

---

Modelos multiestado com fragilidade

*Renata Soares da Costa*

---



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

## Modelos multiestado com fragilidade

**Renata Soares da Costa**

***Orientadora: Profa. Dra. Vera Lúcia Damasceno Tomazella***

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

**USP/UFSCar – São Carlos**  
**Abril de 2016**



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

## Frailty multistate models

**Renata Soares da Costa**

***Advisor:* Profa. Dra. Vera Lúcia Damasceno Tomazella**

Master dissertation submitted to the *Instituto de Ciências Matemáticas e de Computação - ICMC-USP* and to the *Departamento de Estatística – DEs-UFSCar*, in partial fulfillment of the requirements for the degree of the Master Joint Graduate Program in Statistics DEs-UFSCar/ICMC-USP. *FINAL VERSION.*

**USP/UFSCar – São Carlos**  
**April 2016**

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar  
Processamento Técnico  
com os dados fornecidos pelo(a) autor(a)

C837m Costa, Renata Soares da  
Modelos multiestado com fragilidade / Renata  
Soares da Costa. -- São Carlos : UFSCar, 2016.  
79 p.

Dissertação (Mestrado) -- Universidade Federal de  
São Carlos, 2016.

1. Análise de sobrevivência. 2. Modelos  
multiestado. 3. Modelos de fragilidade. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia  
Programa Interinstitucional de Pós-Graduação em Estatística

---

Folha de Aprovação

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata Renata Soares da Costa, realizada em 31/03/2016:

---

Profa. Dra. Vera Lucia Damasceno Tomazella  
UFSCar

---

Profa. Dra. Gleici da Silva Castro Perdoná  
USP

---

Prof. Dr. Vinicius Fernando Calsavara  
CIPE



---

## Agradecimentos

---

Primeiramente agradeço a Deus e ao Universo por permitir minha entrada no programa de pós-graduação, encontrar pessoas especiais e me dar forças para enfrentar os momentos difíceis ao longo desta minha trajetória.

À minha mãe Madalena, que nunca mediu esforços para proporcionar-me a melhor educação possível, seja ela moral ou escolar. Por sempre se preocupar com o meu bem-estar, além do seu apoio em situações difíceis.

Ao meu namorado Bruno e seus familiares por estarem sempre ao meu lado e dar todo o apoio nos momentos difíceis, em especial à Iseuda, Fátima, Climene e Zeudimar.

À Universidade Federal de São Carlos e Universidade de São Paulo e a todos os professores que tive a oportunidade de estudar.

À professora Vera Tomazella, pela orientação neste mestrado.

Aos meus colegas da pós-graduação, em especial aos meus amigos Evelyny, William e Aline, pela amizade e companherismo.



---

# Resumo

---

Frequentemente eventos intermediários fornecem informações mais detalhadas sobre o processo da doença ou recuperação, por exemplo, e permitem uma maior precisão na previsão do prognóstico de pacientes. Tais eventos não fatais durante o curso da doença podem ser vistos como transições de um estado para outro. A ideia básica dos modelos multiestado é que o indivíduo se move através de uma série de estados em tempo contínuo, sendo possível estimar as probabilidades e intensidades de transição entre eles e o efeito das covariáveis associadas a cada transição.

Muitos estudos incluem o agrupamento dos tempos de sobrevivência como, por exemplo, em estudos multicêntricos, e também é de interesse estudar a evolução dos pacientes ao longo do tempo, caracterizando assim dados multiestado agrupados. Devido ao fato de os dados virem de diferentes centros/grupos, os tempos de falha desses indivíduos estarem agrupados e a fatores de risco comuns não observados, é interessante considerar o uso de fragilidades para que possamos capturar a heterogeneidade entre os grupos no risco para os diferentes tipos de transição, além de considerar a estrutura de dependência entre transições dos indivíduos de um mesmo grupo.

Neste trabalho apresentamos a metodologia dos modelos multiestado, dos modelos de fragilidade e, em seguida, a integração dos modelos multiestado com modelos de fragilidade, tratando do seu processo de estimação paramétrica e semiparamétrica. O estudo de simulação realizado mostrou a importância de considerarmos fragilidades em modelos multiestado agrupados, pois sem considerá-las, as estimativas tornam-se viesadas. Além disso,

verificamos as propriedades frequentistas dos estimadores do modelo multiestado com fragilidades aninhadas.

Por fim, como um exemplo de aplicação a um conjunto de dados reais, utilizamos o processo de recuperação de transplante de medula óssea de pacientes tratados em quatro hospitais. Fizemos uma comparação de modelos por meio das medidas de qualidade do ajuste AIC e BIC, chegando à conclusão de que o modelo que considera dois efeitos aleatórios (uma para o hospital e outro para a interação transição-hospital) ajusta-se melhor aos dados. Além de considerar a heterogeneidade entre os hospitais, tal modelo também considera a heterogeneidade entre os hospitais em cada transição. Sendo assim, os valores das fragilidades estimadas da interação transição-hospital revelam o quão frágeis os pacientes de cada hospital são para experimentarem determinado tipo de evento/transição.

**Palavras-chave:** Análise de sobrevivência, Modelos Multiestado, Modelos de fragilidade.

---

# Abstract

---

Often intermediate events provide more detailed information about the disease process or recovery, for example, and allow greater accuracy in predicting the prognosis of patients. Such non-fatal events during the course of the disease can be seen as transitions from one state to another. The basic idea of a multistate models is that the person moves through a series of states in continuous time, it is possible to estimate the transition probabilities and intensities between them and the effect of covariates associated with each transition.

Many studies include the grouping of survival times, for example, in multi-center studies, and is also of interest to study the evolution of patients over time, characterizing grouped multistate data. Because the data coming from different centers/groups, the failure times these individuals are grouped and the common risk factors not observed, it is interesting to consider the use of frailty so that we can capture the heterogeneity between the groups at risk for different types of transition, in addition to considering the dependence structure between transitions of individuals of the same group.

In this work we present the methodology of multistate models, frailty models and then the integration of models with multi-state fragility models, dealing with the process of parametric and semi-parametric estimation. The conducted simulation study showed the importance of considering frailty in grouped multistate models, because without considering them, the estimates become biased. Furthermore, we find the frequentist properties of estimators of multistate model with nested frailty.

Finally, as an application example to a set of real data, we use the process of bone marrow transplantation recovery of patients in four hospitals. We did a comparison of models through

---

quality measures setting AIC and BIC, coming to the conclusion that the model considers two random effects (one for the hospital and another for interaction transition-hospital) fits the data better. In addition to considering the heterogeneity between hospitals, such a model also considers the heterogeneity between hospitals in each transition. Thus, the values of the frailty estimated interaction transition-hospital reveal how fragile patients from each hospital are to experience certain type of event/transition.

**Keywords:** Survival Analysis, Multistate Models, Frailty Models.

---

# Sumário

---

Agradecimentos

Resumo

Abstract

Lista de Figuras

Lista de Tabelas

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Visão geral . . . . .	1
1.2	Objetivos da dissertação . . . . .	4
1.3	Introdução à análise de sobrevivência . . . . .	5
1.3.1	Funções do tempo de sobrevivência . . . . .	7
1.3.2	Modelo de riscos proporcionais de Cox . . . . .	8
1.4	Organização dos capítulos . . . . .	9
<b>2</b>	<b>Modelos multiestado</b>	<b>11</b>
2.1	Modelagem . . . . .	13
2.1.1	Representação dos dados . . . . .	17
2.1.2	Estimação dos parâmetros . . . . .	18
2.2	Considerações finais . . . . .	20

## REFERÊNCIAS BIBLIOGRÁFICAS

---

<b>3 Modelos de fragilidade</b>	<b>22</b>
3.1 Modelo de fragilidade univariada . . . . .	23
3.2 Modelo de fragilidade compartilhada . . . . .	25
3.3 Modelo de fragilidade aninhada . . . . .	26
3.4 Considerações finais . . . . .	28
<b>4 Modelo multiestado com fragilidade</b>	<b>29</b>
4.1 Modelagem . . . . .	30
4.2 Estimação dos parâmetros . . . . .	32
4.2.1 Estimação paramétrica . . . . .	34
4.2.2 Estimação semiparamétrica . . . . .	39
4.3 Estudo de simulação . . . . .	41
4.3.1 Estudo de simulação 1 . . . . .	43
4.3.2 Estudo de simulação 2 . . . . .	50
4.3.3 Comparação de modelos . . . . .	52
4.4 Considerações finais . . . . .	53
<b>5 Aplicação</b>	<b>55</b>
5.1 Descrição do conjunto de dados . . . . .	55
5.2 Resultados da Aplicação . . . . .	59
5.2.1 Interpretação dos resultados do modelo 3 . . . . .	62
5.2.2 Interpretação dos resultados do modelo 4 . . . . .	66
5.3 Considerações finais . . . . .	71
<b>6 Considerações finais e propostas futuras</b>	<b>72</b>
<b>Referências Bibliográficas</b>	<b>74</b>

---

## Lista de Figuras

---

2.1	- Estruturas multiestado comuns (Hougaard, 2000). . . . .	12
2.2	- Modelo de riscos competitivos para pacientes com transplante de medula óssea. . . . .	12
2.3	- Modelo multiestado para pacientes com transplante de medula óssea. . . . .	13
2.4	- Modelo de deficiência para representação do conjunto de dados. . . . .	17
4.1	- Exemplo de aplicação (Liquet <u>et al.</u> , 2012). . . . .	29
4.2	- Casos particulares do modelo multiestado com fragilidades aninhadas (4.1). . . . .	31
4.3	- Esquema multiestado para o estudo de simulação. . . . .	42
4.4	- Comportamento do estudo de simulação 1A. Linha contínua com ●: modelo (4.13) com fragilidade; Linha pontilhada com +: modelo (4.13) sem fragilidade. . . . .	46
4.5	- Comportamento do estudo de simulação 1B. Linha contínua com ●: modelo (4.13) com fragilidade; linha pontilhada com +: modelo (4.13) sem fragilidade. . . . .	48
4.6	- Comportamento dos estudos de simulação 1A e 1B com relação ao modelo (4.13) com fragilidade. Linha contínua com ●: modelo com variâncias da interação transição-grupo iguais; linha pontilhada com ×: modelo com variâncias da interação transição-grupo diferentes. . . . .	49
4.7	- Comportamento dos cenários A, B e C no estudo de simulação 2. . . . .	52
5.1	- Processo de recuperação de um transplante de medula óssea. . . . .	56
5.2	- Fragilidades estimadas do modelo 3. . . . .	65

---

5.3 - Fragilidades estimadas a nível de hospital. . . . .	66
5.4 - Fragilidades estimadas a nível de transição. . . . .	67

---

## Lista de Tabelas

---

2.1	- Representação do conjunto de dados. . . . .	17
2.2	- Representação do conjunto de dados transformado. . . . .	18
4.1	- Cenários do estudo de simulação 1. . . . .	44
4.2	- Parâmetros do estudo de simulação 1. . . . .	44
4.3	- Resultados do estudo de simulação 1A. . . . .	45
4.4	- Resultados do estudo de simulação 1B. . . . .	47
4.5	- Cenários do estudo de simulação 2. . . . .	50
4.6	- Resultados do estudo de simulação 2. . . . .	51
4.7	- Valores médios de AIC e BIC para 500 simulações, considerando uma amostra de $n = 150$ indivíduos ( $G = 5; n_i = 30$ ). . . . .	53
5.1	- Total de pacientes em cada hospital. . . . .	57
5.2	- Primeiras 2 linhas do conjunto de dados. . . . .	57
5.3	- Medidas descritivas das variáveis contínuas. . . . .	58
5.4	- Frequências das transições. . . . .	59
5.5	- Proporção das transições. . . . .	59
5.6	- Primeiros dois pacientes do conjunto de dados transformado. . . . .	60
5.7	- Resultados dos modelos ajustados. . . . .	61
5.8	- Valores de AIC, BIC dos modelos 1, 2 e 4. . . . .	62
5.9	- Intervalos de confiança de 95% dos parâmetros de regressão do modelo 3. . . . .	63
5.10	- Variâncias dos efeitos aleatórios do modelo 3. . . . .	63

---

5.11 - Fragilidades estimadas do modelo 3. . . . .	64
5.12 - Fragilidades estimadas do modelo 4 a nível de hospital $v_h$ ( $h = 1, 2, 3, 4$ ). . . . .	66
5.13 - Fragilidades estimadas do modelo 4 a nível de transição. . . . .	67
5.14 - Intervalos de confiança do modelo 4. . . . .	68

---

# 1

## Introdução

---

### 1.1 Visão geral

A análise de sobrevivência é um ramo da Estatística em que o tempo até a ocorrência de um evento de interesse é a variável resposta, (morte de um paciente, por exemplo), sendo mais conhecido como tempo de falha. A característica principal de dados de sobrevivência é a presença de censura, que é a observação parcial da resposta. Apesar de censuradas, todas as observações de um estudo devem ser utilizadas na análise, já que mesmo incompletas, elas fornecem informações sobre o tempo de falha e a retirada delas pode gerar conclusões viesadas.

Um problema muito comum em análise de sobrevivência são os eventos ditos competitivos. Nessas situações existem diversas causas de falha (eventos) presentes, ao mesmo tempo, na vida de um indivíduo ou sistema. Por essa razão, dizemos que essas causas competem entre si para provocar a falha do mesmo. Entretanto, apenas uma dessas causas é a responsável por determinada falha, impedindo assim a ocorrência das demais. Esse tipo de abordagem tem sido bastante desenvolvida desde os trabalhos de [Gray \(1988\)](#) e [Fine & Gray \(1999\)](#), sendo que uma completa revisão sobre este tema está no livro publicado por [Pintilie \(2006\)](#).

O modelo de riscos competitivos é um caso particular dos modelos multiestado, nos quais descrevem a transição entre eventos intermediários. Modelos multiestado são frequentemente utilizados para descrever dados longitudinais e são definidos a partir de um processo estocástico em tempo contínuo, em que os indivíduos movem-se através de um número fi-

nito de estados, sendo possível estimar as probabilidades e o risco de transição entre eles e o efeito das covariáveis associadas a cada transição. A sua complexidade depende do número de estados definidos e do número de transições permitidas entre eles.

Sobre os modelos multiestado destacam-se os trabalhos de [Andersen & Keiding \(2002\)](#) e [Putter et al. \(2007\)](#), além dos livros de [Andersen et al. \(1993\)](#), [Hougaard \(2000\)](#), [Aalen et al. \(2008\)](#) e [Willekens \(2014\)](#). Tal metodologia é aplicável em diversas áreas do conhecimento. Em demografia, tem-se os estudos sobre migração (vide [Rogers \(1975\)](#) e [Rogers \(1995\)](#)), as alterações do estado civil e outros processos do curso da vida (vide [Courgeau & Lelièvre \(1992\)](#), [Willekens \(1999\)](#) e [Lawrence \(2003\)](#)), além de outras aplicações. Uma visão geral dos modelos multiestado aplicados em Bioestatística está em [Hougaard \(2000\)](#), [Commenges \(2002\)](#) e [Putter et al. \(2007\)](#). Em economia, a principal aplicação tem sido na dinâmica da força do trabalho (ver [Flinn & Heckman \(1983\)](#), [Van den Berg \(2001\)](#) e [Fougere & Kamionka \(2008\)](#)), além de ser aplicado em dinâmica da pobreza e reincidência.

O problema da heterogeneidade devido a fatores de risco não observados, tais como fatores ambientais e genéticos, ou que de alguma forma não foram consideradas no planejamento, foi primeiramente abordado por [Clayton \(1978\)](#) e [Vaupel et al. \(1979\)](#). Para explicar essa heterogeneidade, eles propuseram um modelo de efeitos aleatórios, dando origem a uma vasta literatura sobre o que é hoje conhecido como modelo de fragilidade. O modelo de fragilidade é caracterizado pela inclusão de um efeito aleatório, no qual é uma variável aleatória não observável, que representa as informações que não puderam ou não foram observadas ([Wienke \(2011\)](#) e [Duchateau & Janssen \(2008\)](#)). Num modelo de fragilidade compartilhada o efeito aleatório está associado a grupos de indivíduos, ao invés de estar associado a cada indivíduo isoladamente, como é o caso do modelo de fragilidade individual.

A integração de modelos multiestado com modelos de fragilidade pode fornecer modelos de sobrevivência poderosos para estudar o risco de muitos eventos relacionados entre si, enquanto consideram a dependência entre os indivíduos agrupados. Muitas situações práticas podem exemplificar o interesse dessa integração. Dentre elas temos os estudos multicêntricos de câncer, no qual estudamos a evolução da doença ao longo do tempo (com os estados morte,

recidiva e metástase, por exemplo) e, ao mesmo tempo, consideramos a dependência entre os indivíduos de um mesmo hospital/centro, que compartilham fatores de risco não observados, devido às características do hospital, da região e assim por diante.

A questão da inclusão de fragilidades em modelos multiestado é relativamente recente. Muitos trabalhos abordam essa questão no contexto individual. Um dos primeiros foi o de [Pickles & Crouchley \(1995\)](#), que revisa as principais abordagens em dados de sobrevivência multivariados, sendo que tais dados geralmente têm os tempos de falha correlacionados. Os autores do artigo assumem que a correlação entre os tempos de falha pode ser captada por um modelo de efeitos aleatórios aplicado em um modelo de sobrevivência bivariado, no qual é uma estrutura multiestado. Outros autores, como [Bhattacharyya & Klein \(2005\)](#), [Foucher et al. \(2006\)](#), [Putter & van Houwelingen \(2011\)](#), [Bijwaard \(2014\)](#) e [de Castro et al. \(2015\)](#), levam em conta a associação entre as intensidades de transição de um mesmo indivíduo, enquanto consideram os tempos dos eventos de diferentes indivíduos como independentes. Tais modelos estão mais no espírito dos modelos de fragilidade univariados, onde cada indivíduo tem um nível de risco diferente, devido aos seus próprios fatores de risco não observados. Nesse contexto, a variância será maior quanto maior for a heterogeneidade entre os indivíduos e a dependência entre os tempos das diferentes transições para cada um deles.

Por outro lado, [Yen et al. \(2010\)](#) estudou modelos com fragilidade Gama e Binomial, nas quais eram independentes entre as transições, mas compartilhada por grupo de indivíduos, com foco nas probabilidades de transição. [Ma et al. \(2010\)](#) e [Liquet et al. \(2012\)](#) apresentaram um modelo similar com fragilidades Gama, porém formulado em termos das intensidades de transição. Nesses trabalhos, a modelagem assume que as fragilidades são independentes entre os diferentes tipos de transição, mas é compartilhada por grupos de indivíduos. Ou seja, não existe uma associação entre as transições. Portanto, essas suposições estão mais no espírito dos modelos de fragilidade compartilhada, já que ela é responsável pelo agrupamento, embora não para a dependência entre as transições. Para considerar tanto

a dependência entre os indivíduos de um mesmo grupo, como a dependência entre diferentes transições, [Rotolo \(2013\)](#) propôs um modelo multiestado com fragilidades aninhadas.

## 1.2 Objetivos da dissertação

Com o objetivo de captar a heterogeneidade não observada entre os diferentes grupos (já que indivíduos de um mesmo grupo compartilham fatores de risco não observados), além de lidar com a estrutura de dependência entre as transições de indivíduos desse mesmo grupo, iremos considerar um modelo multiestado com duas fragilidades aninhadas. Uma delas irá considerar os fatores de risco não observados que são específicos do grupo, e uma outra irá lidar com o efeito do grupo em cada transição. Portanto, devido aos fatores de risco não observados, os riscos dos indivíduos em cada transição variam entre os grupos. Além disso, é esperado que exista uma variação entre esses riscos dentro de um mesmo grupo, mas que de alguma forma eles sejam dependentes entre si.

Nesta dissertação fazemos um estudo detalhado do modelo multiestado com duas fragilidades aninhadas, por meio de estudos de simulação e aplicação a um conjunto de dados reais. A partir desse modelo, outros casos particulares, tais como modelos multiestado com fragilidades compartilhadas ou independentes, foram também estudados. Desta forma, o modelo geral é mais flexível e torna-se adequado para descrever a dependência em dois níveis, na qual associa os tempos de indivíduos que compartilham fatores de risco não observados para diferentes tipos de eventos/transições.

Para a construção da função de verossimilhança do modelo utilizamos as abordagens paramétrica e semiparamétrica. A abordagem paramétrica é baseada na maximização da verossimilhança marginal, enquanto que na abordagem semiparamétrica a estimação dos parâmetros é baseada na maximização da verossimilhança parcial penalizada. Neste trabalho, para a estimação dos parâmetros nos estudos de simulação e na aplicação a um conjunto de dados reais, foi utilizado apenas o método de estimação semiparamétrica.

Na seção seguinte faremos uma breve introdução à análise de sobrevivência.

## 1.3 Introdução à análise de sobrevivência

A análise de sobrevivência é uma importante área da estatística que experimentou um rápido crescimento durante a última metade do século XX. Essa área tem uma vasta literatura, que além de vários artigos, muitos autores, tais como [Andersen et al. \(1993\)](#), [Hougaard \(2000\)](#), [Kalbfleisch & Prentice \(2002\)](#), [Klein & Moeschberger \(2003\)](#), [Aalen et al. \(2008\)](#), dentre outros, publicaram livros tratando deste assunto, explorando diversas abordagens e níveis de acessibilidade. De maneira geral, a análise de sobrevivência é definida como a análise do tempo até a ocorrência de um evento de interesse, sendo este tempo denominado de tempo de falha. Um exemplo bem natural desse tempo é o tempo do nascimento à morte de um indivíduo. O termo análise de sobrevivência refere-se basicamente a situações médicas envolvendo dados censurados, mas condições similares ocorrem em outras áreas onde se usam essas mesmas técnicas, como em experimentos industriais, estudos demográficos e sócio-econômicos.

Em relação aos dados de sobrevivência, existem quatro tipos de classificação, em que cada tipo requer uma modificação na análise, visando estimar corretamente a distribuição do tempo de vida. São eles:

- **Dados completos:** Nesses tipos de dados há o conhecimento do tempo de falha de todos os indivíduos, pois neste caso todos os indivíduos sob estudo falharam e seus tempos de falha são conhecidos.
- **Dados censurados:** A característica principal de dados de sobrevivência é a presença de censura, que é a observação parcial da resposta, ou seja, por alguma razão diferente da estudada um indivíduo é retirado do estudo de interesse e, desta forma, o seu tempo de falha não é observado. Isso significa que toda informação referente à resposta se resume ao conhecimento de que o tempo de falha é superior àquele observado. Apesar de censuradas, todas as observações de um estudo devem ser utilizadas na análise, já que mesmo incompletas, elas fornecem informações sobre o tempo de falha e a retirada delas pode gerar conclusões viciadas. O mecanismo de censura mais comum em estudos

envolvendo dados de sobrevivência é a censura à direita, onde o tempo de ocorrência do evento de interesse está a direita do tempo registrado. Mais detalhes deste e de outros mecanismos de censuras são encontrados em [Colosimo & Giolo \(2006\)](#).

- **Dados de eventos recorrentes:** São situações em que os eventos de interesse ocorrem mais de uma vez para o mesmo indivíduo.
- **Dados de riscos competitivos:** Neste caso, várias causas de falha estão presentes ao mesmo tempo na vida de um indivíduo, no entanto, somente é possível observar o tempo até a ocorrência da primeira (e única) causa, o que impede que outras aconteçam (por exemplo, se essas causas representam diferentes causas de morte, apenas a primeira a ocorrer é observada). Desta forma, os indivíduos sob estudo podem experimentar apenas uma das diversas causas de falha, e para cada uma delas é observado o tempo e o tipo de falha ocorrida.

