

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

**Modelos de fração de cura com fragilidade inflacionado de zero sob diferentes esquemas de ativação**

**Danillo Magalhães Xavier Assunção**

Tese de Doutorado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

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

**Danillo Magalhães Xavier Assunção**

## Modelos de fração de cura com fragilidade inflacionado de zero sob diferentes esquemas de ativação

Tese 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 Doutor em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientadora: Profa. Dra. Vera Lucia Damasceno Tomazella

Coorientador: Prof. Dr. Jeremias da Silva Leão

**USP – São Carlos**  
**Agosto de 2023**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados inseridos pelo(a) autor(a)

M188m Magalhães Xavier Assunção, Danillo  
Modelos de fração de cura com fragilidade  
inflacionado de zero sob diferentes esquemas de  
ativação / Danillo Magalhães Xavier Assunção;  
orientador Vera Lucia Damasceno Tomazella;  
coorientador Jeremias da Silva Leão. -- São Carlos,  
2023.  
139 p.

Tese (Doutorado - Programa Interinstitucional de  
Pós-graduação em Estatística) -- Instituto de Ciências  
Matemáticas e de Computação, Universidade de São  
Paulo, 2023.

1. Fração de cura. 2. Zero Inflacionado. 3.  
Análise de Sobrevivência. 4. Esquemas de ativação.  
5. Causas Competitivas. I. Tomazella, Vera Lucia  
Damasceno, orient. II. Leão, Jeremias da Silva,  
coorient. III. Título.

**Danillo Magalhães Xavier Assunção**

Frailty-inflated zero-adjusted cure fraction models under  
different activation schemes.

Doctoral dissertation submitted to the Institute of  
Mathematics and Computer Sciences – ICMC-USP  
and to the Department of Statistics – DEs-UFSCar, in  
partial fulfillment of the requirements for the degree of  
the Doctorate Interagency Program Graduate in  
Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Profa. Dra. Vera Lucia  
Damasceno Tomazella

Co-advisor: Prof. Dr. Jeremias da Silva Leão

**USP – São Carlos**

**August 2023**





# UNIVERSIDADE FEDERAL DE SÃO CARLOS

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

---

## Folha de Aprovação

---

Defesa de Tese de Doutorado do candidato Danillo Magalhães Xavier Assunção, realizada em 22/06/2023.

### Comissão Julgadora:

Profa. Dra. Vera Lucia Damasceno Tomazella (UFSCar)

Prof. Dr. Francisco Louzada Neto (USP)

Prof. Dr. Vinicius Fernando Calsavara (CIPE)

Profa. Dra. Gleici da Silva Castro Perdoná (USP)

Prof. Dr. Pedro Luiz Ramos (PUC-Chile)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.



# AGRADECIMENTOS

---

---

Queridos amigos, colegas e familiares,

Estou extremamente grato por ter completado mais uma grande conquista em minha vida. Esta conquista não teria sido possível sem a ajuda e o apoio de muitas pessoas ao longo desses 5 anos. Gostaria de aproveitar esta oportunidade para expressar minha sincera gratidão a todos que me apoiaram.

Primeiramente, gostaria de agradecer a minha orientadora Prof. Dra Vera Lucia Damasceno Tomazella pela oportunidade de trabalhar com ela e por ter acreditado em mim. Mas também agradeço por todo o apoio inabalável, paciência, suporte emocional nos momentos mais difíceis, pela amizade e orientação durante todo o processo de pesquisa. Sem sua ajuda, este trabalho não teria sido possível. Agradeço também ao meu co-orientador Prof. Dr. Jeremias Leão pelo incentivo a pesquisa, esforço e sugestões dadas durante o desenvolvimento do meu doutorado.

Agradeço também aos demais membros que participaram da banca por seus valiosos comentários e sugestões.

Agradeço a todos os colegas de pesquisa, em especial Alex de la Cruz, Gustavo Sabillon, Oilson Alberto, Fabiano Coelho que compartilharam suas ideias e conhecimentos comigo. Sou especialmente grato a Marina Gonzaga quem sempre me apoiou e me ensinou na disciplinas de Probabilidade Avançada e Inferência avançada. Agradeço a Mauro Oliveira Junior, Vinicius Cal-savara e Pedro Ramos por terem compartilhado conhecimentos e códigos para o desenvolvimento dessa tese.

Agradeço a toda minha família que sempre me apoiou em todas as fases do meu percurso acadêmico e me deu a força necessária para superar os obstáculos. Agradeço especialmente aos meus pais, Dalmy Assunção e Cristina Magalhães, por me ensinarem a importância da educação e me apoiarem incondicionalmente.

Agradeço também a Marina Gandolfi, Ana Carolina e Isaac pela amizade e carinho no decorrer do trabalho. Em especial, agradeço a Naiara Caroline em que compartilhamos momentos felizes além de momentos difíceis.

Por fim, o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001



*“O passado é história,  
o futuro é mistério,  
o agora é uma dádiva  
e por isso se chama  
presente.”  
(Kung-Fu-Panda)*



# RESUMO

ASSUNÇÃO, D. M. X. **Modelos de fração de cura com fragilidade inflacionado de zero sob diferentes esquemas de ativação**. 2023. 139 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Nesta tese de doutorado, a metodologia proposta se baseia em dados de sobrevivência inflacionados de zero para lidar com situações as quais existem uma fração de zeros inflacionados (ou ajustados) e de curados considerando diferentes esquemas de ativação. Nessa abordagem assumimos que a ocorrência do evento de interesse é originada por um estrutura latente de ativação: primeiro, último e aleatório, ou seja, permite que diferentes mecanismos de ativação concorrentes expliquem a ocorrência do fenômeno de interesse. Neste contexto, o novo modelo denominado modelo de fração de cura inflacionado de zero sob diferentes esquemas de ativação é uma extensão do modelo inflacionado de zeros proposto por [Jr, Moreira e Louzada \(2017\)](#); do modelo proposto por [Roman \*et al.\* \(2013\)](#) e do modelo de tempo de promoção proposto por [Yakovlev \(1996\)](#) e [Chen, Ibrahim e Sinha \(1999\)](#), pois incorporamos a inflação de zeros e os esquemas de ativações na modelagem. Para as estimativas dos parâmetros do modelo com longa duração e inflação de zeros, nós consideramos as abordagens clássica e Bayesiana. Os modelos de fração de cura ou modelos de fração de cura zero ajustados assumem indiretamente que todos os pacientes expostos ao evento de interesse têm risco homogêneo, porém, a existência de uma heterogeneidade pode ser medida através da inclusão de covariáveis. Sendo assim, é possível medir uma parcela dessa heterogeneidade pelas covariáveis, contudo tem-se a presença de um grau de heterogeneidade induzida por causas não observáveis. Os modelos que incluem essa heterogeneidade não observável entre os sujeitos são conhecidos como modelos de fragilidade. Neste contexto, o termo de fragilidade é incorporado na função de risco da modelagem proposta para controlar a heterogeneidade não observável dos pacientes em que assumimos uma distribuição gama para a variável de fragilidade. Estudos de simulação são realizados e também aplicações de dados reais.

**Palavras-chave:** Fração de Cura, Inflação de zeros, Esquemas de ativação, Análise de sobrevivência, Fragilidade.



# ABSTRACT

ASSUNÇÃO, D. M. X. **Frailty-inflated zero-adjusted cure fraction models under different activation schemes.** 2023. 139 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

In this doctoral thesis, the proposed methodology is based on zero-inflated survival data to deal with situations where there is a fraction of inflated (or adjusted) zeros and cured cases considering different activation schemes. In this approach, we assume that the occurrence of the event of interest is generated by a latent activation structure: first, last, and random, allowing different competing activation mechanisms to explain the occurrence of the phenomenon of interest. In this context, the new model called the zero-inflated cure rate model under different activation schemes is an extension of the zero-inflated model proposed by [Jr, Moreira e Louzada \(2017\)](#); the model proposed by [Roman \*et al.\* \(2013\)](#) the promotion time model proposed by [Yakovlev \(1996\)](#) e [Chen, Ibrahim e Sinha \(1999\)](#), as we incorporate zero-inflation and activation schemes in the modeling. For the estimation of the parameters of the model with long duration and zero-inflation, we consider classical and Bayesian approaches. The cure rate models or zero cure rate models indirectly assume that all patients exposed to the event of interest have homogeneous risk, but the existence of heterogeneity can be measured through the inclusion of covariates. Thus, it is possible to measure a portion of this heterogeneity by covariates, but there is a degree of heterogeneity induced by unobserved causes. Models that include this unobserved heterogeneity among subjects are known as frailty models. In this context, the frailty term is incorporated into the risk function of the proposed modeling to control the unobserved heterogeneity of patients, and we assume a gamma distribution for the frailty variable. Simulation studies are performed, as well as applications to real data.

