censoring mechanisms for the joint frailty model with frailty variance  $\theta = 0.5$  in 200 replications of Scenario I.

$\gamma$	censoring	censoring proportion		
		min	mean	max
1	$C \sim \mathcal{U}[0, 2]$	0.445	0.564	0.650
	$C = 2$	0.040	0.091	0.140
-1	$C \sim \mathcal{U}[0, 2]$	0.365	0.484	0.580
	$C = 2$	0.000	0.032	0.065

## 5.8.2 Performance of the score test

The general design of the simulation study regarding the score test was laid out in Section 5.4.2. In this supplementary material, we present the results for additional scenarios as summarized in Table 5.5. The samples of  $m = 200$  individuals each are drawn from a joint frailty model with a binary covariate with effects  $\alpha = \beta = 1$  on the hazard of death and the rate of recurrence. Independent censoring is taken to occur uniformly over the total follow-up window  $[0, 2]$ . We assess the power or size, respectively, of the score test for the different combinations of the dependence parameter  $\gamma$  varying in  $\Gamma = \{-1, -0.5, 0, 0.5, 1\}$  and the frailty variance  $\theta$  varying in  $\Theta = \{0.25, 0.5, 0.75\}$ .

In Scenario I.A, we extend this to consider stronger dependence,  $\gamma = \pm 5$ , and a smaller frailty variance,  $\theta = 0.05$ , inspired by the estimates we obtained for the real data (see Section 5.5).

To study the effect of the width of the observation intervals for the recurrent event counts, the observation times  $t_{ij}$  are fixed and the same for all individuals in Scenario I, but varying across individuals in Scenario II. Moreover, the test performance based on finer observation intervals is examined for each of the scenarios (Scenarios I.C and II.B).

The baseline rate of recurrence is specified as a piecewise constant function on  $K^R = 10$  or  $K^R = 20$  intervals of equal length.

In Section 5.4.1, two versions of independent censoring were considered: uniform censoring over the total follow-up window or end-of-study censoring at  $t = 2$ . The resulting proportions of censored individuals under the two mechanisms across 1000 replications are given in Table 5.6.

Table 5.5: Simulation settings for the score test with dependence parameter varying in  $\Gamma = \{-1, -0.5, 0, 0.5, 1\}$  and frailty variance varying in  $\Theta = \{0.25, 0.5, 0.75\}$ .

Scenario	$\gamma$	$\theta$	obs. times $t_{ij}$	$K^R$	Table
Varying the frailty variance in Scenario I					
I.A	$\Gamma \cup \{\pm 5\}$	$\Theta \cup \{0.05\}$	$0.2j$	10	5.7
Varying the specification of the baseline rate in Scenario I					
I.B	$\Gamma$	$\Theta$	$0.2j$	20	5.8
Varying the width of the observation intervals in Scenario I					
I.C	$\Gamma$	$\Theta$	$0.1j$	20	5.9
Varying the observation times across individuals (Scenario II)					
II.A	$\Gamma$	$\Theta$	$0.2j + \varepsilon_{ij}$	10	5.10
Varying the width of the observation intervals in Scenario II					
II.B	$\Gamma$	$\Theta$	$0.1j + \varepsilon_{ij}$	10	5.11

The  $\varepsilon_{ij}$  are drawn from a uniform distribution (see Section 5.4.1).

Table 5.6: Proportion of censoring under different censoring mechanisms for the joint frailty model with dependence parameter  $\gamma = 0$  in 1000 replications of Scenario I.

$\theta$	censoring	censoring proportion		
		min	mean	max
0.25	$C \sim \mathcal{U}[0, 2]$	0.365	0.516	0.625
	$C = 2$	0.000	0.019	0.070
0.5	$C \sim \mathcal{U}[0, 2]$	0.400	0.516	0.645
	$C = 2$	0.000	0.020	0.055
0.75	$C \sim \mathcal{U}[0, 2]$	0.405	0.515	0.645
	$C = 2$	0.000	0.019	0.060

Table 5.7: Power and size of the score test, performed at the 5% level, in the joint frailty model with varying dependence parameter  $\gamma \in \{-5, -1, -0.5, 0, 0.5, 1, 5\}$  and frailty variance  $\theta \in \{0.05, 0.25, 0.5, 0.75\}$ , across 1000 replications each.

$\theta$	Dependence $\gamma$						
	-5	-1	-0.5	0	0.5	1	5
0.05	0.938	0.290	0.142	0.081	0.116	0.263	0.931
0.25	1.000	0.960	0.586	0.066	0.514	0.932	1.000
0.5	1.000	1.000	0.905	0.062	0.872	1.000	1.000
0.75	1.000	1.000	0.984	0.091	0.974	1.000	1.000

Table 5.8: Power and size of the score test, performed at the 5% level, in the joint frailty model with varying dependence parameter  $\gamma \in \{-1, -0.5, 0, 0.5, 1\}$ , frailty variance  $\theta \in \{0.25, 0.5, 0.75\}$ , and  $K^R = 20$ , across 1000 replications each.

$\theta$	Dependence $\gamma$				
	-1	-0.5	0	0.5	1
0.25	0.948	0.525	0.058	0.530	0.929
0.5	0.998	0.878	0.070	0.870	1.000
0.75	1.000	0.974	0.087	0.975	1.000

Table 5.9: Power and size of the score test, performed at the 5% level, in the joint frailty model with varying dependence parameter  $\gamma \in \{-1, -0.5, 0, 0.5, 1\}$ , frailty variance  $\theta \in \{0.25, 0.5, 0.75\}$ , up to 20 counts, and  $K^R = 20$ , across 1000 replications each.

$\theta$	Dependence $\gamma$				
	-1	-0.5	0	0.5	1
0.25	0.955	0.556	0.062	0.533	0.940
0.5	0.999	0.890	0.063	0.873	1.000
0.75	1.000	0.979	0.090	0.977	1.000

Table 5.10: Power and size of the score test, performed at the 5% level, in the joint frailty model with varying dependence parameter  $\gamma \in \{-1, -0.5, 0, 0.5, 1\}$ , frailty variance  $\theta \in \{0.25, 0.5, 0.75\}$ , and observation times varying across individuals (Scenario II), across 1000 replications each.

$\theta$	Dependence $\gamma$				
	-1	-0.5	0	0.5	1
0.25	0.943	0.533	0.071	0.543	0.962
0.5	1.000	0.879	0.075	0.875	0.998
0.75	1.000	0.978	0.065	0.971	1.000

Table 5.11: Power and size of the score test, performed at the 5% level, in the joint frailty model with varying dependence parameter  $\gamma \in \{-1, -0.5, 0, 0.5, 1\}$ , frailty variance  $\theta \in \{0.25, 0.5, 0.75\}$ , up to 20 visits, and  $K^R = 10$ , across 1000 replications each.

$\theta$	Dependence $\gamma$				
	-1	-0.5	0	0.5	1
0.25	0.941	0.540	0.079	0.511	0.947
0.5	1.000	0.890	0.074	0.863	1.000
0.75	1.000	0.989	0.069	0.981	1.000

## Supporting information

Source code to reproduce the simulation studies and perform the analysis of Section 5.5 is available as Supporting Information at <https://onlinelibrary.wiley.com/doi/10.1002/bimj.201900367>.

## References

- Balan, T. A., S. E. Boonk, M. H. Vermeer, and H. Putter (2016). Score test for association between recurrent events and a terminal event. *Statistics in Medicine* 35, 3037–3048.
- Cook, R. J. and J. F. Lawless (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* 16, 911–924.
- Cook, R. J. and J. F. Lawless (2007). *The Statistical Analysis of Recurrent Events*. Statistics for Biology and Health. New York: Springer.
- Dańko, A., R. Schaible, and M. J. Dańko (2020). Salinity effects on survival and reproduction of hydrozoan *Eleutheria dichotoma*. *Estuaries and Coasts* 43, 360–374.
- Ghosh, D. and D. Y. Lin (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* 59, 877–885.
- Gilbert, P. and R. Varadhan (2019). *numDeriv: Accurate Numerical Derivatives*. R package version 2016.8-1.1. <https://CRAN.R-project.org/package=numDeriv>.
- Huang, C.-Y. and M.-C. Wang (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* 99, 1153–1165.
- Huang, X., R. A. Wolfe, and C. Hu (2004). A test for informative censoring in clustered survival data. *Statistics in Medicine* 23, 2089–2107.
- Lancaster, T. and O. Intrator (1998). Panel data with survival: Hospitalization of HIV-positive patients. *Journal of the American Statistical Association* 93, 46–53.
- Lawless, J. F. and M. Zhan (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 26, 549–565.
- Liu, L. and X. Huang (2008). The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine* 27, 2665–2683.
- Liu, L., R. A. Wolfe, and X. Huang (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* 60, 747–756.

- Nelson, K. P., S. R. Lipsitz, G. M. Fitzmaurice, J. Ibrahim, M. Parzen, and R. Strawderman (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics* 15, 39–57.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rondeau, V., S. Mathoulin-Pelissier, H. Jacqmin-Gadda, V. Brouste, and P. Soubeyran (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* 8(4), 708–721.
- Sinha, D. and T. Maiti (2004). A Bayesian approach for the analysis of panel-count data with dependent termination. *Biometrics* 60, 34–40.
- Smyth, G. K. (1998). Numerical integration. In P. Armitage and T. Colton (Eds.), *Encyclopedia of Biostatistics*, pp. 3088–3095. London: Wiley.
- Therneau, T. M. and P. M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.
- Zhao, H., Y. Li, and J. Sun (2013). Analyzing panel count data with a dependent observation process and a terminal event. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 41, 174–191.



# 6

## Incorporating delayed entry into the joint frailty model for recurrent events and a terminal event

### Abstract

In studies of recurrent events, joint modeling approaches are often needed to allow for potential dependent censoring by a terminal event such as death. Joint frailty models for recurrent events and death with an additional dependence parameter have been studied for cases in which individuals are observed from the start of the event processes. However, the samples are often selected at a later time, which results in delayed entry. Thus, only individuals who have not yet experienced the terminal event will be included in the study. We propose a method for estimating the joint frailty model from such left-truncated data. The frailty distribution among the selected survivors differs from the frailty distribution in the underlying population if the recurrence process and the terminal event are associated. The correctly adjusted marginal likelihood can be expressed as a ratio of two integrals over the frailty distribution, which may be approximated using Gaussian quadrature. The baseline rates are specified as piecewise constant functions, and the covariates are assumed to have multiplicative effects on the event rates. We assess the performance of the

---

This chapter has been submitted for publication as: M. Böhstedt, J. Gampe, M.A.A. Caljouw, and H. Putter. Incorporating delayed entry into the joint frailty model for recurrent events and a terminal event.



estimation procedure in a simulation study, and apply the method to estimate age-specific rates of recurrent urinary tract infections and mortality in an older population.

## 6.1 Introduction

Repeated occurrences of the same type of event in one individual arise in various applications. Examples of such recurrent event data include incidents of myocardial infarction, recurrent infections, fractures, or tumor relapses.

If the individual is additionally at risk of experiencing a terminal event such as death, which will stop the recurrent event process, this might induce dependent censoring of the recurrence process. Therefore, approaches for jointly modeling the two processes of the recurrent events and the terminal event have been developed. Moreover, studies might explicitly address the question of whether there is an association between the processes by asking, for instance, whether individuals who experience more recurrences also have a higher risk of experiencing the terminal event. Specific joint models can provide additional insights into the direction and the strength of the association between the event processes.

The choice of the time scale  $t$ , along which the recurrence process and the terminal event process are assumed to evolve, depends on the specific application. In clinical studies, the time since randomization is often used as the time scale; whereas in demographic studies of fertility, for instance, the most relevant time scale is age.

For a given time scale, the event processes can, in some cases, be observed from the time origin  $t_0$ . Consider, for example, a medical study in which the time since the disease onset or diagnosis is used as the time scale. If each individual is observed from the disease onset or diagnosis onwards, then all individuals enter the study at time  $t_0 = 0$ .

However, studies are often initiated at a later point in time when the two processes have already started. If in the above setting patients do not participate in the clinical study until some period of time after their diagnosis, individual times of study entry will differ from 0, and might also vary between patients. Similarly, in studies of certain diseases in old-age populations, such as in register-based studies of cardiovascular disease and mortality, age can be considered the natural time scale (see, for instance, Modig et al., 2013). The individuals included in such analyses have already reached a certain advanced age by the beginning of the study period.

In these cases, the data are left-truncated in the sense that the individuals can enter the study only if they have not yet experienced the terminal event. If the recurrence process and the terminal event process are associated, this sample of survivors is not a random sample of the underlying population. Rather, the sample is comprised of individuals who tend to have a lower risk of experiencing the terminal event, and – if there is a positive association between the two processes – to also have a lower risk of experiencing recurrent events. Thus, to obtain valid inferences, correctly adjusting for the truncation is crucial.

The example we use in this study to illustrate our method focuses on recurrent urinary tract infections (UTIs) in older residents of long-term care facilities (LTCF). The original

study by Caljouw et al. (2014) investigated whether cranberry capsules are effective in preventing UTIs. Given that around 34% of the elderly study population died during the follow-up period, recurrent UTIs and mortality had to be modeled jointly. The original study used time since randomization as the time scale. By contrast, we have chosen to use age as the main time scale of the event processes, as mortality and, presumably, UTI recurrences naturally depend on age. Because the participants were between 64 and 102 years old when they entered the study, the observations are then left-truncated. In our analysis, we seek to assess the effects of the cranberry treatment, to estimate the age-specific rates of UTIs and death, and to determine whether there is an association between UTIs and death risks.

While different joint models for recurrent events and death have been proposed, we focus here on the joint frailty model introduced by Liu et al. (2004). This model has been applied repeatedly in medical studies (e.g., to study recurrent cancer events as in Rondeau et al., 2007, or recurrent heart failure hospitalizations as in Rogers et al., 2016), and extended in several directions (e.g., to the setting of nested case-control studies in Jazić et al., 2019). The joint frailty model allows us to examine the shape of the rates of recurrence and death, as well as the potential dependence between the two processes. The dependence is introduced by a shared individual random effect entering both the rate of recurrence and the hazard of death. An additional parameter determines whether the processes are positively or negatively associated, and how strong this association is. In other models, the frailty affects both event rates in the same way (Huang and Wang, 2004), or the dependence between the processes is left completely unspecified (Cook and Lawless, 1997; Ghosh and Lin, 2003). A common approach to estimating the joint frailty model that we consider here is using Gaussian quadrature to approximate the marginal likelihood (Liu and Huang, 2008; Rondeau et al., 2007). We will see that in this framework, the setting with left-truncated data can be handled in a straightforward manner by adapting the likelihood.