A censura é a principal característica dos dados de sobrevivência. Entre os mecanismos de censura, pode-se citar três deles que são bem conhecidos e são definidos por censura à direita, pois o tempo de ocorrência do evento de interesse está à direita do tempo registrado. Eles são definidos a seguir, conforme [Colosimo & Giolo \(2006\)](#):

- **Censura do tipo I:** aquela em que o estudo será terminado após um período pré-estabelecido de tempo. Sendo assim, as observações cujo evento de interesse não foi observado até este tempo são ditas censuradas.
- **Censura do tipo II:** aquela em que o estudo será terminado após ocorrer o evento de interesse em um número pré-estabelecido de indivíduos.
- **Censura do tipo aleatória:** é o mecanismo de censura mais comum em estudos médicos e pode ocorrer se a observação for retirada no decorrer do estudo sem ter ocorrido o evento de interesse. Uma outra forma de ocorrer este tipo de censura é se o evento de interesse ocorrer por uma razão diferente da estudada.

### 1.3.1 Funções do tempo de sobrevivência

Considerando  $T$  uma variável aleatória positiva, geralmente contínua, na qual representa o tempo de falha de um indivíduo e, seja  $C$  uma outra variável aleatória independente de  $T$ , representando o tempo de censura associado a este indivíduo. Então os dados observados consistem em  $t = \min(T, C)$ . Segundo Colosimo & Giolo (2006), os dados de sobrevivência para o  $i$ -ésimo indivíduo sob estudo são representados, em geral, pelo par  $(t_i, \delta_i)$ , em que  $t_i$  é o tempo de falha ou censura e  $\delta_i$  a variável indicadora de falha ou censura, representada da seguinte maneira:

$$\delta_i = \begin{cases} 1, & \text{se } t_i \text{ é um tempo de falha} \\ 0, & \text{se } t_i \text{ é um tempo censurado.} \end{cases}$$

A **função densidade de probabilidade de  $T$** , denominada  $f(t)$ , pode ser interpretada como a probabilidade de um indivíduo experimentar um evento em um intervalo instantâneo de tempo:

$$f(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(t \leq T < t + \Delta t)}{\Delta t},$$

em que  $\Delta t$  é o incremento de tempo infinitamente pequeno.

Sendo uma das principais funções probabilísticas para descrever estudos de sobrevivência, a **função de sobrevivência** é definida como sendo a probabilidade de um indivíduo não falhar até um tempo  $t$ , isto é, a probabilidade de um indivíduo sobreviver por mais do que um certo tempo  $t$ , sendo escrita em termos probabilísticos como

$$S(t) = P(T \geq t).$$

Assumindo  $\Delta t$  bem pequeno,  $\lambda(t)$  representa a **taxa de falha (ou de risco)** instantânea no tempo  $t$  condicional à sobrevivência até o tempo  $t$ , sendo bastante útil para descrever a distribuição do tempo de vida de indivíduos, bem como descrever a forma em que a taxa instantânea de falha muda com o tempo (Colosimo & Giolo, 2006). As taxas de falha são números positivos, mas sem limite superior e a sua função é definida como:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

A **função de risco acumulado** mede o risco de falha até o tempo  $t$ , e assim como a taxa de falha, ela não está restrita ao intervalo  $[0, 1]$ , sendo definida por:

$$\Lambda(t) = \int_0^t \lambda(u) du.$$

As funções apresentadas anteriormente são matematicamente equivalentes e representam diferentes formas de representar o mesmo fenômeno, sendo algumas dessas relações:

$$\begin{aligned} \lambda(t) &= \frac{f(t)}{S(t)} = -\frac{d}{dt} (\log S(t)) \\ \Lambda(t) &= \int_0^t \lambda(u) du = -\log S(t) \\ S(t) &= \exp \{-\Lambda(t)\} = \exp \left\{ -\int_0^t \lambda(u) du \right\}. \end{aligned}$$

### 1.3.2 Modelo de riscos proporcionais de Cox

Em muitos estudos, o interesse é estimar o efeito das covariáveis no tempo de sobrevivência. O modelo de riscos proporcionais proposto por [Cox \(1972\)](#) é bastante utilizado para tal fim, no qual modela dados de sobrevivência na presença de covariáveis por meio da função de risco  $\lambda(t)$ . Este modelo baseia-se na suposição de que o risco, condicionado às covariáveis explicativas, é proporcional a uma dada função de risco basal  $\lambda_0(t)$ . Considerado um vetor  $\mathbf{X} = (X_1, \dots, X_p)'$  com  $p$  covariáveis, o modelo de riscos proporcionais é definido como

$$\lambda(t|\mathbf{X}) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{X}), \quad (1.1)$$

em que  $\lambda_0(t)$  é o risco no caso em que  $\mathbf{X} = (0, \dots, 0)'$  e  $\boldsymbol{\beta}$  é o vetor dos parâmetros de regressão a serem estimados.

O modelo dado em (1.1) é composto pelo produto de dois componentes:  $\exp(\boldsymbol{\beta}'\mathbf{X})$  que

mede o efeito das covariáveis e  $\lambda_0(t)$ , que pode ser paramétrico ou não. Quando não paramétrico,  $\lambda_0(t)$  é uma função não negativa do tempo; quando paramétrico, ele será um componente aleatório que descreve probabilisticamente o comportamento do tempo de sobrevivência, podendo assumir uma distribuição exponencial, log-normal, Weibull, dentre outras.

A razão entre o risco de dois indivíduos é

$$\frac{\lambda(t|\mathbf{X}_i)}{\lambda(t|\mathbf{X}_{i'})} = \exp\{\boldsymbol{\beta}'(\mathbf{X}_i - \mathbf{X}_{i'})\}$$

no qual é constante ao longo do tempo. Portanto, as funções de risco de diferentes indivíduos são proporcionais entre si. A principal contribuição dada por Cox (1972) é a abordagem da estimação semiparamétrica para os parâmetros de regressão. A função de risco basal pode ser estimada de forma não-paramétrica e retirada da função de verossimilhança. Os parâmetros  $\boldsymbol{\beta}$  podem então ser estimados por meio da maximização da verossimilhança parcial

$$L_P(\boldsymbol{\beta}) = \prod_{i=1}^n \left( \frac{\exp(\boldsymbol{\beta}'\mathbf{X}_i)}{\sum_{i' \in R(t_i)} \exp(\boldsymbol{\beta}'\mathbf{X}_{i'})} \right)^{\delta_i},$$

em que  $R(t_i)$  é o conjunto dos índices das observações sob risco no tempo  $t_i$  e  $\delta_i$  é o indicador de censura.

## 1.4 Organização dos capítulos

Este trabalho está organizado em seis capítulos. No Capítulo 2 é apresentada a metodologia dos modelos multiestado, com uma descrição de suas estruturas, tipos de modelos, da modelagem, além da estimação de seus parâmetros. Os modelos de fragilidades são apresentados no Capítulo 3, com uma visão geral da sua necessidade, além de abordar os modelos de fragilidade univariada, compartilhada e aninhada. No Capítulo 4 descrevemos a integração dos modelos multiestado com modelos de fragilidade, tratando de sua modelagem e seus processos de estimação, apresentando um estudo de simulação no final na capítulo. Como

aplicação da metodologia proposta deste trabalho, iremos trabalhar com um conjunto de dados que trata do processo de recuperação do transplante de medula óssea de pacientes oriundos de quatro hospitais, no qual é estruturado em vários estados e há agrupamentos de indivíduos, sendo discutido em detalhes no Capítulo 5. Por fim, no Capítulo 6 temos as considerações finais do trabalho e suas propostas futuras.

---

## 2

# Modelos multiestado

---

Na análise de sobrevivência padrão, o foco é o tempo até a ocorrência de um único evento de interesse, sendo que este tempo pode ser observado ou censurado. A censura ocorre devido a diversas causas, estando relacionadas ou não com o evento de interesse. Além disso, alguns eventos podem ocorrer antes do evento de interesse, mudando o seu risco. Frequentemente, eventos intermediários fornecem informações mais detalhadas sobre o processo da doença ou recuperação, por exemplo, e permitem uma maior precisão na previsão do prognóstico de pacientes. Tais eventos não fatais durante o curso da doença podem ser vistos como transições de um estado para outro, na qual a origem do tempo é caracterizada como um estado inicial, como o início de um tratamento, estados intermediários transientes, e o ponto final é um estado final absorvente, que não permite transições a partir dele, sendo que quando um indivíduo atinge esse estado permanece nele para sempre. A morte é o exemplo mais comum de estado absorvente.

[Hougaard \(2000\)](#) ilustra algumas estruturas de modelos multiestado, representadas por diagramas que indicam o número finito de estados que um indivíduo pode ocupar, tais como os modelos de mortalidade, de riscos competitivos, de estados alternados, de deficiência, bivariado e de eventos recorrentes, como mostrados na Figura 2.1. Percebe-se que a estrutura do modelo de mortalidade corresponde à análise de sobrevivência padrão.

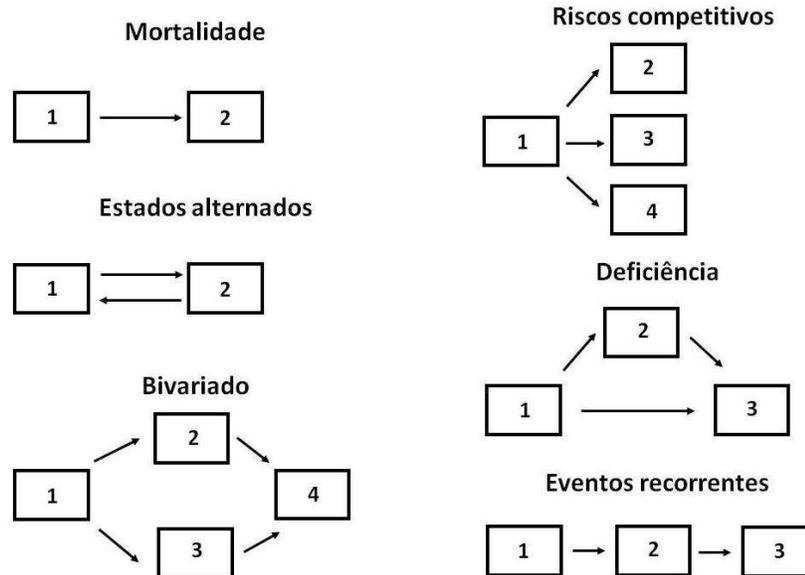


Figura 2.1 - *Estruturas multiestado comuns* (Hougaard, 2000).

Na abordagem de riscos competitivos os indivíduos sob estudo podem experimentar somente um dos tipos distintos de falha, e para cada um deles é observado o tempo de falha e o tipo de falha. Tomando como exemplo o caso de pacientes com transplante de medula óssea, temos um estado inicial (transplante) e dois estados absorventes (recidiva e morte), que atuam como dois eventos competitivos (Figura 2.2).

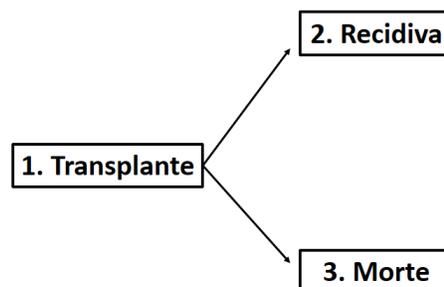


Figura 2.2 - *Modelo de riscos competitivos para pacientes com transplante de medula óssea.*

No entanto, em determinadas situações o interesse é no tempo até um ou mais pontos finais e de que forma ele depende das covariáveis. Um exemplo disso é o modelo multiestado mostrado na Figura 2.3, no qual recuperação de plaquetas é um estado intermediário entre o estado inicial transplante e os estados absorventes recidiva e morte.

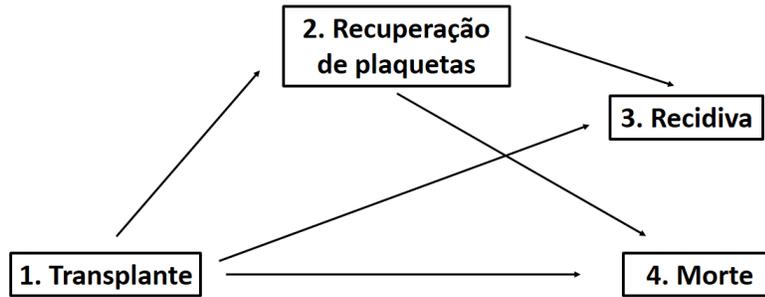


Figura 2.3 - Modelo multiestado para pacientes com transplante de medula óssea.

Modelos multiestado representam a então chamada abordagem *análise de histórico de eventos* para dados longitudinais (Rotolo, 2013). Esta análise lida com dados que são obtidos pela observação de indivíduos ao longo do tempo, com foco nos eventos que ocorrem com eles. Logo, seus dados consistem nos tempos de ocorrência dos eventos e quais tipos de eventos ocorrem (Andersen & Keiding, 2002), em que os eventos podem ser considerados como uma transição de um estado para outro.

## 2.1 Modelagem

Um processo estocástico é um conjunto de variáveis aleatórias  $\mathcal{X}(t)$  que variam ao longo de um período  $T$  pertencente ao conjunto dos inteiros não negativos, em que a variável  $\mathcal{X}(t)$  representa o estado que o processo está no instante  $t$ . Desta forma, um modelo multiestado é um processo estocástico  $(\mathcal{X}(t), t \in [0, \infty))$ , em que  $\mathcal{X}(t)$  é uma variável aleatória que pode assumir valores no conjunto finito de estados  $\varepsilon = \{0, \dots, N\}$ . Tal processo possui a informação das diferentes transições que ocorrem a um indivíduo ao longo do tempo, assim como o tempo de transição.

Quando o indivíduo está em um estado  $m$ , o próximo estado para o qual ele se move e o tempo da mudança são indicados pela intensidade de transição  $\lambda_{ml}(t)$ , para cada par de estados  $m$  e  $l$ , na qual é definida como

$$\lambda_{ml}(t|\mathcal{X}(u), u \in [0, t)) = \lim_{\Delta t \rightarrow 0} \frac{P(\mathcal{X}(t + \Delta t) = l | \mathcal{H}_t^-, \mathcal{X}(t^-) = m)}{\Delta t}, \quad (2.1)$$

em que  $\mathcal{H}_t$  é toda a história do processo até o instante  $t$  e  $\mathcal{H}_{t-} = \{\mathcal{X}(t'), t' \in [0, t)\}$ .

Segundo Hougaard (1999), o risco total fora do estado  $m$  é obtido pela soma sobre todos os outros estados ( $l \neq m$ ), sendo indicado de forma semelhante utilizando apenas um único índice. A partir disso, é mais fácil avaliar a probabilidade da não ocorrência de eventos durante um período, porém avaliar a quantidade e o tipo de um evento específico é muito mais complicado. Para resolver esta questão, as probabilidades de transição devem ser consideradas, sendo então a solução para fazer previsões a longo prazo. Assim, as probabilidades de transição avaliadas no instante  $v$  são definidas como

$$P_l(v, t) = P(\mathcal{X}(t) = l | \mathcal{X}(v) = l, v \in [0, t]), \quad (2.2)$$

que estão condicionadas a todo o desenvolvimento até o instante  $v$ . A equação (2.2) apenas é considerada para  $t \geq v$ , sendo a probabilidade do processo  $\mathcal{X}$  estar no estado  $l$  no instante  $t$ , dado o desenvolvimento até ao momento  $v$  (Hougaard, 1999).

Uma questão muito importante é sobre o que é a escala de tempo para a qual  $t$  se refere. Duas abordagens que estão em frequente uso são a *clock forward* e *clock reset*, abordadas em Putter et al. (2007). Na abordagem *clock forward*, o tempo  $t$  refere-se ao tempo desde que o indivíduo entrou no estado inicial, sendo que o “relógio” continua avançando para este indivíduo também quando ocorrem eventos intermediários. Já na abordagem *clock reset* o tempo  $t$  refere-se ao tempo desde a entrada no estado anterior ao atual, sendo o tempo redefinido a zero toda vez que o indivíduo realiza uma transição. O tempo nesse tipo de abordagem, além de ser chamado de *tempo clock reset*, também é chamado de *gap time*.

Entre os modelos multiestado, podemos diferenciar três:

- **Modelos de Markov:** sua suposição implica que o passado e o futuro são condicionalmente independentes, dado o presente. Ou seja, o passado não tem influência no risco de transição. Logo, para fazer a predição no instante  $t$ , a parte relevante da memória é o estado atual, não sendo assim, necessárias informações de como o processo se desenvolveu antes dele. Quando a função de risco não depende do tempo  $t$  (riscos

constantes), o processo de Markov é chamado de homogêneo. Se for assumido que o tempo de permanência depende da história do processo somente através do estado atual, então o modelo multiestado resultante é um modelo de renovação de Markov (Bijwaard, 2014).

- **Modelos semi-Markov:** o risco de transição depende somente do tempo gasto no estado atual. A diferença está na escolha da escala de tempo, na qual utiliza a abordagem *clock reset*.
- **Modelos de extensão de Markov:** é uma pequena extensão dos modelos de Markov, sendo intermediário entre modelos de Markov e modelos gerais, permitindo que os riscos de transição dependam do tempo de transição para o estado atual, além de depender do tempo atual (Hougaard, 2000).

Sob a suposição de Markov, o termo  $\mathcal{H}_t-$  na equação (2.1) não tem influência. Para as probabilidades de transição, temos que

$$P_{ml}(v, t) = P[\mathcal{X}(t) = l | \mathcal{X}(v) = m],$$

onde  $v \leq t$ . A equação (2.1) é a probabilidade do processo estar no estado  $l$  no tempo  $t$ , dado que o processo estava no estado  $m$  no tempo  $v$ . Esta equação é igual a equação (2.2) quando  $\mathcal{X}(v) = m$ .

Segundo Hougaard (2000), a suposição de Markov geralmente permite uma função de risco separada para cada transição, podendo depender do tempo  $t$ , mas que não depende de qualquer outro aspecto da história, como estados visitados no caminho e os tempos das transições anteriores (exceto na medida em que esta informação é refletida pelo estado atual).

O risco de cada transição pode, então, ser modelado separadamente se cada uma delas têm diferentes parâmetros. Em alguns casos, as transições com risco basal e/ou parâmetros de regressão comuns devem ser consideradas juntas nos procedimentos de estimação, como é ilustrado em detalhes por de Wreede et al. (2011), que em seu artigo, define que os

eventos são do mesmo tipo ou estrato se eles compartilham o mesmo risco basal. Assim, os parâmetros de regressão comuns são modelados por meio de um modelo estratificado, onde os estratos correspondem aos tipos de transição.

Consideramos a seguir um modelo de Cox multiestado, de forma que

$$\lambda_q(t_q; \mathbf{X}_q) = \lambda_{q0}(t_q) \exp\left(\boldsymbol{\beta}'_q \mathbf{X}_q\right), \quad (2.3)$$

em que  $q \in \{1, \dots, Q\}$  é o tipo de transição que ocorre, ou seja,  $q$  pertence ao conjunto que contém todos os pares de transições  $ml$  possíveis;  $\lambda_{q0}(t_q)$  é o risco basal para a transição do tipo  $q$ ;  $\boldsymbol{\beta}_q$  é o vetor de coeficientes de regressão específico de cada transição  $q$ ; e  $\mathbf{X}_q$  é o vetor de covariáveis, podendo ser específico de cada transição ou comum a todas as transições.

O modelo de Cox multiestado (2.3) pode ser representado na forma individual e em formato longo, de modo que

$$\lambda_{qi}(t_{qi}; \mathbf{X}_{qi}) = \lambda_{q0}(t_{qi}) \exp\left(\boldsymbol{\beta}'_q \mathbf{X}_{qi}(t_{qi})\right),$$

com  $i = 1, \dots, n$  representando o indivíduo sob estudo e  $\mathbf{X}_{qi}(t)$  sendo o vetor de covariáveis específicas da transição  $q$ , possivelmente dependentes do tempo, derivado de  $\mathbf{X}_{qi}$ .

O modo como  $\mathbf{X}_{qi}(t)$  é obtida de  $\mathbf{X}_{qi}$  está relacionado com a transformação dos dados para um formato longo, no qual é descrito em detalhes em [Putter \(2011\)](#) e [de Wreede et al. \(2011\)](#). Esse tipo de transformação permite que cada transição tenha seu próprio conjunto de covariáveis ou que cada covariável, em cada transição, tenha diferentes coeficientes. Além disso, é necessário incluir informações relativas às últimas transições. A proporcionalidade dos riscos de diferentes transições pode ser especificada por meio do uso do mesmo risco basal e pela inserção de uma variável *dummy* em  $\mathbf{X}_{qi}(t)$  ([Rotolo, 2013](#)).

### 2.1.1 Representação dos dados

Para a representação de um conjunto de dados multiestado, vamos tomar como exemplo um modelo de deficiência, conforme representado na Figura 2.4, com seus estados e transições identificadas.

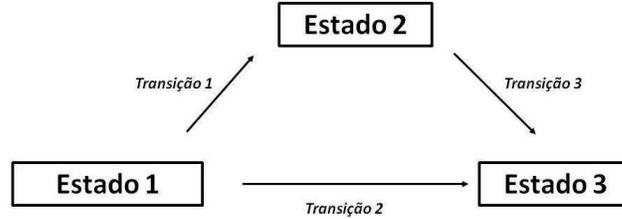


Figura 2.4 - Modelo de deficiência para representação do conjunto de dados.

Sejam  $\mathbf{X} = (X_1, \dots, X_p)'$  o vetor de  $p$  covariáveis para cada indivíduo  $i$  ( $i = 1, \dots, n$ ). A coluna  $T_q$  ( $q = 1, 2, 3$ ) contém os tempos da  $q$ -ésima transição e a coluna  $D_q$  contém os indicadores de falha da transição  $q$ . Os dados para cada indivíduo  $i$  são representados conforme a Tabela 2.1:

Tabela 2.1 - Representação do conjunto de dados.