**Keywords:** cure rate, inflated zero model, activation schemes, survival analysis.



# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Função de Sobrevivência do modelo de taxa de cura inflacionado no zero . . .	33
Figura 2 – Função de distribuição acumulada (imprópria) (FDA) do modelo de fração de cura inflacionado por zero. . . . .	34
Figura 3 – Curvas de sobrevivência de KM dos dados simulados baseado nos valores de parâmetros do cenário 1. . . . .	51
Figura 4 – Curvas de sobrevivência de KM dos dados simulados baseado nos valores de parâmetros do cenário 2. . . . .	52
Figura 5 – Curvas de sobrevivência de KM dos dados simulados baseado nos valores de parâmetros do cenário 3. . . . .	52
Figura 6 – Viés, REQM e probabilidade de cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação máximo . . . . .	54
Figura 7 – Viés, REQM e probabilidade de cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação aleatório . . . . .	55
Figura 8 – Viés, REQM e probabilidade de cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação mínimo . . . . .	56
Figura 9 – Curva de sobrevida estimada obtida pelo método de Kaplan-Meier para dados de insulina . . . . .	62
Figura 10 – Curva de sobrevida estimada obtida por Kaplan-Meier para dados de insulina considerando as variáveis história prévia de GDM, história de tabagismo, macrossomia fetal prévia, hipertensão crônica anterior, IMC pré-gravidez e história familiar de diabetes, e estimativas das respectivas funções de sobrevivência de acordo com ajuste do modelo ZICR sob diferentes de mecanismos de ativação latente . . . . .	65
Figura 11 – Curva de sobrevida estimada obtida pelo método de Kaplan-Meier para dados hemorragia intracraniana . . . . .	72

Figura 12 – Curva de sobrevida estimada obtida por Kaplan-Meier para dados de Hemorragia intracraniana considerando as variáveis e as respectivas estimativas da função de sobrevivência de acordo com ajuste do modelo ZICR sob diferentes de mecanismos de ativações latente . . . . .	73
Figura 13 – Abordagem Bayesiana para estimação dos parâmetros para o Viés, REQM e cobertura empírica para intervalo de credibilidade com 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação máximo . . . . .	84
Figura 14 – Abordagem Bayesiana para estimação dos parâmetros - Viés, REQM e cobertura empírica para intervalos de credibilidade com 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação aleatório . . . . .	85
Figura 15 – Abordagem Bayesiana para estimação dos parâmetros - Viés, REQM e cobertura empírica para intervalos de credibilidade com 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação mínimo	86
Figura 16 – Curva de Sobrevivência estimada pelo modelo ZICR sob os três esquemas de ativação com abordagem bayesiana para a variável Microsomia anterior. . .	90
Figura 17 – Curva de Sobrevivência estimada pelo modelo ZICR sob os três esquemas de ativação com abordagem bayesiana para a variável Histórico tabagismo. . .	91
Figura 18 – Curva de Sobrevivência estimada pelo modelo ZICR sob os três esquemas de ativação com abordagem bayesiana para a variável Histórico Diabetes na família. . . . .	91
Figura 19 – Exemplo de funções não condicional de Sobrevivência (b) e risco (a) para o modelo gamma ZICR-F para diferentes valores de fragilidade ( $\alpha$ ) considerando a distribuição Log-normal para o tempo. . . . .	99
Figura 20 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 1 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo. . . . .	107
Figura 21 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 1 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação aleatório. . . . .	108

Figura 22 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 1 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação máximo. . . . .	109
Figura 23 – Curva de sobrevida estimada obtida por Kaplan-Meier para dados de Insulina e as respectivas estimativas da função de sobrevivência de acordo com ajuste do modelo ZICR-F sob diferentes de mecanismos de ativações latente. . . .	110
Figura 24 – Gráfico QQ plot dos resíduos quantílicos randomizados do modelo final sob esquema de ativação final . . . . .	112
Figura 25 – Curva de sobrevida estimada obtida por Kaplan-Meier para dados de Hemorragia intracraniana considerando as variáveis e as respectivas estimativas da função de sobrevivência de acordo com ajuste do modelo ZICR-F sob diferentes de mecanismos de ativações latente . . . . .	113
Figura 26 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 2 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo . . . . .	128
Figura 27 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 3 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo . . . . .	129
Figura 28 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 2 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo . . . . .	130
Figura 29 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 3 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo . . . . .	131
Figura 30 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 2 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo . . . . .	132
Figura 31 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 3 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo . . . . .	133
Figura 32 – Gráfico traço de média para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo . . .	134

Figura 33 – Gráfico Histórico das Cadeias para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo .	134
Figura 34 – Gráfico traço de média para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação minimo . .	134
Figura 35 – Gráfico Histórico das Cadeias para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação minimo .	135
Figura 36 – Gráfico traço de média para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório . .	135
Figura 37 – Gráfico Histórico das Cadeias para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório .	135
Figura 38 – Gráfico traço de média para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo . .	136
Figura 39 – Gráfico Histórico das Cadeias para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo .	136
Figura 40 – Gráfico traço de média para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação minimo . .	136
Figura 41 – Gráfico Histórico das Cadeias para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação minimo .	137
Figura 42 – Gráfico traço de média para a variável histórico diabetes modelo ZICR com abordagem bayesiana e esquema de ativação aleatório . . . . .	137
Figura 43 – Gráfico Histórico das Cadeias para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório .	137
Figura 44 – Gráfico traço de média para a variável Macrossomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo . .	138
Figura 45 – Gráfico Histórico das Cadeias para Macrossomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo . . . . .	138
Figura 46 – Gráfico traço de média para a variável Macrossomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação minimo . .	138
Figura 47 – Gráfico Histórico das Cadeias para Macrossomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação minimo . . . . .	139
Figura 48 – Gráfico traço de média para a variável Macrossomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório . .	139
Figura 49 – Gráfico Histórico das Cadeias para Macrossomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório . . . . .	139

# LISTA DE TABELAS

---

---

Tabela 1	– Função de sobrevivência $S_p(t)$ , função de densidade $f_p(t)$ e fração de cura considerando algumas distribuições referente ao número de causas latentes, $M$ .	30
Tabela 2	– Proporção de curados, $p_1$ ; Funções de sobrevivência da parcela suscetíveis ao evento $S_F^*(t)$ e função de sobrevivência da população suscetíveis e não suscetíveis $S_p(t)$ considerando o esquema de ativação mínimo. . . . .	39
Tabela 3	– Proporção de curados, $p_1$ ; Funções de sobrevivência da parcela suscetíveis ao evento $S_{max}^*$ e função de sobrevivência da população suscetíveis e não suscetíveis $S_{p_{max}}(t)$ considerando o esquema de ativação máximo. . . . .	40
Tabela 4	– Função de sobrevivência $S_p^*(t)$ e função densidade $f_p^*(t)$ para os esquemas de ativação. . . . .	48
Tabela 5	– Valor dos parâmetros sob diferentes cenários . . . . .	51
Tabela 6	– % de caso com o menor AIC e esquema de ativação mínimo como modelo verdadeiro . . . . .	59
Tabela 7	– % de casos com menor AIC e esquema de ativação máximo como modelo verdadeiro . . . . .	60
Tabela 8	– % de casos com menor AIC e esquema de ativação aleatório como modelo verdadeiro . . . . .	61
Tabela 9	– Frequência observada das variáveis presente no banco de dados da insulina .	63
Tabela 10	– EMV, EP, IC com 95%, AIC e BIC das variáveis Família diabetes Macrosonomia prévia obtidos por ZICR sob diferentes mecanismos de ativação considerando os dados de insulina . . . . .	66
Tabela 11	– EMV, EP, IC com 95%, AIC e BIC das variáveis Hipertensão prévia e fumante obtidos por ZICR sob diferentes mecanismos de ativação considerando os dados de insulina . . . . .	67
Tabela 12	– EMV, EP, IC com 95% e AIC, BIC das variáveis primigiesta, IMC pré-gravides obtidos por ZICR sob diferentes mecanismos de ativação considerando os dados de insulina . . . . .	68
Tabela 13	– EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR sob diferentes mecanismos de ativação considerando conjunto de dados de insulina . . . . .	69
Tabela 14	– EMV, EP, IC com 95%, AIC e BIC obtidos pelo modelo ZICR sob diferentes mecanismos de ativação ajustados para o conjunto de dados insulina . . . . .	70
Tabela 15	– Frequência observada das variáveis presente no banco de dados dos pacientes com câncer e hemorragia intracraniana . . . . .	71