Incorporating left truncation, which is also called delayed entry, has received varying levels of attention in studies based on different frailty models and joint models. Several studies have discussed handling left truncation in shared frailty models for clustered survival data (e.g., Jensen et al., 2004; van den Berg and Drepper, 2016). In a recurrent event setting, Balan et al. (2016) considered event dependent selection; i.e., individuals were included in the study only if they had experienced at least one recurrent event in a given time period. Recurrent event studies with selection dependent on survival were briefly discussed in Cook and Lawless (2007), but not specifically in the context of frailty models. It has been argued that for the joint frailty model for recurrent events and death, as specified by Liu et al. (2004), delayed entry can be easily incorporated (Rondeau et al., 2007). However, to our knowledge, no detailed account of this approach has previously been provided. At the time of writing, the R package `frailtypack`, which can be used for fitting a variety of frailty models, does not provide functionality for estimating the joint frailty model from left-truncated data (according to the manual of version 3.3.2, date 2020-10-07, Rondeau et al., 2020). Extensions to left-truncated data have been considered

in some other joint models for recurrent events and death. Emura et al. (2017) introduced a joint frailty-copula model for two event times that can be adapted to accommodate left truncation and recurrent event data. Outside of the class of shared frailty models, Cai et al. (2017) proposed a model for longitudinal measurements, recurrent events, and a terminal event with inferences based on estimation equations that can be generalized to allow for left-truncated data. Liu et al. (2012) presented an estimating equation procedure for a partial marginal model of recurrent events in the presence of a terminal event with left truncation. In the context of joint models for longitudinal data and death, estimation procedures based on left-truncated data have been derived for different models with shared random effects (see, for example, van den Hout and Muniz-Terrera, 2016; Crowther et al., 2016; Piulachs et al., 2021).

In this chapter, we propose a method for estimating the joint frailty model for recurrent events and a terminal event, as introduced by Liu et al. (2004), when data are left-truncated. Our approach adapts the method of Liu and Huang (2008), who use Gaussian quadrature to approximate the marginal likelihood of the joint frailty model.

The chapter is structured as follows. Section 6.2 first presents the joint frailty model and the corresponding likelihood in the usual setting without truncation, and then shows the adjustments for left-truncated data. The method of estimation is detailed in Section 6.3, and its performance is assessed via simulation studies in Section 6.4. We illustrate the approach using the data set on recurrent UTIs in Section 6.5, and conclude with a discussion in Section 6.6.

## 6.2 Joint frailty model and left truncation

The joint frailty model for recurrent events and a terminal event has been applied most frequently in situations in which the time of the terminal event is subject to independent right-censoring only. In the following, we will first present the model and the corresponding likelihood for such right-censored data. Then, we will lay out how certain assumptions and, in particular, the likelihood function are adjusted to the case of left-truncated data. Throughout, we will often refer to the terminal event as death for the sake of simplicity.

### 6.2.1 Joint frailty model

We consider independent individuals  $i$ ,  $i = 1, \dots, m$ , who can experience recurrent events between time  $t_0 = 0$  and the time  $D_i$  of the terminal event. Let  $C_i$  denote a censoring time, which is assumed to be independent of the recurrence and terminal event processes. An individual can then be observed only up to his or her follow-up time  $X_i = \min(C_i, D_i)$ , and  $\delta_i = \mathbb{1}\{D_i \leq C_i\}$  will indicate whether the terminal event occurred before censoring, with the indicator function  $\mathbb{1}\{\cdot\}$ . The at-risk indicator at time  $t \geq 0$  is given by  $Y_i(t) = \mathbb{1}\{t \leq X_i\}$ , if individuals enter the study at  $t_0 = 0$ .

The number of recurrent events experienced by individual  $i$  up to time  $t$  is recorded by the actual recurrent event process  $N_i^{R^*}(t)$ . Similarly, we define the actual counting process

of the terminal event as  $N_i^{D^*}(t) = \mathbb{1}\{D_i \leq t\}$ . However, due to right-censoring, we can only observe the processes  $N_i^D(t) = \mathbb{1}\{X_i \leq t, \delta_i = 1\}$  and  $N_i^R(t) = \int_0^t Y_i(s) dN_i^{R^*}(s) = N_i^{R^*}(\min(t, X_i))$ . Here,  $dN_i^{R^*}(t) = N_i^{R^*}((t + dt)^-) - N_i^{R^*}(t^-)$  is the increment of the recurrence process, equal to the number of events in the small interval  $[t, t + dt)$ , with  $t^-$  as the left-hand limit.

The observed data of individual  $i$  up to time  $t$  are given by  $\mathcal{O}_i(t) = \{Y_i(s), N_i^R(s), N_i^D(s), 0 \leq s \leq t; \mathbf{z}_i\}$ , including the observed time-fixed covariate vector  $\mathbf{z}_i$ . Individual risks will depend on the covariates as well as on the unobservable frailty value  $u_i$ , where the frailties  $u_i$  are independent realizations of a positive random variable  $U$ .

In the joint frailty model introduced by Liu et al. (2004), the observed recurrence process is assumed to have the intensity  $Y_i(t)\lambda_i(t|u_i)$  with

$$\begin{aligned} \mathbb{P}(dN_i^R(t) = 1 \mid \mathcal{F}_{t^-}, D_i \geq t) &= Y_i(t)\lambda_i(t|u_i)dt \\ \lambda_i(t|u_i)dt &= d\Lambda_i(t|u_i) = \mathbb{P}(dN_i^{R^*}(t) = 1 \mid \mathbf{z}_i, u_i, D_i \geq t). \end{aligned} \quad (6.1)$$

Here,  $\mathcal{F}_t = \sigma\{\mathcal{O}_i(s), 0 \leq s \leq t, u_i; i = 1, \dots, m\}$  denotes the  $\sigma$ -algebra generated by the frailty and the observed data. The terminal event process is characterized by the intensity  $Y_i(t)h_i(t|u_i)$  with

$$\begin{aligned} \mathbb{P}(dN_i^D(t) = 1 \mid \mathcal{F}_{t^-}) &= Y_i(t)h_i(t|u_i)dt \\ h_i(t|u_i)dt &= dH_i(t|u_i) = \mathbb{P}(dN_i^{D^*}(t) = 1 \mid \mathbf{z}_i, u_i, D_i \geq t). \end{aligned} \quad (6.2)$$

The first lines in (6.1) and (6.2) follow from the assumption that the censoring mechanism is conditionally independent of the two event processes given the process history.

Following Liu et al. (2004), we specify the intensities as

$$\begin{aligned} \lambda_i(t|u_i) &= u_i e^{\boldsymbol{\beta}^\top \mathbf{z}_i} \lambda_0(t), \\ h_i(t|u_i) &= u_i^\gamma e^{\boldsymbol{\alpha}^\top \mathbf{z}_i} h_0(t). \end{aligned} \quad (6.3)$$

The baseline rates of recurrence and death,  $\lambda_0(t)$  and  $h_0(t)$ , are affected by the covariates  $\mathbf{z}_i$  through a multiplicative model with effects  $\boldsymbol{\beta}$  and  $\boldsymbol{\alpha}$ , respectively. The inclusion of the frailty  $u$  in the recurrence rate accommodates heterogeneity across individuals and dependence between the recurrences within one individual. The association between the recurrent events and death results from the fact that the shared frailty  $u$  also enters the hazard of death. Due to the additional parameter  $\gamma$ , the model can capture associations of variable magnitudes and in different directions. For positive  $\gamma > 0$ , individuals with a higher rate of recurrence will also be subject to a higher hazard of death. For  $\gamma < 0$ , a higher rate of recurrence implies a lower hazard of death. If  $\gamma = 0$ , the intensities in (6.3) do not share any parameters, and the censoring of the recurrence process by death is non-informative.

A common choice for the distribution of the frailty  $U$  is a gamma distribution with a mean of one and a variance of  $\theta$ . We will more generally consider that the frailties  $u_i$  follow a distribution with a density of  $g_\theta(u)$  and a corresponding distribution function  $G_\theta(u)$

with parameter  $\theta$ . This assumption refers to the initial distribution of frailties in the population at time  $t_0 = 0$ . However, if  $\gamma \neq 0$ , the distribution of frailties in the population will change over time due to selection effects, which will cause the population at time  $t$  to be composed of survivors with lower mortality risks.

We now formulate the likelihood of the joint frailty model (6.3) when individuals are observed from time  $t_0 = 0$ . Let  $t_{ij}$ ,  $j = 1, \dots, J_i$ , be the observed recurrence times of individual  $i$ . Based on the arguments stated in Liu et al. (2004), the conditional likelihood contribution of individual  $i$  given his or her frailty value  $u_i$  can be written as

$$\begin{aligned} L_i^{(c)}(u_i) &= \left[ \prod_{j=1}^{J_i} \lambda_i(t_{ij}|u_i) \right] \exp \left\{ - \int_0^\infty Y_i(s) \lambda_i(s|u_i) ds \right\} \\ &\quad \cdot h_i(x_i|u_i)^{\delta_i} \exp \left\{ - \int_0^\infty Y_i(s) h_i(s|u_i) ds \right\} \\ &= \left[ \prod_{j=1}^{J_i} u_i e^{\beta^\top z_i} \lambda_0(t_{ij}) \right] \exp \left\{ - \int_0^{x_i} u_i e^{\beta^\top z_i} \lambda_0(s) ds \right\} \\ &\quad \cdot \left[ u_i^\gamma e^{\alpha^\top z_i} h_0(x_i) \right]^{\delta_i} \exp \left\{ - \int_0^{x_i} u_i^\gamma e^{\alpha^\top z_i} h_0(s) ds \right\}. \end{aligned} \quad (6.4)$$

The marginal likelihood  $L_i$  of the observed data of individual  $i$  is obtained by integrating the above expression over the frailty distribution,

$$L_i = \int_0^\infty L_i^{(c)}(u) dG_\theta(u) = \int_0^\infty L_i^{(c)}(u) g_\theta(u) du. \quad (6.5)$$

## 6.2.2 Adjusting for left truncation

We will now extend the above framework to allow for left truncation, that is, individuals entering the study at times that may be later than  $t_0 = 0$ . Before deriving the likelihood for the left-truncated data, we introduce some additional notations and assumptions.

A sample of  $m_V$  independent individuals  $i$ ,  $i = 1, \dots, m_V$ , is left-truncated if the individuals  $i$  enter the study only at times  $V_i \geq t_0$ , with strict inequality for at least some individuals. Then, the observation of individual  $i$  is conditional on his or her survival up to the entry time,  $D_i > V_i$ , and events can only be observed in the interval  $[V_i, X_i]$ . Hence, the at-risk indicator  $Y_i(t)$  of Section 6.2.1 is replaced by  ${}_V Y_i(t) = \mathbb{1}\{V_i \leq t \leq X_i\}$ .

As a consequence, the observed recurrent event process  ${}_V N_i^R(t) = \int_0^t {}_V Y_i(s) dN_i^{R*}(s) = [N_i^{R*}(\min(t, X_i)) - N_i^{R*}(V_i)] \mathbb{1}\{t > V_i\}$  in this setting records only the recurrences after study entry at  $V_i$ . Analogously, the left-truncated counting process of the terminal event is given by  ${}_V N_i^D(t) = \mathbb{1}\{V_i \leq X_i \leq t, \delta_i = 1\}$ . The observed data for individual  $i$  are then  ${}_V \mathcal{O}_i(t) = \{{}_V Y_i(s), {}_V N_i^R(s), {}_V N_i^D(s), V_i \leq s \leq t; z_i; V_i\}$ , and the  $\sigma$ -algebra is modified as  ${}_V \mathcal{F}_t = \sigma\{{}_V \mathcal{O}_i(s), 0 \leq s \leq t, u_i; i = 1, \dots, m_V\}$ .

In addition to the assumption of conditionally independent censoring already made in Section 6.2.1, we further assume that the truncation times  $V_i$  are conditionally independent of the recurrence and terminal event processes given the process history. Hence, the intensity of the observed recurrence process is given by  ${}_vY_i(t)\lambda_i(t|u_i)$  and (6.1) is adapted as

$$P(d{}_vN_i^R(t) = 1 \mid {}_v\mathcal{F}_{t-}, D_i \geq t) = {}_vY_i(t)\lambda_i(t|u_i). \quad (6.1')$$

The intensity of the observed terminal event process is, correspondingly,  ${}_vY_i(t)h_i(t|u_i)$ , such that (6.2) is modified as

$$P(d{}_vN_i^D(t) = 1 \mid {}_v\mathcal{F}_{t-}) = {}_vY_i(t)h_i(t|u_i). \quad (6.2')$$

Based on this, we can develop the likelihood of the joint frailty model (6.3) for left-truncated data. The conditional likelihood contribution of individual  $i$  given  $u_i$  is constructed in analogy to (6.4), with  $Y_i(s)$  replaced by  ${}_vY_i(s)$ , and appropriately restricting to the observed recurrence times  $t_{ij} \geq v_i$ ; that is,

$$\begin{aligned} {}_vL_i^{(c)}(u_i) &= \left[ \prod_{t_{ij} \geq v_i} \lambda_i(t_{ij}|u_i) \right] \exp \left\{ - \int_0^\infty {}_vY_i(s)\lambda_i(s|u_i) ds \right\} \\ &\quad \cdot h_i(x_i|u_i)^{\delta_i} \exp \left\{ - \int_0^\infty {}_vY_i(s)h_i(s|u_i) ds \right\} \\ &= \left[ \prod_{t_{ij} \geq v_i} u_i e^{\beta^\top z_i} \lambda_0(t_{ij}) \right] \exp \left\{ - \int_{v_i}^{x_i} u_i e^{\beta^\top z_i} \lambda_0(s) ds \right\} \\ &\quad \cdot \left[ u_i^\gamma e^{\alpha^\top z_i} h_0(x_i) \right]^{\delta_i} \exp \left\{ - \int_{v_i}^{x_i} u_i^\gamma e^{\alpha^\top z_i} h_0(s) ds \right\}. \end{aligned} \quad (6.6)$$

The marginal likelihood contribution is again obtained by integrating out the frailty. However, as the frailty distribution in the sample of survivors differs from the frailty distribution at time  $t_0$ , we need to integrate over the conditional frailty distribution given survival to the time of entry into the study. This point has previously been discussed in the context of clustered survival data by van den Berg and Drepper (2016) and Eriksson et al. (2015) and for general state duration models by Lawless and Fong (1999). More formally, the marginal likelihood contribution of individual  $i$  is thus

$${}_vL_i = \int_0^\infty {}_vL_i^{(c)}(u) dG_\theta(u \mid D_i > v_i, V_i = v_i, z_i) = \int_0^\infty {}_vL_i^{(c)}(u) dG_\theta(u \mid D_i > v_i, z_i), \quad (6.7)$$

under the assumption that the truncation time  $V_i$  is independent of  $u$ .