$i$	$T_1$	$D_1$	$T_2$	$D_2$	$T_3$	$D_3$	$X_1$	$\dots$	$X_p$
1	$t_{11}$	$\delta_{11}$	$t_{21}$	$\delta_{21}$	$t_{31}$	$\delta_{31}$	$x_{11}$	$\dots$	$x_{p1}$
2	$t_{12}$	$\delta_{12}$	$t_{22}$	$\delta_{22}$	$t_{32}$	$\delta_{32}$	$x_{12}$	$\dots$	$x_{p2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n$	$t_{1n}$	$\delta_{1n}$	$t_{2n}$	$\delta_{2n}$	$t_{3n}$	$\delta_{3n}$	$x_{1n}$	$\dots$	$x_{pn}$

Sendo assim, o tempo e o indicador de falha da  $q$ -ésima transição do indivíduo  $i$  são representados por  $t_{qi}$  e  $\delta_{qi}$ , respectivamente, em que

$$\delta_{qi} = \begin{cases} 1, & \text{se } t_{qi} \text{ é um tempo de falha na transição } q \text{ para o indivíduo } i \\ 0, & \text{se } t_{qi} \text{ é um tempo de censura na transição } q \text{ para o indivíduo } i. \end{cases}$$

Vamos supor que o indivíduo 1 experimentou a transição 1 (e por consequência experimentou a transição 3); o indivíduo 2 também experimentou a transição 1, mas seu tempo na transição 3 foi censurado; e o indivíduo  $n$  experimentou a transição 2. Ao trans-

formarmos o conjunto de dados representado na Tabela 2.1 para um formato longo, ele fica da seguinte forma:

Tabela 2.2 - Representação do conjunto de dados transformado.

i	de	para	trans	status	Tinício	Tparada	tempo	$X_1$	$\dots$	$X_p$
1	1	2	1	1	0	$t_{11}$	$t_{11} - 0$	$x_{111}$	$\dots$	$x_{p11}$
1	1	3	2	0	0	$t_{21} = t_{11}$	$t_{21} - 0$	$x_{121}$	$\dots$	$x_{p21}$
1	2	3	3	1	$t_{21}$	$t_{31}$	$t_{31} - t_{21}$	$x_{131}$	$\dots$	$x_{p31}$
2	1	2	1	1	0	$t_{12}$	$t_{12} - 0$	$x_{112}$	$\dots$	$x_{p12}$
2	1	3	2	0	0	$t_{22} = t_{12}$	$t_{22} - 0$	$x_{122}$	$\dots$	$x_{p22}$
2	2	3	3	0	$t_{12}$	$t_{32}$	$t_{32} - t_{12}$	$x_{132}$	$\dots$	$x_{p32}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$						
$n$	1	2	1	0	0	$t_{1n}$	$t_{1n} - 0$	$x_{11n}$	$\dots$	$x_{p1n}$
$n$	1	3	2	1	0	$t_{2n} = t_{1n}$	$t_{2n} - 0$	$x_{12n}$	$\dots$	$x_{p2n}$
$n$	2	3	3	0	0	$t_{3n} = t_{1n}$	$t_{3n} - 0$	$x_{13n}$	$\dots$	$x_{p3n}$

A coluna *de* representa o estado de origem, *para* o estado de destino, *trans* o tipo de transição, *status* o status da transição, *Tinício* o tempo desde a entrada no estado, *Tparada* o tempo de parada no estado e *tempo* o tempo de permanência no estado. Se utilizarmos o modelo multiestado de Markov, o tempo utilizado será o *Tparada*. Caso utilizemos o modelo de multiestado de semi-Markov, o tempo utilizado será o *tempo*.

Em posse de um conjunto de dados multiestado, podemos efetuar a transformação para um formato longo com o auxílio do pacote `mstate` (Putter [et al.](#) (2007) e Putter (2011)) disponível no sistema R (R Core Team, 2013).

### 2.1.2 Estimação dos parâmetros

De acordo com uma estrutura multiestado, para cada indivíduo  $i$  ( $i = 1, \dots, n$ ), existem  $Q$  tipos de transições possíveis. Iremos considerar um total de  $n$  indivíduos, onde o tempo  $t_{qi}$  do indivíduo  $i$  na transição  $q$  ( $q = 1, \dots, Q$ ) pode ser observado ou censurado.

Conforme Rotolo (2013), os indivíduos só estarão sob risco para determinado tipo de evento somente a partir do momento que eles entram no estado inicial associado a este evento. Os tempos destes indivíduos são registrados como tempos de truncamento à esquerda,

denotado por  $\tau_{qi}$ , que corresponde ao tempo da transição anterior. Caso um indivíduo nunca estiver sob risco para um determinado tipo de transição, o tempo de truncamento à esquerda será definido como infinito.

Portanto, para a estimação dos parâmetros do modelo (2.3), utilizaremos o método da máxima verossimilhança. A sua função de verossimilhança é dada por:

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)) = \prod_{i=1}^n \prod_{q=1}^Q [\lambda_{qi}(t_{qi})]^{\delta_{qi}} \frac{S_{qi}(t_{qi})}{S_{qi}(\tau_{qi})}, \quad (2.4)$$

em que  $\boldsymbol{\lambda}_0(\cdot) = (\lambda_{10}(t), \dots, \lambda_{Q0}(t))'$  é o vetor das funções de risco basal. Para o indivíduo  $i$  na transição  $q$ ,  $\lambda_{qi}(t_{qi})$  é a sua função de risco,  $\delta_{qi}$  é o seu indicador de falha,  $S_{qi}(t_{qi})$  e  $S_{qi}(\tau_{qi})$  é a sua função de sobrevivência que considera o tempo de falha/censura na transição e tempo de truncamento à esquerda, respectivamente.

A função de sobrevivência do indivíduo  $i$  na transição  $q$  é dada por

$$\begin{aligned} S_{qi}(t_{qi}) &= \exp \left\{ - \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qi}) \Lambda_{q0}(t_{qi}) \right\} \\ S_{qi}(\tau_{qi}) &= \exp \left\{ - \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qi}) \Lambda_{q0}(\tau_{qi}) \right\}. \end{aligned} \quad (2.5)$$

Substituindo as expressões (2.5) na função de verossimilhança (2.4), temos a seguinte forma:

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)) &= \prod_{i=1}^n \prod_{q=1}^Q \left[ \lambda_{q0}(t_{qi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qi}) \right]^{\delta_{qi}} \frac{\exp \left\{ - \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qi}) \Lambda_{q0}(t_{qi}) \right\}}{\exp \left\{ - \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qi}) \Lambda_{q0}(\tau_{qi}) \right\}} \\ &= \prod_{i=1}^n \prod_{q=1}^Q \left[ \lambda_{q0}(t_{qi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qi}) \right]^{\delta_{qi}} \exp \left\{ - \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qi}) [\Lambda_{q0}(t_{qi}) - \Lambda_{q0}(\tau_{qi})] \right\} \end{aligned} \quad (2.6)$$

e sua respectiva função de log-verossimilhança é dada por

$$\ell(\boldsymbol{\beta}, \boldsymbol{\lambda}_0(\cdot)) = \sum_{i=1}^n \sum_{q=1}^Q \left\{ \delta_{qi} \log \left[ \lambda_{q0}(t_{qi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qi}) \right] - \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qi}) [\Lambda_{q0}(t_{qi}) - \Lambda_{q0}(\tau_{qi})] \right\}. \quad (2.7)$$

O modelo (2.3) pode ser paramétrico ou semiparamétrico, pois é composto pelo produto de dois componentes: o paramétrico  $\exp(\beta'_q \mathbf{X}_q)$ , que mede o efeito das covariáveis, e  $\lambda_{q0}(t_q)$ , podendo ser paramétrico ou não. Quando não paramétrico, o componente  $\lambda_{q0}(t_q)$  deve ser uma função não negativa do tempo. Quando paramétrico, ele assume uma distribuição de probabilidade, tais como a exponencial, Weibull, log-normal entre outras. Sendo assim, se a forma paramétrica  $\lambda_{q0}(\cdot; \xi_q)$  é assumida para os riscos basais  $\lambda_0(t)$ , a função de verossimilhança completa (2.7) pode ser maximizada com respeito a todos os parâmetros  $(\xi', \beta)'$ , onde  $\xi = \xi_1, \dots, \xi_q$ , com  $\xi_q$  sendo o parâmetro do risco basal na transição  $q$ .

Uma formulação da função de verossimilhança em processos de contagem é algumas vezes mais conveniente. Nesse tipo de representação, as transições são registradas como alterações do valor de um processo estocástico. Para mais detalhes desta formulação, ver [Andersen et al. \(1993\)](#), [Kalbfleisch & Prentice \(2002\)](#) e [de Wreede et al. \(2010\)](#).

## 2.2 Considerações finais

Neste capítulo apresentamos a metodologia dos modelos multiestado, com uma representação de suas estruturas, nas quais indicam o número finito de estados que um indivíduo pode ocupar. Falamos também de sua modelagem, definindo funções importantes, tais como a intensidade de transição e as probabilidades de transição, além de comentar sobre a escala de tempo para qual o tempo  $t$  se refere, sendo a *clock forward* e *clock reset* duas abordagens muito utilizadas para este fim. Entre os tipos de modelos multiestado, comentamos brevemente dos modelos de Markov, semi-Markov e extensão de Markov.

Devido à suposição de Markov, a função de risco pode ser definida para cada transição, podendo depender do tempo  $t$ , mas que não depende de outros aspectos da história, tais como os estados visitados no caminho e os tempos das transições anteriores. Sendo assim, definimos um modelo de Cox multiestado, no qual é definido para cada transição, e o seu processo de estimação. Além disso, mostramos como seria um conjunto de dados multiestado, tomando como base uma estrutura com três estados definidos. Depois disso, mostramos

como seria esse mesmo conjunto após a transformação para um formato longo. Esse tipo de transformação é muito útil quando temos uma estrutura mais complexa, com um número maior de estados e transições, podendo ser efetuada com o auxílio do pacote `mstate` ([Putter et al. \(2007\)](#) e [Putter \(2011\)](#)) disponível no sistema R ([R Core Team, 2013](#)).

---

# 3

## Modelos de fragilidade

---

Em muitos estudos assumimos que os efeitos dos tratamentos ou riscos nos indivíduos são condicionalmente independentes dadas as covariáveis observadas, ou seja, a população sob estudo é homogênea. No entanto, existem casos em que essa suposição passa a ser questionada, como quando há eventos múltiplos sob um mesmo indivíduo (eventos recorrentes) ou quando os tempos entre os indivíduos estão relacionados (mesma família, por exemplo). Temos como exemplo um estudo da incidência de determinada doença, onde pessoas de uma mesma família têm um maior risco de sofrer o evento de interesse, já que pode existir uma predisposição genética que não é observada e que não pode ser medida diretamente.

A falta de covariáveis importantes, tais como fatores ambientais e genéticos, ou que de alguma forma não foram consideradas no planejamento, limita a capacidade preditiva do modelo, refletindo assim na variabilidade das observações, na qual é muito maior que a esperada e afeta a estimação. Desta forma, os indivíduos apresentam grande heterogeneidade ou diferentes fragilidades não atribuíveis a qualquer característica medida. O modelo de fragilidade é caracterizado pela inclusão de um efeito aleatório, em que é uma variável aleatória não observável que representa as informações que não puderam ou não foram observadas. Uma das formas para introduzir esse efeito aleatório, chamado de variável de fragilidade, é introduzi-lo na modelagem da função de risco com o objetivo de controlar a heterogeneidade não observável das unidades sob estudo, inclusive a dependência das unidades que partilham os mesmos fatores de risco ([Tomazella, 2003](#)).

Os modelos de fragilidade podem ser considerados como modelos de efeitos aleatórios com

duas fontes de variação: a variabilidade natural explicada pela função de risco, causada por covariáveis individuais não observadas que não foram incluídas no planejamento do estudo; e a variabilidade comum para os indivíduos pertencentes ao mesmo grupo que, quando as covariáveis não são observadas, geram dependência entre os tempos de eventos.

As fragilidades podem ser introduzidas de forma multiplicativa ou de forma aditiva na função de risco, visando assim responder as diferentes formas de avaliar a influência da heterogeneidade entre as unidades na função de risco, ou intensidade nos processos de contagem (dos Santos, 2010). Aqui será tratada a forma multiplicativa. Para mais detalhes dos modelos de fragilidade aditivos ver Rocha (1996), Silva (2001) e Tomazella (2003).

Clayton (1978) propôs uma extensão do modelo de Cox (1972) introduzindo multiplicativamente um efeito aleatório (fragilidade) na função de risco para analisar a tendência familiar na incidência de famílias crônicas. Geralmente o modelo de fragilidade é expresso da forma

$$\lambda(t; \mathbf{X}|u) = u\lambda_0(t) \exp(\beta' \mathbf{X}).$$

Vaupel et al. (1979) primeiramente nomeou  $u$  de termo de fragilidade, devido ao fato dele ser um fator multiplicativo que aumenta ou diminui o nível de risco em todos os momentos, descrevendo as diferentes predisposições para o evento de interesse.

A seguir mostraremos os modelos de fragilidade univariada, compartilhada e aninhada. Para mais detalhes destes, ver Wienke (2011) e Duchateau & Janssen (2008).

### 3.1 Modelo de fragilidade univariada

O modelo de fragilidade univariada, tratado em detalhes por Wienke (2011), incorpora o efeito de todos os fatores de risco não observados que são relevantes numa variável de fragilidade específica para cada indivíduo. Considerando inicialmente um modelo de fragilidade univariada sem covariáveis, a função de risco de um indivíduo depende somente da variável aleatória não observável e não negativa  $U$ , na qual é independente do tempo e atua

multiplicativamente da função de risco basal

$$\lambda_i(t|u_i) = u_i\lambda_0(t), \quad (3.1)$$

em que  $U$  varia por toda a população, refletindo a fragilidade do indivíduo, visando controlar a heterogeneidade não observada da unidade sob estudo (heterogeneidade individual).

De acordo com [Wienke \(2011\)](#), todos os indivíduos têm a mesma chance de falha, porém a variável de fragilidade individual  $u_i$  muda o risco individual. O modelo de fragilidade em (3.1) representa uma visão bastante simplificada de como a heterogeneidade pode agir e esse modelo simples mostra uma maneira de entender as consequências da heterogeneidade.

Ao introduzirmos covariáveis, o modelo (3.1) fica da forma

$$\lambda_i(t; \mathbf{X}_i|u_i) = u_i\lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{X}_i), \quad (3.2)$$

em que  $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$  é o vetor de covariáveis e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$  o vetor dos parâmetros de regressão. A variável  $U$  tem média igual a 1 e variância  $\sigma^2$ . Segundo [Elbers & Geert \(1982\)](#), é necessário que a distribuição de fragilidade tenha média 1 para que o modelo seja identificável, sendo sua variância interpretada como uma medida de heterogeneidade da população. Portanto, altos valores de  $\sigma^2$  indicam grande heterogeneidade entre indivíduos e baixos valores indicam pouca heterogeneidade, ou seja, é desnecessário incluir esse termo aleatório no modelo.

Como a fragilidade atua de forma multiplicativa na função de risco no modelo (3.2), é esperado que quanto maior for o valor da fragilidade, maior será o risco de falha. Sendo assim, quanto maior o valor de  $u_i$ , mais “frágil” o indivíduo  $i$  está para falhar.

## 3.2 Modelo de fragilidade compartilhada

Uma outra causa da violação da suposição de independência é o agrupamento, ou seja, a presença de grupos de indivíduos quem têm fatores de riscos não observáveis em comum. Um exemplo bem comum são os ensaios clínicos multicêntricos, onde pacientes de um mesmo hospital têm um risco possivelmente mais similares em relação a pacientes de outro hospital, devido a fatores sociais, climáticos e geográficos, por exemplo.

O conceito de fragilidade compartilhada remonta a Clayton (1978) e foi bastante estudada nos livros de Hougaard (2000) e Duchateau & Janssen (2008). A abordagem do modelo de fragilidade compartilhada demonstrou ser uma extensão útil e popular do modelo de Cox, quando as observações dos indivíduos não são estatisticamente independentes entre si (Wienke, 2011).

Num modelo de fragilidade compartilhada, o efeito aleatório está associado a grupos de indivíduos, ao invés de estar associada a cada indivíduo isoladamente. Inicialmente, considera-se que existam  $H$  grupos e que cada grupo  $h$  tenham  $n_h$  observações, associadas com a variável de fragilidade  $u_h$  ( $1 \leq h \leq H$ ). O vetor  $\mathbf{X}_{hi}$  ( $1 \leq h \leq H, 1 \leq i \leq n_h$ ) contém as informações das covariáveis do tempo de falha da  $i$ -ésima observação no  $h$ -ésimo grupo. Condicionado ao termo de fragilidade  $u_h$ , os tempos de sobrevivência no grupo  $h$  são considerados independentes e a função de risco para o  $i$ -ésimo indivíduo no  $h$ -ésimo grupo é da forma

$$\lambda_{hi}(t; \mathbf{X}_{hi}|u_h) = u_h \lambda_0(t) \exp(\boldsymbol{\beta}' \mathbf{X}_{hi}), \quad (3.3)$$

em que  $\lambda_0(t)$  denota a função de risco basal e  $\boldsymbol{\beta}$  o vetor de parâmetros de regressão associado ao vetor das covariáveis observadas  $\mathbf{X}_{hi}$ . As fragilidades  $U_h$  são variáveis aleatórias independentes e identicamente distribuídas com função densidade de probabilidade conhecida.

A principal suposição do modelo de fragilidade compartilhada é de que todos os indivíduos de um mesmo grupo  $h$  partilham da mesma distribuição de fragilidade  $U_h$ . Os tempos de vida são considerados condicionalmente independentes com relação à fragilidade compartilhada,

sendo ela a causa da dependência entre os tempos de vida dentro dos grupos. Desta forma, há dependência dos tempos de falha dentro dos grupos e independência desses tempos entre os grupos.

No caso em que todos os grupos são de tamanho um, o modelo (3.3) reduz-se ao modelo de fragilidade univariada (3.2). Portanto, o modelo de fragilidade univariada é um caso particular do modelo de fragilidade compartilhada. No entanto a interpretação é muito diferente para esses modelos em termos de informações não observadas: a variância da variável de fragilidade é interpretada como uma medida de dependência no caso de informações compartilhadas, enquanto que é considerado como um índice de superdispersão no caso univariado (Rotolo, 2013).

### 3.3 Modelo de fragilidade aninhada

O modelo de fragilidade aninhada torna-se apropriado em problemas nos quais os dados são naturalmente agrupados em diferentes níveis hierárquicos ou gerados por um projeto de amostragem hierárquica. Em estudos sobre a relação entre fatores ambientais, tais como poluição do ar e mortalidade, por exemplo, os indivíduos são agrupados em cidades e, além disso, podem ser divididos em mais dois subgrupos, como bairros e famílias. Um outro exemplo de agrupamento hierárquico é o caso de famílias que residem em comunidades de uma determinada cidade ou região. Temos, então, os grupos (comunidades) e os subgrupos (famílias). As famílias de uma mesma comunidade compartilham informações que não puderam ou não foram observadas, tais como questões socioeconômicas e geográficas, por exemplo. É interessante considerar uma possível heterogeneidade existente entre essas comunidades, incorporando assim um efeito aleatório, num primeiro nível, que considere essas questões. Num segundo nível, temos os indivíduos de uma mesma família compartilhando fatores de risco não observáveis, tais como fatores genéticos e sociais, gerando assim uma estrutura de dependência entre as observações desses indivíduos. A hipótese de independência das observações torna-se questionável nesses contextos, então um modelo de sobrevivência com efeitos aleatórios aninhados a nível de grupo e a nível de subgrupos parece apropriado.

As primeiras publicações nesse assunto trataram de modelos de fragilidade aditivos (Vaupel & Yashin (1985), Parner (1998), Petersen (1998) e Hougaard (2000)), mas a partir de Sastry (1997) a abordagem multiplicativa tornou-se predominante, destacando-se os trabalhos de Yau (2001), Rondeau *et al.* (2006), Horny (2009) e Shih & Lu (2009).

Sastry (1997) considerou que os tempos de sobrevivência para uma amostra de indivíduos agrupados de forma hierárquica são condicionais a dois efeitos aleatórios independentes  $v_h$  e  $w_{hj}$  que atuam de forma multiplicativa na função de risco. Desta forma, a função de risco condicional para o indivíduo  $i = 1, \dots, n_{hj}$  que está no subgrupo  $j = 1, \dots, J_h$  do grupo  $h = 1, \dots, H$  é

$$\lambda_{hji}(t|v_h, w_{hj}) = v_h w_{hj} \lambda_0(t) \exp(\beta' \mathbf{X}_{hji}),$$

em que  $v_h$  é o efeito aleatório do grupo  $h = 1, \dots, H$ ,  $w_{hj}$  o efeito aleatório do subgrupo  $j = 1, \dots, J_h$  e  $\mathbf{X}_{hji}$  é o vetor de covariáveis observadas do indivíduo  $i$  no subgrupo  $j$  do grupo  $h$ .

Seja  $U_{hj} = V_h W_{hj}$  o termo de fragilidade grupo-subgrupo e assumindo que as fragilidades  $V_1, \dots, V_H, W_{11}, \dots, W_{HJH}$  são mutuamente independentes, temos então independência entre todos os grupos e uma dependência positiva entre todos os subgrupos de forma que

$$U_{hj} \perp U_{h'j'} \quad (h \neq h'),$$

e

$$\begin{aligned} \text{Cov}(U_{hj}, U_{h'j'}) &= \mathbb{E}(V_h^2 W_{hj} W_{h'j'}) - \mathbb{E}(V_h W_{hj}) \mathbb{E}(V_{h'} W_{h'j'}) \\ &= \text{Var}(V_h) \mathbb{E}(W_{hj}) \mathbb{E}(W_{h'j'}) > 0. \end{aligned}$$

Devido a forma como as variáveis de fragilidade  $V_h$  e  $W_{hj}$  atuam na função de risco, suas distribuições devem ser não negativas e usualmente contínuas. A distribuição Gama é a mais utilizada, devido à sua simplicidade algébrica. Logo, ambas as variáveis aleatórias independentes  $V_h$  e  $W_{hj}$  seguem uma distribuição Gama com médias iguais a 1 e variâncias  $1/\theta_V$  e  $1/\theta_W$ , respectivamente, por razões de identificabilidade (Wienke, 2011).

## 3.4 Considerações finais

Neste capítulo apresentamos a metodologia dos modelos de fragilidade, nos quais são tratados em detalhes nos livros de [Wienke \(2011\)](#) e [Duchateau & Janssen \(2008\)](#). O modelo de fragilidade compartilhada pode ser empregado em situações onde há uma possível heterogeneidade não observada entre os indivíduos agrupados, sendo o efeito aleatório associado a grupos. Quando o efeito aleatório está associado ao indivíduo, temos o modelo de fragilidade univariada. O efeito aleatório é introduzido multiplicativamente na função de risco com o objetivo de controlar a heterogeneidade não observável dos indivíduos sob estudo, inclusive a dependência dos indivíduos que compartilham os mesmos fatores de risco.

O modelo de fragilidade aninhada é uma extensão do modelo de fragilidade compartilhada, sendo apropriado em casos onde existem níveis de agrupamento, como famílias que vivem em comunidades de uma determinada região, por exemplo. Desta forma, teremos um efeito aleatório a nível de grupo (comunidades) e um outro a nível de subgrupo (famílias). Tais efeitos aleatórios, que são mutualmente independentes, também são introduzidos de forma multiplicativa na função de risco, com o objetivo de controlar a heterogeneidade não observada entre os grupos e captar a correlação entre as observações dos indivíduos que compartilham os mesmo fatores de risco num subgrupo específico.

No próximo capítulo apresentaremos como é feita a integração dos modelos multiestado com o modelo de fragilidade, de forma que essa integração possa considerar a heterogeneidade entre os grupos no risco de diferentes tipos de transição e também considere a estrutura de dependência entre essas transições.

---

# 4

## Modelo multiestado com fragilidade

---

Muitos estudos incluem o agrupamento dos tempos de sobrevivência como, por exemplo, o caso de pacientes que estão internados em Unidades de Terapia Intensiva (UTI) (Liquet *et al.*, 2012). Nesse exemplo podemos estudar a evolução dos pacientes ao longo do tempo através dos estados representados na Figura 4.1, onde o estado 1 (VAP) representa uma infecção hospitalar relacionada à pneumonia. Como os dados vêm de diferentes UTI's (ou centros), e pelo motivo dos tempos de falha dos pacientes estarem agrupados por UTI, os resultados observados não podem ser considerados independentes. Além disso, temos também uma estrutura multiestado dos dados. Desta forma, um modelo multiestado com efeitos aleatórios (fragilidades) torna-se necessário para que possamos captar essa heterogeneidade não observada entre os diferentes grupos (UTI's), considerando também a dependência entre pacientes de um mesmo grupo, bem como a dependência entre os eventos de diferentes tipos de transição dentro desse mesmo grupo.