Tabela 16 – EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR das variáveis Gênero, e Metástase sob diferentes mecanismos de ativação considerando os dados de HI.	74
Tabela 17 – EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR das variáveis Not CNS nor Meta e Others Primary Site sob diferentes mecanismos de ativação considerando os dados de HI.	75
Tabela 18 – EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR das variáveis CSNS Metastasi, CSNS Primary Site e Idade sob diferentes mecanismos de ativação considerando os dados de HI.	76
Tabela 19 – Valor dos parâmetros sob diferentes cenários	82
Tabela 20 – Resumos a <i>posteriori</i> dos parâmetros do modelo ZICR sob os mecanismos de ativação para a variável histórico diabetes	88
Tabela 21 – Resumos a <i>posteriori</i> dos parâmetros do modelo ZICR sob os mecanismos de ativação para a variável macrosomia a priori	89
Tabela 22 – Resumos a <i>posteriori</i> dos parâmetros do modelo ZICR sob os mecanismos de ativação para a variável histórico fumante	90
Tabela 23 – Função de sobrevivência populacional de T e densidade para cada cenário do modelo ZICR	100
Tabela 24 – Função de sobrevivência populacional com fragilidade (ZICR-F) gama para cada esquema de ativação	100
Tabela 25 – Função de densidade com fragilidade gama para cada esquema de ativação	101
Tabela 26 – Função de Sobrevivência para os indivíduos não curados ( $S^*(t)$ ) do modelo ZICR	101
Tabela 27 – Expressões para $S_p^*(t)$ para o modelo ZICR-F considerando cada esquema de ativação	102
Tabela 28 – Expressões para $f_p^*(t)$ do modelo ZICR-F considerando cada esquema de ativação	102
Tabela 29 – Valor dos parâmetros sob diferentes cenário	105
Tabela 30 – EMV, EP, IC com 95%, AIC e BIC obtidos pelo modelo ZICR-F sob diferentes mecanismos de ativação ajustados para o conjunto de dados insulina	111
Tabela 31 – EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR das variáveis CSNS Metastasi, CSNS Primary Site e Idade sob diferentes mecanismos de ativação considerando os dados de HI.	114

# SUMÁRIO

---

---

1	INTRODUÇÃO . . . . .	21
1.1	Objetivo . . . . .	26
1.2	Organização do trabalho . . . . .	26
2	REFERENCIAL TEÓRICO . . . . .	27
2.1	Modelo de Fração de Cura . . . . .	27
2.2	Modelo de fração de cura zero inflacionado . . . . .	31
2.2.1	<i>Modelo de regressão de fração de cura zero inflacionado</i> . . . . .	34
2.3	Esquemas de ativação . . . . .	35
2.3.1	<i>Ativação pelo tempo mínimo</i> . . . . .	38
2.3.2	<i>Ativação pelo tempo máximo</i> . . . . .	39
2.3.3	<i>Ativação aleatória</i> . . . . .	40
2.4	Conclusão . . . . .	41
3	MODELOS DE FRAÇÃO DE CURA ZERO INFLACIONADO CON- SIDERANDO DIFERENTES ESQUEMAS DE ATIVAÇÃO . . . . .	43
3.1	Introdução . . . . .	43
3.2	Modelos de fração de cura no cenário $FR_{max}$ . . . . .	44
3.3	Modelos de fração de cura no cenário $FR_{rand}$ . . . . .	45
3.4	Modelos de fração de cura no cenário $FR_{min}$ . . . . .	46
3.5	Modelo de fração de cura zero inflacionado na presença de covariável . . . . .	47
3.5.1	<i>Inferência</i> . . . . .	48
3.6	Estudos de simulação . . . . .	49
3.6.1	<i>Propriedades assintóticas</i> . . . . .	49
3.6.2	<i>Especificação incorreta do modelo</i> . . . . .	56
3.7	Aplicações para dados reais . . . . .	61
3.7.1	<i>Dados sobre insulina</i> . . . . .	61
3.7.2	<i>Hemorragia intracraniana em pacientes com câncer</i> . . . . .	70
3.8	Conclusão . . . . .	77
4	ABORDAGEM BAYESIANA DO MODELO ZICR SOB DIFEREN- TES ESQUEMAS DE ATIVAÇÃO . . . . .	79
4.1	Introdução . . . . .	79
4.2	Distribuições <i>a priori</i> e <i>a posteriori</i> . . . . .	80

4.3	Estudos de Simulação . . . . .	82
4.4	Aplicação de dados sobre insulina . . . . .	87
4.5	Conclusão . . . . .	92
5	<b>MODELOS DE FRAÇÃO DE CURA ZERO INFLACIONADO SOB DIFERENTES ESQUEMAS DE ATIVAÇÃO COM FRAGILIDADE .</b>	<b>93</b>
5.1	Introdução . . . . .	93
5.2	Modelo de fragilidade . . . . .	94
5.3	Modelo de fração de cura zero ajustado com fragilidade gama . . . . .	97
5.3.1	<i>Modelo ZICR-F log-normal com Fragilidade gama</i> . . . . .	<i>98</i>
5.4	Especificação do Modelo . . . . .	99
5.5	Estudo de simulação . . . . .	104
5.6	Aplicação em dados reais . . . . .	109
5.6.1	<i>Insulina Data set</i> . . . . .	<i>110</i>
5.6.2	<i>Hemorragia intracraniana</i> . . . . .	<i>112</i>
5.7	Considerações finais . . . . .	115
6	<b>CONCLUSÕES . . . . .</b>	<b>117</b>
6.1	Observações finais . . . . .	117
6.2	Propostas futuras . . . . .	118
	<b>REFERÊNCIAS . . . . .</b>	<b>119</b>
<b>ANEXO A</b>	<b>FIGURAS . . . . .</b>	<b>128</b>
A.1	Figuras - Capítulo 3 . . . . .	128
A.2	Figuras - Aplicação: Capítulo 4 . . . . .	134
A.2.1	<i>Variável Histórico Fumante</i> . . . . .	<i>134</i>
A.2.2	<i>Variável Histórico Familiar Diabetes</i> . . . . .	<i>136</i>
A.2.3	<i>Variável Macrosomia fetal</i> . . . . .	<i>138</i>

---

# INTRODUÇÃO

---

A análise de dados de tempo de vida foca no estudo do tempo até a ocorrência de um evento, conhecido como tempo de falha, e é relevante em economia, educação, medicina, epidemiologia, engenharia, demografia e outras áreas. Estudos realizados por diversos pesquisadores têm contribuído nessa área, incluindo os trabalhos de [Giovannetti, Ricchiuti e Velucchi \(2011\)](#), [Kaminski e Geisler \(2012\)](#), [Strouse \*et al.\* \(2005\)](#), [Nunes e Sarmiento \(2012\)](#), [Kyriakopoulos \*et al.\* \(2018\)](#), entre outros. Na análise de sobrevivência, a principal característica é a possibilidade da presença de observações incompletas, que são denominados de censuras ou truncamento em que cada uma destas possuem características particulares.

Nesse contexto, duas funções principais são utilizadas: a função de sobrevivência e a de risco. A primeira é denotada por  $S_0(t)$  a qual representa a probabilidade de um indivíduo sobreviver acima de um certo tempo  $t$  enquanto que a segunda, chamada de  $h_0(t)$ , representa a taxa instantânea de ocorrência do evento ao longo do tempo, dado que o paciente sobreviveu até o tempo  $t$ . A função  $S_0(t)$  é não decrescente, ou seja, a probabilidade de que um indivíduo não apresente o evento até o tempo atual é sempre maior ou igual a essa probabilidade em tempos posteriores. O estimador de Kaplan-Meier, proposto por [Kaplan e Meier \(1958\)](#), é um método não-paramétrico amplamente utilizado na literatura para estimação da curva de sobrevivência.

O modelo de riscos proporcionais, proposto por [Cox \(1972\)](#), é outra abordagem amplamente utilizada na análise de sobrevivência. Esse modelo permite avaliar a associação entre variáveis explicativas e o tempo para um evento, além de lidar de forma eficiente com desafios técnicos como censura e truncamento. Essa metodologia é comumente utilizada devido à facilidade no tratamento de dificuldades técnicas, permitindo a inclusão de covariáveis para descrever suas respectivas influências e modelar diferentes níveis de risco para diferentes subgrupos. Exemplos de estudos utilizando essa abordagem podem ser encontrados em [Lane, Looney e Wansley \(1986\)](#), [Fisher e Lin \(1999\)](#), [Fan e Li \(2002\)](#), [Bender, Augustin e Blettner \(2005\)](#), [Benítez-Parejo, Águila e Pérez-Vicente \(2011\)](#), entre outros. A interpretação atraente da taxa de risco como

um risco que muda com o tempo torna esse modelo uma ferramenta poderosa em análises de sobrevivência.