In particular, if  $\gamma > 0$  such that the recurrence process and the mortality process are positively associated, individuals who survived up to time  $v$  will tend to have lower frailty values than individuals who died before time  $v$ , for given  $z_i$ . Hence, the frailty

distribution among survivors beyond time  $v$ ,  $G_\theta(u \mid D > v)$ , will tend to have more probability mass at lower values  $u$  than the frailty distribution  $G_\theta(u)$  in the underlying population at time  $t_0$ . Consequently, neglecting the effect of the survivor selection on the frailty distribution in the sample and constructing a marginal likelihood as

$${}_vL_i^{\text{naive}} = \int_0^\infty {}_vL_i^{(c)}(u) dG_\theta(u) = \int_0^\infty {}_vL_i^{(c)}(u) g_\theta(u) du, \quad (6.8)$$

would lead to an invalid inference if  $\gamma \neq 0$ . We will illustrate the resulting biases in the parameter estimates in a simulation study in Section 6.4.

For computing the correct marginal likelihood (6.7), we first apply Bayes' theorem to find that

$$g_\theta(u \mid D_i > v_i, \mathbf{z}_i) = \frac{\text{P}(D_i > v_i \mid u, \mathbf{z}_i) g_\theta(u)}{\text{P}(D_i > v_i \mid \mathbf{z}_i)} = \frac{\exp\left\{-\int_0^{v_i} h_i(s \mid u) ds\right\} g_\theta(u)}{\int_0^\infty \text{P}(D_i > v_i \mid u) g_\theta(u) du}, \quad (6.9)$$

where we suppress the dependence on the covariates  $\mathbf{z}_i$  in the last expression for notational convenience. Combining equations (6.6), (6.7), and (6.9), we can express the marginal likelihood contribution  ${}_vL_i$  of individual  $i$  with left-truncated data as

$$\frac{\int_0^\infty \left[ \prod_{t_{ij} \geq v_i} \lambda_i(t_{ij} \mid u) \right] \exp\left\{-\int_{v_i}^{x_i} \lambda_i(s \mid u) ds\right\} h_i(x_i \mid u)^{\delta_i} \exp\left\{-\int_0^{x_i} h_i(s \mid u) ds\right\} g_\theta(u) du}{\int_0^\infty \text{P}(D_i > v_i \mid u) g_\theta(u) du}. \quad (6.10)$$

Interestingly, the formula for  ${}_vL_i$  in (6.10) could have been equivalently derived as the marginal (with respect to the frailty) probability of the recurrence and follow-up data on individual  $i$ , conditional on individual  $i$  being included in the study,  $D_i > v_i$ . To see this, let us denote by  $E_i$  the event that individual  $i$  has follow-up time  $x_i$  with indicator  $\delta_i$  and the observed recurrence times  $t_{ij}$  over  $[v_i, x_i]$ , and consider

$$\text{P}(E_i \mid D_i > v_i) = \frac{\text{P}(E_i \cap \{D_i > v_i\})}{\text{P}(D_i > v_i)} = \frac{\text{P}(E_i)}{\text{P}(D_i > v_i)} = \frac{\int_0^\infty \text{P}(E_i \mid u) g_\theta(u) du}{\int_0^\infty \text{P}(D_i > v_i \mid u) g_\theta(u) du}.$$

As the integrals over the frailty distribution in (6.10) will not, in general, have closed-form expressions, we will use numerical integration in the following.

### 6.3 Estimation of the joint frailty model

Liu and Huang (2008) proposed using Gaussian quadrature to approximate the marginal likelihood of frailty proportional hazards models, including the joint frailty model (6.3). In combination with a piecewise constant specification of the baseline rates, this approach allows for the direct maximization of the approximated likelihood.

The aim of Gaussian quadrature is to replace the integral of a function with a weighted sum of function values. The Gauss-Hermite quadrature rule gives an approximation for a specific integral of a function  $f(x)$ ,

$$\int_{-\infty}^{\infty} f(x)e^{-x^2} dx \approx \sum_{q=1}^Q w_q f(x_q).$$

The quadrature points  $x_q$  need to be determined as the roots of the  $Q^{\text{th}}$ -order Hermite polynomial, while  $w_q$  specify corresponding weights. As the marginal likelihood of a model with normal random effects is easily rewritten in the above form, it follows that such a likelihood can be approximated as

$$\int_{-\infty}^{\infty} L^{(c)}(b) \phi(b) db \approx \sum_{q=1}^Q \tilde{w}_q L^{(c)}(\tilde{x}_q) \phi(\tilde{x}_q), \quad (6.11)$$

where  $\phi(\cdot)$  denotes the standard normal density, while  $\tilde{x}_q = \sqrt{2}x_q$  and  $\tilde{w}_q = \sqrt{2}w_q e^{x_q^2}$  are modified quadrature points and weights, respectively. Marginal likelihoods integrated over non-normal random effects can be expressed as an integral of the form in (6.11) by applying the probability integral transformation (see Nelson et al., 2006; Liu and Huang, 2008). We show in Section 6.7.1 how this quadrature approach can be used to derive an approximation of the marginal likelihood of the joint frailty model with left truncation.

The evaluation of the approximate marginal likelihood depends on the specific form of the baseline rates  $\lambda_0(t)$  and  $h_0(t)$ . As suggested by Liu and Huang (2008), we adopt a piecewise constant model for these functions,

$$\lambda_0(t) = \sum_{k=1}^{K^R} \lambda_{0k} \mathbb{1}\{t \in I_k^R\} \quad \text{and} \quad h_0(t) = \sum_{k=1}^{K^D} h_{0k} \mathbb{1}\{t \in I_k^D\}.$$

with intervals  $I_k^R = (t_{k-1}^R, t_k^R]$ ,  $k = 1, \dots, K^R$ , and  $I_k^D = (t_{k-1}^D, t_k^D]$ ,  $k = 1, \dots, K^D$ . Specifications with a moderate number of up to 10 intervals and the cut-points  $t_k$ ,  $k \geq 1$ , which have been determined based on the quantiles of the observed event times, are generally expected to lead to good results in practice (see Cook and Lawless, 2007; Liu and Huang, 2008). In the setting with left truncation, appropriate choices have to be made for the starting points of the first intervals,  $t_0^R$  and  $t_0^D$ . Depending on the study design, they might be set equal to the lowest study entry time,  $\min_i v_i$ , or a lower value  $t^* \geq t_0$ .

The direct maximization of the marginal likelihood would also be possible if the baseline rates were assumed to follow a simple parametric model, such as the Weibull model. Nonetheless, we recommend the use of the more flexible piecewise constant rate models, unless prior knowledge allows for an informed choice of a specific parametric model.

Finally, the parameter estimates in the joint frailty model with left-truncated data are obtained by maximizing the approximate marginal log-likelihood. The calculation of the



standard errors is based on the inverse of the negative Hessian matrix of the approximate marginal log-likelihood. We give additional details on the implementation in Section 6.7.2.

## 6.4 Simulation study

To evaluate the performance of the proposed method for estimating the parameters of the joint frailty model in case of left-truncated data, we conducted a simulation study. We will also demonstrate which biases can arise if the likelihood is not correctly adjusted to the survivor selection, in particular, to the selection effects on the frailty distribution.

Estimator performance will depend on various aspects of the observation scheme. One aspect is the distribution of the study entry times  $V_i$ , in which both the range of the distribution and its shape matter. Furthermore, the censoring mechanism – that is, the length of the individual follow-up periods and the number of additional drop-outs – will influence the performance of the method. To study these issues, we will first present a base scenario, and will then assess how different observational settings affect the results.

### 6.4.1 Settings

In the base scenario, we generated data from a joint frailty model (6.3). The time scale  $t$  is the age of the individual. The hazard of death and the rate of recurrence are each affected by a single binary covariate, which is drawn from a Bernoulli distribution with parameter 0.5. The regression coefficients are  $\alpha = 0.5$  (death) and  $\beta = 0.5$  (recurrence), respectively. The frailty values are realizations of a gamma distribution with a mean of one and a variance of  $\theta = 0.5$ . The values of the dependence parameter  $\gamma$  were chosen to cover a positive ( $\gamma = 0.5$ ) and a negative ( $\gamma = -0.5$ ) association between the recurrence process and death, as well as the case in which the recurrence rate does not affect the mortality risk ( $\gamma = 0$ ).

The baseline rates  $h_0(t)$  and  $\lambda_0(t)$  were designed to mimic a study in an older population among whom the death rates as well as the recurrent event rates increase exponentially with age. Hence, we chose for both baseline rates the Gompertz-Makeham form,  $ae^{bt} + c$ , where  $t = 0$  corresponds to age 75. By setting  $a = 0.984$ ,  $b = 0.045$ , and  $c = 0$  for the recurrence process ( $\lambda_0(t)$ ), and  $a = 0.108$ ,  $b = 0.07$ , and  $c = 0.12$  for the survival process ( $h_0(t)$ ), the baseline rates were comparable to the estimated rates for the high risk group in the data example in Section 6.5.

To arrive at the left-truncated samples, the following steps were combined. For each individual, a survival time  $D$  (i.e., age  $> 75$ ) and an entry time into study  $V$  were simulated. Only those individuals who survived beyond his or her entry time – that is, for whom  $D > V$  – were included in the final sample (i.e., were ‘observed’). Therefore, the distribution of entry times  $V$  that are observed in the final sample depends on both the mortality model in (6.3) and the initial distribution of the truncation times before selection.

In the base scenario, our aim was to have entry times in the final sample that were distributed across the total age range – here, ages 75 to 95 – with higher numbers of study entries at the younger ages than at the older ages. This scheme will be referred to as truncation pattern A in the following.

To obtain a final observed sample with such characteristics, the entry times  $V$  were drawn from a truncated normal distribution defined on the age range 75 to 95. More specifically, the truncated normal distribution was specified to have a mode equal to the maximum age of 95 with parameter values chosen so that the distribution of the observed study entry times in the truncated sample had the desired shape (the left panel of Figure 6.1 illustrates this procedure). The initial number of generated survival times was chosen such that the final truncated samples had an average size of about  $m_V = 500$  individuals.

An independent censoring mechanism was imposed in the following way. For most individuals, the censoring times were the end of a planned individual follow-up period of  $t_C = 4$  years. However, some of the follow-up times were longer than four years, and some premature random drop-outs occurred. Again, this was done in response to the situation that we observed in the data application of Section 6.5. Accordingly, we generated random durations from a mixture distribution with an 85% point mass at  $t_C$ , a 10% uniform distribution on  $[0, t_C]$ , and a 5% uniform distribution on  $[t_C, t_C + 0.5]$ , with the latter two covering the drop-outs before  $t_C$  and the longer follow-up periods, respectively. These random durations were added to the individual  $V_i$ , and the individual censoring time  $C_i$  was the minimum of this sum and age 95.

The right panel of Figure 6.1 illustrates how the mechanisms of truncation and censoring jointly determine the number of individuals at risk at any time  $t$  across the age range  $[75, 95]$ . Truncation pattern A causes the number of individuals at risk to increase steeply at the early ages, and then to decrease only gradually across the age range. However, due to the relatively short individual follow-up times, the number of individuals at risk across ages is considerably smaller than the total sample size of about 500.

In the setting with a positive association between the recurrence and mortality process ( $\gamma = 0.5$ ), additional simulation scenarios were set up by varying the censoring and truncation patterns.

First, we considered the effect of changing the planned individual follow-up times to  $t_C = 1$  year or  $t_C = 8$  years, respectively. Longer individual follow-up times increase the number of individuals who are under observation at a certain time  $t$ , and are therefore expected to improve the estimator performance.

Second, we explored a scenario with a more unimodal distribution of the study entry times in which relatively few individuals entered the study at the youngest and the oldest ages (see Figure 6.1). This is truncation pattern B. To obtain a final sample with these characteristics, we simulated the initial truncation times again from a truncated normal distribution on the age range 75 to 95. However, in this scenario, the distribution had a mode equal to 90, that is, within the above age range.

Finally, we examined a setting with a wider age range of  $[64, 105]$ . If  $t = 0$  was now expected to correspond to age 64, but the Gompertz-Makeham rates were expected

to agree with the rates of the base scenario over  $[75, 95]$ , the parameters needed to be adapted. This was achieved by maintaining the values of  $b$  and  $c$ , but setting  $a = 0.6$  or  $a = 0.05$  for the recurrence and death processes, respectively. The initial distributions of the study entry times were adapted to produce truncation patterns A or B on the wider age range  $[64, 105]$ . In all of the additional scenarios, the truncated samples again had a target size of  $m_V = 500$  individuals. The parameter values for the distributions of the study entry times and the initial sample sizes for the different scenarios can be found in Section 6.7.3.

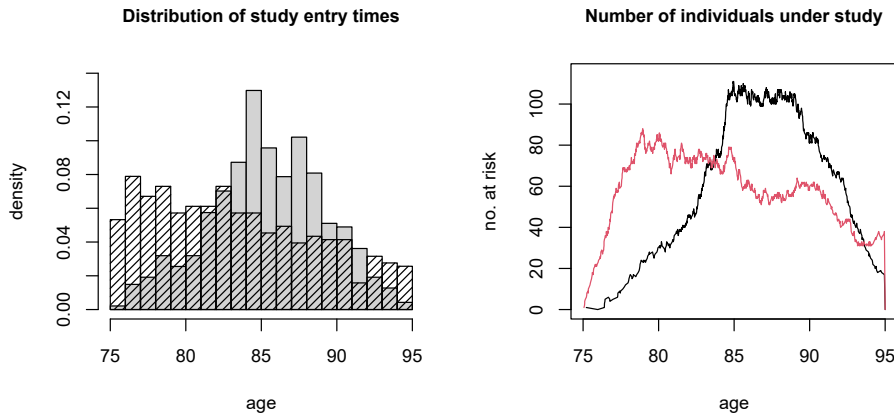


Figure 6.1: Distribution of the ages at study entry (left) and the number of individuals at risk across the age range  $[75, 95]$  (right) for one simulated sample from the base scenario with truncation pattern A (shaded bars, red line) or truncation pattern B (gray bars, black line), both with planned individual follow-ups of four years.