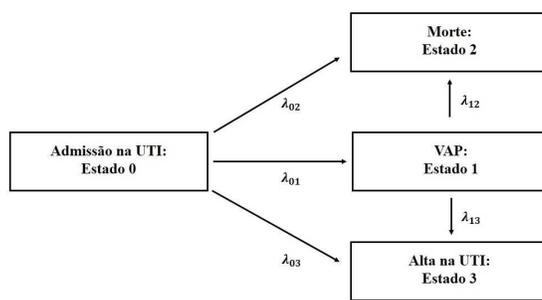


Figura 4.1 - Exemplo de aplicação (Liquet *et al.*, 2012).

## 4.1 Modelagem

A fim de considerar a heterogeneidade entre os grupos no risco para os diferentes tipos de transição, devido a fatores de risco comuns não observados, além de considerar a estrutura de dependência entre transições dos indivíduos de um mesmo grupo, um modelo multiestado com fragilidade aninhada é definido através da função de risco condicional específica da transição do tipo  $q$  ( $q = 1, \dots, Q$ ). Inicialmente, vamos considerar que existam  $H$  grupos, sendo que eles são independentes entre si, porém há dependência entre os indivíduos dentro de cada grupo  $h$  ( $h = 1, \dots, H$ ) e, dentro de cada grupo  $h$ , há dependência entre as transições de seus indivíduos. Portanto, para um indivíduo  $i$  ( $i = 1, \dots, n_h$ ) no grupo  $h$ , com vetor de covariáveis  $\mathbf{X}_{qhi}$  específico da transição  $q$ , sua função de risco específica da transição  $q$ , condicionada à fragilidade  $u_{qh}$  é

$$\lambda_{qhi}(t|u_{qh}) = u_{qh}\lambda_{q0}(t) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}). \quad (4.1)$$

em que, para a transição  $q$ ,  $\lambda_{q0}(t)$  é a função de risco basal e  $\boldsymbol{\beta}'_q$  é o vetor de parâmetros de regressão.

Assume-se que o termo de fragilidade  $u_{qh}$  para a transição  $q$  dos indivíduos no grupo  $h$  é uma realização não observada da variável aleatória

$$U_{qh} = V_h W_{qh} \quad (4.2)$$

com

$$\begin{aligned} V_h &\stackrel{iid}{\sim} f_V(v, \theta_v), \quad h = 1, \dots, H \\ W_{qh} &\stackrel{iid}{\sim} f_W(w, \theta_q), \quad h = 1, \dots, H \\ V_h &\perp W_{qh}, \quad \forall (h, q), \end{aligned} \quad (4.3)$$

em que  $V_h$  representa o efeito global do grupo e  $W_{qh}$  representa o efeito aleatório a nível de

transição no grupo  $h$ , considerando a interação entre o tipo de transição e o grupo (Rotolo, 2013).

Em situações com uma estrutura multiestado e de agrupamento, como mostrada na Figura 4.1, é natural assumir que entre as fragilidades transição-grupo  $U_{qh}$  exista uma dependência positiva dentro de um mesmo grupo e independência entre os grupos. Para fragilidades de diferentes grupos, temos que

$$U_{qh} \perp U_{qh'} \quad (h \neq h'),$$

na qual decorre das suposições (4.3). Se considerarmos o coeficiente de correlação de Pearson como uma medida de dependência linear, temos que dentro de um mesmo grupo (Rotolo, 2013)

$$\text{Corr}(U_{qh}, U_{q'h}) \in [0, 1], \quad q \neq q'.$$

No Capítulo 5 iremos considerar quatro modelos multiestado, sendo três deles casos particulares do modelo (4.1). Tais casos particulares estão descritos no diagrama 4.2:

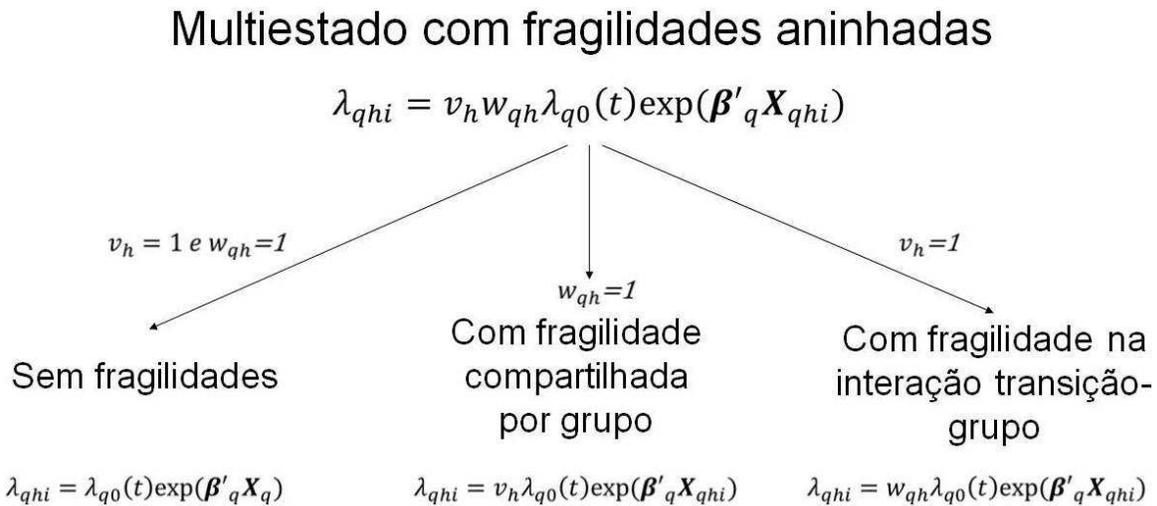


Figura 4.2 - Casos particulares do modelo multiestado com fragilidades aninhadas (4.1).

Conforme o diagrama 4.2, a partir do modelo com fragilidades aninhadas, obtém-se os seguintes modelos (da esquerda para a direita):

- **Modelo multiestado sem fragilidades:** ocorre quando  $v_h \equiv 1$  e  $w_{qh} \equiv 1$ ,  $\forall h$  e  $q$ .

- **Modelo multiestado com fragilidade compartilhada apenas para o efeito do grupo:** no caso em que  $w_{qh} \equiv 1$ , fragilidades compartilhadas são obtidas, resultando em  $u_{qh} = u_{q'h} = v_h$ , isto é,  $\text{Corr}(U_{qh}, U_{q'h}) = 1$ . Desta forma, temos o modelo com um efeito aleatório apenas para o grupo, que é o caso do modelo de fragilidade compartilhada apresentado na Seção 3.2.
- **Modelo multiestado com fragilidade compartilhada apenas para o efeito da interação transição-grupo:** quando  $v_h \equiv 1$ , as fragilidades são consideradas independentes entre as transições e o modelo reduz-se ao caso em que  $\text{Corr}(U_{qh}, U_{q'h}) = 0$  ( $U_{qh} \perp U_{q'h}$ ), similar aos modelos propostos por [Ma et al. \(2010\)](#), [Yen et al. \(2010\)](#) e [Liquet et al. \(2012\)](#). Essa suposição de fragilidades independentes dentro dos grupos ignora a estrutura de correlação de diferentes eventos que são influenciados pelos mesmos fatores de risco não observados para os indivíduos dentro de um mesmo grupo.

A seguir mostraremos a estimação dos parâmetros para o modelo (4.1).

## 4.2 Estimação dos parâmetros

Conforme [Rotolo \(2013\)](#), os indivíduos só estarão sob risco para determinado tipo de evento somente a partir do momento que eles entram no estado inicial associado a este evento. Os tempos destes indivíduos são registrados como tempos de truncamento à esquerda, denotado por  $\tau_{qhi}$ , que corresponde ao tempo da transição anterior. Se um indivíduo nunca estiver sob risco para um determinado tipo de transição, o tempo de truncamento à esquerda será definido como infinito.

De acordo com a estrutura multiestado, para cada indivíduo  $i$  ( $i = 1, \dots, n_h$ ), existem  $Q$  tipos de transições possíveis. Iremos considerar um total de  $n$  indivíduos agrupados em  $H$  grupos de tamanhos  $n_1, \dots, n_H$ , com  $n = \sum_{h=1}^H n_h$ . O tempo do tipo de transição,  $t_{qhi}$ , para o indivíduo  $i$  no grupo  $h$  pode ser observado ou censurado, sendo o indicador de

evento/censura  $\delta_{ghi}$  definido como:

$$\delta_{ghi} = \begin{cases} 1, & \text{se o tempo na transição } q \text{ para o indivíduo } i \text{ dentro do grupo } h \text{ é observado.} \\ 0, & \text{se o tempo na transição } q \text{ para o indivíduo } i \text{ dentro do grupo } h \text{ é censurado.} \end{cases}$$

Temos então os seguintes vetores de parâmetros a serem estimados:

- $\boldsymbol{\theta} = (\theta_V, \theta_1, \dots, \theta_Q)'$ : vetor de todos os parâmetros de fragilidade
- $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_Q)'$ : vetor de parâmetros de regressão.
- $\boldsymbol{\lambda}_0(\cdot) = (\lambda_{10}(\cdot), \dots, \lambda_{Q0}(\cdot))'$ : vetor das funções de risco basal.
- $\boldsymbol{\xi} = (\boldsymbol{\xi}'_1, \dots, \boldsymbol{\xi}'_Q)'$ : vetor de parâmetros da função de risco basal.
- $\boldsymbol{\zeta} = (\boldsymbol{\theta}, \boldsymbol{\xi}, \boldsymbol{\beta})'$ : vetor de todos os parâmetros.

Devido às suposições de independência condicional dadas em (4.3), a função de verossimilhança condicional para o modelo (4.1) pode ser obtida como o produto de todos os grupos e das contribuições dos indivíduos somente nas transições para as quais eles estão sob risco, ou seja, aqueles que têm o tempo de truncamento à esquerda finito. Desta forma, a função de verossimilhança condicional é definida como:

$$\begin{aligned} L_C(\boldsymbol{\xi}, \boldsymbol{\beta}) &= \prod_{h=1}^H L_{c,h}(\boldsymbol{\xi}, \boldsymbol{\beta}) \\ &= \prod_{h=1}^H \prod_{i=1}^{n_h} \prod_{q=1}^Q [\lambda_{ghi}(t_{ghi}|v_h w_{qh})]^{\delta_{ghi}} \frac{S(t_{ghi}|v_h w_{qh})}{S(\tau_{ghi}|v_h w_{qh})} \\ &= \prod_{h=1}^H \prod_{i=1}^{n_h} \prod_{q=1}^Q [v_h w_{qh} \lambda_{q0}(t_{ghi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi})]^{\delta_{ghi}} \\ &\quad \times \exp \left\{ -v_h w_{qh} [\Lambda_{q0}(t_{ghi}) - \Lambda_{q0}(\tau_{ghi})] \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right\}. \end{aligned} \quad (4.4)$$

Pelo fato das variáveis aleatórias não serem observáveis, a função de verossimilhança (4.4) não pode ser maximizada diretamente. A seguir mostraremos o processo de estimação dos parâmetros do modelo (4.1) por meio das abordagens paramétrica e semiparamétrica, respectivamente.

### 4.2.1 Estimação paramétrica

Na abordagem paramétrica devemos maximizar a função de verossimilhança marginal e, para encontrá-la, devemos integrar as fragilidades não observadas da função de verossimilhança condicional, de modo que

$$L_M(\zeta) = \int_0^\infty \dots \int_0^\infty L_c(\xi, \beta) f_{\mathbf{V}, \mathbf{W}}(\mathbf{v}, \mathbf{w}; \zeta | t_{qhi} > \tau_{qhi}) dw_{11} \dots dw_{Q1} \dots dw_{1H} \dots dw_{QH} dv_1 \dots dv_H. \quad (4.5)$$

Devido às suposição de independências dadas em (4.3), a integração da verossimilhança condicional com respeito às fragilidades pode ser feita grupo por grupo. Temos então que  $L_M(\zeta) = \prod_{h=1}^H L_{M,h}(\zeta)$  e a contribuição de cada grupo é

$$L_{M,h}(\zeta) = \int_0^\infty \dots \int_0^\infty L_{c,h}(\xi, \beta) f_{V_h, \mathbf{W}_h}(v_h, \mathbf{w}_h; \zeta | t_{qhi} > \tau_{qhi}) dw_{1h} \dots dw_{Qh} dv_h, \quad (4.6)$$

com  $\mathbf{W}_h = (W_{1h}, \dots, W_{Qh})'$  o vetor da interação transição x grupo para o grupo  $h$  e  $\mathbf{w}_h$  o vetor de suas realizações.

A distribuição de fragilidade conjunta condicional do grupo  $h$ , que faz parte da equação (4.6), é condicional aos tempos de truncamento à esquerda, sendo definida como

$$f_{v_h, \mathbf{w}_h}(v_h, \mathbf{w}_h; \zeta | t_{qhi} > \tau_{qhi}) = \underbrace{\frac{S_h(\tau_{qhi}; \beta, \xi | v_h w_{qh})}{S_h(\tau_{qhi}; \zeta)}}_B \times \underbrace{f_{V_h, \mathbf{W}_h}(v_h, \mathbf{w}_h; \theta)}_C,$$

em que A é a distribuição de sobrevivência conjunta condicional:

$$\begin{aligned} A &= \prod_{i=1}^{n_h} \prod_{q=1}^Q \exp \left\{ -v_h w_{qh} \Lambda_{q0}(\tau_{qhi}) \exp(\beta'_q \mathbf{X}_{qhi}) \right\} \\ &= \prod_{q=1}^Q \exp \left\{ -v_h w_{qh} \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\beta'_q \mathbf{X}_{qhi}) \right\}, \end{aligned}$$

B é a função de sobrevivência marginal:

$$B = \int_0^\infty \dots \int_0^\infty \prod_{q=1}^Q \exp \left\{ -v_h w_{qh} \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right\} \\ \times f_{V_h, \mathbf{w}_h}(v_h, \mathbf{w}_h; \boldsymbol{\theta}) dw_{1h} \dots dw_{Qh} dv_h,$$

que, devido às suposições de independência (4.3), pode ser reescrita como

$$B = \int_0^\infty \prod_{q=1}^Q \left[ \exp \left\{ -v_h w_{qh} \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right\} f_W(w_{qh}; \boldsymbol{\theta}_q) dw_{qh} \right] f_V(v_h; \theta_V) dv_h \\ = \int_0^\infty \prod_{q=1}^Q \mathcal{L}_q \left( v_h \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right) f_V(v_h; \theta_V) dv_h,$$

em que  $\mathcal{L}(s)$  é a transformada de Laplace de  $f_W(w_{qh}; \boldsymbol{\theta}_q)$  no ponto  $s = v_h \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi})$ , relativa à transição  $q$ . Devido à suposição de que os efeitos aleatórios são independentes,

$$C = f_V(v_h, \theta_v) \prod_{q=1}^Q f_W(w_{qh}; \boldsymbol{\theta}_q).$$

Desta forma, a função de distribuição de probabilidade conjunta das fragilidades de um grupo  $h$ , condicionada aos tempos de truncamento à esquerda é

$$f_{v_h, \mathbf{w}_h}(v_h, \mathbf{w}_h; \boldsymbol{\zeta} | t_{qhi} > \tau_{qhi}) = \frac{\prod_{i=1}^{n_h} \prod_{q=1}^Q \exp \left\{ -v_h w_{qh} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right\}}{\int_0^\infty \prod_{q=1}^Q \mathcal{L}_q \left( \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right) f_V(v_h; \theta_V) dv_h} \\ \times f_V(v_h, \theta_v) \prod_{q=1}^Q f_W(w_{qh}; \boldsymbol{\theta}_q).$$

Logo, devido às suposições de independência (4.3), a função de verossimilhança marginal (4.6) para o grupo  $h$  pode ser reescrita como

$$L_{M,h}(\boldsymbol{\zeta}) = \int_0^\infty \prod_{q=1}^Q \left\{ \int_0^\infty L_{C,h}(\boldsymbol{\beta}, \boldsymbol{\xi}) f_{v_h, \mathbf{w}_h}(v_h, \mathbf{w}_h; \boldsymbol{\zeta} | t_{qhi} > \tau_{qhi}) f_W(w_{qh}; \boldsymbol{\theta}_q) dw_{qh} \right\} f_V(v_h; \theta_V) dv_h \\ = \int_0^\infty \prod_{q=1}^Q \left\{ \int_0^\infty \prod_{i=1}^{n_h} \prod_{q=1}^Q \left[ [v_h w_{qh} \lambda_{q0}(t_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi})] \right]^{\delta_{qhi}} \right.$$

$$\begin{aligned}
& \times \frac{\exp \left\{ -v_h w_{qh} \lambda_{q0}(t_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right\}}{\exp \left\{ -v_h w_{qh} \lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right\}} \Bigg] \\
& \times \frac{\prod_{i=1}^{n_h} \prod_{q=1}^Q \exp \left\{ -v_h w_{qh} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right\}}{\int_0^\infty \prod_{q=1}^Q \mathcal{L}_q \left( \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right) f_V(v_h; \theta_V) dv_h} \\
& \times f_W(w_{qh}; \boldsymbol{\theta}_q) dw_{qh} \Bigg\} f_V(v_h; \theta_V) dv_h \\
= & \int_0^\infty \prod_{q=1}^Q \left\{ \int_0^\infty [\lambda_{q0}(t_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi})]^{\delta_{qhi}} v_h^{d_h} \right. \\
& \times \prod_{q=1}^Q \left[ w_{qh}^{d_{qh}} \exp \left\{ -v_h w_{qh} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right\} \right] f_W(w_{qh}; \boldsymbol{\theta}_q) \\
& \times \frac{1}{\int_0^\infty \prod_{q=1}^Q \mathcal{L}_q \left[ v_h \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right] f_V(v_h; \theta_V) dv_h} \Bigg\} f_V(v_h; \theta_V) dv_h \\
= & \prod_{i=1}^{n_h} \prod_{q=1}^Q \left\{ [\lambda_{q0}(t_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi})]^{\delta_{qhi}} \right. \\
& \times \frac{\int_0^\infty v_h^{d_h} \prod_{q=1}^Q (-1)^{d_{qh}} \mathcal{L}_q^{d_{qh}} \left[ v_h \sum_{i=1}^{n_h} \Lambda_{q0}(t_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right] f_V(v_h; \theta_V) dv_h}{\int_0^\infty \prod_{q=1}^Q \mathcal{L}_q \left[ v_h \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right] f_V(v_h; \theta_V) dv_h},
\end{aligned}$$

em que  $d_h = \sum_{i=1}^{n_h} \sum_{q=1}^Q \delta_{qhi}$  é a soma de todos os eventos no grupo  $h$  e  $d_{qh} = \sum_{i=1}^{n_h} \delta_{qhi}$  é o número de eventos do tipo  $q$  dentro do grupo  $h$ . Portanto, a função de verossimilhança marginal completa é o produto da função de verossimilhança marginal sobre todos os grupos:

$$\begin{aligned}
L_M(\boldsymbol{\zeta}) &= \prod_{h=1}^H \left\{ \prod_{i=1}^{n_h} \prod_{q=1}^Q [\lambda_{q0}(t_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi})]^{\delta_{qhi}} \right\} \\
& \times \int_0^\infty v_h^{d_h} \prod_{q=1}^Q (-1)^{d_{qh}} \mathcal{L}_q^{d_{qh}} \left[ v_h \sum_{i=1}^{n_h} \Lambda_{q0}(t_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right] f_V(v_h; \theta_V) dv_h \\
& \times \left\{ \int_0^\infty \prod_{q=1}^Q \mathcal{L}_q \left[ v_h \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right] f_V(v_h; \theta_V) dv_h \right\}^{-1}. \quad (4.7)
\end{aligned}$$

E sua respectiva função log-verossimilhança marginal de é

$$\begin{aligned}
\ell_M(\zeta) &= \sum_{h=1}^H \left\{ \sum_{i=1}^{n_h} \sum_{q=1}^Q \delta_{qhi} \log [\lambda_{q0}(t_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi})] \right\} \\
&\quad + \log \int_0^\infty v_h^{d_h} \prod_{q=1}^Q (-1)^{d_{qh}} \mathcal{L}_q^{d_{qh}} \left[ v_h \sum_{i=1}^{n_h} \Lambda_{q0}(t_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right] f_V(v_h; \theta_V) dv_h \\
&\quad - \log \int_0^\infty \prod_{q=1}^Q \mathcal{L}_q \left( v_h \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\boldsymbol{\beta}'_q \mathbf{X}_{qhi}) \right) f_V(v_h; \theta_V) dv_h. \tag{4.8}
\end{aligned}$$

Para obtermos a função log-verossimilhança marginal (4.8), é preciso integrá-la em dois níveis. Caso a integração relativa às interações  $w_{qh}$  possa ser feita analiticamente e for expressa em termos da derivada da transformada de Laplace de sua distribuição, a integração analítica já não é possível para as fragilidades  $v_h$ . Com base nisso, integração com aproximação numérica torna-se necessária para obtermos a verossimilhança marginal.

Em modelos de fragilidade é muito comum utilizar a distribuição Gama para os efeitos aleatórios, devido à sua facilidade algébrica. A seguir consideraremos distribuição Gama para os efeitos aleatórios para obtermos a log-verossimilhança marginal do modelo (4.1).

### Modelo multiestado com fragilidades Gama

A função log-verossimilhança marginal (4.8) depende da distribuição de  $V_h$  e da transformada de Laplace de  $W_{qh}$  e sua derivada. Assumindo distribuições Gama com médias iguais a 1 e variâncias  $\theta_V$  e  $\theta_q$  para as variáveis  $V_h$  e  $W_{qh}$ , temos que

$$\begin{aligned}
f_V &\stackrel{iid}{\sim} \text{Gama}(1/\theta_V, 1/\theta_V) \\
f_W &\stackrel{iid}{\sim} \text{Gama}(1/\theta_q, 1/\theta_q),
\end{aligned}$$

onde

$$f_V(v_h; \theta_V) = \frac{(1/\theta_V)^{1/\theta_V}}{\Gamma(1/\theta_V)} v_h^{1/\theta_V - 1} \exp(-v_h/\theta_V)$$

$$f_W(w_{qh}; \theta_q) = \frac{(1/\theta_q)^{1/\theta_q}}{\Gamma(1/\theta_q)} w_{qh}^{1/\theta_q - 1} \exp(-w_{qh}/\theta_q).$$

Com relação à distribuição de  $W_{qh}$ , sua transformada de Laplace e sua  $d_{qh}$ -ésima derivada da transformada de Laplace no ponto  $k$  na transição  $q$  são, respectivamente,

$$\begin{aligned} \mathcal{L}_q(k) &= \mathbb{E}[\exp(-W_{qh}k)] = (1 + \theta_q k)^{-1/\theta_q} \\ \mathcal{L}_q^{d_{qh}}(k) &= (-1)^{d_{qh}} \mathbb{E}[W_{qh} \exp(-W_{qh}k)] = \frac{(-1)^{d_{qh}} \Gamma(d_{qh} + 1/\theta_q)}{\Gamma(1/\theta_q) \theta_q^{1/\theta_q} (1/\theta_q + k)^{d_{qh} + 1/\theta_q}}. \end{aligned}$$

Sendo assim, assumindo distribuição Gama para os termos aleatórios, a função log-verossimilhança (4.8) é dada por

$$\begin{aligned} \ell_M(\zeta) &= \sum_{h=1}^H \left\{ \sum_{i=1}^{n_h} \sum_{q=1}^Q \delta_{qhi} \log [\lambda_{q0}(t_{qhi}) \exp(\beta'_q \mathbf{X}_{qhi})] \right. \\ &\quad + \log \int_0^\infty v_h^{d_h} \prod_{q=1}^Q \frac{(-1)^{d_{qh}} (-1)^{d_{qh}} \Gamma(d_{qh} + 1/\theta_q)}{\Gamma(1/\theta_q) \theta_q^{1/\theta_q} [1/\theta_q + v_h \sum_{i=1}^{n_h} \Lambda_{q0}(t_{qhi}) \exp(\beta'_q \mathbf{X}_{qhi})]^{d_{qh} + 1/\theta_q}} \\ &\quad \times \frac{(1/\theta_V)^{1/\theta_V}}{\Gamma(1/\theta_V)} v_h^{1/\theta_V - 1} \exp(-v_h/\theta_V) dv_h \\ &\quad \left. - \log \int_0^\infty \prod_{q=1}^Q \left[ 1 + \theta_q v_h \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\beta'_q \mathbf{X}_{qhi}) \right]^{-1/\theta_q} \right. \\ &\quad \left. \times \frac{(1/\theta_V)^{1/\theta_V}}{\Gamma(1/\theta_V)} v_h^{1/\theta_V - 1} \exp(-v_h/\theta_V) dv_h \right\} \\ &= \sum_{h=1}^H \left\{ \sum_{i=1}^{n_h} \sum_{q=1}^Q \delta_{qhi} \log [\lambda_{q0}(t_{qhi}) \exp(\beta'_q \mathbf{X}_{qhi})] \right. \\ &\quad + \log \int_0^\infty \frac{v_h^{d_h + 1/\theta_V - 1} \exp(-v_h/\theta_V)}{\prod_{q=1}^Q [1/\theta_q + v_h \sum_{i=1}^{n_h} \Lambda_{q0}(t_{qhi}) \exp(\beta'_q \mathbf{X}_{qhi})]^{d_{qh} + 1/\theta_q}} dv_h \\ &\quad - \log \int_0^\infty \frac{v_h^{1/\theta_V - 1} \exp(-v_h/\theta_V)}{\prod_{q=1}^Q [\theta_q v_h \sum_{i=1}^{n_h} \Lambda_{q0}(\tau_{qhi}) \exp(\beta'_q \mathbf{X}_{qhi}) + 1]^{1/\theta_q}} dv_h \\ &\quad \left. + \sum_{q=1}^Q \log \left[ \frac{\Gamma(d_{qh} + 1/\theta_q)}{\Gamma(1/\theta_q) \theta_q^{1/\theta_q}} \right] \right\}. \end{aligned} \tag{4.9}$$

Com distribuições já definidas para as variáveis de fragilidade, a função de verossimilhança marginal (4.9) pode ser maximizada. Para isso, assumimos uma distribuição para as  $Q$  funções de risco de base, e maximizamos  $\ell_M(\zeta)$  para estimarmos o vetor de parâmetros  $\zeta = (\theta, \xi, \beta)'$ . As distribuições mais comuns para o risco de base são as distribuições exponencial, Weibull e log-normal.