Por outro lado, em outras áreas, principalmente na de saúde, é comum existir eventos de interesse os quais os indivíduos em estudo não sejam suscetíveis, essa parcela é chamada de imunes ou curados. Por causa desses dois grupos observados, curados e não curados, é considerar uma modelagem que capte tal característica. Deste modo, considera-se para a construção de um modelo a mistura de duas distribuições, em que uma caracteriza a distribuição dos tempos da ocorrência do evento de interesse, ou seja, representando a sobrevivência dos não curados enquanto que a outra relaciona a uma distribuição degenerada que admita tempos de sobrevida para pacientes não suscetíveis a falha (MALLER; ZHOU, 1996).

A partir dos dados coletados em ensaios clínicos, foram desenvolvidos e incorporados na literatura modelos que consideram os indivíduos que não são suscetíveis ao fenômeno de interesse. Estes podem ser usados para modelar dados cuja a variável de interesse são os tempos até o evento para vários tipos de cânceres para os quais uma parcela significativa de pacientes pode vir a ser considerada “curada”, incluindo câncer de mama, linfomas, leucemia, câncer de próstata, melanoma e entre outros (CHEN; IBRAHIM; SINHA, 1999). A modelagem utilizada para esse tipo de dados é denominada modelos de fração de cura, em que o mais conhecido é o modelo de mistura padrão, proposto por Boag (1949) e Berkson e Gage (1952). Eles incorporaram a fração de cura para lidar com situações as quais os indivíduos têm imunidade ou se curam. Yakovlev *et al.* (1993), Tsodikov, Yakovlev e Asselain (1996) e Yakovlev (1996) propuseram outra possibilidade para o modelo de taxa de cura, que é chamado de modelo de cura de promoção e foi motivado por uma doença específica (célula tumoral). A abordagem clássica tem sido amplamente utilizada para estimação de parâmetros em modelos com fração de cura. Dentre os estudos que adotaram essa abordagem, podemos citar Goldman (1984); Greenhouse e Wolfe (1984); Gray e Tsiatis (1989); Kuk e Chen (1992); Ewell e Ibrahim (1997); Yin e Ibrahim (2005); Masud, Tu e Yu (2018); Milani *et al.* (2021).

A abordagem bayesiana é uma alternativa (a clássica) que também tem sido amplamente utilizada na modelagem de sobrevivência com fração de cura. Essa abordagem permite incorporar informações prévias dos parâmetros do modelo e realizar inferências probabilísticas mais robustas. Ambas as abordagens apresentam vantagens e limitações, sendo importante escolher a abordagem mais adequada para cada contexto de pesquisa. Chen, Ibrahim e Sinha (1999) realizaram um estudo com modelo de fração de cura com abordagem bayesiana. Outros estudos utilizando essa modelagem também são encontradas na literatura como, por exemplo: Ibrahim *et al.* (2001).

Nesse contexto, os autores consideram causas competitivas para a ocorrência do evento de interesse. Porém em certas situações, quando se está investigando as circunstâncias que influenciam a ocorrência do evento de interesse, é possível que exista mais de uma causa competindo para a ocorrência do evento o qual estamos interessados. Considere, por exemplo,

um paciente com câncer de próstata que morre devido a ocorrência de algum outro tipo de câncer ou, ainda, por uma causa distinta que não esteja relacionada à doença. Essas outras possíveis razões são nomeadas como eventos competitivos os quais são causas latentes (não observáveis). Como mencionado em [Aalen \(1978\)](#), essa abordagem foi estudada inicialmente nas áreas de ciências atuariais, demografia e estatísticas médicas por [Nelson \(1969\)](#), [Hoem \(1971\)](#), [Peterson \(1975\)](#), [Aalen \(1976\)](#) e [Altshuler \(1970\)](#). Eles desenvolveram uma análise considerando as causas competitivas.

Outros autores desenvolveram trabalhos utilizando modelos de fração de cura com causas (ou riscos) competitivas, por exemplo, [Rodrigues et al. \(2009\)](#) propuseram uma extensão do modelo proposto por [Chen, Ibrahim e Sinha \(1999\)](#) através da função geradora de uma sequência real, chamado de modelo unificado de fração de cura. [Leão et al. \(2020\)](#) realizaram um estudo considerando um modelo de fração de cura em que as causas competitivas seguia uma distribuição geométrica modificada no zero. Outros exemplos podem ser visto em [Beyersmann et al. \(2009\)](#), [Nascimento et al. \(2019\)](#) e [Viana et al. \(2019\)](#). [Delgado \(2014\)](#) estudou modelos de fração de cura considerando os três mecanismos de ativação em que o número de fatores latentes seguia distribuição de série de potências. Essas abordagens são aplicadas com sucesso em problemas onde o evento de interesse não é exposto a tempo de sobrevivência zero, ou seja, no tempo zero, a função de sobrevivência é igual a um.

Entretanto, em algumas situações, a ocorrência do evento de interesse nos indivíduos está sujeita a acontecer no início do estudo, ou seja, os tempos de sobrevivência seriam iguais a zero. Por exemplo, iniciar um determinado medicamento para aumentar o número de plaquetas sanguíneas ou medidas educativas para reduzir o consumo excessivo de sal na dieta de um paciente. Uma vez adotada a medida, será necessário um certo período de tempo até que o número de plaquetas se normalize, reduzindo assim a ocorrência de eventos hemorrágicos, ou até que a redução da ingestão de sal reflita uma queda na pressão arterial de uma população. Antes desse período de tempo, a incidência do evento não reflete a realidade do efeito da medida adotada. O conceito de zero ajustado é utilizado para encontrar o momento adequado no tempo a partir do qual podemos observar e justificar a ocorrência de eventos como consequência da medida adotada. No entanto, em muitos casos, os modelos usuais não se ajustam bem a essa situação, o que leva os pesquisadores a descartarem essas informações na prática

A ocorrência de zeros em uma amostra pode ser explicada por diversos aspectos, considerando por exemplo o trabalho de [Martin et al. \(2005\)](#), os autores comentam sobre quatro situações: os valores zeros “verdadeiros” associados a baixa frequência de ocorrência do evento ou a inexistência do evento, os zeros aleatórios, decorrente de erro de amostragem ou ainda os zeros falsos, decorrente de erro humano. Então, o surgimento de dados com zeros nos dados de sobrevivência é frequentemente encontrado em aplicações reais.

Particularmente, modelos de sobrevivência que permitem a modelagem de tempos iguais a zero ainda são escassos. Mais recentemente, [Jr, Moreira e Louzada \(2017\)](#); [Louzada, Moreira e](#)

[Oliveira \(2018\)](#) propuseram modelo de sobrevivência inflacionados de zero aplicados em dados financeiros. Seu objetivo principal era investigar os fatores de risco associados à ocorrência de empréstimos fraudulentos. Nesses estudos, a distribuição Weibull foi adotada para analisar o tempo decorrido até a ocorrência do evento de interesse. Por outro lado [Souza et al. \(2022\)](#) optaram por utilizar a distribuição Log-normal para estudar o tempo de gestação de mulheres na África. Esse trabalho contribuiu para uma melhor compreensão dos fatores que influenciam a duração da gestação em populações específicas. [Calsavara et al. \(2019b\)](#) propuseram modelos defeituosos com presença de tempos ajustados a zero para o tempo decorrido até a oclusão em pacientes com câncer de pâncreas e o tempo decorrido até o início do uso da insulina em mulheres com diabetes gestacional.

Dessa forma, ao investigar as possíveis causas para a ocorrência de um evento de interesse, como a morte do paciente ou a recidiva de uma doença, é possível que as causas sejam provenientes de diferentes formas,. Estas são consideradas latentes devido ao fato de não haver qualquer informação sobre a origem das causas, isto é, quantas são ou foram ativadas, ou dentre elas, qual(is) foi(ram) a(s) responsável(is) pela surgimento do evento. Ainda no exemplo de um tumor, a recidiva dele pode estar associado as células cancerígenas que se separam do tumor principal e entram na corrente sanguínea, este processo, segundo a [Oncology et al. \(2008\)](#), é chamado de metástase. Ainda segundo a [Oncology et al. \(2008\)](#), estas células podem ou não formar novos tumores, isto é, não se tem a certeza se foram realmente elas as causas da origem da recorrência do tumor. E portanto temos um cenário de múltiplas causas ou riscos.