## 6.4.2 Estimation and results

The estimation of the joint frailty model was carried out under the assumptions of gamma distributed frailties with a mean of one and piecewise constant models for the two baseline rates  $\lambda_0(t)$  and  $h_0(t)$ . For both rates, 10 intervals were used that were denoted by  $I_k^R$  (recurrence process) and  $I_k^D$  (mortality),  $k = 1, \dots, 10$ . The intervals were determined by the deciles of the observed recurrence and survival times, respectively. We set  $t_0^R = t_0^D = 75$  (or  $t_0^R = t_0^D = 64$ ) equal to the starting point of the respective age range and  $t_{10}^R = t_{10}^D$  equal to the maximum follow-up time in the sample. The marginal likelihood was approximated using non-adaptive Gauss-Hermite quadrature with  $Q = 30$  quadrature points. We ran 200 replications in each setting. All computations were performed in R (R Core Team, 2020). Further details on the implementation are provided in Section 6.7.2.

Figures 6.2 and 6.3 illustrate the results of the base scenario with different underlying associations,  $\gamma \in \{-0.5, 0, 0.5\}$ . The top panels of Figure 6.2 show that the covariate effects  $\alpha$  and  $\beta$ , the dependence parameter  $\gamma$ , and the frailty variance  $\theta$  are estimated without significant bias. The estimated standard errors of these parameters in the bottom panels of Figure 6.2 are largely in line with the empirical standard deviations of the respective parameter estimates across the replications. Nevertheless, we notice that the estimator performance varies for different true values of the dependence parameter.

This pattern can be explained to some extent by different survivor selection effects. The truncated sample consists of survivors, who tend to have lower mortality risks. If the recurrence process and the mortality process are positively associated, this implies that the frailty values and the recurrence rates are lower in the sample of survivors. In the current setting, this lower frailty variance in the sample is favorable for the estimation of  $\theta$ ; whereas the low recurrence rate, which is associated with higher probabilities of having no observed recurrent event, increases the variability in the corresponding estimated covariate effect  $\hat{\beta}$ . The opposite effects are observed if the event processes are negatively associated. If the recurrence rate has no effect on survival ( $\gamma = 0$ ), the method still yields reliable results, and the parameters exclusively affecting survival are estimated with higher levels of precision.

The estimates of the baseline rates, displayed for the base scenario with positive dependence  $\gamma = 0.5$  in Figure 6.3, also perform satisfactorily.

It is instructive to look at how the results change if the effects of survivor selection on the frailty distribution in the sample are not taken into account correctly. As Figures 6.4 and 6.5 show, if the inference is based on the naively constructed marginal likelihood (6.8), biases can be seen in all parameter estimates in case the recurrence process and the mortality process are associated. Moreover, as the estimated standard errors for the covariate effects are substantially smaller than those obtained using the correct likelihood, they do not adequately reflect the uncertainty in the parameter estimates. The baseline rates of recurrence and death are increasingly underestimated for advancing age in the base scenario with positive dependence ( $\gamma = 0.5$ ), as depicted in Figure 6.5. This is because in a setting with a positive association, the distribution of frailty among the survivors tends to be concentrated at lower values. Accordingly, for negative associations, the recurrence rate will be overestimated at the older ages, while the hazard of death will again be underestimated at the older ages. Hence, failing to construct the marginal likelihood based on the correct distribution of the frailty, see (6.7), introduces marked biases in the estimates and the standard errors. Only if the event rates are not associated ( $\gamma = 0$ ) is the distribution of frailty among survivors equal to the initial frailty distribution  $G_\theta$ , such that the naive marginal likelihood coincides with the correct marginal likelihood (6.7) and yields valid inferences.

Lastly, we want to examine the results for the additional simulation scenarios with modified censoring and truncation patterns. The figures illustrating these results can be found in Section 6.7.3. In the scenario with a planned individual follow-up of only  $t_C = 1$  year, we find increased variability in all parameter estimates (see Figures 6.8 and 6.9).

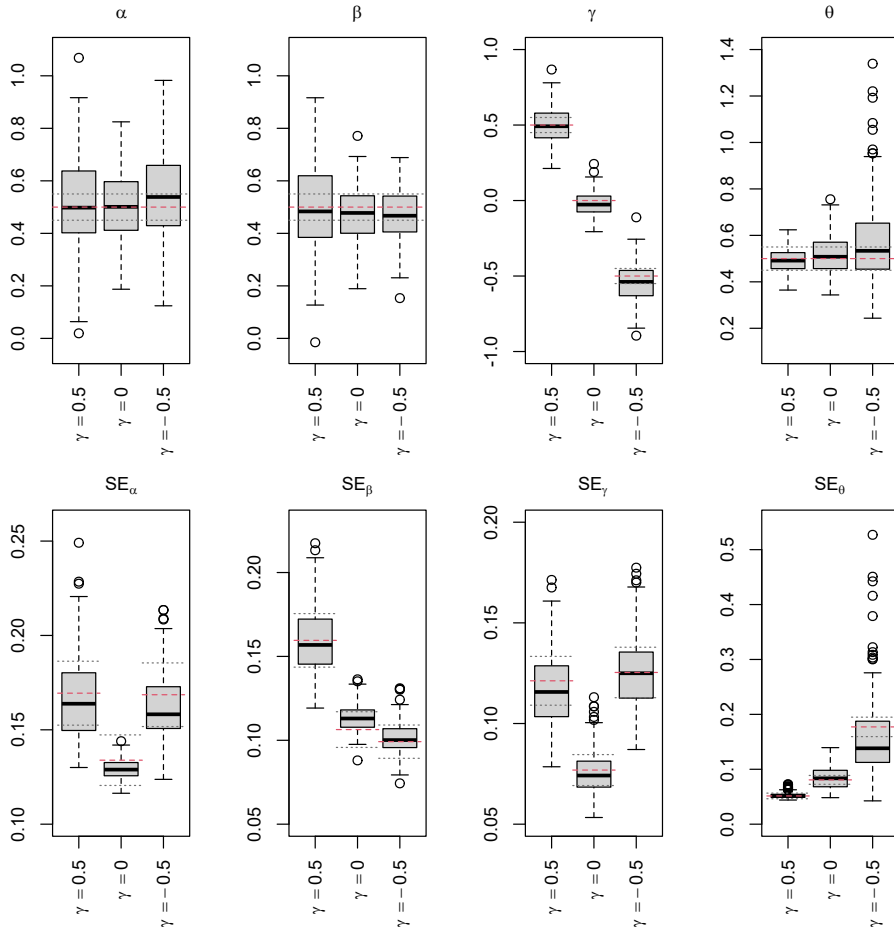


Figure 6.2: Box plots of the parameter estimates (top) and estimated standard errors (bottom) in the joint frailty model for positive ( $\gamma = 0.5$ ), no ( $\gamma = 0$ ), or negative ( $\gamma = -0.5$ ) dependence under the base scenario. Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ) based on 200 truncated samples with a target size of 500. The red dashed line marks the true parameter value (top) or empirical standard deviation (bottom); the gray dotted lines mark 10% deviations from the respective value.

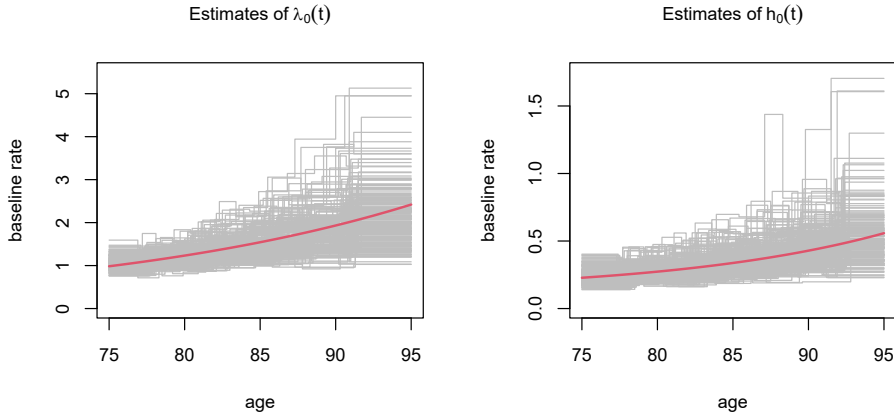


Figure 6.3: Estimates (gray) of the baseline rate of recurrence (left) and of death (right) based on 200 truncated samples with a target size of 500 generated from a joint frailty model with positive dependence ( $\gamma = 0.5$ ) under the base scenario. The red bold line gives the true baseline rate.

This is expected, because with shorter individual follow-up times, fewer individuals are observed at a given age  $t$  than in the base scenario. Further extending the planned individual follow-up times of the base scenario from  $t_C = 4$  to  $t_C = 8$  years does not lead to considerable improvements, apart from some reduced variability in the estimates of the frailty variance and the baseline rates.

A change in the distribution of the study entry times can markedly influence the estimation results. In the modified base scenario with truncation pattern B, the estimated covariate effects  $\hat{\alpha}$  and  $\hat{\beta}$  are more variable than under truncation pattern A, occasionally with negative estimates (see Figure 6.10). In addition, the first piece of each of the baseline rates shows an upward bias (cf. top panels of Figure 6.11) because few individuals entered the study at the younger ages. Although the intervals for the rate pieces were constructed to contain roughly equal numbers of observed events, the first intervals cover a relative large age range with few individuals under study at a given age  $t$  due to the delayed entry.

The last scenario combines a wider age range [64, 105] and truncation times spread across the whole age range according to pattern A or B, with individual follow-ups planned for  $t_C = 4$  years. This setting is more demanding because the amount of information available at a given age  $t$  is considerably smaller than it is in the scenarios with age range [75, 95]. Therefore, the variability in the estimates tends to increase, and the estimates of the dependence parameter and the frailty variance exhibit a small downward bias (see Figure 6.10).

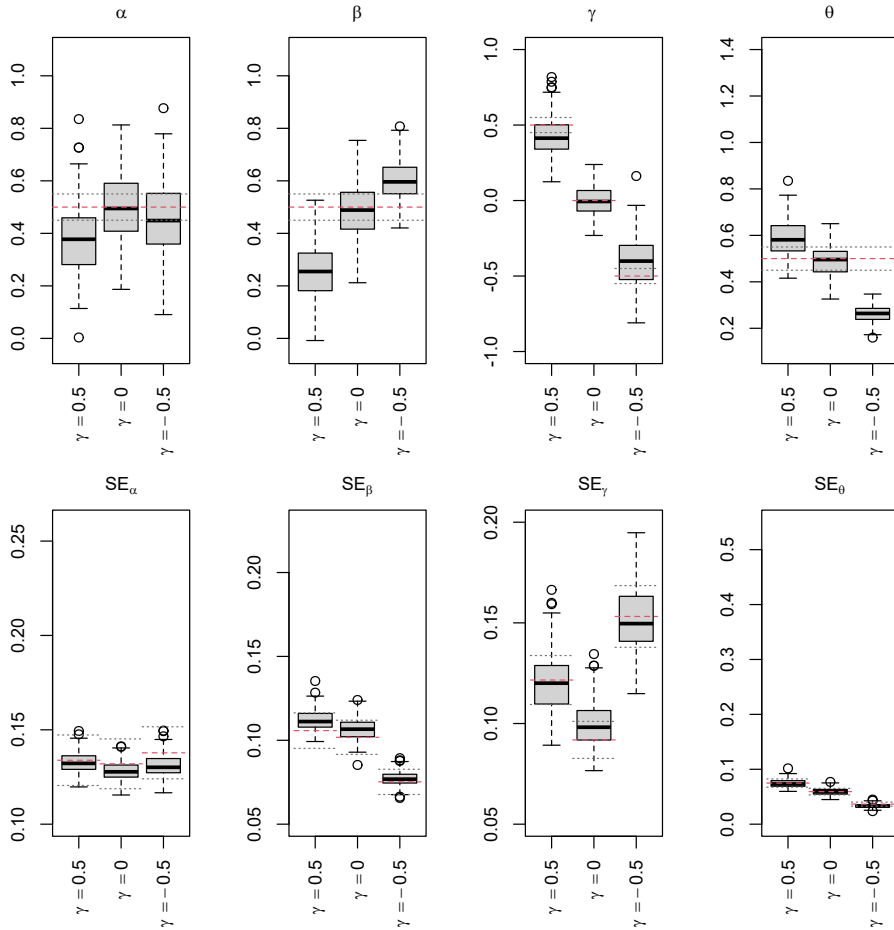


Figure 6.4: Box plots of the parameter estimates (top) and estimated standard errors (bottom) based on the naive likelihood of the joint frailty model for positive ( $\gamma = 0.5$ ), no ( $\gamma = 0$ ), or negative ( $\gamma = -0.5$ ) dependence under the base scenario. Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ) based on 200 truncated samples with a target size of 500. The red dashed line marks the true parameter value (top) or empirical standard deviation (bottom); the gray dotted lines mark 10% deviations from the respective value.

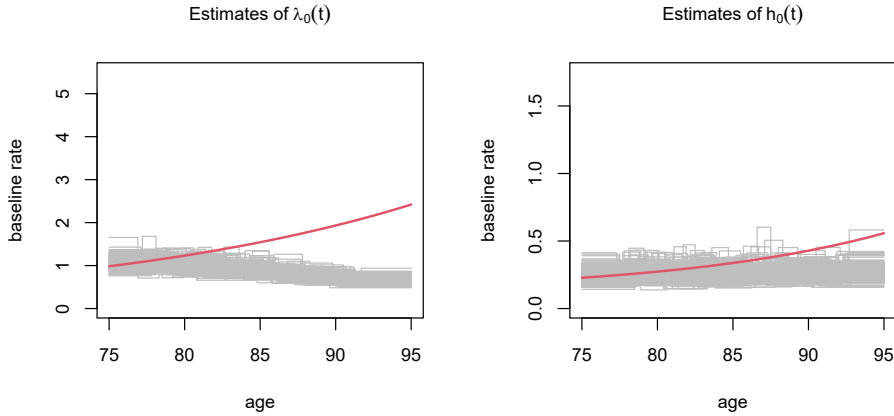


Figure 6.5: Estimates (gray) of the baseline rate of recurrence (left) and of death (right) based on the naive likelihood for 200 truncated samples with a target size of 500 generated from a joint frailty model with positive dependence ( $\gamma = 0.5$ ) under the base scenario. The red bold line gives the true baseline rate.