### 4.2.2 Estimação semiparamétrica

Na estimação semiparamétrica, não assumimos uma distribuição para os tempos de sobrevivência. Sendo assim, a presença do componente não-paramétrico  $\lambda_{q0}(t)$  na função de risco (4.1) torna o método da máxima verossimilhança inapropriado. Uma forma de contornar essa situação é condicionar a construção da função de verossimilhança ao conhecimento da história passada de falhas e censuras para eliminar esta função de perturbação da verossimilhança (Colosimo & Giolo, 2006). Essa metodologia foi proposta por Cox (1975), sendo conhecida como método da verossimilhança parcial.

Para a versão semiparamétrica de modelos multiestado com fragilidade também é preciso considerar a contribuição das fragilidades não observadas na função de verossimilhança. Para isso, efetuamos uma modificação no risco condicional da transição específica do modelo (4.1), aplicando  $\exp \{ \log [\lambda_{ghi}(t|v_h w_{qh})] \}$ , de forma que

$$\lambda_{ghi}(t|v_h w_{qh}) = \lambda_{q0}(t) \exp \left\{ \log v_h + \log w_{qh} + \beta'_q \mathbf{X}_{ghi} \right\}.$$

Para construirmos a função de verossimilhança parcial, vamos considerar que para cada indivíduo  $i$  ( $i = 1, \dots, n_h$ ), existem  $Q$  ( $q = 1, \dots, Q$ ) tipos de transições possíveis. No total, temos uma amostra de  $n = \sum_{h=1}^H n_h$  indivíduos agrupados em  $H$  grupos de tamanhos  $n_1, \dots, n_H$ . A probabilidade condicional do  $i$ -ésimo indivíduo do grupo  $h$  vir a falhar na transição do tipo  $q$  no tempo  $t_{ghi}$ , conhecendo quais observações estão sob risco em  $t_{ghi}$  é:

$$\begin{aligned} & P[\text{indivíduo do grupo } h \text{ falhar na transição } q \text{ em } t_{ghi} \mid \text{uma falha em } t_{ghi} \text{ e história até } t_{ghi}] \\ &= \frac{P[\text{indivíduo do grupo } h \text{ falhar na transição } q \text{ em } t_{ghi} \mid \text{sobreviveu a } t_{ghi} \text{ e história até } t_{ghi}]}{P[\text{uma falha em } t_{ghi} \mid \text{história até } t_{ghi}]} \end{aligned}$$

$$\begin{aligned}
&= \frac{\lambda_{qhi}(t|u_{qh})}{\sum_{h'i' \in R_q(t_{qhi})} \lambda_{qh'i'}(t|u_{qh'})} \\
&= \frac{\lambda_{q0}(t) \exp \left\{ \log v_h + \log w_{qh} + \boldsymbol{\beta}'_q \mathbf{X}_{qhi} \right\}}{\sum_{h'i' \in R_q(t_{qhi})} \lambda_{q0}(t) \exp \left\{ \log v'_h + \log w_{qh'} + \boldsymbol{\beta}'_q \mathbf{X}_{qh'i'} \right\}} \\
&= \frac{\exp \left\{ \log v_h + \log w_{qh} + \boldsymbol{\beta}'_q \mathbf{X}_{qhi} \right\}}{\sum_{h'i' \in R_q(t_{qhi})} \exp \left\{ \log v'_h + \log w_{qh'} + \boldsymbol{\beta}'_q \mathbf{X}_{qh'i'} \right\}}, \tag{4.10}
\end{aligned}$$

em que  $R_q(y) = \{(h', i') \in [\tau_{qh'i'}, t_{qh'i'}]\}$  é o conjunto das transições do tipo  $q$  que estão sob risco no tempo  $y$ , na qual considera as informações dadas pelos tempos de truncamento à esquerda  $\tau_{qhi}$ . Portanto, após eliminarmos  $\lambda_{q0}$ , a função de verossimilhança parcial condicional do modelo (4.1) é formada pelo produto de todos os termos representados em (4.10), associados aos tempos distintos de falha  $\delta_{qhi}$ , sobre os indivíduos de todos os grupos, de forma que

$$\begin{aligned}
L_P(\boldsymbol{\beta}) &= \prod_{h=1}^H \prod_{i=1}^{n_h} \prod_{q=1}^Q \left( \frac{\exp \left\{ \log v_h + \log w_{qh} + \boldsymbol{\beta}'_q \mathbf{X}_{qhi} \right\}}{\sum_{h'i' \in R_q(t_{qhi})} \exp \left\{ \log v'_h + \log w_{qh'} + \boldsymbol{\beta}'_q \mathbf{X}_{qh'i'} \right\}} \right)^{\delta_{qhi}} \\
&= \prod_{h=1}^H \left\{ v_h^{d_h} \prod_{q=1}^Q \left[ w_{qh}^{d_{qh}} \prod_{i=1}^{n_h} \left( \frac{\exp \left\{ \boldsymbol{\beta}'_q \mathbf{X}_{qhi} \right\}}{\sum_{h'i' \in R_q(t_{qhi})} v'_h w_{qh'} \exp \left( \boldsymbol{\beta}'_q \mathbf{X}_{qh'i'} \right)} \right)^{\delta_{qhi}} \right] \right\}, \tag{4.11}
\end{aligned}$$

em que  $d_h = \sum_{i=1}^{n_h} \sum_{q=1}^Q \delta_{qhi}$  é a soma de todos os eventos no grupo  $h$  e  $d_{qh} = \sum_{i=1}^{n_h} \delta_{qhi}$  é o número de eventos do tipo  $q$  dentro do grupo  $h$ . Portanto, condicionada às fragilidades, o logaritmo da função de verossimilhança parcial (4.11) é

$$\begin{aligned}
\ell_P(\boldsymbol{\beta}) &= \sum_{h=1}^H \left\{ d_h \log v_h + \sum_{q=1}^Q \left[ d_{qh} \log w_{qh} + \sum_{i=1}^{n_h} \left( \delta_{qhi} \boldsymbol{\beta}'_q \mathbf{X}_{qhi} \right. \right. \right. \\
&\quad \left. \left. \left. - \log \sum_{h'i' \in R_q(t_{qhi})} v_{h'} w_{qh'} \exp \left( \boldsymbol{\beta}'_q \mathbf{X}_{qhi} \right) \right) \right] \right\}.
\end{aligned}$$

A versão perfilada da função de verossimilhança completa  $\ell(\boldsymbol{\zeta}) = \ell_C(\boldsymbol{\xi}, \boldsymbol{\beta}) + \log f_{\mathbf{v}, \mathbf{w}}(\mathbf{v}, \mathbf{w}; \boldsymbol{\theta})$

é

$$\ell_{PP}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \ell_P(\boldsymbol{\beta}) + \log f_{\mathbf{v}, \mathbf{w}}(\mathbf{v}, \mathbf{w}; \boldsymbol{\theta}), \quad (4.12)$$

na qual será chamada de função log-verossimilhança parcial penalizada (com a sigla PPL, em inglês), em analogia à abordagem PPL para modelos de fragilidade semiparamétricos. Sob a restrição de que todas as interações transição-grupo  $W_{gh}$  têm a mesma variância ( $\theta_W := \theta_1 = \dots = \theta_Q$ ), (4.12) pode ser maximizada através do algoritmo EM-PL proposto por [Horny \(2009\)](#), no qual alterna entre um algoritmo EM e PPL para ajustar modelos de riscos proporcionais semiparamétricos com qualquer quantidade de níveis de agrupamento, sendo válido apenas para distribuições que admitem uma transformada de Laplace. No pacote `mlfm` ([Rotolo & Horny, 2012](#)) é implementado o método de estimação EM-PL proposto por [Horny \(2009\)](#) para modelos de fragilidade multinível, podendo ser utilizado para ajustar um modelo multiestado com fragilidade aninhada sob a restrição de que todas as interações  $W_{gh}$  têm a mesma variância.

### 4.3 Estudo de simulação

Por meio das propriedades frequentistas dos parâmetros do modelo multiestado com fragilidade, apresentamos um primeiro estudo de simulação para investigar como a incorporação de fragilidades no modelo pode melhorar a estimação dos parâmetros, e um segundo estudo para avaliar o modelo multiestado com fragilidades aninhadas.

Consideramos um esquema multiestado de pacientes agrupados por hospital com três estados definidos como **saúde**, **doença** e **morte**, conforme representado na Figura 4.3.



Figura 4.3 - Esquema multiestado para o estudo de simulação.

Nos dois estudos os tempos de cada transição foram gerados a partir de uma distribuição Weibull, em que no estudo 1 o efeito aleatório da interação transição-grupo segue uma distribuição Gama  $(1/\theta_q, 1/\theta_q)$  e no estudo 2, os dois efeitos aleatórios seguem uma distribuição Gama  $(1/\theta_V, 1/\theta_V)$  e  $(1/\theta_W, 1/\theta_W)$ , respectivamente. Consideramos também o efeito de uma covariável binária em cada transição.

Em resumo, para  $q = 1, 2$  e  $3$ , a geração dos dados consiste em:

- Fixar os valores dos parâmetros  $\beta_q$  e  $\theta_q$  nos estudos 1 e 2. No estudo 2, fixar os parâmetros  $\theta_V$  (variância a nível de hospital) e  $\theta_W$  (variância a nível de transição).
- Na transição do tipo  $q$  para cada indivíduo  $i$  ( $i = 1, \dots, n_h$ ) dentro do grupo  $h$  ( $h = 1, \dots, H$ ):

- Gerar  $u_{qhi} \sim U(0, 1)$ .
- Gerar os tempos  $t_{1hi}$ ,  $t_{2hi}$  e  $t_{3hi}$  representando, respectivamente, seus tempos de falha/censura na transição 1 (saúde→doença), na transição 2 (saúde→morte) e na transição 3 (doença→morte), em que

$$t_{qhi} = -b_q \left[ \frac{\log(1 - u_{qhi})}{v_h w_q \exp(\beta'_q X_{qhi})} \right]^{1/a_q},$$

sendo  $a_q$  e  $b_q$  os parâmetros da distribuição Weibull específicos da transição  $q$ .

- Se  $t_{1hi} = \min(t_{1hi}, t_{2hi}, t_{3hi})$ , então o indivíduo  $i$  realizou a transição 1 e, como consequência, irá realizar a transição 3 no tempo  $t_{1hi} + t_{3hi}$ . Caso  $t_{2hi} = \min(t_{1hi}, t_{2hi}, t_{3hi})$ , então o indivíduo realizou a transição 2.

–  $X_{ghi} \sim \text{Bernoulli}(0, 5)$ .

- Os modelos foram ajustados semiparametricamente com o auxílio do pacote `mlfm` (Rotolo & Horny, 2012) e o risco da transição depende apenas do tempo de permanência no estado atual. Ou seja, iremos utilizar a abordagem *clock reset* na escala de tempo.

Em cada cenário foram geradas  $N$  amostras aleatórias e, dessas  $N$  amostras aleatórias, foram calculadas as médias das estimativas de máxima verossimilhança (EMV), as médias dos vícios e as raízes quadradas do erro quadrático médio ( $\sqrt{\text{EQM}}$ ) das EMV. Para  $\beta_q$ , temos que

$$\sqrt{\text{EQM}}(\hat{\beta}_q) = \sqrt{\frac{1}{N} \sum_{j=1}^N (\hat{\beta}_q^{(j)} - \beta_q)^2} \quad (q = 1, 2, 3),$$

em que  $\hat{\beta}_q^{(j)}$  representa a estimativa de  $\beta_q$  na  $j$ -ésima simulação, sendo  $\beta_q$  o valor fixado.

Após gerarmos dados multiestado com fragilidades, iremos considerar a estimação semi-paramétrica dos parâmetros de regressão em cada transição, que na prática é o interesse principal em estudos clínicos. Os resultados dos estudos de simulação são detalhados a seguir.

### 4.3.1 Estudo de simulação 1

Neste estudo de simulação verificamos como a incorporação da fragilidade no modelo multiestado melhora as estimativas dos parâmetros, comparando os modelos com e sem fragilidade através das propriedades frequentistas dos estimadores dos parâmetros, vício e erro quadrático médio.

Duas situações serão consideradas para o valor da variância da interação transição-grupo  $\theta_q$ : no estudo 1A essas variâncias serão iguais ( $\theta_q = \theta_1 = \theta_2 = \theta_3$ ) e no estudo 1B as variâncias serão diferentes, ou seja, cada transição terá variância  $\theta_q$  ( $q = 1, 2, 3$ ).

Cada intensidade de transição é modelada através do modelo (4.1), com a condição de que  $V_h \equiv 1$ , ou seja, iremos gerar dados de um modelo multiestado com fragilidade compartilhada apenas para o efeito da interação transição-grupo, onde as fragilidades são consideradas

independentes entre as transições. Desta forma, os dados são gerados do modelo

$$\lambda_{ghi}(t|w_{gh}) = w_{gh}\lambda_{q0}(t) \exp(\beta'_q \mathbf{X}_{ghi}). \quad (4.13)$$

São reproduzidos três cenários, com  $N = 500$  réplicas cada um, conforme descrito na Tabela 4.1.

Tabela 4.1 - *Cenários do estudo de simulação 1.*

Cenário	I	II	III
Nº de grupos	10	10	10
Nº de indivíduos no grupo	10	50	100
Nº total de indivíduos	100	500	1000

Desta forma, fixamos o número de grupos, variando apenas o total de indivíduos em cada grupo, totalizando tamanhos amostrais de 100, 500 e 1000 indivíduos, respectivamente, em cada cenário. Os parâmetros da distribuição Weibull utilizados para a geração dos dados multiestado, os parâmetros de regressão e a variância da interação transição-grupo serão os mesmos em cada cenário e estão na Tabela 4.2.

Tabela 4.2 - *Parâmetros do estudo de simulação 1.*

Parâmetros	Transição		
	1	2	3
$(a_q ; b_q)$	(1,3 ; 15)	(1,3 ; 35)	(1,25 ; 35)
$\beta_q$	0,8	0,6	0,3
$\theta_q$ (estudo 1A)	2	2	2
$\theta_q$ (estudo 1B)	0,7	1,2	2

### Estudo de simulação 1A

Os resultados do estudo de simulação 1A, em que considera variâncias iguais para a interação transição-hospital (Tabela 4.2), encontram-se na Tabela 4.3, na qual apresenta o valor fixado de  $\beta_q$  ( $q = 1, 2, 3$ ), a média, o vício e raiz quadrada do erro quadrático médio das

estimativas de máxima verossimilhança do modelo (4.13). Para cada cenário, é comparado os resultados para o modelo (4.13) ajustado com e sem fragilidade compartilhada apenas para a interação transição-grupo, respectivamente.

Tabela 4.3 - Resultados do estudo de simulação 1A.

$\beta_q$ (valor real)	Média	Vício	$\sqrt{\text{EQM}}$
<b>Cenário I: <math>G = 10, n_i = 10</math> (<math>n = 100</math>)</b>			
Com fragilidades			
$\beta_1$ (0, 8)	0,7774	-0,0226	0,3334
$\beta_2$ (0, 6)	0,5209	-0,0791	0,3555
$\beta_3$ (0, 3)	0,3174	0,0174	0,3438
Sem fragilidades			
$\beta_1$ (0, 8)	0,4256	-0,3744	0,4835
$\beta_2$ (0, 6)	0,3710	-0,2990	0,3993
$\beta_3$ (0, 3)	0,1317	-0,1683	0,3517
<b>Cenário II: <math>G = 10, n_i = 50</math> (<math>n = 500</math>)</b>			
Com fragilidades			
$\beta_1$ (0, 8)	0,7922	-0,0078	0,1394
$\beta_2$ (0, 6)	0,5209	-0,0791	0,1699
$\beta_3$ (0, 3)	0,3022	0,0022	0,1395
Sem fragilidades			
$\beta_1$ (0, 8)	0,4422	-0,3578	0,3913
$\beta_2$ (0, 6)	0,3644	-0,2356	0,2867
$\beta_3$ (0, 3)	0,1248	-0,1782	0,2243
<b>Cenário III: <math>G = 10, n_i = 100</math> (<math>n = 1000</math>)</b>			
Com fragilidades			
$\beta_1$ (0, 8)	0,7863	-0,0137	0,0941
$\beta_2$ (0, 6)	0,5161	-0,0839	0,1408
$\beta_3$ (0, 3)	0,3038	0,0038	0,0916
Sem fragilidades			
$\beta_1$ (0, 8)	0,4382	-0,3618	0,3850
$\beta_2$ (0, 6)	0,3652	-0,2348	0,2629
$\beta_3$ (0, 3)	0,1253	-0,1747	0,2023

De acordo com os resultados apresentados na Tabela 4.3, no modelo multiestado (4.13) ajustado com fragilidades os parâmetros de regressão  $\beta_q$ , em geral, mostram-se adequados, com os vícios e  $\sqrt{\text{EQM}}$  diminuindo à medida que o tamanho da amostra aumenta, além das médias das estimativas ficarem mais próximas dos valores fixados. Entretanto, se omitirmos

o termo de fragilidade no ajuste do modelo, as médias das estimativas ficam muito distantes dos valores fixados. E ao compararmos os vícios e  $\sqrt{\text{EQM}}$ , vemos que eles são maiores no modelo ajustado sem fragilidades. Os gráficos na Figura 4.4 mostram o comportamento dos modelos com fragilidade (linha contínua com  $\bullet$ ) e sem fragilidade (linha pontilhada com  $+$ ). Concluimos que, em geral, a incorporação de fragilidades no modelo multiestado (4.13) melhora a estimativa dos parâmetros.

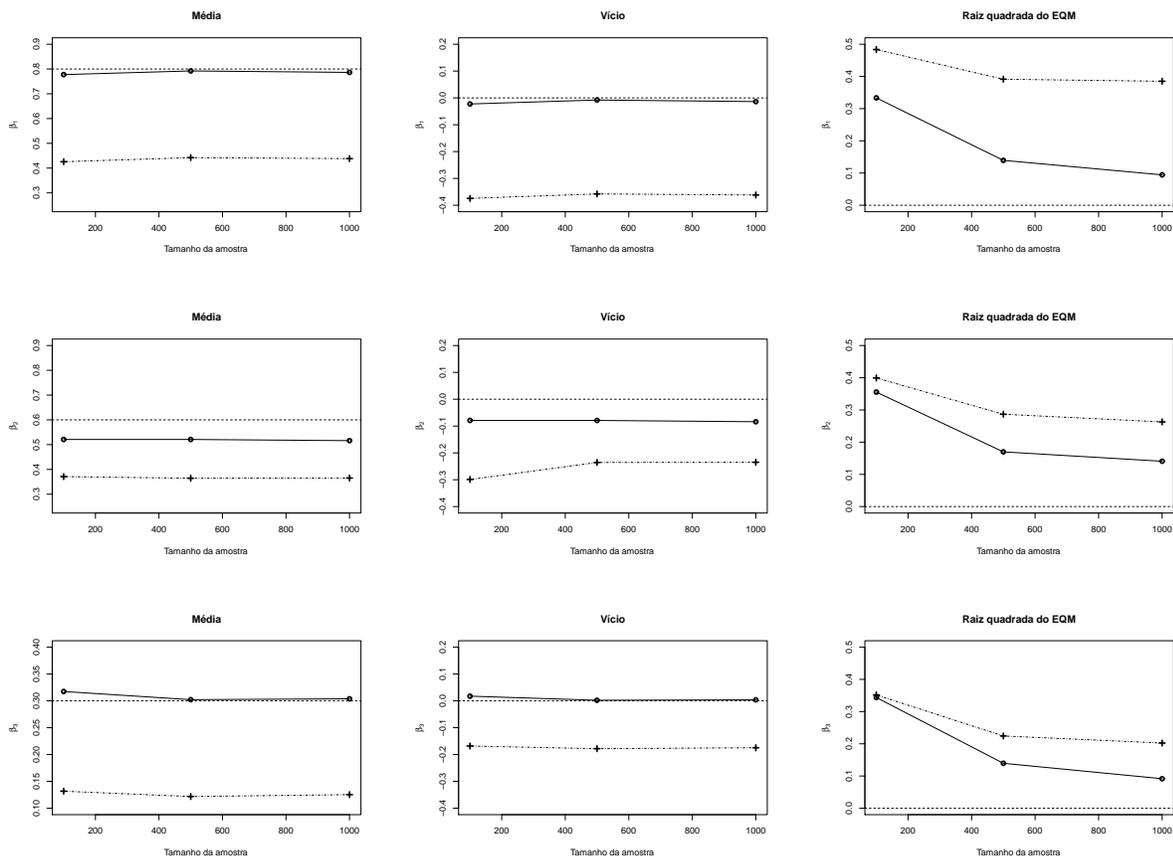


Figura 4.4 - Comportamento do estudo de simulação 1A. Linha contínua com  $\bullet$ : modelo (4.13) com fragilidade; Linha pontilhada com  $+$ : modelo (4.13) sem fragilidade.

## Estudo de simulação 1B

Os resultados do estudo de simulação 1B, no qual considera diferentes variâncias para a interação transição-grupo (Tabela 4.2), encontram-se na Tabela 4.4.