Nesse contexto, é difícil distinguir qual dos cenários originou o fenômeno de interesse. [Minayo \(1988\)](#) afirma que a etiologia das doenças indica que não existe uma única causa para a doença e também observa que é consideravelmente difícil definir qual é a causa mais importante ou quais são irrelevantes. Esse fato gera controvérsias na comunidade científica, e há grandes questionamentos, por exemplo, sobre os agentes etiológicos envolvidos no aparecimento da maioria das neoplasias malignas. Enquanto as ciências médicas caminham para desvendar as possíveis causas das doenças, a modelagem estatística precisa inserir essas informações nos modelos estatísticos propostos, acomodando essas novas situações. Quando a possibilidade de ocorrência do evento é com ativação do primeiro fator de risco, nos referimos ao tempo mínimo de ocorrência de cada um dos fatores de risco envolvidos. Outros autores consideraram esta suposição, incluindo: [Adamidis e Loukas \(1998\)](#), [Louzada-Neto \(1999\)](#), [Lu e Tsiatis \(2001\)](#) e [Lu e Tsiatis \(2005\)](#). Uma outra possibilidade é considerar que o evento de interesse é causado pela ativação do último fator de risco envolvido. Assim, todos os fatores de risco devem ser ativados para que a falha aconteça, neste caso assume-se que o tempo observado se refere ao máximo dos tempos entre os tempos. Considerando esta abordagem [Kuş \(2007\)](#) e [Cancho, Louzada-Neto e Barriga \(2011\)](#) são exemplos. No entanto, ainda persiste a possibilidade de o evento de interesse ter sido desencadeado por uma situação intermediária às apresentadas, conforme discutido por [Cooner, Banerjee e McBean \(2006\)](#). Neste caso, o tempo observado refere-se ao tempo entre o mínimo e o máximo, que é chamado de ativação aleatória.

Outros estudos considerando os esquemas de ativação foram realizados por diversos autores, como por exemplo: [Cooner, Banerjee e McBean \(2006\)](#), em que eles modelaram dados de sobrevivência geograficamente referenciados (incorporaram informações geográficas dos pacientes) com fração de cura sob diferentes esquemas de ativação, nesse caso a primeira e a última. [Cooner et al. \(2007\)](#) propuseram uma classe de modelos hierárquicos em que consideraram uma modelagem estocástica do processo de manifestação da doença sob diferentes esquemas de ativação (primeira e última). [Cancho et al. \(2013\)](#) estudaram um modelo de fração de cura considerando os esquemas de ativação aleatório, primeiro e último. Eles aplicaram a casos em que o evento de interesse pode conter um processo destrutivo dos fatores de risco iniciais, nesse caso mostra-se a parte que está danificada do número inicial de fatores de risco. Existem outros trabalhos relevantes na literatura sobre o tema, que foram desenvolvidos por autores renomados, tais como: [Yiqi et al. \(2011\)](#) em que eles realizaram um estudo sobre erros de especificação considerando o modelo Weibull Binomial negativa na presença de dois esquemas de ativação, a primeira e última e [Roman et al. \(2013\)](#) realizaram um estudo considerando o modelo de fração de cura sob diferentes esquemas de ativação.

Nos últimos anos, as propostas relativas aos modelos de longa duração têm como objetivo abranger outras características de heterogeneidade dos dados. Os modelos de fração de cura assumem indiretamente que todas os pacientes suscetíveis ao evento de interesse têm risco homogêneo, porém, a existência de uma heterogeneidade pode ser medida através da inclusão de covariáveis. Sendo assim, é possível medir uma parcela dessa heterogeneidade pelas covariáveis, contudo há a presença de um grau de heterogeneidade induzida por causas não observáveis. Os modelos que incluem essa heterogeneidade não observável entre os sujeitos são conhecidos como modelos de fragilidade. A modelagem com fragilidade em análise de sobrevivência, foi introduzida por [Vaupel, Manton e Stallard \(1979\)](#) em que eles estudaram modelos de sobrevivência univariado. Posteriormente, [Clayton \(1978\)](#) e [Oakes \(1982\)](#) propuseram o primeiro modelo de fragilidade para dados multivariados. Outros autores, como [Rondeau et al. \(2013\)](#) e [Gonzales, Tomazella e Taconelli \(2013\)](#), apresentam propostas que conjugam informações de fragilidade e covariáveis. [Leão et al. \(2018\)](#) incorporaram essa heterogeneidade, através da fragilidade, em um modelo de fração de cura com base na distribuição de Birnbaum - Saunders. [Morita et al. \(2020\)](#) realizaram um estudo incluindo essa heterogeneidade na abordagem bayesiana enquanto que [Cancho et al. \(2019\)](#) realizaram um estudo considerando fragilidade discreta [Souza \(2020\)](#).

O aparecimento de dados com zeros no tempo de sobrevida é frequentemente encontrado em aplicações reais, porém, existem poucos estudos na literatura considerando dados de sobrevida inflacionados a zero na presença de taxa de cura com fragilidade. [Souza et al. \(2022\)](#) considerou um estudo nessa temática com apenas o esquema de ativação mínima.

## 1.1 Objetivo

O objetivo principal deste trabalho de tese é incorporar diferentes esquemas de ativação (mínima, aleatória e máxima) no modelo de fração de cura zero inflacionado proposto por [Jr, Moreira e Louzada \(2017\)](#). Assumimos que as causas competitivas  $M$  segue uma distribuição Poisson e o tempo de promoção de cada causa segue uma distribuição Weibull.

Propomos uma extensão do modelo de taxa de cura com inflação zero, a fim de permitir a estimativa da heterogeneidade devido ao presença de variáveis não observadas sob diferentes esquemas de ativação. Supomos distribuição Log-Normal para o tempo de vida dos pacientes em risco e fragilidade gama. Adicionalmente, supomos que as causas competitivas segue uma distribuição Poisson.

Outro objetivo é o desenvolvimento de processos inferenciais nas perspectivas clássica e Bayesiana. Realizamos estudos de simulação de Monte Carlo para avaliar as propriedades dos estimadores dos parâmetros do modelo proposto.

Os modelos propostos são fundamentados pela análise de dois conjuntos de dados reais: um relacionado a mulheres no período gestacional que foram diagnosticadas com diabetes mellitus gestacional (DMG) antes de 24 semanas de gestação e que realizaram o pré-natal no Hospital das Clínicas da Universidade de São Paulo, Brasil, entre 2012 e 2015. O segundo sobre um estudo retrospectivo de todos os procedimentos cirúrgicos para o tratamento de Hemorragia intracraniana (em inglês ICH) em pacientes com câncer admitidos a partir de 2008 no Instituto do Câncer do Estado de São Paulo – Universidade de São Paulo, Brasil(ICESP).

## 1.2 Organização do trabalho

O restante do trabalho está organizado da seguinte forma: no Capítulo 2, realizamos uma revisão bibliográfica além de apresentação dos conceitos introdutórios referentes ao modelo de tempo de promoção e esquemas de ativação. No Capítulo 3, formulamos nosso modelo proposto, apresentamos a abordagem para estimativa de parâmetros, realizamos um estudo baseado em simulações de Monte Carlo com uma variedade de parâmetros além de fazer uma aplicação para um conjunto de dados reais.

No Capítulo 4, apresentamos uma abordagem Bayesiana para os modelos de fração de cura zero inflacionados sob diferentes esquemas de ativação obtidos no Capítulo 3. No Capítulo 5, apresentamos o modelo de fração de cura zero inflacionado incorporando um termo aleatório não observável, chamado de fragilidade, na função de risco. Adicionalmente, um estudo de simulação foi realizado para verificar o comportamento dos estimadores do modelo proposto. Uma aplicação em dados reais foi utilizada com o objetivo de mostrar a aplicabilidade do modelo. Por último, no Capítulo 6, apresentamos as considerações finais e as propostas futuras.

---

## REFERENCIAL TEÓRICO

---

### 2.1 Modelo de Fração de Cura

Com os avanços da tecnologia, tem se tornado cada vez mais frequente em muitos estudos a identificação de pacientes que respondem bem ao tratamento, ou seja, indivíduos que não são suscetíveis ou não apresentaram o evento de interesse. Esses pacientes adquirem imunidade aos sinais e sintomas da enfermidade estudada, sendo considerados como curados. Nesse contexto, é essencial estudar e estimar a proporção de curados. Quando isso acontece, é necessário o uso de modelos que levem em conta a fração de cura, também conhecidos como modelos de taxa de cura. Esses modelos consideram a existência de uma população heterogênea, dividida em duas subpopulações: imunes (curados) e suscetíveis (não curados).

Ao empregar tais modelos, torna-se possível uma melhor compreensão e análise dos resultados, proporcionando uma abordagem mais precisa para avaliar os efeitos do tratamento e sua eficácia. Isso é fundamental para o avanço da pesquisa médica e aprimoramento dos procedimentos terapêuticos.