Overall, the simulation studies suggest that the proposed method for the estimation of the joint frailty model based on left-truncated data performs satisfactorily. The parameter estimates are largely unbiased if the study design ensures that a reasonable number of individuals are under observation across time  $t$ . Including a relatively large number of individuals early on and a preferably stable number of study entries across the remaining time range benefits the estimation. In addition, the individual follow-up times should be sufficiently long given the total time window and the sample size. As expected, the patterns of censoring and truncation that cause more information to be lost negatively affect the estimator performance.

## 6.5 Recurrent infections and mortality in an older population

We use the proposed method to analyze recurrent urinary tract infections and mortality in an institutionalized elderly population. The data come from a double-blind, randomized, placebo-controlled trial in long-term care facilities that aimed to assess the effect of cranberry capsules on the occurrence of UTIs in vulnerable older persons (Caljouw et al., 2014). At baseline, the participants were stratified into two groups of high or low UTI risk depending on whether they had diabetes mellitus, a urinary catheter, or at least one treated UTI in the preceding year. Within these two strata, the participants were randomly assigned to the treatment or the control group. The participants took cranberry or placebo capsules twice a day over a period of one year. Occurrences of UTIs were recorded by



the treating physicians according to a clinical definition based on international practice guidelines for LTCF residents, and a strict definition based on scientific criteria. We focus here on the occurrence of the more broadly defined clinical UTIs.

The final study population consisted of 928 individuals, most of whom were women (703; 75.8%). Of these individuals, 516 were considered to be at high baseline UTI risk, while 412 were considered to be at low baseline UTI risk. Individuals entered the study between ages 64 and 102, as shown in the left panel of Figure 6.6, and were followed on average for about a year (mean: 332 days, median: 372 days). A total of 317 participants (34.2%) died during the study period. The number of observed UTIs per individual ranged from zero to 10, with 62.2% of the individuals having no UTIs, 20.8% having one UTI, and 17.0% experiencing two or more UTIs during the follow-up period.

Unlike in the original study, we modeled recurrent UTIs and mortality to evolve with age, where  $t_0 = 0$  corresponded to age 64. Because of the specific distribution of the ages at study entry in conjunction with the short individual follow-up times, relatively few individuals were under observation at a given age, in particular at the youngest and oldest ages, as the right panel of Figure 6.6 shows.

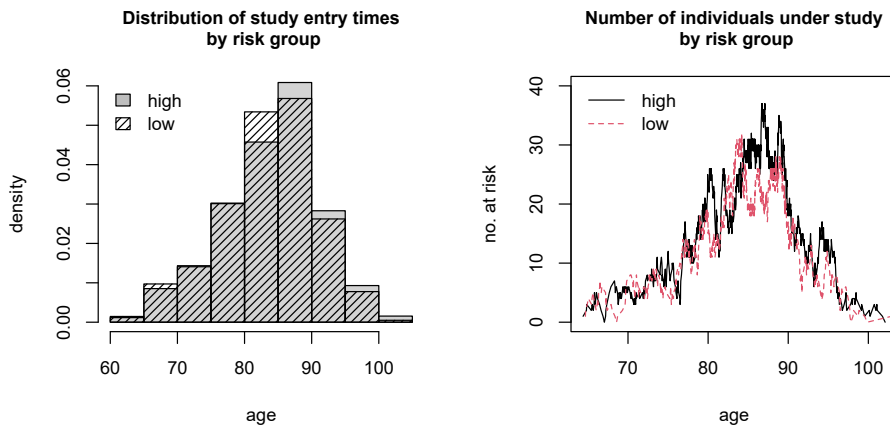


Figure 6.6: Distribution of ages at study entry (left) and number of individuals under observation across the age range (right) in the cranberry data set, separately for the groups with high baseline UTI risk (gray bars, black solid line) and low baseline UTI risk (shaded bars, red dashed line).

We estimated the joint frailty model for UTIs and overall mortality with age as the time scale separately for the groups with high and low baseline UTI risk. Two binary covariates for treatment and gender were included, and frailties were assumed to follow a gamma distribution with a mean of one. The baseline rate of UTI recurrence and the hazard of death were specified as piecewise constant functions with 10 intervals over the age range 64 to 103 in the high risk group and 64 to 104 in the low risk group.

Separately for the two risk groups, the cut-points for the intervals were determined based on the deciles of the observed recurrence or death times, respectively. The likelihood was approximated using non-adaptive Gaussian quadrature with 30 nodes.

The parameter estimates for both risk groups are reported in Table 6.1. In the group with a high baseline UTI risk, the infection rates varied between participants with an estimated frailty variance of  $\hat{\theta} = 0.380$  (SE: 0.086). In particular, individuals with a higher rate of recurrent infections tended to also experience higher mortality risks, as indicated by the positive estimate of the dependence parameter,  $\hat{\gamma} = 0.181$  (SE: 0.084). The participants in the low risk group seemed to be more heterogeneous ( $\hat{\theta} = 1.122$ , SE: 0.316), but the analysis did not detect an association between the occurrence of UTIs and survival ( $\hat{\gamma} = 0.058$ , SE: 0.044). The results suggest that the cranberry capsules did not have a noticeable effect on the occurrence of UTIs irrespective of the baseline UTI risk. When we look at gender differences, we see that males and females experienced similar infection rates, while males had higher mortality levels than females in both groups.

Table 6.1: Parameter estimates (with standard errors) for the joint frailty model fitted to the cranberry data set, separately by risk group.

	High baseline UTI risk	Low baseline UTI risk
Recurrent UTIs		
Treatment (cranberry)	0.000 (0.161)	0.189 (0.217)
Gender (male)	0.061 (0.218)	-0.384 (0.381)
Mortality		
Treatment (cranberry)	0.107 (0.152)	-0.001 (0.197)
Gender (male)	0.396 (0.178)	0.787 (0.210)
Association		
Dependence $\gamma$	0.181 (0.084)	0.058 (0.044)
Frailty variance $\theta$	0.380 (0.086)	1.122 (0.316)

The estimated baseline rates displayed in Figure 6.7 demonstrate nicely the age dependence of the recurrence rate and the hazard of death. For the individuals with a high baseline UTI risk, both the rate of recurrent infection and the mortality risk showed a general tendency to increase with age, although the small number of observations leads to considerable uncertainty at the highest ages. In addition, the individuals with a high baseline UTI risk tended to experience higher rates of recurrent infection and death than the individuals with a low baseline UTI risk.

The original study, which used a different time scale, reported a positive treatment effect of the intake of cranberry capsules only in the group with a high baseline UTI risk and only for the outcome of UTI incidence (first infection during follow-up). When all recurrent UTIs were analyzed in a gamma-frailty model, no treatment effect was detected, which is in line with the findings presented here.

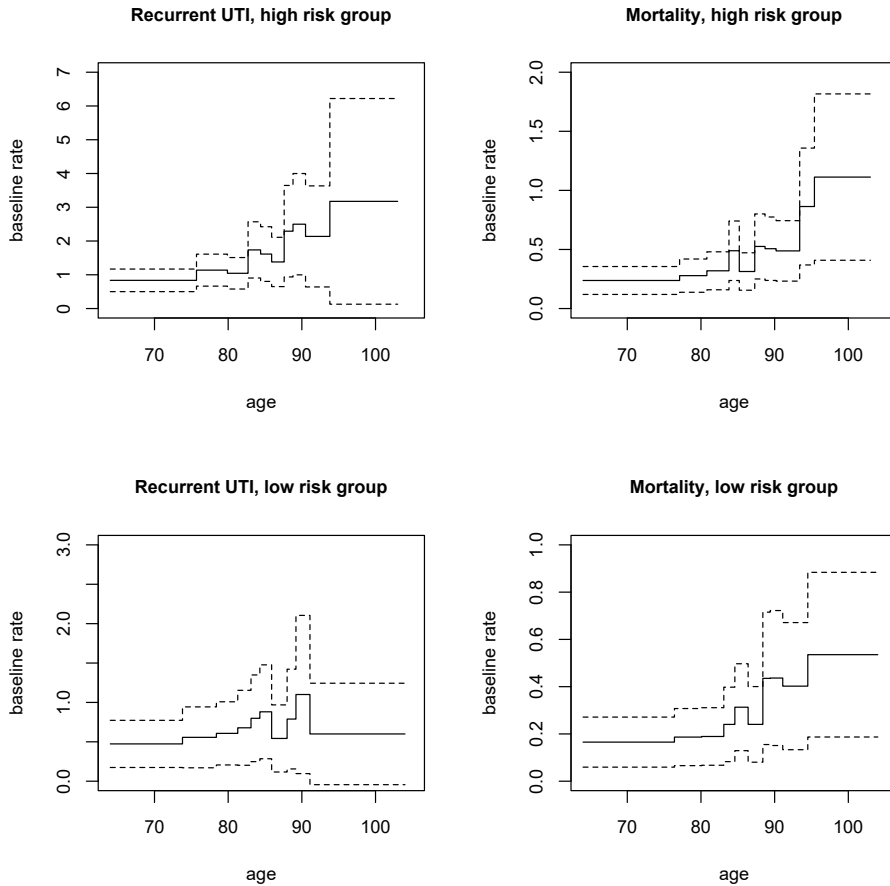


Figure 6.7: Estimated baseline rates (solid) of recurrence (left) and mortality (right) with  $\pm 2$  SE-confidence bounds (dashed) for the cranberry data, separately for the groups with high baseline UTI risk (top) and low baseline UTI risk (bottom).

## 6.6 Discussion

We have proposed a method for estimating the joint frailty model for recurrent events and a terminal event based on left-truncated data. The marginal likelihood of the model can be expressed as a ratio of two integrals over the frailty distribution, each of which is approximated using Gauss-Hermite quadrature. The direct maximization of the approximate marginal likelihood is possible if the baseline rates are specified as piecewise constant functions.

The simulation studies presented here have shown that the estimation procedure performs satisfactorily in general, and have demonstrated how different observation schemes affect the estimator performance. While any pattern of truncation or censoring results in incomplete information, study designs should still aim to provide enough information to meet the needs of a model as complex as the joint frailty model. Having a sufficient number of individuals under observation across most of the time range, and especially at the start of the process, seems to be crucial for the method to yield reliable results.

Allowing for left truncation in frailty models requires us to consider carefully how the frailty distribution in the sample of survivors may differ from the frailty distribution in the underlying population due to selection effects. We illustrated through simulations the biases that can arise in the parameter estimates of the joint frailty model if this difference in the frailty distributions is ignored.

Extending the framework of the joint frailty model to incorporate delayed entry allowed us to study age-specific rates of recurrent urinary tract infections and death in an older population. Similarly, the proposed approach enables researchers to use the joint frailty model in a wider variety of contexts in which subjects are included in a study only if they have not yet experienced the terminal event. Apart from clinical studies with delayed entry, these contexts may include register-based studies of event processes evolving with age as the main time scale, with individuals entering at different ages.

For a complete specification of the model and the approximate likelihood function, we need to choose a frailty distribution as well as the number of quadrature points and the intervals for the baseline rates. Although the simulation study and the application covered only the common choice of a gamma distribution for the frailties, the quadrature approach can be employed with other frailty distributions that have a closed-form inverse distribution function, or a log-normal distribution. The number of quadrature points then determines the accuracy of the integral approximations in the marginal likelihood, as well as the computation time. In line with previous recommendations for gamma frailty models (see Liu and Huang, 2008), we used  $Q = 30$  quadrature points, which produced good results in a reasonable period of time in our settings. Regarding the baseline rate functions, the number of intervals corresponds to the number of parameters for the rates, and should thus be selected to allow for sufficient flexibility of the shape of the rates, while retaining numerical stability and the computational feasibility of the method. Adequate results can often be obtained with moderate numbers of up to 10 intervals.

Nevertheless, in some applications, it may seem appealing to aim for smooth rate estimates, such as through the use of penalized splines. However, automatic smoothing parameter selection in joint frailty models is an issue that needs further investigation.

Moreover, the current approach is limited to applications in which there is heterogeneity in recurrence rates. Due to the specific dependence structure in the joint frailty model considered here, the association between the recurrence process and the terminal event process cannot be assessed if the frailty variance is close to zero.

Finally, in some applications, it might be of interest to extend the proposed method to use information on recurrences before entry into the study. These additional event

times can be included in the marginal likelihood, and are expected to lead to increased precision in the estimation of the model for the recurrence process. However, when using such an approach, researchers should reflect critically on the quality of the retrospectively collected data, as recollections by study participants may be less reliable than data drawn from other sources, such as registries.

## 6.7 Supplementary material

### 6.7.1 Approximation of the marginal likelihood using Gaussian quadrature

In this section, we will elaborate on the use of Gauss-Hermite quadrature for approximating the marginal likelihood of the joint frailty model. We will first recap the quadrature approach as proposed by Liu and Huang (2008) in the setting with right-censoring only, and then show how to adapt the method to the setting with left truncation.

An approximation to marginal likelihoods integrated over normal random effects based on Gauss-Hermite quadrature was already presented in Section 6.3. But the marginal likelihoods of the joint frailty model given in Section 6.2 involve integrals over non-normal random effects. Thus, we use the probability integral transformation (Nelson et al., 2006; Liu and Huang, 2008) to rewrite the integrals over the random effect  $u$  with distribution function  $G_\theta(u)$  as integrals over standard normal random effects. This relies on the fact that the  $G_\theta(u)$  have a standard uniform distribution, such that their transformations  $a = \Phi^{-1}[G_\theta(u)]$  follow a standard normal distribution, if  $\Phi(\cdot)$  denotes the standard normal distribution function.