Tabela 4.4 - Resultados do estudo de simulação 1B.

$\beta_q$ (valor real)	Média	Vício	$\sqrt{\text{EQM}}$
<b>Cenário I: <math>G = 10, n_i = 10</math> (<math>n = 100</math>)</b>			
Com fragilidades			
$\beta_1$ (0, 8)	0,8275	0,0275	0,2678
$\beta_2$ (0, 6)	0,6239	0,0239	1,1964
$\beta_3$ (0, 3)	0,2967	-0,0033	0,2790
Sem fragilidades			
$\beta_1$ (0, 8)	0,5859	-0,2141	0,3268
$\beta_2$ (0, 6)	0,5706	-0,0294	1,1030
$\beta_3$ (0, 3)	0,1267	-0,1733	0,3018
<b>Cenário II: <math>G = 10, n_i = 50</math> (<math>n = 500</math>)</b>			
Com fragilidades			
$\beta_1$ (0, 8)	0,8055	0,0055	0,1106
$\beta_2$ (0, 6)	0,5397	-0,0603	0,2477
$\beta_3$ (0, 3)	0,2971	-0,0029	0,1001
Sem fragilidades			
$\beta_1$ (0, 8)	0,5643	-0,2357	0,2710
$\beta_2$ (0, 6)	0,4861	-0,1139	0,2703
$\beta_3$ (0, 3)	0,1312	-0,1688	0,1997
<b>Cenário III: <math>G = 10, n_i = 100</math> (<math>n = 1000</math>)</b>			
Com fragilidades			
$\beta_1$ (0, 8)	0,8065	0,0065	0,0749
$\beta_2$ (0, 6)	0,5422	-0,0578	0,1823
$\beta_3$ (0, 3)	0,3041	0,0041	0,0759
Sem fragilidades			
$\beta_1$ (0, 8)	0,5605	-0,2395	0,2607
$\beta_2$ (0, 6)	0,4799	-0,1201	0,2218
$\beta_3$ (0, 3)	0,1304	-0,1696	0,1853

Assim como no estudo de simulação 1A (Tabela 4.3), os parâmetros de regressão  $\beta_q$  são bem estimados e o modelo (4.13) com fragilidade mostrou-se mais adequado, já que a média das EMV aproximam-se mais do valor fixado, além do vício e a raiz quadrada do EQM

serem mais próximos de zero quando comparamos com o modelo sem fragilidade. A Figura 4.5 ilustra essa comparação dos modelo com fragilidade (linha contínua com  $\bullet$ ) e sem fragilidade (linha pontilhada com  $+$ ).

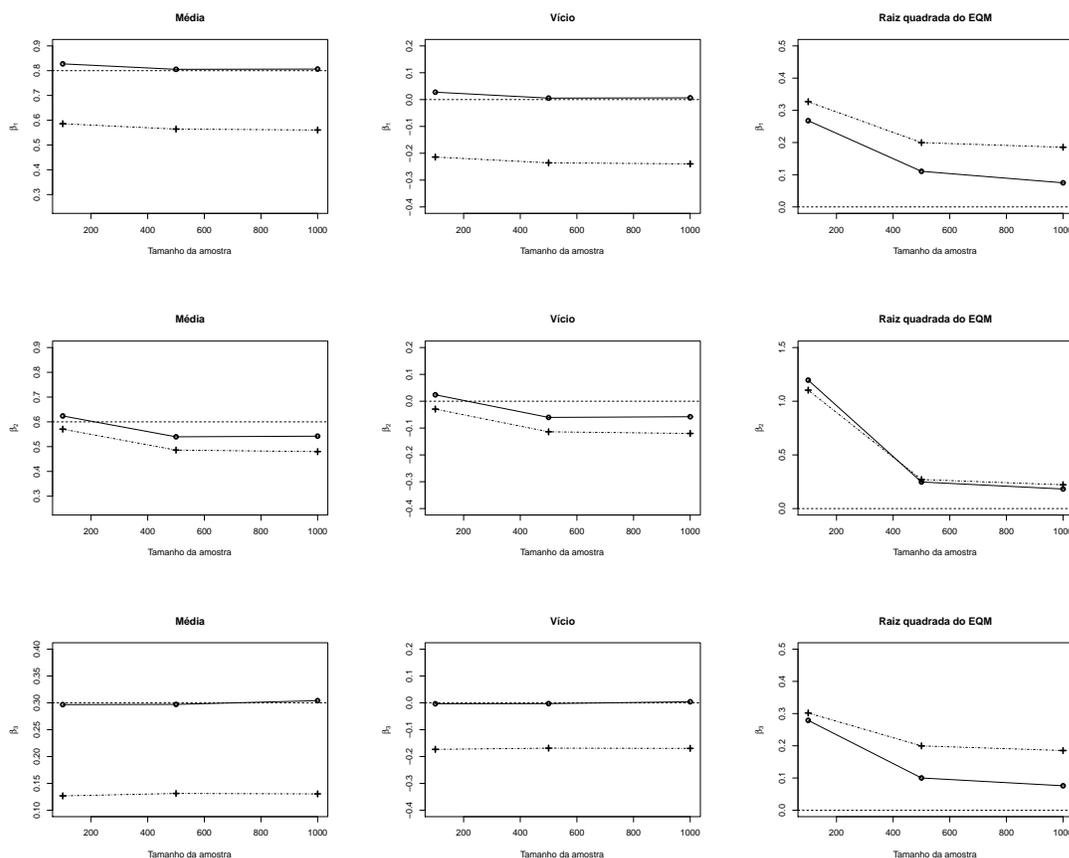


Figura 4.5 - *Comportamento do estudo de simulação 1B. Linha contínua com  $\bullet$ : modelo (4.13) com fragilidade; linha pontilhada com  $+$ : modelo (4.13) sem fragilidade.*

No entanto, comparando os resultados dos estudos de simulação 1A e 1B nas Tabelas 4.3 e 4.4, nota-se que os parâmetros do modelo são mais bem estimados quando as variâncias da interação transição-grupo são diferentes. Os gráficos na Figura 4.6 ilustram essa comparação, em que a linha contínua com  $\bullet$  representa o modelo com variâncias iguais para todas as transições e a linha pontilhada com  $\times$  representa o modelo com variâncias diferentes.

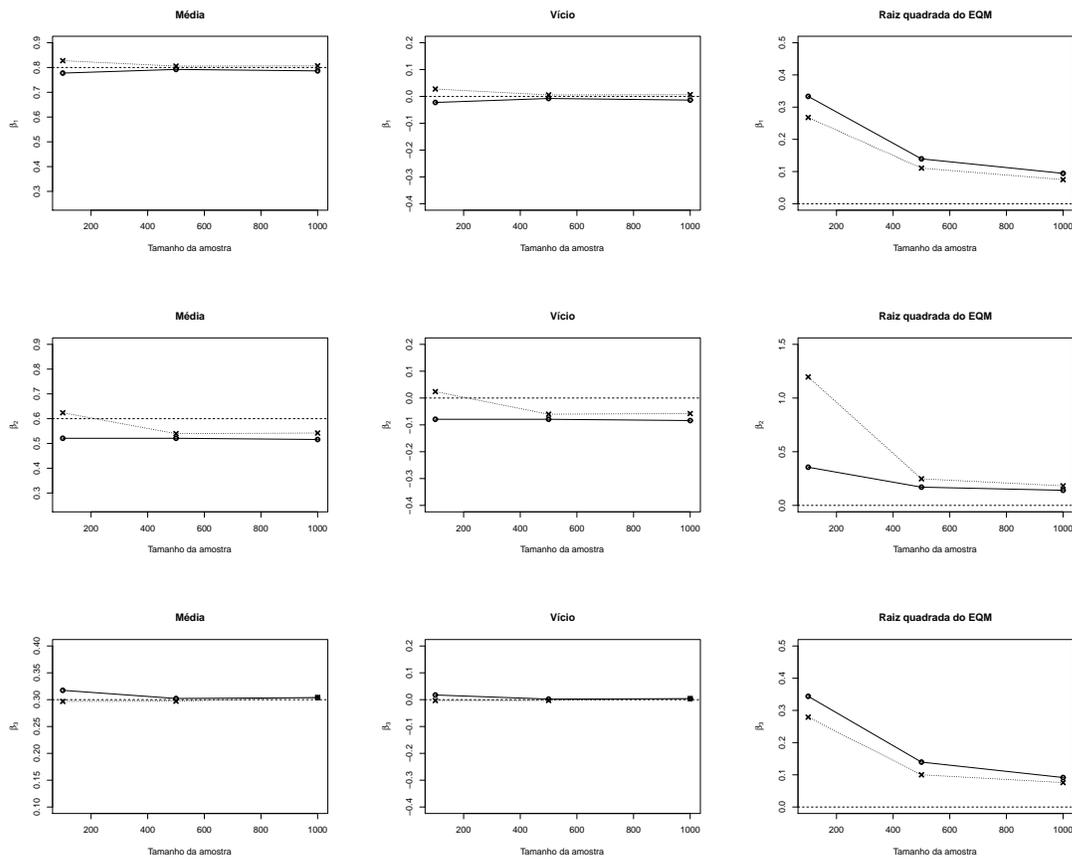


Figura 4.6 - Comportamento dos estudos de simulação 1A e 1B com relação ao modelo (4.13) com fragilidade. Linha contínua com ●: modelo com variâncias da interação transição-grupo iguais; linha pontilhada com ×: modelo com variâncias da interação transição-grupo diferentes.

### 4.3.2 Estudo de simulação 2

Neste estudo de simulação verificamos as propriedades frequentistas dos estimadores do modelo multiestado com fragilidades aninhadas por meio do vício, erro quadrático médio, erro padrão, erro padrão empírico e amplitude do intervalo de confiança dos parâmetros.

Cada intensidade de transição é modelada através do modelo (4.1) sob a restrição de que todas as interações transição-grupo  $W_{qh}$  têm a mesma variância ( $\theta_W := \theta_1 = \dots = \theta_Q$ ). Logo  $W_{qh}$  será denotada por  $W_q$ . Então, os dados são gerados a partir do modelo

$$\lambda_{ghi}(t|v_h w_q) = v_h w_q \lambda_{q0}(t) \exp(\beta'_q \mathbf{X}_{ghi}). \quad (4.14)$$

São reproduzidos três cenários, com  $N = 500$  réplicas cada um, conforme detalhado na Tabela 4.5. Os valores para os parâmetros da distribuição Weibull e de regressão para cada transição em todos os cenários são os mesmos da Tabela 4.2.

Tabela 4.5 - *Cenários do estudo de simulação 2.*

Cenário	A	B	C
Nº de grupos	5	5	5
Nº de indivíduos no grupo	10	20	100
Nº total de indivíduos	50	100	500

O agrupamento é gerado por duas fragilidades Gama, em que

$$V_h \stackrel{iid}{\sim} \text{Gama}(1/\theta_V, 1/\theta_V), \quad h = 1, \dots, 5$$

$$W_q \stackrel{iid}{\sim} \text{Gama}(1/\theta_W, 1/\theta_W), \quad q = 1, 2, 3$$

$$V_h \perp W_q, \quad \forall (h, q),$$

sendo  $V_h$  o efeito global do grupo e  $W_q$  o termo de fragilidade a nível de transição no grupo, no qual será igual para todos os grupos na transição  $q$ , devido à restrição de que a variância é

igual em todas as interações transição-grupo. Os valores das variâncias dos efeitos aleatórios foram fixados como  $\theta_V = 0,8$  e  $\theta_W := \theta_1 = \theta_2 = \theta_3 = 0,4$ .

Em cada cenário foram geradas  $N = 500$  amostras aleatórias e, dessas 500 amostras aleatórias, foram calculadas as médias, as raízes quadradas do erro quadrático médio ( $\sqrt{\text{EQM}}$ ) das EMV, além dos erros padrão assintóticos (EPA), os erros padrão empíricos (EPE) e as médias das amplitudes dos intervalos de confiança de 95% (AmpIC) das estimativas. Os erros padrão empíricos são obtidos a partir da raiz quadrada da variância das  $N$  simulações, de forma que

$$\text{Var}(\hat{\beta}_q) = \frac{1}{N-1} \sum_{j=1}^N (\hat{\beta}_q^{(j)} - \tilde{\beta}_q)^2,$$

em que  $\tilde{\beta}_q = \sum_{j=1}^N \hat{\beta}_q^{(j)} / N$ , sendo  $\hat{\beta}_q^{(j)}$  a estimativa de  $\beta_q$  na  $j$ -ésima simulação.

Os resultados deste estudo de simulação encontram-se na Tabela 4.6. Portanto, à medida que o tamanho amostral aumenta, as médias aproximam-se mais do valores fixados e os valores de  $\sqrt{\text{EQM}}$ , EPA, EPE e AmpIC(95%) diminuem, sugerindo que o modelo atende às propriedades frequentistas. Nas Figuras 4.7 estão os gráficos que mostram o comportamento desses valores.

Tabela 4.6 - Resultados do estudo de simulação 2.

$\beta_q$ (valor real)	Média	$\sqrt{\text{EQM}}$	EPA	EPE	AmpIC (95%)
<b>Cenário A:</b> $G = 5, n_i = 10$ ( $n = 50$ )					
$\beta_1$ (0, 8)	0,8120	0,4464	0,4264	0,9261	1,3954
$\beta_2$ (0, 6)	0,4884	0,5780	0,5422	0,7481	2,9619
$\beta_3$ (0, 3)	0,3031	0,5069	0,4494	0,5903	1,3561
<b>Cenário B:</b> $G = 5, n_i = 20$ ( $n = 100$ )					
$\beta_1$ (0, 8)	0,8028	0,3241	0,2964	0,8653	1,1489
$\beta_2$ (0, 6)	0,4864	0,4218	0,3673	0,6335	1,1859
$\beta_3$ (0, 3)	0,2937	0,3264	0,3048	0,4389	1,1343
<b>Cenário C:</b> $G = 5, n_i = 100$ ( $n = 500$ )					
$\beta_1$ (0, 8)	0,8092	0,1372	0,1270	0,8203	0,4088
$\beta_2$ (0, 6)	0,5009	0,2020	0,1595	0,5306	0,8563
$\beta_3$ (0, 3)	0,3091	0,1330	0,1262	0,3362	0,4062

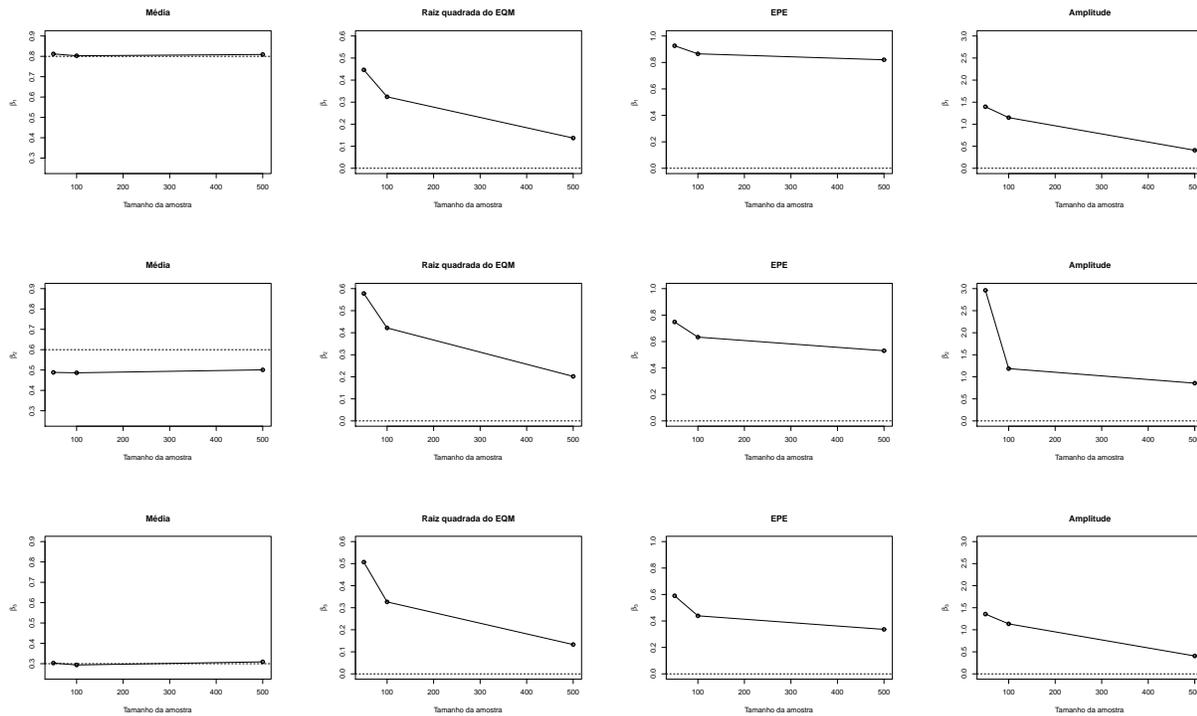


Figura 4.7 - *Comportamento dos cenários A, B e C no estudo de simulação 2.*

### 4.3.3 Comparação de modelos

Com base na estrutura multiestado apresentada na Figura 4.3, foram simuladas  $N = 500$  réplicas com a seguinte configuração: dados multiestado com um total de 5 grupos, com 30 indivíduos cada, totalizando 150 indivíduos na amostra. Iremos gerar dados multiestado dos modelos multiestado sem fragilidades (MSF), multiestado com fragilidade compartilhada apenas para o efeito do grupo (MF1) e multiestado com fragilidades aninhadas (MFA). Para cada um desses dados gerados, iremos ajustar os modelos MSF, MF1 e MFA, de forma que possamos comparar como os ajustes se comportam, dependendo de como os dados foram gerados. Essa comparação será feita por meio dos valores do AIC e do BIC, conforme mostrados na Tabela 4.7.

Tabela 4.7 - Valores médios de AIC e BIC para 500 simulações, considerando uma amostra de  $n = 150$  indivíduos ( $G = 5; n_i = 30$ ).

Modelo no qual os dados foram gerados						
Modelo ajustado	MSF		MF1		MFA	
	AIC	BIC	AIC	BIC	AIC	BIC
MSF	1859,336	1871,227	1837,592	1849,531	2180,576	2192,445
MF1	1868,208	1900,050	1736,875	1768,710	1680,874	1712,526
MFA	2222,416	2234,357	2135,275	2147,213	1671,262	1683,131

Com base nos resultados apresentados na Tabela 4.7, vemos que os ajustes dos modelos são coerentes com os dados gerados, de forma que, se os dados foram gerados de um modelo multiestado com fragilidades aninhadas, por exemplo, espera-se que os valores do AIC e BIC sejam menores quando ajustarmos um modelo com fragilidades aninhadas.

## 4.4 Considerações finais

Neste capítulo mostramos como é feita a integração dos modelos multiestado com o modelo de fragilidade. Essa integração permite lidar com a heterogeneidade não observada entre os grupos de indivíduos, além de considerar a estrutura de dependência entre transições dos indivíduos de um mesmo grupo. Definimos um modelo multiestado com fragilidade aninhada através da função de risco condicional específica da transição do tipo  $q$  ( $q = 1, \dots, Q$ ), em seguida mostramos como é feito o processo de estimação dos seus parâmetros através de duas abordagens: a paramétrica e a semiparamétrica.

No estudo de simulação, mostramos como a incorporação de fragilidades no modelo multiestado pode melhorar suas estimativas ao compararmos com o modelo sem fragilidades. Verificamos também que as propriedades frequentistas dos estimadores do modelo multiestado com fragilidades aninhadas foram atendidas. Além disso, fizemos uma comparação entre modelos multiestado, de forma que púdessemos avaliar se o ajuste do modelo seria coerente com a forma que seus dados foram gerados, por meio dos valores do AIC e do BIC.

No próximo capítulo apresentaremos uma aplicação com um conjunto de dados, extraído

---

do livro de [Klein & Moeschberger \(2003\)](#), que trata do processo de recuperação de medula óssea de pacientes tratados em quatro hospitais. Nesse processo todos pacientes foram submetidos ao transplante de medula óssea, sendo descrito o histórico de eventos que eles podem experimentar. Assim, iremos ajustar um modelo multiestado com fragilidades aninhadas que considere a heterogeneidade não observável entre os hospitais e capte a estrutura de dependência entre as transições que ocorrem para indivíduos de um mesmo hospital, avaliando assim o efeito das covariáveis em cada transição.

---

# 5

## Aplicação

---

### 5.1 Descrição do conjunto de dados

Iremos utilizar o conjunto de dados `bmt` do livro de [Klein & Moeschberger \(2003\)](#), que também está disponível no pacote `mstate` ([de Wreede et al. \(2010\)](#) e [de Wreede et al. \(2011\)](#)) do sistema R ([R Core Team, 2013](#)). Tal conjunto trata do processo de recuperação do transplante de medula óssea, que é um tratamento padrão para leucemia aguda. A recuperação após o transplante é um processo complexo, já que o prognóstico da recuperação pode depender de fatores de risco conhecidos no momento do transplante, tais como a idade do paciente e a idade do doador, além de outros fatores. O prognóstico final também pode mudar à medida que o histórico de pós-transplante do paciente desenvolve-se com a ocorrência de eventos durante o processo de recuperação, tais como o desenvolvimento da doença enxerto versus hospedeiro aguda (AGvHD), a recuperação da contagem de plaquetas a níveis normais, o retorno de granulócitos para níveis normais ou o desenvolvimento de infecções. O transplante pode ser considerado um fracasso quando há a recidiva da leucemia no paciente ou quando o paciente morre enquanto a doença está em remissão, isto é, a morte está relacionada com o tratamento.

Conforme [Klein & Moeschberger \(2003\)](#), o processo de recuperação do paciente é baseado em dois eventos intermediários, nos quais podem ocorrer antes de dois eventos finais (morte e recidiva). Os eventos intermediários são a possibilidade de desenvolver AGvHD, que geralmente ocorre dentro dos primeiros cem dias após o transplante, e a recuperação da contagem

de plaquetas para um nível auto sustentável que seja maior ou igual a  $40 \times 10^9/1$  (sendo chamado de recuperação de plaquetas na sequência do texto). Imediatamente após o transplante, os pacientes têm uma queda da contagem de plaquetas e estão livres de AGvHD. Porém, em algum momento do processo, eles podem ter AGvHD ou ter suas plaquetas recuperadas, o que pode mudar o prognóstico deles. Esses eventos podem ocorrer em qualquer ordem, podendo acontecer um seguido do outro e, na sequência, os pacientes podem morrer ou ter a recidiva da doença. Os pacientes também podem morrer ou ter a recidiva da doença sem experimentar qualquer um desses eventos intermediários.

A Figura 5.1 mostra o esquema desse processo de recuperação, com seus estados e suas possíveis transições. Todos os pacientes entram no estudo após serem transplantados, sendo o transplante o estado inicial, AGvHD e recuperação de plaquetas (Rec.P) os estados intermediários, e os estados recidiva e morte os estados absorventes, ou seja, a partir do momento que os pacientes entram em um deles não há mais transições. O paciente também pode desenvolver AGvHD após a recuperação de plaquetas (AGvHD/Rec.P), sendo este também um estado intermediário. Como forma de identificar qual o tipo de transição que ocorre, enumeramos todas elas, conforme mostrado na Figura 5.1, sendo um total de 10 tipos de transições possíveis.