O modelo proposto, inicialmente, por [Boag \(1949\)](#) e [Berkson e Gage \(1952\)](#) possui as seguintes expressões para as funções de sobrevivência e densidade, respectivamente:

$$S(t) = p_1 + (1 - p_1)S_0(t), \quad t \geq 0, \quad (2.1)$$

$$f(t) = (1 - p_1)f_0(t), \quad t \geq 0,$$

em que  $p_1 \in (0, 1)$  representa os sobreviventes de longo prazo,  $S_0(\cdot)$  e  $f_0(\cdot)$  são, respectivamente, as funções de sobrevivência e densidade de probabilidade adequadas para os indivíduos suscetíveis ao evento de interesse. Na literatura, as distribuições Exponencial ou Weibull são as mais utilizadas para  $S_0(\cdot)$ , mas distribuições como Lognormal, Log-logística, entre outras, podem ser especificadas. Além disso,  $S_0(t)$  é uma função própria e  $f_0(t)$  é a função de densidade de probabilidade. Ao contrário de  $S_0(t)$ ,  $S(t)$  não é uma função própria pois  $\lim_{t \rightarrow \infty} S(t) = p_1 > 0$ .

Porém o modelo proposto pela Equação (2.1) possui algumas limitações em que segundo [Chen, Ibrahim e Sinha \(1999\)](#), no contexto da recidiva do câncer, o modelo proposto por [Berkson e Gage \(1952\)](#) e [Boag \(1949\)](#) não descrevia o processo biológico subjacente que gera o tempo de falha em que modelos de fração de cura são usados com frequência. Dessa forma, [Chen, Ibrahim e Sinha \(1999\)](#) propuseram uma classe alternativa de modelos de fração cura chamado de modelo de cura de tempo de promoção (*promotion time cure model*). Eles assumiram que uma determinada fração,  $p_1$ , da população é considerada curada, enquanto os demais  $(1 - p_1)$  são não curados. Além disso, eles também supuseram que a ocorrência do evento de interesse nos indivíduos acontece devido a causas competitivas, representadas por uma variável aleatória  $M$ . Em outras palavras,  $M$  denota o número de células cancerígenas para o  $i$ -ésimo indivíduo.

No estudo realizado por eles, a variável aleatória  $Z_i$ ,  $i = 1, 2, \dots$  denota o tempo aleatório para a  $i$ -ésima célula produzir uma massa de câncer detectável. As variáveis  $Z_i$ , são independentes de  $M$  e identicamente distribuídas (i.i.d) com função de distribuição comum dada por  $F(t) = 1 - S(t)$ . Neste trabalho considera-se que  $M$  segue uma distribuição de Poisson com parâmetro  $(\theta)$ .

O modelo de fração de cura pode ser definido de uma forma mais abrangente, isto é, como sendo um modelo fundamentado em um processo de dois estágios ou etapas para o acontecimento do evento em estudo. Em que o primeiro estágio, chamado de estágio de iniciação, considera uma variável aleatória  $M$  que corresponde a quantidades de causas ou riscos competindo para que ocorra o evento em estudo. A causa responsável pela ocorrência do evento é considerada latente, isto é, ela é desconhecida e portanto a variável  $M$  é não-observada. A distribuição de probabilidade,  $p_m$ , que representa esta variável e sua calda é dada, respectivamente, por:

$$p_m = \mathbb{P}[M = m] \quad \text{e} \quad q_m = \mathbb{P}[M > m],$$

com  $m = 0, 1, 2, \dots$

Com relação ao segundo estágio, chamado de etapa de maturação, segue que dado  $M = m$ , considere  $Z_k$  com  $k = 1, \dots, m$  como sendo variáveis aleatórias contínuas (não-negativas) independentes com sua função de distribuição acumulada dada por  $F(z) = 1 - S(z)$  e seja  $Z$  e  $M$  independentes. Então  $Z_k$  correspondem ao tempo de ocorrência do fenômeno de interesse de acordo à  $k$ -ésima causa.

Para incorporar a parcela de pacientes não suscetíveis ao evento de interesse, o tempo de ocorrência é definido como sendo:

$$T = \min(Z_0, Z_1, Z_2, \dots, Z_M),$$

em que  $P[Z_0 = \infty] = 1$  representa a chance de que uma parcela da população, denotada como  $p_1$ , não experimente o evento de interesse. Nessa definição,  $T$  é uma variável aleatória que pode ser observada ou censurada, e  $Z_j$  e  $M$  são variáveis latentes.

Agora considere  $\{a_n\}$  como sendo uma sequência de números reais. Então se  $s$  pertence ao intervalo  $[0, 1]$ , podemos concluir que:

$$A(s) = a_0 + a_1s + a_2s^2 + \dots,$$

converge, e portanto  $A(s)$  é definida como a função geradora da sequência  $\{a_n\}$ .

Seja uma função de sobrevivência da variável aleatória  $T$ , dada por  $S_p(t) = P[T > t]$ . Dada a função de sobrevivência própria,  $S(t)$ , a função de sobrevivência da variável aleatória  $T$  será dada por

$$\begin{aligned} S_p(t) &= \mathbb{P}[M = 0] + \mathbb{P}[Z_1 > t, Z_2 > t, \dots, Z_M > t, M \geq 1] \\ &= \mathbb{P}[M = 0] + \sum_{m=1}^{\infty} \mathbb{P}[M = m] \mathbb{P}[Z_1 > t, Z_2 > t, \dots, Z_M > t] \\ &= p_1 + \sum_{m=1}^{\infty} p_m S(t)^m \\ &= A[S(t)], \end{aligned} \tag{2.2}$$

em que  $A(\cdot)$  é a função geradora da sequência  $p_n$  e  $S_p(t)$  é uma função de sobrevivência imprópria pois  $\lim_{t \rightarrow \infty} S(t) \neq 0$ , isto é,  $\lim_{t \rightarrow \infty} S(t) = p_1$ , em que  $p_1$  é a proporção dos indivíduos imunes ao evento de interesse na população, ou seja, a fração de curados. A seguir será mostrado as propriedades da função de sobrevivência populacional  $S_p(t)$ :

- Se  $p_1 = 0$ , temos que  $S_p(t) = S(t)$ ;
- $S_p(0) = 1$ ;
- $S_p(t)$  é decrescente;
- $\lim_{t \rightarrow \infty} S_p(t) = p_1$ .

As Equações (2.3) e (2.4) representam, respectivamente, as funções de densidade e de risco relacionadas à função de sobrevivência de longa duração

$$f_p(t) = f(t) \frac{dA(s)}{ds} \Big|_{s=S(t)}, \tag{2.3}$$

e

$$h_p(t) = \frac{f_p(t)}{S_p(t)} = f(t) \frac{dA(s)}{ds} \Big|_{s=S(t)}. \tag{2.4}$$

Nesse contexto associado a funções geradoras de probabilidade, existem distribuições bem conhecidas e bastante utilizadas na literatura, como por exemplo Bernoulli, Binomial, Poisson, Binomial Negativa e Geométrica. Mas para aplicarmos os resultados mostrados até agora, consideraremos  $M$  seguindo a distribuição Poisson.

#### a) Poisson

Seja  $M$ , o número de causas do evento de interesse e que esta segue uma distribuição de probabilidade Poisson. Deste modo, temos que  $M$  tem sua função de probabilidade dada por:

$$\mathbb{P}(M = m) = \frac{e^{-\theta} \theta^m}{m!},$$

com  $M = 0, 1, \dots$ ,  $\theta > 0$ . A esperança e variância é definida, respectivamente, como:  $\mathbb{E}(M) = \theta$   $\text{Var}(M) = \theta$ . Calculando a sua função geradora de probabilidades, temos que:

$$\begin{aligned} A(s) &= \sum_{n=0}^{\infty} s^n p_n \\ &= \sum_{n=0}^{\infty} s^n \frac{\theta^n}{n!} e^{-\theta} \\ &= e^{-\theta} \sum_{n=0}^{\infty} \frac{(\theta s)^n}{n!} \\ &= e^{-\theta(1-s)}, \quad s \in \mathbb{R}. \end{aligned} \quad (2.5)$$

Desse modo, considerando a função geradora de probabilidade dada por (2.5), conseguimos obter, respectivamente, as suas funções de sobrevivência e densidade associadas a população, são elas:

$$S_p(t) = A(S(t)) = \exp[-\theta F(t)] \quad \text{e} \quad f_p(t) = -\frac{dS_p(t)}{dt} = \theta f(t) \exp[-\theta F(t)]. \quad (2.6)$$

Da Equação (2.6), obtemos a função de risco associada a esta distribuição, que é dada por:

$$h_p(t) = \theta f(t). \quad (2.7)$$

A fração de cura na população é dada por

$$p_1 = \lim_{t \rightarrow \infty} S_p(t) = \exp(-\theta).$$

Analogamente, estes mesmos cálculos podem ser feitos considerando outras distribuições. A seguir, na Tabela 1, mostramos um resumo referente a função de sobrevivência, a densidade imprópria e fração de cura correspondentes aos modelos de longa duração mais utilizados.