For the marginal likelihood contribution (6.5) of individual  $i$  in the joint frailty model without truncation, the substitution  $u = G_\theta^{-1}[\Phi(a)]$  yields

$$L_i = \int_0^\infty L_i^{(c)}(u) g_\theta(u) du = \int_{-\infty}^\infty L_i^{(c)}(G_\theta^{-1}[\Phi(a)]) \phi(a) da.$$

We can then employ Gauss-Hermite quadrature as in (6.11) to arrive at the approximate marginal likelihood contribution of the form

$$L_i \approx \sum_{q=1}^Q L_i^{(c)}(G_\theta^{-1}[\Phi(\tilde{x}_q)]) \phi(\tilde{x}_q) \tilde{w}_q.$$

In the setting with left-truncated data, the marginal likelihood of the joint frailty model has a slightly more complex structure. In (6.10), the marginal likelihood contribution  ${}_vL_i$  of individual  $i$  is expressed as a ratio of two integrals over the density  $g_\theta(u)$ . Hence, we will approximate the likelihood by applying the above approach separately to the two integrals,

$${}_vL_i = \frac{\int_0^\infty {}_vL_i^{(c)}(u) \exp\{-H_i(v_i|u)\} g_\theta(u) du}{\int_0^\infty \exp\{-H_i(v_i|u)\} g_\theta(u) du}$$

$$\approx \frac{\sum_{q=1}^Q vL_i^{(c)}(G_\theta^{-1}[\Phi(\tilde{x}_q)]) \exp\{-H_i(v_i|G_\theta^{-1}[\Phi(\tilde{x}_q)])\} \phi(\tilde{x}_q) \tilde{w}_q}{\sum_{q=1}^Q \exp\{-H_i(v_i|G_\theta^{-1}[\Phi(\tilde{x}_q)])\} \phi(\tilde{x}_q) \tilde{w}_q},$$

with  $vL_i^{(c)}(u)$  given in (6.6). The approximate marginal likelihood of the joint frailty model with left truncation is then given by

$$\prod_{i=1}^{m_V} \frac{\sum_{q=1}^Q vL_i^{(c)}(G_\theta^{-1}[\Phi(\tilde{x}_q)]) \exp\{-H_i(v_i|G_\theta^{-1}[\Phi(\tilde{x}_q)])\} \phi(\tilde{x}_q) \tilde{w}_q}{\sum_{q=1}^Q \exp\{-H_i(v_i|G_\theta^{-1}[\Phi(\tilde{x}_q)])\} \phi(\tilde{x}_q) \tilde{w}_q}.$$

In the naive likelihood (6.8), there is only one integral over the frailty distribution, and hence only one approximation is required,

$$vL_i^{\text{naive}} = \int_0^\infty vL_i^{(c)}(u) g_\theta(u) du \approx \sum_{q=1}^Q vL_i^{(c)}(G_\theta^{-1}[\Phi(\tilde{x}_q)]) \phi(\tilde{x}_q) \tilde{w}_q.$$

## 6.7.2 Computational details

We used R (R Core Team, 2020) to implement the estimation procedure. The quadrature points and weights were calculated using function `gauss.quad()` from package `statmod` (Smyth, 1998). For numerical optimization of the approximate marginal log-likelihood, we applied function `nlm()`, which performs minimization based on a Newton-type algorithm, to the negative log-likelihood. The Hessian of the marginal log-likelihood was approximated numerically using function `hessian()` from package `numDeriv` (Gilbert and Varadhan, 2019).

As the frailty variance and the parameters of the piecewise constant baseline rates are restricted to be non-negative, the log-likelihood was maximized with respect to the log-transform of these parameters, which guaranteed non-negative estimates. The delta-method was then applied to derive the respective standard errors.

However, when specifying the baseline rates as piecewise constant functions, the numerically computed Hessian of the log-likelihood may in some cases not be invertible. To overcome this issue, one can add to the log-likelihood small, fixed ridge penalties (e.g., with penalty parameter  $10^{-6}$ ) on the logarithm of the rate parameters and derive the Hessian matrix from this penalized log-likelihood.

## 6.7.3 Supplement to the simulation studies

### Generation of different truncation patterns

A brief description of how the truncation patterns A and B were generated was already given in Section 6.4. For both patterns, truncation times were simulated from a truncated normal distribution defined on the corresponding age range [75, 95] or [64, 105]. However, the parameters  $\mu$  and  $\sigma^2$  of the underlying normal distribution were chosen such that

the density of the resulting truncated normal distribution was either increasing over the whole age range with mode equal to the maximum age or unimodal with a mode within the age range. These two distinct shapes yielded the desired truncation patterns A (TrA) or B (TrB) in the final samples. Table 6.2 reports, for the different settings, the parameter values for the distribution of the truncation times as well as the initial sample size  $M$  needed to obtain truncated samples with an average size of  $m_V = 500$ .

When implementing the simulation study in R, we used function `rttruncnorm()` from package `truncnorm` (Mersmann et al., 2018) for drawing random numbers from a truncated normal distribution.

Table 6.2: Initial sample size  $M$  and parameter values for the distribution of the truncation times.

$\gamma$	age range	pattern	$M$	$\mu$	$\sigma^2$
0.5	[75, 95]	TrA	$1.07 \cdot 10^4$	109	124
0.5	[75, 95]	TrB	$1.20 \cdot 10^4$	90	18
0.5	[64, 105]	TrA	$7.65 \cdot 10^4$	120	225
0.5	[64, 105]	TrB	$4.75 \cdot 10^4$	93	52
0	[75, 95]	TrA	$3.50 \cdot 10^4$	115	110
-0.5	[75, 95]	TrA	$3.44 \cdot 10^4$	115	109

### Additional figures for the simulation results

The following figures illustrate the impact of different censoring and truncation patterns on the performance of the estimation procedure. All settings are modifications of the base scenario with positive dependence  $\gamma = 0.5$ , that was presented in Section 6.4. In particular, samples with a target size of  $m_V = 500$  were generated from a joint frailty model with covariate effects  $\alpha = \beta = 0.5$ , frailty variance  $\theta = 0.5$ , and Gompertz-Makeham baseline rates.

- In Figures 6.8 and 6.9, the effect of different censoring mechanisms is studied by comparing the estimation results for different lengths of the planned individual follow-up  $t_C \in \{1, 4, 8\}$  years.
- Changes in the distribution of truncation times are examined in Figures 6.10 and 6.11. The displayed settings assume different distributions of the study entry times in the final sample, both in terms of the shape (TrA: truncation pattern A, or TrB: truncation pattern B) and the support of the distribution (75+: ages 75 to 95, or 64+: ages 64 to 105).

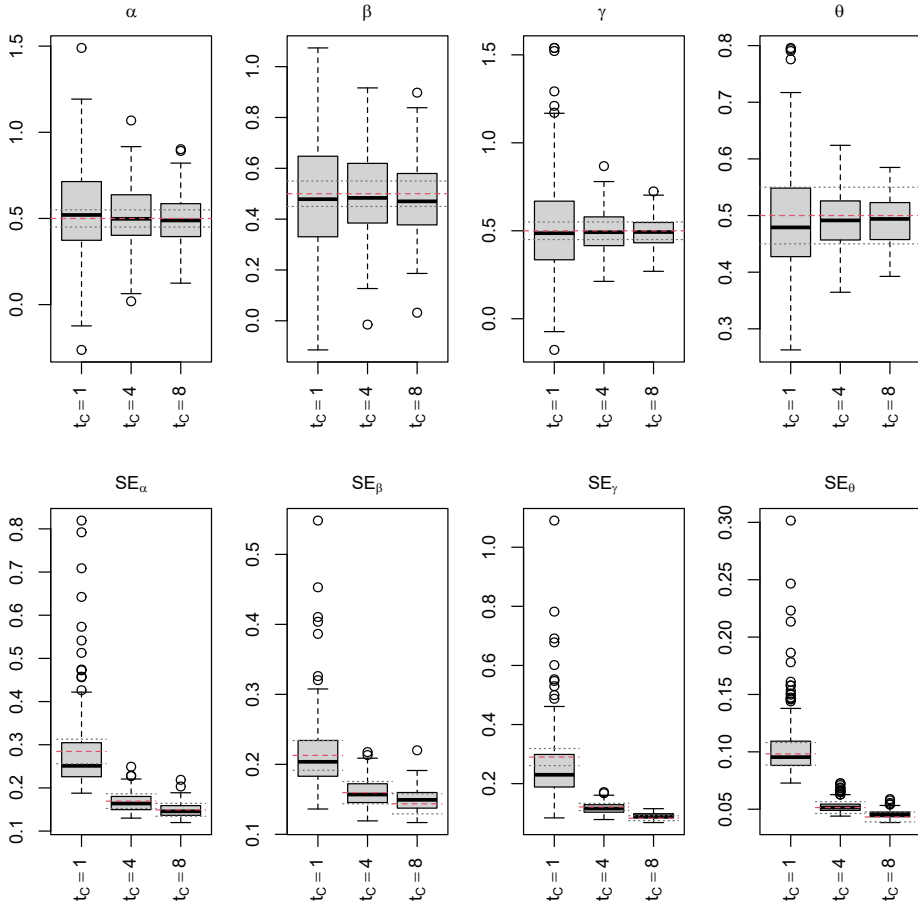


Figure 6.8: Box plots of the parameter estimates (top) and estimated standard errors (bottom) in the joint frailty model with positive dependence ( $\gamma = 0.5$ ) as in the base scenario with truncation pattern A for ages 75+, but varying the planned individual follow-up  $t_C \in \{1, 4, 8\}$ . Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ) based on 200 truncated samples with a target size of 500. The red dashed line marks the true parameter value (top) or empirical standard deviation (bottom); the gray dotted lines mark 10% deviations from the respective value.



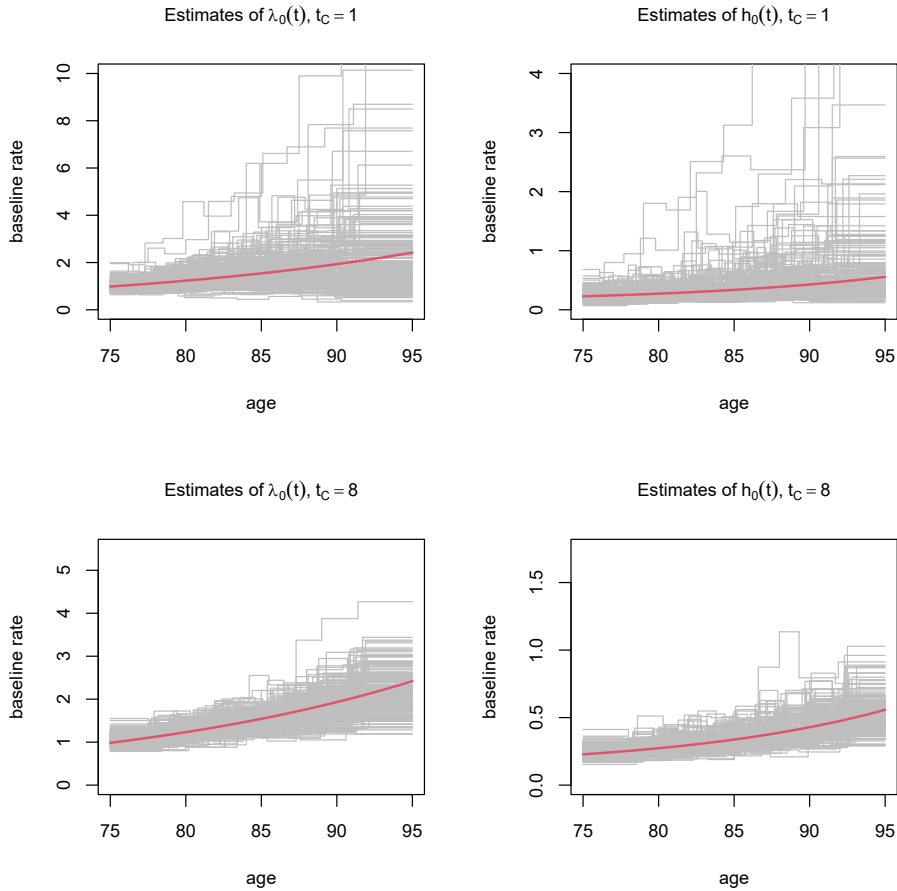


Figure 6.9: Estimates (gray) of the baseline rate of recurrence (left) and of death (right) based on 200 truncated samples with a target size of 500 generated from a joint frailty model with positive dependence ( $\gamma = 0.5$ ). As in the base scenario truncation follows pattern A for ages 75+, but planned individual follow-up is  $t_C = 1$  (top) or  $t_C = 8$  (bottom) year(s). The red bold line gives the true baseline rate. (Note the different scales of the vertical axes.)

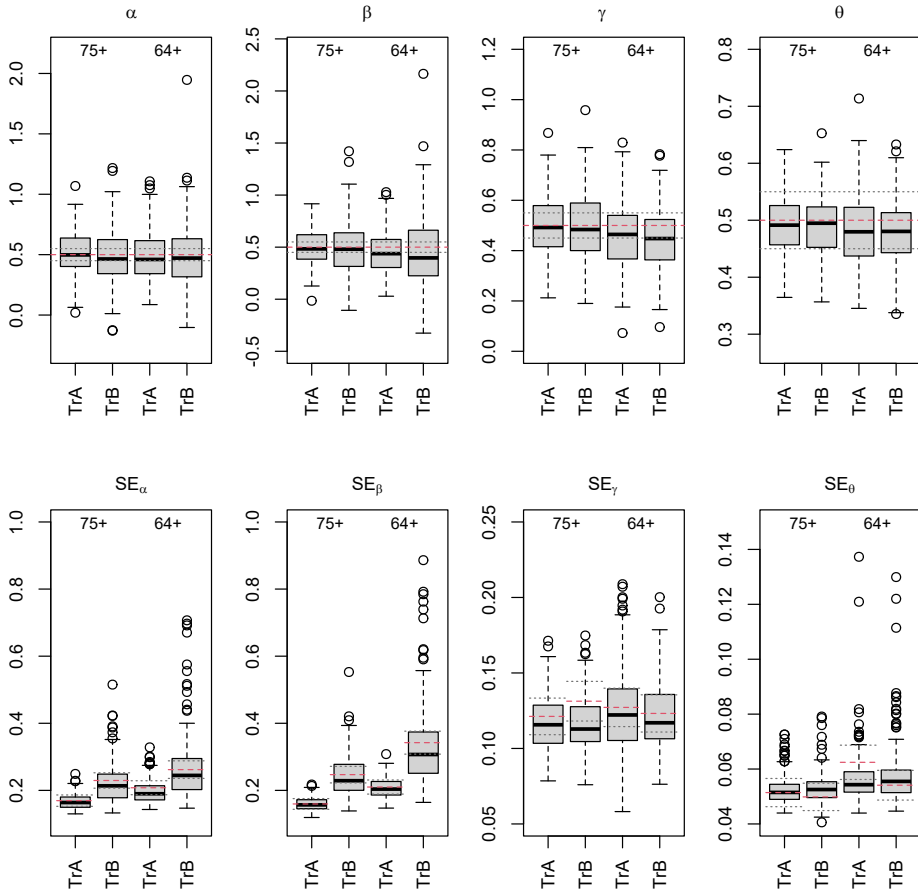


Figure 6.10: Box plots of the parameter estimates (top) and estimated standard errors (bottom) in the joint frailty model with positive dependence ( $\gamma = 0.5$ ) as in the base scenario with  $t_C = 4$ , but truncation times distributed with different shapes (TrA: truncation pattern A, TrB: truncation pattern B) and across different age ranges (75+: [75, 95], 64+: [64, 105]). Left to right: covariate effect on mortality ( $\alpha$ ) and on recurrences ( $\beta$ ), dependence parameter ( $\gamma$ ), and frailty variance ( $\theta$ ) based on 200 truncated samples with a target size of 500. The red dashed line marks the true parameter value (top) or empirical standard deviation (bottom); the gray dotted lines mark 10% deviations from the respective value.