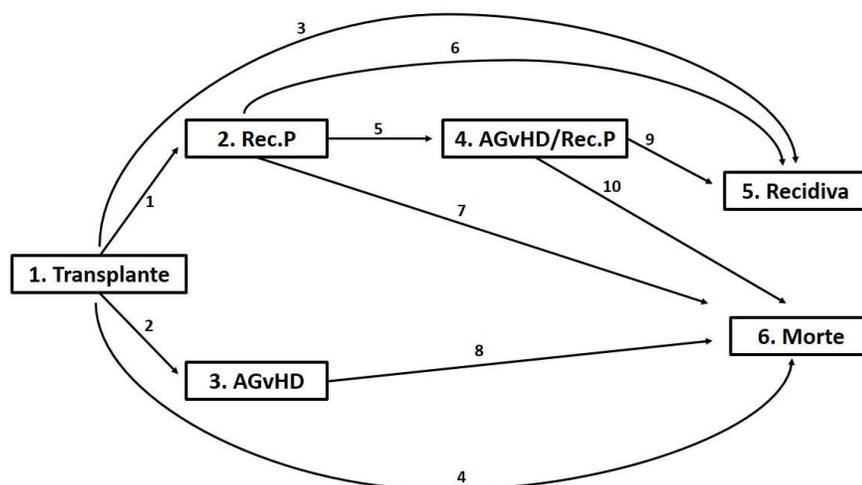


Figura 5.1 - Processo de recuperação de um transplante de medula óssea.

Foi considerado um estudo multicêntrico de pacientes preparados para o transplante com um regime de radiação livre condicionada. No total, 137 pacientes foram tratados em um dos quatro hospitais: 76 nos Ohio State University Hospitals (OSU) em Colombo; 17 na Hahnemann University (HU) em Filadélfia; 23 no St. Vincent's Hospital (SVH) em Sydney; e 21 no Alfred Hospital (AH) em Melbourne. De forma mais clara, o total de pacientes em cada hospital está descrito na Tabela 5.1. Para mais detalhes sobre esse estudo, vide [Copelan et al. \(1991\)](#).

Tabela 5.1 - Total de pacientes em cada hospital.

Hospital	1 (OSU)	2 (HU)	3 (SVH)	4 (AH)
Nº de pacientes	76	17	23	21

A Tabela 5.2 mostra os dois primeiros pacientes do conjunto de dados:

Tabela 5.2 - Primeiras 2 linhas do conjunto de dados.

paciente	hospital	t1	t2	d1	d2	ta	da	tp	dp	X <sub>1</sub>	X <sub>2</sub>
1	1	2081	2081	0	0	67	1	13	1	26	33
2	1	1602	1602	0	0	1602	0	18	1	21	37

O conjunto é descrito da seguinte forma:

- **paciente**: identificação do paciente.
- **hospital**: hospital no qual o paciente fez o tratamento (1: OSU; 2: HU; 3: SVH; 4: AH).
- **t1**: tempo (em dias) para a morte ou último acompanhamento.
- **t2**: tempo (em dias) de sobrevivência livre de doenças (tempo para recidiva, para a morte ou o último acompanhamento).
- **d1**: indicador de morte (1: sim; 0: não).
- **d2**: indicador de recidiva (1: sim; 0: não).

- **ta**: tempo (em dias) para o desenvolvimento de AGvHD.
- **da**: indicador de AGvHD (1: sim; 0: não).
- **tp**: tempo (em dias) para a recuperação das plaquetas.
- **dp**: indicador de recuperação de plaquetas (1: plaquetas voltaram ao normal; 0: plaquetas não voltaram ao normal).
- **X<sub>1</sub>**: idade do paciente (em anos) no momento do transplante.
- **X<sub>2</sub>**: idade do doador (em anos).

Na Tabela 5.3 estão as medidas descritivas das duas covariáveis contínuas que serão consideradas para o ajuste.

Tabela 5.3 - *Medidas descritivas das variáveis contínuas.*

Covariáveis	Mínimo	1º quartil	Mediana	Média	3º quartil	Máximo	Desvio padrão
X <sub>1</sub>	7	21	28	28,36	35	52	9,56
X <sub>2</sub>	2	21	28	28,33	35	56	10,18

A partir daqui os estados **transplante**, **recuperação de plaquetas**, **AGvHD**, **recidiva** e **morte** serão representados como **T**, **Rec.P**, **AGvHD**, **R** e **M**, respectivamente. Se um paciente desenvolver AGvHD depois da contagem da recuperação de plaquetas, esse estado será representado como **AGvHD/Rec.P**. Na Tabela 5.4 temos o total de eventos para cada tipo de transição e na Tabela 5.5 a proporção dos eventos. Sendo assim, de acordo com os resultados apresentados na Tabela 5.4, temos que, dos 137 pacientes pós-transplantados, 116 tiveram a contagem de suas plaquetas recuperadas (transição **T**→**Rec.P**), 8 desenvolveram a doença enxerto versus hospedeiro aguda (transição **T**→**AGvHD**), 3 tiveram a recidiva da doença sem experimentar eventos intermediários (transição **T**→**R**) e 10 morreram sem também experimentar os eventos intermediários (transição **T**→**M**). Dos 116 pacientes que tiveram a contagem de suas plaquetas recuperadas após o transplante, 18 deles desenvolve-

ram AGvHD (transição Rec.P $\rightarrow$ AGvHD), 20 morreram sem desenvolver AGvHD (transição Rec.P $\rightarrow$ M) e a maioria deles (34) tiveram a recidiva da doença (transição Rec.P $\rightarrow$ R).

Tabela 5.4 - *Frequências das transições.*

De/Para	T	Rec.P	AGvHD	AGvHD/Rec.P	R	M
T	-	116	8	-	3	10
Rec.P	-	-	-	18	34	20
AGvHD	-	-	-	-	0	7
AGvHD/Rec.P	-	-	-	-	3	6
R	-	-	-	-	-	-
M	-	-	-	-	-	-

Tabela 5.5 - *Proporção das transições.*

De/Para	T	Rec.P	AGvHD	AGvHD/Rec.P	R	M
T	-	0,8467	0,0584	-	0,0219	0,0730
Rec.P	-	-	-	0,1552	0,2931	0,1724
AGvHD	-	-	-	-	0	0,8750
AGvHD/Rec.P	-	-	-	-	0,1677	0,3333
R	-	-	-	-	-	-
M	-	-	-	-	-	-

## 5.2 Resultados da Aplicação

Iremos considerar quatro modelos semiparamétricos para o conjunto de dados de transplante de medula óssea, ajustados de acordo com a estrutura mostrada na Figura 5.1. São os modelos:

- **Modelo 1 (MSF):** Modelo multiestado sem fragilidades.
- **Modelo 2 (MF1):** Modelo multiestado com fragilidade compartilhada apenas para o efeito do grupo.
- **Modelo 3 (MF2):** Modelo multiestado com fragilidade compartilhada apenas para o efeito da interação transição-grupo.

- **Modelo 4 (MFA):** Modelo multiestado com fragilidades aninhadas.

Antes de ajustar os modelos citados acima, devemos efetuar uma transformação dos dados para um formato longo, conforme mencionado na Seção 2.1. Com o auxílio do pacote `mstate` fazemos essa transformação dos dados apresentados na Tabela 5.2, que ficam como apresentados na Tabela 5.6.

Tabela 5.6 - *Primeiros dois pacientes do conjunto de dados transformado.*

de	para	trans	Tinício	Tparada	tempo	status	paciente	hospital	X <sub>1</sub>	X <sub>2</sub>
1	2	1	0	13	13	1	1	1	26	33
1	3	2	0	13	13	0	1	1	26	33
1	6	3	0	13	13	0	1	1	26	33
1	7	4	0	13	13	0	1	1	26	33
2	4	5	13	67	54	1	1	1	26	33
2	6	6	13	67	54	0	1	1	26	33
2	7	7	13	67	54	0	1	1	26	33
4	6	11	67	2081	2014	0	1	1	26	33
4	7	12	67	2081	2014	0	1	1	26	33
<hr/>										
1	2	1	0	18	18	1	2	1	21	37
1	3	2	0	18	18	0	2	1	21	37
1	6	3	0	18	18	0	2	1	21	37
1	7	4	0	18	18	0	2	1	21	37
2	4	5	18	1602	1584	0	2	1	21	37
2	6	6	18	1602	1584	0	2	1	21	37
2	7	7	18	1602	1584	0	2	1	21	37

Após a transformação do conjunto de dados, ajustamos os modelos de forma semiparamétrica. As estimativas dos parâmetros dos modelos 1, 2 e 3 foram obtidas com o auxílio do pacote `mstate`, enquanto que o modelo 4 foi ajustado com o auxílio do pacote `mlfm` (Rotolo & Horny, 2012), sob a restrição de que todas as interações transição-grupo  $W_{qh}$  têm a mesma variância ( $\theta_W := \theta_1 = \dots = \theta_{10}$ ), sendo essa interação denotada agora por  $W_q$ . Na Tabela 5.7 encontram-se as estimativas dos parâmetros de regressão  $\beta_q = (\beta_1, \beta_2)$  ( $q = 1, \dots, 10$ ) dos quatro modelos, com seus respectivos erros padrão.

Tabela 5.7 - Resultados dos modelos ajustados.

	Modelo 1			Modelo 2			Modelo 3			Modelo 4		
	$\hat{\beta}_q$	$\exp(\hat{\beta}_q)$	$EP(\hat{\beta}_q)$									
$\beta_{1,1}$	0,0211	1,0213	0,0156	0,0278	1,0282	0,0163	0,0398	1,0406	0,0170	0,0272	1,0276	0,0159
$\beta_{1,2}$	0,0933	1,0978	0,0624	0,0978	1,1027	0,0632	0,0953	1,0999	0,0628	0,0974	1,1023	0,0631
$\beta_{1,3}$	0,0840	1,0876	0,1079	0,1003	1,1055	0,1108	0,0840	1,0876	0,1079	0,0987	1,1037	0,1101
$\beta_{1,4}$	-0,0610	0,9409	0,0552	-0,0535	0,9479	0,0561	-0,1062	0,8992	0,0611	-0,0541	0,9474	0,0560
$\beta_{1,5}$	0,0114	1,0115	0,0373	0,0184	1,0186	0,0378	0,0114	1,0115	0,0373	0,0181	1,0182	0,0377
$\beta_{1,6}$	0,0356	1,0362	0,0267	0,0431	1,0441	0,0271	0,0356	1,0362	0,0267	0,0426	1,0436	0,0270
$\beta_{1,7}$	0,0135	1,0136	0,0364	0,0217	1,0220	0,0371	0,0244	1,0247	0,0377	0,0212	1,0214	0,0370
$\beta_{1,8}$	-0,0430	0,9579	0,1001	-0,0675	0,9347	0,1004	-0,0430	0,9579	0,1001	-0,0657	0,9364	0,1000
$\beta_{1,9}$	-0,1586	0,8533	0,1202	-0,1758	0,8388	0,1291	-0,1586	0,8533	0,1202	-0,1742	0,8401	0,1282
$\beta_{1,10}$	0,0413	1,0422	0,0517	0,0462	1,0473	0,0526	0,1269	1,1353	0,0808	0,0463	1,0474	0,0522
$\beta_{2,1}$	-0,0134	0,9867	0,0136	-0,0150	0,9851	0,0141	-0,0217	0,9785	0,0149	-0,0148	0,9854	0,0138
$\beta_{2,2}$	0,0016	1,0016	0,0508	0,0017	1,0017	0,0522	0,0027	1,0027	0,0515	0,0017	1,0017	0,0520
$\beta_{2,3}$	-0,0133	0,9868	0,0869	-0,0196	0,9806	0,0885	-0,0133	0,9868	0,0869	-0,0188	0,9814	0,0880
$\beta_{2,4}$	0,0972	1,1021	0,0468	0,0967	1,1015	0,0481	0,1122	1,1187	0,0417	0,0967	1,1015	0,0479
$\beta_{2,5}$	0,0198	1,0200	0,0392	0,0175	1,0176	0,0400	0,0198	1,0200	0,0392	0,0176	1,0177	0,0399
$\beta_{2,6}$	-0,0205	0,9798	0,0268	-0,0241	0,9762	0,0274	-0,0205	0,9797	0,0268	-0,0239	0,9764	0,0273
$\beta_{2,7}$	-0,0265	0,9738	0,0363	-0,0311	0,9694	0,0369	-0,0319	0,9686	0,0374	-0,0308	0,9697	0,0368
$\beta_{2,8}$	0,0808	1,0842	0,0856	0,0791	1,0824	0,0839	0,0808	1,0842	0,0856	0,0794	1,0827	0,0840
$\beta_{2,9}$	0,1483	1,1599	0,1770	0,1618	1,1756	0,1743	0,1483	1,1599	0,1770	0,1606	1,1742	0,1743
$\beta_{2,10}$	-0,0320	0,9685	0,1034	-0,0367	0,9640	0,1037	-0,1037	0,9015	0,1260	-0,0367	0,9640	0,1036
Variância dos efeitos aleatórios												
Hospital	-				0,0471			-			0,0401	
Transição	-				-			*			2,1381	

Com relação aos resultados da Tabela 5.7, as estimativas dos parâmetros de regressão permanecem positivos ou negativos em todos os modelos, sendo, em geral, seus valores muito próximos um do outro. Na tabela 5.8 estão os valores de AIC e BIC dos modelos 1, 2 e 4. Os valores do modelo 3 foram omitidos porque foi estimado um modelo multiestado para cada transição, sendo isto possível graças à suposição de Markov (Hougaard, 2000).

Tabela 5.8 - Valores de AIC, BIC dos modelos 1, 2 e 4.

Modelo	AIC	BIC
Modelo 1	1781.157	1878.075
Modelo 2	1778.006	1894.307
Modelo 4	1770.495	1867.413

A partir dos resultados apresentados na Tabela 5.8, observamos que o AIC e BIC do modelo com fragilidades aninhadas (modelo 4) têm valores menores em relação aos modelos 1 e 2, o que sugere ser o melhor modelo dentre os ajustados para estudar o processo de recuperação do transplante de medula óssea. A seguir iremos interpretar os resultados obtidos para o modelos 3 e 4.

### 5.2.1 Interpretação dos resultados do modelo 3

Em resumo, a idade do paciente aumenta o risco de, após o transplante, recuperar a contagem de suas plaquetas (transição 1), desenvolver AGvHD (transição 2) e ter a recidiva da doença (transição 3). No entanto, quanto mais velho for o paciente, menor será o risco dele morrer após o transplante (transição 4). Ou seja, para cada ano de idade do paciente, o risco de morte após o transplante sem experimentar eventos intermediários diminui 10,08%.

O limite inferior (LI) e limite superior (LS) dos intervalos de confiança dos parâmetros de regressão do modelo 3 encontram-se na Tabela 5.9.

Tabela 5.9 - Intervalos de confiança de 95% dos parâmetros de regressão do modelo 3.

Parâmetros	1 (T→Rec.P)			2 (T→AGvHD)			3 (T→R)		
	LI	$\exp(\beta_q)$	LS	LI	$\exp(\beta_q)$	LS	LI	$\exp(\beta_q)$	LS
$\beta_1$	1,0066	1,0406	1,0758	0,9726	1,0999	1,2440	0,8804	1,0876	1,3440
$\beta_2$	0,9504	0,9785	1,0074	0,9065	1,0027	1,1090	0,8322	0,9868	1,170
Parâmetros	4 (T→M)			5 (Rec.P→AGvHD/Rec.P)			6 (Rec.P→R)		
	LI	$\exp(\beta_q)$	LS	LI	$\exp(\beta_q)$	LS	LI	$\exp(\beta_q)$	LS
$\beta_1$	0,7978	0,8992	1,0140	0,9401	1,0115	1,0880	0,9834	1,0362	1,0920
$\beta_2$	1,0309	1,1187	1,2140	0,9445	1,0200	1,1010	0,9296	0,9797	1,0330
Parâmetros	7 (Rec.P→M)			8 (AGvHD→M)			9 (AGvHD/Rec.P→R)		
	LI	$\exp(\beta_q)$	LS	LI	$\exp(\beta_q)$	LS	LI	$\exp(\beta_q)$	LS
$\beta_1$	0,9518	1,0247	1,1030	0,7873	0,9579	1,1660	0,6742	0,8533	1,0800
$\beta_2$	0,9002	0,9686	1,0420	0,9167	1,0842	1,2820	0,8199	1,1599	1,641
Parâmetros	10 (AGvHD/Rec.P→M)								
	LI	$\exp(\beta_q)$	LS						
$\beta_1$	0,9691	1,1353	1,3300						
$\beta_2$	0,7042	0,9015	1,1540						

As variâncias dos efeitos aleatórios relativos às transições do modelo 3 (\*) estão na Tabela 5.10. O valor de  $\hat{\theta}_q$  ( $q = 1, \dots, 10$ ) no modelo 3 reflete a heterogeneidade entre os hospitais para cada transição. Quanto maior for a estimativa de  $\hat{\theta}_q$ , maior será a heterogeneidade entre os hospitais com relação à transição  $q$ .

Tabela 5.10 - Variâncias dos efeitos aleatórios do modelo 3.

Transição	$\hat{\theta}_q$
1 (T→Rec.P)	0,3126
2 (T→AGvHD)	0,0823
3 (T→R)	$5 \times 10^{-9}$
4 (T→M)	0,7434
5 (Rec.P→AGvHD/Rec.P)	$5 \times 10^{-7}$
6 (Rec.P→R)	$5 \times 10^{-7}$
7 (Rec.P→M)	0,2824
8 (AGvHD→M)	$5 \times 10^{-9}$
9 (AGvHD/Rec.P→R)	$5 \times 10^{-7}$
10 (AGvHD/Rec.P→M)	1,0190

As fragilidades estimadas da interação transição-hospital estão na Tabela 5.11 e seus respectivos gráficos estão na Figura 5.2. Tais estimativas representam um efeito aleatório que descrevem o risco comum para um determinado evento/transição, isto é, a fragilidade

compartilhada por pacientes dentro de um mesmo hospital para aquela transição específica. Sendo assim, o valor da estimativa da fragilidade da interação transição-hospital reflete o quão frágeis são os pacientes de um hospital para experimentar determinado tipo de evento/transição. Logo, devido a fragilidade atuar de forma multiplicativa no modelo, os pacientes com valores grandes de fragilidades nessa interação deverão experimentar determinado tipo de evento em tempos menores do que aqueles com valores pequenos dessa fragilidade.

Tabela 5.11 - *Fragilidades estimadas do modelo 3.*

Hospital	1 (T→Rec.P)	2 (T→AGvHD)	3 (T→R)
1 (OSU)	1,4525	1,1713	1
2 (HU)	0,4236	0,9999	1
3 (SVH)	0,7617	0,8940	1
4 (AH)	1,3622	0,9349	1
Hospital	4 (T→M)	5 (Rec.P→AGvHD/Rec.P)	6 (Rec.P→R)
1 (OSU)	0,2911	1	1
2 (HU)	2,0707	1	1
3 (SVH)	0,5655	1	1
4 (AH)	1,0726	1	1
Hospital	7 (Rec.P→M)	8 (AGvHD→M)	9 (AGvHD/Rec.P→R)
1 (OSU)	1,4115	1,0842	1
2 (HU)	1,806	1	1
3 (SVH)	0,8434	1	1
4 (AH)	0,5645	1	1
Hospital	10 (AGvHD/Rec.P→M)		
1 (OSU)	1,0556		
2 (HU)	1,2784		
3 (SVH)	1,5490		
4 (AH)	0,1170		

Na transição 1 (T→Rec.P), por exemplo, vemos que a fragilidade do hospital 1 (OSU) é maior que a fragilidade dos demais, sugerindo que os pacientes deste hospital irão recuperar a contagem de suas plaquetas em um tempo menor do que os pacientes dos outros hospitais. Já em relação à transição 4 (T→M), os pacientes do hospital 2 (HU) deverão morrer após o transplante, sem experimentar os eventos intermediários, em um tempo menor que os pacientes dos outros hospitais, pois o valor de sua fragilidade é o maior.

Nas transições 3, 5, 6, 8 e 9, as fragilidades estimadas dos 4 hospitais foram iguais/próximas

a 1, devido às estimativas da variância da fragilidade dessas transições serem muito próximas de zero (Tabela 5.10). Tal resultado implica que não há heterogeneidade entre os hospitais em relação à essas transições. No entanto, quanto maior for o valor dessa estimativa, maior será essa heterogeneidade.

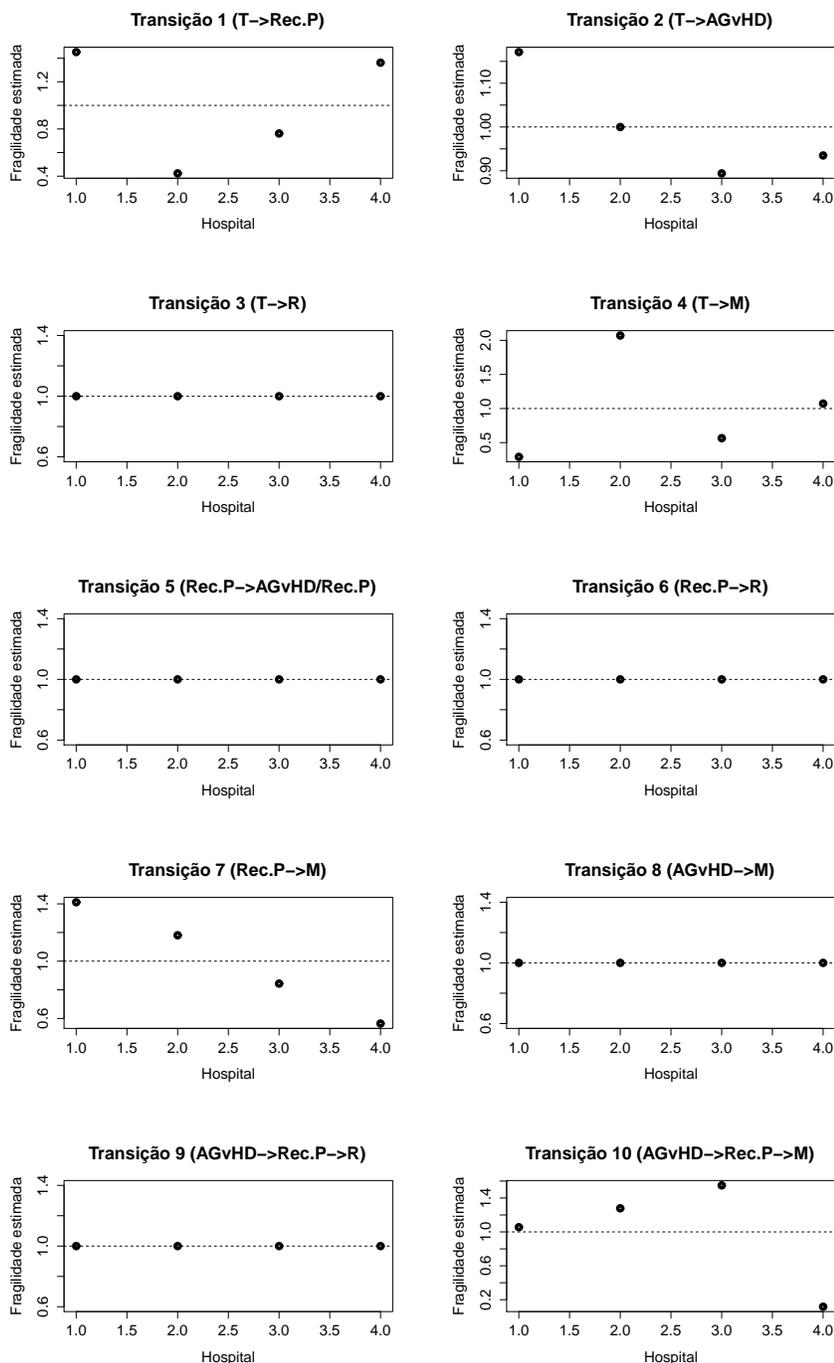


Figura 5.2 - *Fragilidades estimadas do modelo 3.*

### 5.2.2 Interpretação dos resultados do modelo 4

No modelo multiestado com fragilidade aninhada, quanto maior a variância a nível de hospital ( $\theta_V$ ), maior será a heterogeneidade entre eles e quanto maior a variância a nível de transição no hospital ( $\theta_q$ ), mais alta será a correlação entre as transições dentro de um mesmo hospital, ou seja, haverá uma dependência maior entre os tempos das transições de pacientes de um mesmo hospital.