Tabela 1 – Função de sobrevivência  $S_p(t)$ , função de densidade  $f_p(t)$  e fração de cura considerando algumas distribuições referente ao número de causas latentes,  $M$ .

Distribuição	$S_p(t)$	$f_p(t)$	$p_1$
Bernoulli( $p$ )	$1 - p_1 + pS(t)$	$pf(t)$	$1 - p$
Binomial( $K, p^*$ )	$(1 - p^* + p^*S(t))^K$	$Kp^*f(t)(1 - p^* + p^*S(t))^{K-1}$	$(1 - p^*)^K$
Poisson( $\theta$ )	$\exp(-\theta F(t))$	$\theta f(t) \exp(-\theta F(t))$	$\exp(-\theta)$
Geométrica( $p$ )	$\{1 + pF(t)\}^{-1}$	$pf(t)\{1 + pF(t)\}^{-2}$	$1/(1 + p)$
Binomial Negativa( $\tau, \theta$ )	$\{1 + \eta\theta F(t)\}^{-1/\eta}$	$\theta f(t)\{1 + \eta\theta F(t)\}^{-1-1/\eta}$	$(1 + \eta\theta)^{-1/\eta}$

em que  $p^*$  é nova reparametrização, isto é,  $p^* = p/(1 + p)$ .

Quando nós analisamos os resultados da Tabela 1 com relação a  $S_p(\infty)$ , percebemos que este resultado é maior do que zero, por exemplo, se  $M$  segue uma distribuição Poisson então  $S_p(\infty)$  é igual à  $\exp(-\theta)$  e portanto  $S_p$  não é uma função própria. Dessa forma, [Chen, Ibrahim e Sinha \(1999\)](#) argumenta que a expressão para a função de sobrevivência para toda a população (2.6) tem uma função de risco atrativa (2.7) uma vez que essa estrutura obtida para a função de risco tem  $\theta$  e  $f(t)$  agindo de forma multiplicativa e portanto a expressão (2.6) tem a estrutura de riscos proporcionais com as covariáveis sendo modelada por  $\theta$ . Segundo [Chen, Ibrahim e Sinha \(1999\)](#), a expressão da função de sobrevivência para a parcela de pacientes não curados da população em estudo é dada por:

$$S^*(t) = P(T > t | M \leq 1) = \frac{\exp(-\theta)F(t)}{1 - \exp(-\theta)} \quad (2.8)$$

Diferentemente da função de Sobrevivência ( $S_p(t)$ ) imprópria dada na Tabela 1, nós temos que  $S^*(t)$  dada na expressão da equação (2.8) é uma função própria pois  $S^*(0) = 1$  e  $S^*(\infty) = 0$ . A relação matemática entre os modelos propostos pelas expressões (2.1) e (2.8) em que a equação (2.1) pode ser reescrita da seguinte forma

$$S_p(t) = \exp(-\theta) + (1 - \exp(-\theta))S^*(t),$$

em que  $S^*(t)$  é dada pela equação (2.8), portanto, o modelo de mistura padrão com taxa de cura igual a  $\exp(-\theta)$  e função de sobrevivência para a população não curada dada por  $S^*(t)$ . Então, [Chen, Ibrahim e Sinha \(1999\)](#) mostram que todo modelo definido pela expressão (2.2) pode ser escrito como um modelo de mistura padrão de [Berkson e Gage \(1952\)](#) e [Boag \(1949\)](#).

## 2.2 Modelo de fração de cura zero inflacionado

O modelo de sobrevivência de fração de cura assume que todos os indivíduos que experimentaram o evento de interesse, mesmo os casos em que aconteceram logo no início do estudo, foi devido ao evento de interesse. Porém em algumas situações a ocorrência do evento de interesse, principalmente logo no início do estudo pode acontecer devido a outros fatores. Essa situação é comum em estudos de coorte retrospectivo ou em estudos longitudinais em que alguns indivíduos já estão doentes no momento da inclusão no estudo. Sendo assim, quando existe essa fração de indivíduos que experimentaram o fenômeno de interesse antes do começo do estudo ou logo no início, o modelo de fração de cura pode resultar em estimativas tendenciosas para a função de sobrevivência e outras quantidades de interesse.

Os modelos estatísticos zeros inflacionados (ou ajustados) são uma alternativa para essas situações pois permitem que a análise leve em consideração situações as quais o evento de interesse ocorre antes ou logo no início. Nesse tipo de modelagem, utiliza-se uma mistura de

duas distribuições com dois processos subjacentes, um que trata do excesso de zeros, e outro que trata da parte diferente de zero (MARTIN *et al.*, 2005).

Essa modelagem tornou-se comum em diversas áreas de estudo, como produção (número de defeitos por unidade), ciências médicas (contagem de doenças), ciências biológicas (número de organismos por unidade de volume) e espacial (número de meteoritos que colidem com um satélite durante uma órbita) (SHANKER; HAGOS, 2016). Na área de análise de sobrevivência existem poucos trabalhos que utilizam essa abordagem de modelos inflacionados no zero. Diante disso, recentemente, uma extensão do modelo de Berkson e Gage (1952) que permite a adição de uma proporção de tempos iguais a zero (ou zero ajustado) foi proposta, no contexto de dados financeiros, por Jr, Moreira e Louzada (2017).

O modelo proposto por estes autores permite que a função de sobrevivência seja modelada para os indivíduos que já experimentaram o evento de interesse (fração de zeros) no início do estudo, para os indivíduos que estão em risco e os que não estarão suscetíveis ao evento de interesse (fração de cura). Com isso, o modelo de fração de cura zero inflacionado pode fornecer estimativas mais precisas e menos tendenciosas da função de sobrevivência e outras quantidades de interesse em populações heterogêneas. A função de sobrevivência de todos os tempos possíveis, é dada por:

$$S(t) = p_1 + (1 - p_0 - p_1)S_0(t), \quad t \geq 0, \quad (2.9)$$

em que  $S_0(\cdot)$  é a função de sobrevivência relacionada à proporção  $(1 - p_0 - p_1)$  de pessoas suscetíveis à falha,  $p_1$  é proporção de indivíduos imunes ou curados na população e  $p_0$  é a proporção de tempos de sobrevivência inflacionados com zeros e esta parcela é o que diferencia para o modelo de mistura padrão proposto pela equação (2.1). Além disso, note que o modelo dado pela equação (2.9) tem as seguintes propriedades.

$$\lim_{t \rightarrow \infty} S(t) = p_1 > 0. \quad \text{e} \quad S(0) = 1 - p_0 < 1.$$

Além do mais, se  $p_0 = 0$ , ou seja, sem a inflação de zeros, obtemos o modelo de mistura padrão dado por Berkson e Gage (1952).

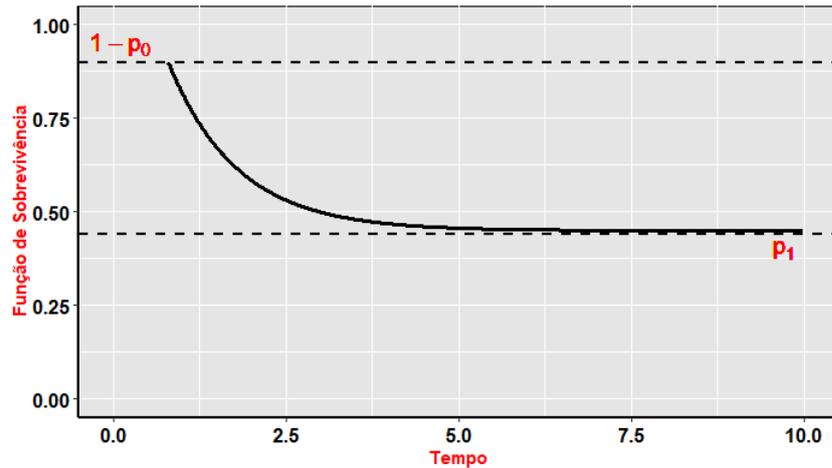


Figura 1 – Função de Sobrevivência do modelo de taxa de cura inflacionado no zero

Na Figura 1, temos o comportamento da função de sobrevivência deste modelo e podemos observar que existem duas características nos dados as quais devem ser contempladas adequadamente nos modelos de sobrevivência: a proporção de zeros e a fração de cura. Os modelos que contemplam tais proporções simultaneamente são chamados de modelos de sobrevivência zero-inflacionados (ou zero-ajustados) com fração de cura, em inglês: *zero-inflated cure rate models*.