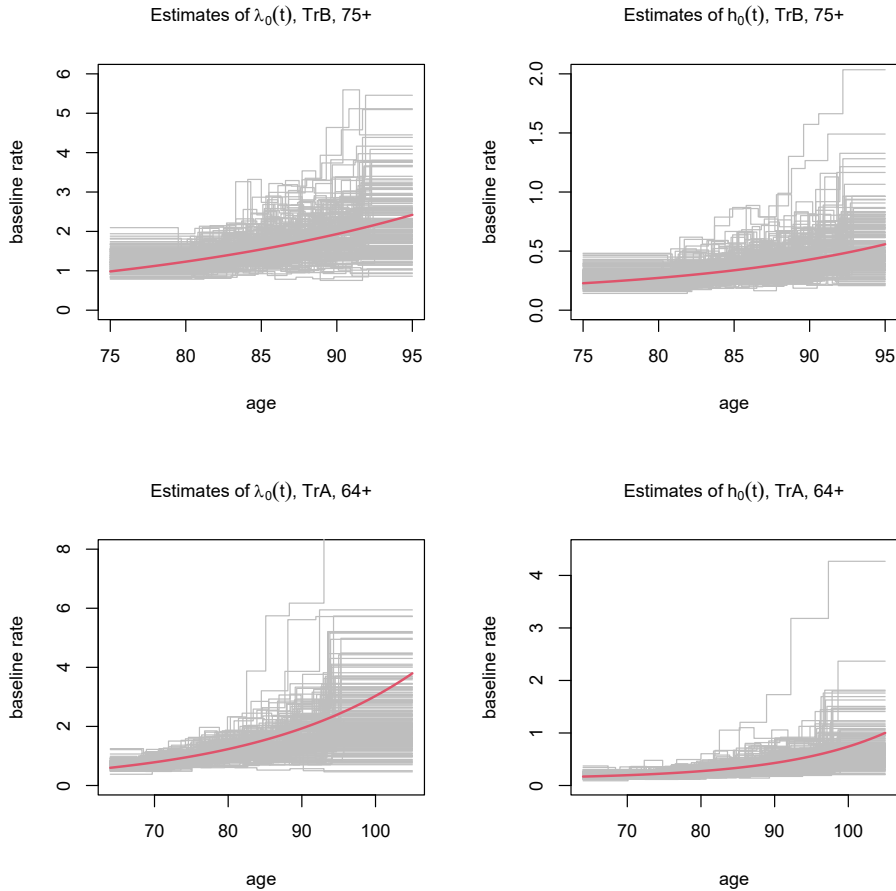


Figure 6.11: Estimates (gray) of the baseline rate of recurrence (left) and of death (right) based on 200 truncated samples with a target size of 500 generated from a joint frailty model with positive dependence ( $\gamma = 0.5$ ) and planned individual follow-up of  $t_C = 4$  years as in the base scenario. Truncation times are distributed according to pattern B across ages [75, 95] (top; TrB, 75+) or according to pattern A across ages [64, 105] (bottom; TrA, 64+). The red bold line gives the true baseline rate. (Note the different scales of the horizontal and vertical axes.)

## References

- Balan, T. A., M. A. Jonker, P. C. Johannesma, and H. Putter (2016). Ascertainment correction in frailty models for recurrent events data. *Statistics in Medicine* 35, 4183–4201.
- Cai, Q., M.-C. Wang, and K. C. G. Chan (2017). Joint modeling of longitudinal, recurrent events and failure time data for survivor's population. *Biometrics* 73, 1150–1160.
- Caljouw, M. A. A., W. B. van den Hout, H. Putter, W. P. Achterberg, H. J. M. Cools, and J. Gussekloo (2014). Effectiveness of cranberry capsules to prevent urinary tract infections in vulnerable older persons: A double-blind randomized placebo-controlled trial in long-term care facilities. *Journal of the American Geriatrics Society* 62, 103–110.
- Cook, R. J. and J. F. Lawless (1997). Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* 16, 911–924.
- Cook, R. J. and J. F. Lawless (2007). *The Statistical Analysis of Recurrent Events*. Statistics for Biology and Health. New York: Springer.
- Crowther, M. J., T. M.-L. Andersson, P. C. Lambert, K. R. Abrams, and K. Humphreys (2016). Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Statistics in Medicine* 35, 1193–1209.
- Emura, T., M. Nakatochi, K. Murotani, and V. Rondeau (2017). A joint frailty-copula model between tumour progression and death for meta-analysis. *Statistical Methods in Medical Research* 26, 2649–2666.
- Eriksson, F., T. Martinussen, and T. H. Scheike (2015). Clustered survival data with left-truncation. *Scandinavian Journal of Statistics* 42, 1149–1166.
- Ghosh, D. and D. Y. Lin (2003). Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics* 59, 877–885.
- Gilbert, P. and R. Varadhan (2019). *numDeriv: Accurate Numerical Derivatives*. R package version 2016.8-1.1. <https://CRAN.R-project.org/package=numDeriv>.
- Huang, C.-Y. and M.-C. Wang (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* 99, 1153–1165.
- Jazić, I., S. Haneuse, B. French, G. MacGrogan, and V. Rondeau (2019). Design and analysis of nested case-control studies for recurrent events subject to a terminal event. *Statistics in Medicine* 38(22), 4348–4362.

- Jensen, H., R. Brookmeyer, P. Aaby, and P. K. Andersen (2004). *Shared Frailty Model for Left-Truncated Multivariate Survival Data*. Biostatistisk Afdeling: Museum Tusulanum.
- Lawless, J. F. and D. Y. T. Fong (1999). State duration models in clinical and observational studies. *Statistics in Medicine* 18, 2365–2376.
- Liu, D., D. E. Schaubel, and J. D. Kalbfleisch (2012). Computationally efficient marginal models for clustered recurrent event data. *Biometrics* 68, 637–647.
- Liu, L. and X. Huang (2008). The use of Gaussian quadrature for estimation in frailty proportional hazards models. *Statistics in Medicine* 27, 2665–2683.
- Liu, L., R. A. Wolfe, and X. Huang (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* 60, 747–756.
- Mersmann, O., H. Trautmann, D. Steuer, and B. Bornkamp (2018). *truncnorm: Truncated Normal Distribution*. R package version 1.0-8. <https://CRAN.R-project.org/package=truncnorm>.
- Modig, K., T. Andersson, S. Drefahl, and A. Ahlbom (2013). Age-specific trends in morbidity, mortality and case-fatality from cardiovascular disease, myocardial infarction and stroke in advanced age: Evaluation in the Swedish population. *PLOS ONE* 8. doi: 10.1371/journal.pone.0064928.
- Nelson, K. P., S. R. Lipsitz, G. M. Fitzmaurice, J. Ibrahim, M. Parzen, and R. Strawderman (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics* 15(1), 39–57.
- Piulachs, X., E.-R. Andrinopoulou, M. Guillén, and D. Rizopoulos (2021). A Bayesian joint model for zero-inflated integers and left-truncated event times with a time-varying association: Applications to senior health care. *Statistics in Medicine* 40(1), 147–166.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rogers, J. K., A. Yaroshinsky, S. J. Pocock, D. Stokar, and J. Pogoda (2016). Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. *Statistics in Medicine* 35, 2195–2205.
- Rondeau, V., J. R. Gonzalez, Y. Mazroui, A. Mauguen, A. Diakite, A. Laurent, M. Lopez, A. Król, C. L. Sofeu, J. Dumerc, and D. Rustand (2020). *frailtypack: General Frailty Models: Shared, Joint and Nested Frailty Models with Prediction; Evaluation of Failure-Time Surrogate Endpoints*. R package version 3.3.2. <https://CRAN.R-project.org/package=frailtypack>.

- Rondeau, V., S. Mathoulin-Pelissier, H. Jacqmin-Gadda, V. Brouste, and P. Soubeyran (2007). Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* 8(4), 708–721.
- Smyth, G. K. (1998). Numerical integration. In P. Armitage and T. Colton (Eds.), *Encyclopedia of Biostatistics*, pp. 3088–3095. London: Wiley.
- van den Berg, G. J. and B. Drepper (2016). Inference for shared-frailty survival models with left-truncated data. *Econometric Reviews* 35, 1075–1098.
- van den Hout, A. and G. Muniz-Terrera (2016). Joint models for discrete longitudinal outcomes in aging research. *Journal of the Royal Statistical Society, Series C* 65, 167–186.



# Summary

Frailty models continue to play an important role in the analysis of time-to-event data. Introducing a positive random effect, the frailty, that affects the risk of experiencing the event of interest can account for unobserved heterogeneity between individuals. In studies of multivariate survival data, frailties can also be used to account for the dependence between observations on the same individual or on different units belonging to a cluster. A variety of frailty models have been proposed in different contexts, ranging from frailty proportional hazards models for univariate survival data, to shared frailty models for clustered survival data, to joint frailty models in which two or more event processes are modeled simultaneously. Although considerable advances have been made in this field in recent decades, further developments of the methodology are often needed to handle specific applications.

In this thesis, we extend the available statistical techniques for inference in two distinct frailty models. In the first part, we focus on a frailty proportional hazards model for studying human adult mortality. With the aim of identifying methods that allow for a reliable assessment of the mortality trajectory at advanced ages, we discuss aspects of parameter estimation, hypothesis testing, and study design in this model, as well as model selection. The second part is concerned with inference in a joint frailty model for recurrent events and a terminal event in settings in which there is delayed entry, or in which the recurrence process is observed only intermittently. In the following, we provide a more detailed overview of the individual chapters.

Chapter 1 gives a brief introduction to frailty modeling of time-to-event data in general, and provides some background information on the models and the methods we study in the two following parts of the thesis.

In the first part of the thesis, which consists of Chapters 2 to 4, we examine different aspects of statistical studies on a demographic phenomenon known as mortality deceleration. This slowing down of the death rates at advanced ages is analyzed in the framework of the gamma-Gompertz model. In this frailty proportional hazards model, the deceleration arises as an effect of selection, because the group of survivors to older ages is made up of the more robust individuals with lower mortality risks. The competing situation without mortality deceleration occurs if the frailty variance takes the value of zero. The model is then simplified to the Gompertz model. However, in this case, the frailty variance lies on the boundary of its parameter space. This non-standard condition of a ‘boundary parameter’ requires adjustments of the statistical methods used for detecting the potential



deceleration of the death rates. Moreover, conducting empirical investigations of old-age mortality can be challenging due to the scarcity of data at the tail of the survival distribution. In addition, as a scientific validation of the ages at death is often compulsory, but is naturally subject to time and cost constraints, samples may be restricted to cover only survivors beyond a certain age. In light of these challenges, we study and compare different approaches for assessing mortality deceleration in the first part of the thesis.

Chapter 2 examines the asymptotic properties of likelihood inference in the gamma-Gompertz model, as well as the Akaike information criterion (AIC) for this model. Due to the boundary parameter, the standard results of likelihood theory for regular problems do not hold in this setting. Our derivations are based on a framework of local alternatives. We find that the maximum likelihood estimator in the gamma-Gompertz model asymptotically follows a mixture distribution if the underlying frailty variance is small. Similarly, the test statistic of a likelihood ratio test for a zero frailty variance asymptotically has a mixture distribution. An explicit formula for the approximate local power of this test to detect a positive frailty variance is presented. Moreover, we prove that the standard formula for the AIC does not apply to the gamma-Gompertz model, but gives a biased estimate of the corresponding Akaike information.

In Chapter 3, we discuss issues of study design for old-age mortality studies based on the gamma-Gompertz model, and introduce information measures for evaluating such designs. The design aspects we consider include the choice of the age range covered by a sample and the sample size. Drawing on the theory of optimal design, we define different information measures as scalar functions of the Fisher information matrix of the gamma-Gompertz model. Interpretations of these measures are provided, including with regard to the non-standard condition of the boundary parameter. We demonstrate how the proposed measures can be used to appraise different study designs. The approach is complemented by calculations of the power of a likelihood ratio test to detect a deceleration in the death rates in specific design settings. We find that in the scenarios we consider, the information carried by the data and the power of the test are markedly reduced if only survivors to increasingly high ages are included in a study. In an application of the proposed methods to mortality data of French-Canadians born in the late 19<sup>th</sup> century, we investigate the potential benefits for the statistical analysis if information on deaths at ages 85-89 could be added to an existing data set on survivors beyond age 90.

In Chapter 4, we introduce focused model selection as a new method for assessing mortality deceleration. A focused information criterion (FIC) evaluates the model performance for a specific parameter of interest, the focus parameter, and therefore allows us to address model performance at the advanced ages for which mortality deceleration is most evident. We derive formulas of a FIC for choosing between two candidate models that differ only by a boundary parameter. In a simulation study, we investigate how this new version of the FIC performs when selecting between the gamma-Gompertz model and the Gompertz model depending on the choice of the focus parameter. We find that compared to a modified version of the AIC that includes a bias correction term based on the results of Chapter 2, the FIC is a more reliable tool for detecting a deceleration in the

death rates. Model selection using the new information criteria is also illustrated with the French-Canadian mortality data.

The second part of the thesis, which consists of Chapters 5 and 6, deals with the study of recurrent events in the presence of a terminal event. A joint frailty model is applied to account for the potential dependent censoring of the recurrence process by the terminal event. The model incorporates an additional association parameter that governs whether and, if so, to what extent higher recurrence rates imply that there is a higher or a lower hazard of experiencing the terminal event. Up to now, inference in this joint frailty model has been studied mainly in ideal situations in which the recurrence times are known exactly, and individuals are observed from the moment they become at risk of experiencing the events of interest. We adapt the statistical methods to two other common observational settings. In the first setting, only interval counts of the recurrent events, which give the numbers of recurrences in specific time intervals, are available. In the second setting, individuals are included in the study only after the processes have started, resulting in left truncation. In both settings, parameter estimation is based on the marginal likelihood, which is approximated using Gaussian quadrature. For the baseline rates, we adopt piecewise constant models.

In Chapter 5, we present the studies with interval counts of recurrent events in a general setting in which the observation intervals can vary across individuals. Apart from the estimation method, we also adjust a score test to the setting with interval counts, which allows us to assess the association between the recurrent events and the terminal event before estimating the joint model. Both the estimation and the testing procedure perform well on simulated data. We apply the methods to determine whether asexual reproduction and mortality in the marine organism *Eleutheria dichotoma* are related.