O ajuste do modelo 4 foi feito sob a restrição de que todas as interações transição-grupo  $W_{qh}$  têm a mesma variância ( $\theta_W := \theta_1 = \dots = \theta_{10}$ ). Nas tabelas 5.12 e 5.13 temos as fragilidades estimadas do modelo a nível de hospital e a nível de transição no hospital, respectivamente, e nas Figuras 5.3 e 5.4 seus respectivos gráficos.

Tabela 5.12 - *Fragilidades estimadas do modelo 4 a nível de hospital  $v_h$  ( $h = 1, 2, 3, 4$ ).*

Hospital	Estimativa
1 (OSU)	1,2406
2 (HU)	0,8530
3 (SVH)	0,8514
4 (AH)	1,0550

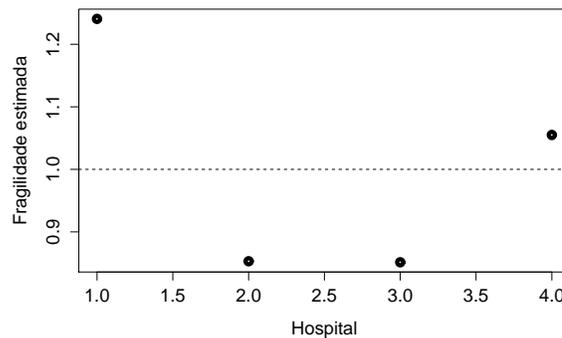


Figura 5.3 - *Fragilidades estimadas a nível de hospital.*

Devido à restrição de que todas as interações transição-grupo têm a mesma variância  $\theta_W$ , o valor de sua estimativa de fragilidade  $\hat{w}_q$  na transição  $q$  será igual para todos os hospitais. Sendo assim, quanto maior for o valor desta estimativa, mais frágeis serão os pacientes de

todos os hospitais para experimentar a transição  $q$ . O que irá aumentar/diminuir ainda mais o risco do paciente experimentar a transição  $q$  será o valor de  $\hat{v}_h$ , pois  $\hat{u}_{qh} = \hat{v}_h \hat{w}_q$  atua de forma multiplicativa na função de risco. Temos então que, com relação à transição 1 (T→Rec.P), os pacientes do hospital 1 (OSU) irão recuperar a contagem de suas plaquetas mais rápido do que os pacientes de outros hospitais, enquanto que os pacientes do hospital 2 (HU) irão efetuar essa transição num tempo maior.

Caso não houvesse essa restrição, a estimativa de fragilidade da interação transição-grupo ( $\hat{w}_{qh}$ ) refletiria quão frágeis são os pacientes do hospital  $h$  para experimentar a transição  $q$ , além de considerar, também, o risco comum de todos os pacientes de um mesmo hospital ( $\hat{v}_h$ ), o que aumentaria/diminuiria ainda mais sua fragilidade, dependendo do seu valor.

Tabela 5.13 - *Fragilidades estimadas do modelo 4 a nível de transição.*

Transição	Estimativa
1 (T→Rec.P)	6,8848
2 (T→AGvHD)	0,0318
3 (T→R)	0,0250
4 (T→M)	0,2061
5 (Rec.P→AGvHD/Rec.P)	0,0991
6 (Rec.P→R)	0,2924
7 (Rec.P→M)	0,3806
8 (AGvHD→M)	1,3713
9 (AGvHD/Rec.P→R)	0,2836
10 (AGvHD/Rec.P→M)	0,4254

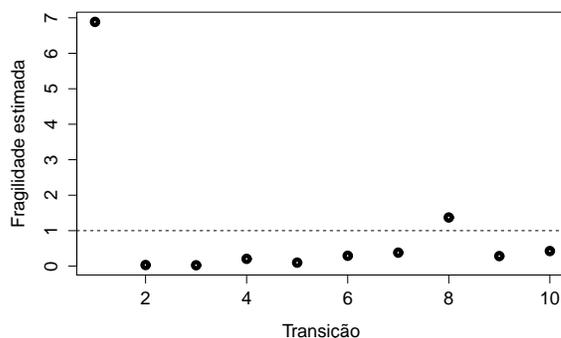


Figura 5.4 - *Fragilidades estimadas a nível de transição.*

Os intervalos de confiança dos parâmetros de regressão do modelo 4 encontram-se na Tabela 5.14.

Tabela 5.14 - *Intervalos de confiança do modelo 4.*

Parâmetros	LI	exp( $\beta$ )	LS
$\beta_{1.1}$	0,9960	1,0280	1,0600
$\beta_{1.2}$	0,9741	1,1020	1,2470
$\beta_{1.3}$	0,8894	1,1040	1,3700
$\beta_{1.4}$	0,8490	0,9474	1,0570
$\beta_{1.5}$	0,9458	1,0180	1,0960
$\beta_{1.6}$	0,9898	1,0440	1,1000
$\beta_{1.7}$	0,9500	1,0210	1,0980
$\beta_{1.8}$	0,7697	0,9364	1,1390
$\beta_{1.9}$	0,6535	0,8401	1,0800
$\beta_{1.10}$	0,9454	1,0470	1,1600
$\beta_{2.1}$	0,9590	0,9854	1,0120
$\beta_{2.2}$	0,9047	1,0020	1,1090
$\beta_{2.3}$	0,8259	0,9814	1,1660
$\beta_{2.4}$	1,0030	1,1020	1,2100
$\beta_{2.5}$	0,9411	1,0180	1,1010
$\beta_{2.6}$	0,9255	0,9764	1,0300
$\beta_{2.7}$	0,9022	0,9697	1,0420
$\beta_{2.8}$	0,9184	1,0830	1,2760
$\beta_{2.9}$	0,8345	1,1740	1,6520
$\beta_{2.10}$	0,7869	0,9640	1,1810

Sobre as intensidades de transição do modelo 4, de forma mais detalhada, temos que:

### 1. Transição 1 ( $T \rightarrow \text{Rec.P}$ )

- Para um incremento de uma unidade na idade, o risco dele ter sua contagem de plaquetas recuperadas logo após o transplante (transição 1) aumenta em 2,76%. Sendo assim, quanto mais velho for o paciente, maior será o risco dele recuperar a contagem de suas plaquetas.
- Já, para cada ano a mais na idade do doador, o risco do paciente recuperar suas plaquetas diminui em 1,46% ( $1-0,9854$ ).

### 2. Transição 2 ( $T \rightarrow \text{AGvHD}$ )

- A idade também tem um efeito positivo nesta transição, aumentando em 10,23% o risco do paciente desenvolver AGvHD logo após o transplante.

- Para cada ano a mais na idade do doador, o risco do paciente desenvolver AGvHD aumenta em 0,17%.

### 3. Transição 3 (T→R)

- Para cada ano da idade do paciente, o risco dele ter uma recidiva da doença, sem ter a contagem de suas plaquetas recuperadas e ter desenvolvido AGvHD, aumenta 10,37%.
- Para cada ano da idade do doador, o risco do paciente ter a recidiva da doença, sem experimentar os eventos intermediários, diminui 1,86%.

### 4. Transição 4 (T→M)

- Para cada ano de idade do paciente, o risco de morte após o transplante sem experimentar os eventos intermediários, diminui 5,26%.
- Para cada ano de idade do doador, o risco desta transição aumenta 10,15%.

### 5. Transição 5 (Rec.P→AGvHD/Rec.P)

- Para cada ano de idade do paciente, o risco de um paciente, que após o transplante teve a recuperação de plaquetas e, em seguida, desenvolver AGvHD, aumenta em 1,82%.
- A idade do doador também tem um efeito positivo nesta transição, aumentando seu risco em 1,77% para cada ano de idade do doador.

### 6. Transição 6 (Rec.P→R)

- O risco de um paciente que teve a contagem de suas plaquetas recuperadas após o transplante e, em seguida, ter a recidiva da doença, aumenta 4,36% para cada ano de sua idade.
- No entanto, o risco desta transição diminui em 2,36% para cada ano de idade do doador.

**7. Transição 7 (Rec.P→M)**

- O risco de um paciente que teve a contagem de suas plaquetas recuperadas após o transplante e, em seguida, morrer, aumenta 2,14% para cada ano de sua idade.
- O risco desta transição diminui em 3,03% para cada ano de idade do doador.

**8. Transição 8 (AGvHD→M)**

- Para cada ano de idade do paciente, o risco dele desenvolver AGvHD, dado que ele teve a recuperação da contagem de plaquetas logo após o transplante, diminui em 6,36%.
- Já o risco da transição aumenta em 8,27% para cada ano de idade do doador.

**9. Transição 9 (AGvHD/Rec.P→R)**

- O risco de um paciente desenvolver AGvHD logo após o transplante diminui cerca de 16% para cada ano de sua idade.
- O risco da transição também aumenta para cada ano de idade do doador, sendo esse aumento de 17,42%.

**10. Transição 10 (AGvHD/Rec.P→M)**

- O risco de um paciente recidir, dado que ele desenvolveu AGvHD após a recuperação de suas plaquetas, aumenta em 4,74% para cada ano de sua idade.
- Porém, o risco para essa transição diminui em 3,6% para cada ano de idade do doador.

## 5.3 Considerações finais

Efetuamos o ajuste de quatro modelos diferentes, no entanto demos ênfase aos modelos multiestado com fragilidade compartilhada para a interação transição-grupo (modelo 3) e multiestado com fragilidades aninhadas (modelo 4). A diferença básica entre esses dois modelos é que o modelo 3 considera apenas o efeito da interação transição-grupo, enquanto que o modelo 4, além de considerar essa interação, ainda considera o efeito do grupo.

A restrição do modelo 4 no nosso ajuste, foi de que todas as interações transição-hospital têm a mesma variância, implicando que quanto maior for a estimativa das fragilidades para essas interações mais frágeis serão os pacientes de todos os hospitais para experimentar um determinado evento/transição. O que irá aumentar/diminuir ainda mais o risco do paciente experimentar uma transição será a estimativa de fragilidade para o hospital, já que  $\hat{u}_{qh} = \hat{v}_h \hat{w}_q$  atua de forma multiplicativa na função de risco.

Em ambos os modelos, em resumo, o risco do paciente recuperar a contagem de suas plaquetas (transição 1), desenvolver AGvHD (transição 2) e ter a recidiva da doença (transição 3) aumenta para cada ano de idade do paciente. Porém, quanto mais velho for o paciente, menor será o risco dele morrer após o transplante (transição 4), sem experimentar eventos intermediários. Caso o paciente desenvolva AGvHD (transição 2), o risco de recidiva (transição 9) e morte (transição 10) diminui a cada ano de idade do paciente.

Quanto maior for a idade do doador, menor será o risco de um paciente ter a contagem de suas plaquetas recuperadas logo após o transplante (transição 1) e menor, também, será o risco do paciente recidir ou morrer depois que teve a recuperação de plaquetas após o transplante (transições 6 e 7, respectivamente).

O valor das fragilidades estimadas da interação transição-hospital (modelo 3) poderia ser usado como um indicador da qualidade do tratamento que é oferecido aos seus pacientes. Se percebemos que os pacientes de determinado hospital estão mais propensos a experimentar eventos não favoráveis, que irão diminuir sua qualidade de vida ou até causar sua morte, por exemplo, isso poderia servir de alerta para que haja uma melhoria no tratamento que ele oferece a seus pacientes.

---

# 6

## Considerações finais e propostas futuras

---

Inicialmente fizemos uma revisão da literatura sobre a metodologia base do tema proposto e, em seguida, uma breve introdução à análise de sobrevivência. Nos dois capítulos seguintes, abordamos a metodologia dos modelos multiestado e dos modelos de fragilidade, mostrando o por quê de sua necessidade, para que no Capítulo 4 possamos tratar da integração desses dois modelos, além de mostrarmos estudos de simulação para comparação de modelos multiestado com e sem fragilidade e para verificar as propriedades frequentistas dos estimadores do modelo multiestado com fragilidades aninhadas.

A aplicação a um conjunto de dados reais sobre o processo de recuperação de medula óssea mostrou a importância do uso de fragilidades em modelos multiestado, onde tomamos como base o modelo de fragilidades aninhadas, e a partir dele, ajustamos também seus casos particulares (conforme mostrado no Capítulo 4), para que pudéssemos fazer uma comparação entre os modelos. Feita a comparação entre os modelos ajustados, fizemos a interpretação dos resultados dos modelos 3 (multiestado com fragilidade compartilhada para o efeito da interação transição-hospital) e 4 (multiestado com fragilidades aninhadas). Ambos os modelos lidam com o efeito do hospital em cada transição, porém o modelo 3 ignora a estrutura de correlação de diferentes eventos que são influenciados pelos mesmos fatores de risco não observados para pacientes de um mesmo hospital, pois as fragilidades da interação são consideradas independentes dentro do mesmo hospital. Para contornar essa situação, o modelo 4 assume que, dentro de um mesmo hospital, exista dependência entre as fragilidades da interação transição-grupo, implicando numa estrutura de correlação entre os eventos

dentro do mesmo hospital e independência entre eventos de diferentes hospitais, além de considerar a heterogeneidade não observada entre os hospitais na sua função de risco.

Os valores das fragilidades estimadas da interação transição-grupo revelam o quão frágeis os pacientes de cada hospital são para experimentarem determinado tipo de evento/transição. Portanto, tais estimativas podem auxiliar na melhoria do tratamento ofertado aos pacientes de cada hospital, pois se percebermos que os pacientes de determinado hospital estão mais propensos a experimentarem eventos não favoráveis do que os demais, que irão diminuir sua qualidade/expectativa de vida, por exemplo, isso serviria de alerta para que haja uma melhoria de seus serviços prestados.

Para trabalhos futuros, propomos considerar, no modelo multiestado com fragilidades aninhadas, que as variâncias das interações transição-grupo sejam diferentes, ou seja, cada interação  $W_{qh}$  tenha variância  $\theta_q$ . Além disso, propomos a estimação paramétrica dos parâmetros do modelo multiestado com fragilidades aninhadas.

---

## Referências Bibliográficas

---

- Aalen, O. O., Borgan, & Gjessing, H. K. (2008). Survival and event history analysis: a process point of view. Springer Science & Business Media. [2](#), [5](#)
- Andersen, P. & Keiding, N. (2002). Multi-state models for event history analysis. Statistical Methods in Medical Research, **11**, 91–115. [2](#), [13](#)
- Andersen, P., Gill, R., & Keiding, N. (1993). Statistical models based on counting process. Springer Series in Statistics. Springer, 2nd edition. [2](#), [5](#), [20](#)
- Bhattacharyya, M. & Klein, J. (2005). A random effects model for multistate survival analysis with application to bone marrow transplants. Mathematical biosciences, **194**, 37–48. [3](#)
- Bijwaard, G. (2014). Multistate event history analysis with frailty. Demography Research, **30**(58), 1591–1620. [3](#), [15](#)
- Clayton, D. (1978). A model for association in bivariate life tables and its application in epidomological studies of familial tendency in chronic disease incidence. Biometrika, **65**(1), 141–151. [2](#), [23](#), [25](#)
- Colosimo, E. & Giolo, S. (2006). Análise de sobrevivência aplicada. ABE - Projeto Fisher. [6](#), [7](#), [39](#)
- Commenges, D. (2002). Inference for multi-state models from interval-censored data. Statistical Methods in Medical Research, **11**(2), 167–182. [2](#)

- 
- Copelan, E., Biggs, J., Thompson, J., Crilley, P., Szer, J., Klein, J., Kapoor, N., Avalos, B., Cunningham, I., Atkinson, K., Downs, K., Harmon, G., Daly, M., Brodsky, I., Bulova, S., & Tutschka, P. (1991). Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with bu/cy. Blood, **78**, 838–843. [57](#)
- Courgeau, D. & Lelièvre, E. (1992). Event history analysis in demography. Clarendon Press. [2](#)
- Cox, D. (1972). Regression model and life-tables (with discussion). Journal of the Royal Statistical Society, **34**(2), 187–220. [8](#), [9](#), [23](#)
- Cox, D. (1975). Partial likelihood. Biometrika, **62**(2), 269–276. [39](#)
- de Castro, M., Chen, M., & Zhang, Y. (2015). Bayesian path specific frailty models for multi-state survival data with applications. Biometrics. [3](#)
- de Wreede, L., Fiocco, M., & Putter, H. (2010). The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. Computer Methods and Programs in Biomedicine, **99**(3), 261–74. [20](#), [55](#)
- de Wreede, L., Fiocco, M., & Putter, H. (2011). mstate: An r package for the analysis of competing risks and multi-state models. Journal of Statistical Software, **38**(7), 1–30. [15](#), [16](#), [55](#)
- dos Santos, C. (2010). Dados de sobrevivência multivariados na presença de covariáveis e observações censuradas: Uma abordagem bayesiana. Tese (doutorado em estatística), Universidade Federal de São Carlos, São Carlos, Brasil. [23](#)
- Duchateau, L. & Janssen, P. (2008). The frailty model. Springer Science & Business Media. [2](#), [23](#), [25](#), [28](#)
- Elbers, C. & Geert, R. (1982). True and spurious duration dependence: The identifiability of the proportional hazard model . The Review of Economic Studies, **49**(3), 403–409. [24](#)

- 
- Fine, J. & Gray, R. (1999). A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association, **94**(446), 496–509. [1](#)
- Flinn, C. & Heckman, J. (1983). Are unemployment and out of the labor force behaviorally distinct labor force states? Journal of Labor Economics, **1**(1), 28–42. [2](#)
- Foucher, Y., Saint-Pierre, P., Daures, J., & Durand, J. (2006). A semi-markov frailty model for multistate survival data: illustration on hiv disease. Far East Journal of Theoretical Statistics, **19**(2), 185–201. [3](#)
- Fougere, D. & Kamionka, T. (2008). Econometrics of individual labor market transitions. Springer. [2](#)
- Gray, R. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. The Annals of Statistics, **16**(3), 1141–1154. [1](#)
- Horny, G. (2009). Inference in mixed proportional hazard models with k random effects. Statistical Papers, **50**(3), 481–499. [27](#), [41](#)
- Hougaard, P. (1999). Multi-state model: A review. Lifetime Data Analysis, **5**, 239–264. [14](#)
- Hougaard, P. (2000). Analysis of multivariate survival data. Springer. , [2](#), [5](#), [11](#), [12](#), [15](#), [25](#), [27](#), [62](#)
- Kalbfleisch, J. & Prentice, R. (2002). The statistical analysis of failure time data. Series in Probability and Statistics. John Wiley & Sons, 2nd edition. [5](#), [20](#)
- Klein, J. & Moeschberger, M. (2003). Survival analysis: Techniques for censored and truncated data. Springer, 2nd edition. [5](#), [54](#), [55](#)
- Lawrence, L. (2003). Handbook of the life course, chapter Event History Models for Life Course Analysis, pages 477–502. Kluwer Academic/Plenum Publisher. [2](#)
- Liquet, B., Timsit, J., & Rondeau, V. (2012). Investigating hospital heterogeneity with a

- 
- multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. BMC Medical Research Methodology, **12**(1), 79. , [3](#), [29](#), [32](#)
- Ma, T., Joly, I., & Raux, C. (2010). A shared frailty semi-parametric markov renewal model for travel and activity time-use pattern analysis. [3](#), [32](#)
- Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. The Annals of Statistics, **26**(1), 183–214. [27](#)
- Petersen, J. (1998). An additive frailty model for correlated life times. Biometrics, **54**(2), 646–661. [27](#)
- Pickles, A. & Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. Statistics in Medicine, **14**(13), 1447–1461. [3](#)
- Pintilie, M. (2006). Competing risks: a practical perspective. Wiley. [1](#)
- Putter, H. (2011). Tutorial in biostatistics: Competing risks and multi-state models analyses using the mstate package. [16](#), [18](#), [21](#)
- Putter, H. & van Houwelingen, H. (2011). Frailties in multi-state models: Are they identifiable? do we need them? Statistical Methods in Medical Research, **0**(0), 1–18. [3](#)
- Putter, H., Fiocco, M., & Geskus, R. (2007). Tutorial in biostatistics: Competing risks and multi-state models. Statistics in Medicine, **26**, 2389–2430. [2](#), [14](#), [18](#), [21](#)
- R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. [18](#), [21](#), [55](#)
- Rocha, C. (1996). Modelos com fragilidade em análise de sobrevivência. Tese de doutorado, Universidade de Lisboa, Lisboa, Portugal. [23](#)
- Rogers, A. (1975). Introduction to multiregional mathematical demography. Wiley. [2](#)
- Rogers, A. (1995). Multiregional demography: Principles, methods and extensions. Wiley. [2](#)

- 
- Rondeau, V., Filleul, L., & Joly, P. (2006). Nested frailty models using maximum penalized likelihood estimation. Statistics in Medicine, **25**(23), 4036–4052. [27](#)
- Rotolo, F. (2013). Frailty multi-state models for the analysis of survival data from multicenter clinical trials. Tese (doutorado em ciências estatísticas), Università degli Studi di Padova, Pádua - Itália. [4](#), [13](#), [16](#), [18](#), [26](#), [31](#), [32](#)
- Rotolo, F. & Horny, G. (2012). mlfm: Semiparametric multilevel frailty models, estimated via empl method. In R package. [41](#), [43](#), [60](#)
- Sastry, N. (1997). A nested frailty model for survival data, with an application to the study of child survival in northeast brazil. Journal of the American Statistical Association, **92**(438), 426–435. [27](#)
- Shih, J. & Lu, S. (2009). Semiparametric estimation of a nested random effects model for the analysis of multi-level clustered failure time data. Computational Statistics & Data Analysis, **53**(11), 3854–3871. [27](#)
- Silva, G. (2001). Análise bayesiana de modelos de sobrevivência com fragilidade. Tese de doutorado, Universidade de Lisboa, Lisboa, Portugal. [23](#)
- Tomazella, V. (2003). Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade. Tese (doutorado em ciências da computação e matemática computacional), Universidade Federal de São Carlos, São Carlos, SP. [22](#), [23](#)
- Van den Berg, G. (2001). Duration models: specification, identification and multiple durations. Handbook of econometrics, **5**, 3381–3460. [2](#)
- Vaupel, J. & Yashin, A. (1985). The deviant dynamics of death in heterogeneous populations. Sociological Methodology, **15**, 179–211. [27](#)
- Vaupel, J., Manton, K., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography, **16**(3), 439–454. [2](#), [23](#)

Wienke, A. (2011). Frailty models in survival analysis. CRC Press. [2](#), [23](#), [24](#), [25](#), [27](#), [28](#)

Willekens, F. (1999). Population issues: An interdisciplinary focus, chapter The Life Course: Models and Analysis, pages 23–51. The Plenum Series on Demographic Methods and Population Analysis. Plenum Press. [2](#)

Willekens, F. (2014). Multistate analysis of life histories with R. Springer. [2](#)

Yau, K. (2001). Multilevel models for survival analysis with random effects. Biometrics, **57**(1), 96–102. [27](#)

Yen, A., Chen, T., Duffy, S., & Chen, C.-D. (2010). Incorporating frailty in a multi-state model: application to disease natural history modelling of adenoma-carcinoma in the large bowel. Statistical Methods in Medical Research, **19**(5), 529–546. [3](#), [32](#)