In Chapter 6, we address the issue of delayed entry into recurrent event studies with a terminal event. In such cases, individuals can be included in the study only if the terminal event has not yet occurred for them at the start of the study. If the recurrence rates and the hazard of the terminal event are associated, this selection can lead to differences between the frailty distributions in the sample and the underlying population. In simulation studies, we demonstrate that neglecting the selection effects on the frailty distribution in the construction of the likelihood can bias the estimation results. Furthermore, we show how different observation schemes, and corresponding patterns of incomplete information, affect the performance of the estimation procedure. The proposed method enables us to study recurrent urinary tract infections in an elderly population while using age as the main time scale.



# Samenvatting

Frailty modellen blijven een belangrijke rol spelen in de analyse van survival data. Het introduceren van een positief random effect, de frailty, dat het risico om de betreffende gebeurtenis te ervaren beïnvloedt, kan ongeobserveerde heterogeniteit tussen individuen verklaren. In studies met multivariate survival data kunnen frailties ook gebruikt worden om de afhankelijkheid tussen observaties van hetzelfde individu of van verschillende eenheden van een cluster te verklaren. Een verscheidenheid van frailty modellen zijn voorgesteld in verschillende contexten, variërend van frailty proportional hazards modellen voor univariate survival data, tot shared frailty modellen voor geclusterde survival data, tot joint frailty modellen waarin twee of meer processen tegelijkertijd worden gemodelleerd. Hoewel er aanzienlijke vooruitgang is geboekt op dit gebied de afgelopen decennia, is verdere ontwikkeling van de methodologie vaak nodig om specifieke toepassingen aan te pakken.

In dit proefschrift breiden we de beschikbare statistische technieken voor schatten en toetsen in twee verschillende frailty modellen. In het eerste deel richten we ons op een frailty proportional hazards model om mortaliteit in volwassen mensen te bestuderen. Met het doel om methoden te identificeren die een goede beoordeling van het mortaliteitsproces op gevorderde leeftijd toestaan, bespreken we aspecten van parameterschatting, hypothesetoetsen, en onderzoeksopzet in dit model, evenals modelselectie. Het tweede deel betreft schatten en toetsen in een joint frailty model voor recurrenente gebeurtenissen en een terminale gebeurtenis in situaties waarin individuen pas later onder observatie komen (vertraagde binnenkomst), of waarin het recurrenente proces slechts met tussenpozen wordt geobserveerd. Nu volgt een gedetailleerder overzicht van de afzonderlijke hoofdstukken.

Hoofdstuk 1 geeft een korte introductie tot frailty modellering van overlevingsduur data in het algemeen, en biedt achtergrondinformatie over de modellen en de methoden die we bestuderen in de twee volgende delen van het proefschrift.

In het eerste deel van het proefschrift, dat bestaat uit hoofdstuk 2 tot en met 4, onderzoeken we verschillende aspecten van statistische studies over een demografisch fenomeen dat bekend staat als mortaliteitsvertraging. Deze vertraging van de sterftecijfers op gevorderde leeftijd wordt geanalyseerd in het kader van het gamma-Gompertz model. In dit frailty proportional hazards model ontstaat de vertraging als een effect van selectie, omdat de groep overlevenden tot op hogere leeftijd bestaat uit de sterkere individuen met een lager sterfterisico. De concurrerende situatie zonder sterftevertraging doet zich voor als de frailty variantie de waarde nul aanneemt. Het model wordt dan vereenvoudigd

tot het Gompertz model. In dit geval ligt de frailty variantie echter op de grens van zijn parameter ruimte. Deze niet-standaard situatie van een ‘parameter op de grens’ vereist aanpassingen van de statistische methoden die worden gebruikt voor het ontdekken van de potentiële vertraging van de sterftcijfers. Het uitvoeren van empirische onderzoeken naar mortaliteit op hoge leeftijd kan bovendien lastig zijn vanwege het gebrek aan data in de staart van de survival verdeling. Daarnaast kunnen steekproeven, aangezien een wetenschappelijke bevestiging van de leeftijd bij overlijden vaak verplicht is, maar van nature onderhevig is aan tijds- en kostenlimieten, worden begrensd om alleen overlevenden voorbij een zekere leeftijd te dekken. In het licht van deze uitdagingen bestuderen en vergelijken we verschillende benaderingen om mortaliteitsvertraging vast te stellen in het eerste deel van het proefschrift.

Hoofdstuk 2 onderzoekt de asymptotische eigenschappen van likelihood-schattingen in het gamma-Gompertz model, evenals het Akaike informatie criterium (AIC) voor dit model. Vanwege de grensparameter zijn de standaard resultaten van likelihood theorie niet geldig in dit geval. Onze afleidingen zijn gebaseerd op een kader van lokale alternatieven. Een bevinding van ons onderzoek is dat de maximum likelihood schatter in het gamma-Gompertz model asymptotisch een mengsel van twee verdelingen volgt als de onderliggende frailty variantie klein is. Evenzo heeft de toetsingsgrootheid van een likelihood ratio toets voor een frailty variantie van nul asymptotisch een mengsel van twee verdelingen. Een expliciete formule voor de geschatte lokale power van deze toets om een positieve frailty variantie te ontdekken wordt gepresenteerd. Bovendien bewijzen we dat de standaard formule voor de AIC niet van toepassing is op het gamma-Gompertz model, maar een onzuivere schatting geeft van de overeenkomstige Akaike informatie.

In hoofdstuk 3 bespreken we problemen van onderzoeksdesign voor onderzoeken naar mortaliteit op hoge leeftijd gebaseerd op het gamma-Gompertz model, en introduceren we informatiematen om dergelijke designs te evalueren. De design aspecten die we in overweging nemen omvatten de keuze van de leeftijdscategorie van een steekproef en de steekproefomvang. Op basis van de theorie van optimaal design definiëren we verschillende informatiematen als functies van de Fisher informatie matrix van het gamma-Gompertz model. Er worden interpretaties van deze maten gegeven, onder andere met betrekking tot de niet-standaard toestand van de grensparameter. We laten zien hoe de voorgestelde maten gebruikt kunnen worden om verschillende onderzoeksdesigns te beoordelen. De aanpak wordt aangevuld met berekeningen van het onderscheidend vermogen van een likelihood ratio toets om een vertraging in de sterftcijfers in specifieke design situaties te detecteren. Wij vinden dat de informatie die de data bevat en het onderscheidend vermogen van de toets, in de scenario's die wij bekijken, duidelijk verminderd zijn als alleen overlevenden tot toenemende hoge leeftijden in een studie worden opgenomen. In een toepassing van de voorgestelde methoden op sterftedata van Frans-Canadezen geboren eind 19e eeuw, onderzoeken we de mogelijke voordelen voor de statistische analyse als informatie over sterftes op de leeftijden 85-89 toegevoegd zou kunnen worden aan een bestaande dataset over overlevenden voorbij de leeftijd van 90.

In hoofdstuk 4 introduceren we “focused model selectie” als een nieuwe methode om sterftevertraging vast te stellen. Een focused informatie criterium (FIC) evalueert de model prestatie voor een specifieke parameter van belang, de focus parameter, en stelt ons daarom in staat om model prestatie op hoge leeftijden, waarvoor sterftevertraging het opvallendst is, op te lossen. We leiden formules van een FIC af om tussen twee kandidaat-modellen te kiezen die slechts met een grensparameter verschillen. In een simulatiestudie onderzoeken we hoe deze nieuwe versie van de FIC presteert bij het selecteren tussen het gamma-Gompertz model en het Gompertz model, afhankelijk van de keuze van de focus parameter. We zien dat, vergeleken met een aangepaste versie van de AIC die een bias correctie term bevat gebaseerd op de resultaten van hoofdstuk 2, de FIC een betrouwbaarder hulpmiddel is om een vertraging in de sterftecijfers te detecteren. Modelselectie met behulp van de nieuwe informatiecriteria wordt ook geïllustreerd met de Frans-Canadese sterftedata.

Het tweede deel van het proefschrift, dat bestaat uit hoofdstukken 5 en 6, gaat over de studie van recurrente gebeurtenissen in de aanwezigheid van een terminale gebeurtenis. Een joint frailty model wordt gebruikt om de potentiële afhankelijke censurering van het recurrente proces door de terminale gebeurtenis te verklaren. Het model bevat een aanvullende associatie parameter die bepaald of, en zo ja, in hoeverre hogere recurrence rates impliceren dat er een hoger of lager risico is om de terminale gebeurtenis te ondergaan. Tot nu toe werden schattingsmethoden in dit joint frailty model voornamelijk bestudeerd in ideale situaties, waarin de recurrentietijden precies bekend zijn en individuen geobserveerd worden vanaf het moment dat ze gevaar lopen om de betreffende gebeurtenissen mee te maken. Wij passen de statistische methoden aan aan twee andere veelvoorkomende situaties. In de eerste situatie zijn alleen interval tellingen van de recurrente gebeurtenissen, die de aantallen recurrenties in specifieke tijdsintervallen geven, beschikbaar. In de tweede situatie worden individuen pas in de studie opgenomen nadat de processen zijn begonnen, wat resulteert in truncatie van links. In beide situaties is parameter schatting gebaseerd op de marginale waarschijnlijkheid, die benaderd wordt met behulp van Gaussian quadrature. Voor de baseline rates nemen we stuksgewijs constante modellen aan.

In hoofdstuk 5 presenteren we de studies met interval tellingen van recurrente gebeurtenissen in een algemene situatie waarin de observatie-intervallen kunnen variëren tussen individuen. Naast de schattingsmethode passen we ook de scoretoets aan aan de situatie met interval tellingen, wat ons in staat stelt vóór het schatten van het joint model de associatie tussen de recurrente gebeurtenissen en de terminale gebeurtenis vast te stellen. Zowel de schattings- als de toetsingsprocedure presteren goed bij gesimuleerde data. We passen de methodes toe om te bepalen of asexuele reproductie en sterfte in het zeeorganisme *Eleutheria dichotoma* gerelateerd zijn.

In hoofdstuk 6 behandelen we het probleem van vertraagde binnenkomst in studies met recurrente gebeurtenissen met een terminale gebeurtenis. In zulke gevallen kunnen individuen alleen worden opgenomen in de studie als de terminale gebeurtenis voor hen nog niet heeft plaatsgevonden aan het begin van de studie. Als de recurrence rates en de

hazard van de terminale gebeurtenis geassocieerd zijn, kan deze selectie leiden tot verschillen tussen de frailty verdelingen in de steekproef en de onderliggende populatie. In simulatiestudies demonstreren we dat het negeren van de selectie effecten op de frailty verdeling in de constructie van de likelihood de schattingsresultaten kan vertekenen. Verder laten we zien hoe verschillende observatieschema's, en overeenkomstige patronen van incomplete informatie, de prestatie van de schattingsprocedure kunnen beïnvloeden. De voorgestelde methoden stellen ons in staat recurrenente urineweginfecties in een oudere bevolking te bestuderen met het gebruik van leeftijd als de belangrijkste tijdsschaal.

## List of publications

M. Böhnstedt and J. Gampe (2019). Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model. *Statistics and Probability Letters* 150, 68–73. doi: 10.1016/j.spl.2019.02.013.

M. Böhnstedt, H. Putter, A. Daňko, M. J. Daňko, and J. Gampe (2021). Joint modeling of interval counts of recurrent events and death. *Biometrical Journal* 63, 323–340. doi: 10.1002/bimj.201900367.

M. Böhnstedt, J. Gampe, and H. Putter (2021). Information measures and design issues in the study of mortality deceleration: findings for the gamma-Gompertz model. *Lifetime Data Analysis* 27, 333–356. doi: 10.1007/s10985-021-09518-4.

M. Böhnstedt, H. Putter, N. Ouellette, G. Claeskens, and J. Gampe. Shifting attention to old age: Detecting mortality deceleration using focused model selection. *Manuscript submitted for publication*.

M. Böhnstedt, J. Gampe, M. A. A. Caljouw, and H. Putter. Incorporating delayed entry into the joint frailty model for recurrent events and a terminal event. *Manuscript submitted for publication*.





# Acknowledgements

I would like to give a word of thanks to everyone who, one way or another, contributed to the completion of this thesis.

First of all, I would like to express my gratitude to my supervisors Prof. dr. Hein Putter and Dr. Jutta Gampe for giving me the opportunity to embark on my PhD studies and for providing valuable guidance throughout. It was great to work with the two of you and learn from such brilliant researchers.

Several people were involved in the research that resulted in this thesis. I would like to thank Nadine Ouellette, Gerda Claeskens, Maciej Dańko, Aleksandra Dańko, Monique Caljouw, and Paul Eilers for the pleasant collaborations and their helpful comments and contributions.

During my visits to the Leiden University Medical Center, I met a number of people who were all very welcoming and always happy to help, which I very much appreciated.

I also want to express my thanks to the many colleagues at the Max Planck Institute for Demographic Research – scientific and non-scientific staff – who created such a positive work environment. To all the former and current fellow PhD students, I was happy to share this experience with you, and I always enjoyed the friendly and supportive atmosphere in our group.

Heartfelt thanks go to my friends, especially to Angela Carollo and László Németh, for all – work-related or other – conversations, pieces of advice, and encouragements.

Last but not least, I am deeply grateful to my family, in particular to my mum, for the continuing love and support.



# Curriculum vitae

Marie Böhnstedt was born on 2<sup>nd</sup> August 1990 in Stralsund, Germany. She completed her secondary studies at Hansa-Gymnasium Stralsund in 2009.

Afterwards, she started her studies at the University of Rostock, Germany, where she obtained a bachelor's degree in Mathematics with a minor in Economics in 2012. She continued her studies as MSc student with a major in Mathematics and a minor in Economics and Demography. In 2014, she graduated with a MSc in Business Mathematics from the University of Rostock. For the work on her MSc thesis entitled "Statistical Assessment of Mortality Deceleration", she joined the Max Planck Institute for Demographic Research (MPIDR) in Rostock.

From 2014 to 2017, she was employed as research scientist in the Laboratory of Statistical Demography at the MPIDR, before taking up a PhD position at the same institute. In 2018, she enrolled as PhD student in the department of Biomedical Data Sciences at the Leiden University Medical Center under the supervision of Prof. dr. Hein Putter. She conducted her PhD research at the MPIDR, where she was supervised by Dr. Jutta Gampe. Her work, which resulted in this thesis, focused on the development of statistical methodology for specific frailty models for time-to-event data in demographic and biomedical applications. She presented her research at different statistical conferences.