O modelo de fração de cura zero inflacionado, definido pela equação (2.9) com função densidade  $f(t|\omega)$ , utiliza o vetor de parâmetros  $\omega$  para representar a distribuição dos dados, permitindo sua escolha de acordo com o estudo em questão. A função de distribuição cumulativa associada (imprópria) (FDA) e a função de densidade de probabilidade (imprópria) (fdp) são dadas por

$$F(t | p_0, p_1, \omega) = p_0 + (1 - p_0 - p_1)F_0(t | \omega), \quad t \geq 0, \quad (2.10)$$

e

$$f(t | p_0, p_1, \omega) = \begin{cases} p_0, & \text{se } t = 0 \\ (1 - p_0 - p_1)f_0(t | \omega), & \text{se } t > 0, \end{cases} \quad (2.11)$$

em que  $F_0(\cdot)$  e  $f_0(\cdot)$  são, respectivamente, a FDA e fdp acomodando a fração de indivíduos suscetíveis a falha do evento de interesse  $(1 - p_0 - p_1)$ .

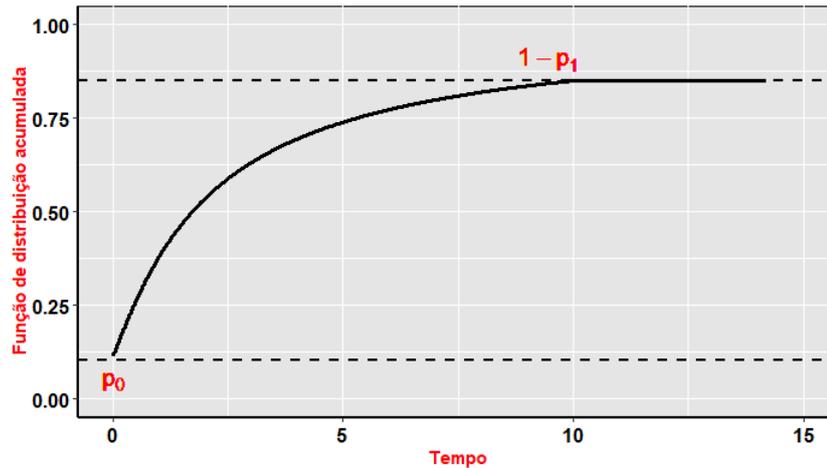


Figura 2 – Função de distribuição acumulada (imprópria) (FDA) do modelo de fração de cura inflacionado por zero.

Além disso, a FDA imprópria dos não suscetíveis, representado por  $F(t|p_0, p_1, \omega)$  tem a propriedade de lidar com proporção excessiva de zeros ( $p_0$ ) desde que satisfaça  $F(t|p_0, p_1, \omega) = p_0$  e o modelo também contabiliza a fração de curados,  $p_1$ , pois também satisfaz:

$$\lim_{t \rightarrow \infty} F(t | p_0, p_1, \omega) = 1 - p_1.$$

### 2.2.1 Modelo de regressão de fração de cura zero inflacionado

Nessa subseção, nós mostramos uma função de ligação para relacionar as covariáveis com os parâmetros do modelo de taxa de cura com inflação de zeros. Dessa forma, é possível determinar como as covariáveis afetam os tempos com inflação de zeros, a fração de cura e os eventos observáveis. Para isso, Júnior (2016) estabelece uma relação entre os parâmetros referente a proporção de zeros ( $p_0$ ), proporção de cura ( $p_1$ ) e o vetor parâmetros da distribuição de risco de base ( $\omega$ ) e um conjunto de vetores de covariáveis  $x_1, x_2, \dots, x_{k+2}$ .

O modelo de regressão de fração de cura zero inflacionado pode ser definida considerando as equações (2.9), (2.10), (2.11), (3.14) e (3.15) e através dos componentes sistemáticos dada por

$$\begin{cases} H(p_{0k}, p_{1k}) = (\xi_{0k}, \xi_{1k}), \\ g_j(\omega_{j,k}) = \eta_{jk}, \quad j = 1, \dots, i, \end{cases} \quad (2.12)$$

em que os preditores lineares são representados por  $\xi_{0k} = x_{1i}^\top \beta_1$ ,  $\xi_{1k} = x_{2i}^\top \beta_2$  e  $\eta_{jk} = x_{(j+2)}^\top i \beta_{(j+2)}$ . Além disso, o vetor  $\beta_j$  é composto por  $p + 1$  coeficientes de regressão desconhecidos que precisam ser estimados. As funções de ligações  $H$  e  $g_j$  proporcionam a relação entre os preditores lineares e os parâmetros da função de distribuição. Utilizando o mesmo conceito de Pereira, Botter e Sandoval (2013), a função de ligação para  $H$  é definida através de uma regressão

logística multinomial (Jr, Lemeshow e Sturdivant (2013), p. 261). Desse forma, a função é dada por:

$$H(p_{0k}, p_{1k}) = \left( \log \left( \frac{p_{0k}}{1 - p_{0k} - p_{1k}} \right), \log \left( \frac{p_{1k}}{1 - p_{0k} - p_{1k}} \right) \right). \quad (2.13)$$

Na prática, a escolha da função de ligação depende do suporte da distribuição McCullagh e Nelder (1989). Nesta Tese, nós atribuímos a distribuição Weibull para o tempo de sobrevivência dos pacientes, e por esta distribuição ter parâmetros estritamente positivos, então uma escolha para a função de ligação  $g_1$  e  $g_2$  é dada por  $g_1(\omega_{1i}) = \log(\omega_{1i})$  e  $g_2(\omega_{2i}) = \log(\omega_{2i})$ . Dessa forma,  $\omega_{1i} = \exp\{x_{3i}^\top \beta_3\}$  e  $\omega_{2i} = \exp\{x_{4i}^\top \beta_4\}$ . Essas funções de ligações são convenientes, pois as funções de ligação  $g_1(\cdot)$  e  $g_2(\cdot)$  são estritamente monótonas e duas vezes diferenciáveis, o que permite que elas abranjam o conjunto de números reais positivos em o conjunto de números reais.

Observe que a função de ligação estabelecida para o componente  $H$  garante o cumprimento das seguintes desigualdades:  $0 < v_{0i} < 1$ ,  $0 < \theta_{1i} < 1$  e  $0 < 1 - v_{0i} - \theta_{1i} < 1$ , em que isto sempre acontece desde que:

$$(p_{0i}, p_{1i}) = \left( \frac{\exp\{x_{1i}^\top \beta_1\}}{1 + \exp\{x_{1i}^\top \beta_1\} + \exp\{x_{2i}^\top \beta_2\}}, \frac{\exp\{x_{2i}^\top \beta_2\}}{1 + \exp\{x_{1i}^\top \beta_1\} + \exp\{x_{2i}^\top \beta_2\}} \right). \quad (2.14)$$

Além disso,  $H$  é uma função de ligação bijetora e duas vezes diferenciáveis que mapeia  $C$  em  $R^2$ , onde  $C$  é um subespaço de  $R^2$  o qual é definido como  $C = \{(p_{0i}, p_{1i}) | 0 < p_{0i} < 10 < p_{1i} < 1 - p_{0i}\}$  (mais detalhes ver Pereira, Botter e Sandoval (2013) p. 128).

## 2.3 Esquemas de ativação

No contexto em que o evento de interesse pode acontecer devido a diferentes causas, é difícil distinguir qual dos cenários deu origem ao fenômeno de interesse. Essa dificuldade pode estar relacionada ao fato de que a quantidade de fatores de risco é uma variável latente, ou seja, não observável. Considere então mecanismos que estão associados há alguns FR (ROMAN *et al.*, 2013). A ocorrência do evento de interesse é causado pela ativação de uma, algumas ou todas as causas competitivas, mas a quantidade destas são desconhecidas, ou seja, nós consideramos que houve a ocorrência de  $k$  ( $0 < k \leq M$ ) fatores de risco para desencadear o fenômeno de interesse. Caso não haja a ocorrência do que está sendo estudado, então teremos a presença de censura. Além do mais, existe a possibilidade de, mesmo com a ativação de algum fator de risco, não ocorrer o evento de interesse.

Assim os modelos com causas ou riscos latentes dispõem de uma estrutura a qual FR latentes explicam como o evento em estudo é desencadeado, ou seja, quantas causas são necessárias para que ocorra a ativação dessa estrutura, implicando no acontecimento do fenômeno

de interesse; representado pela variável aleatória  $T$ . O tempo até o acontecimento do evento de interesse fica condicionado a ocorrência de pelo menos um dos FR's.

A teoria relacionada aos esquemas de ativação é baseada em estudos desenvolvidos por Cooner *et al.* (2007) e Cooner, Banerjee e McBean (2006). Segundo estes autores, temos as seguintes suposições para o modelo com causas competitivas:

- (S<sub>1</sub>) Defina  $M$  como sendo uma variável aleatória discreta, com  $M \geq 0$ , em que esta caracteriza a quantidade de FR's relacionados a ocorrência do fenômeno de interesse. Assumi-se que  $p_1 = P(M = 0)$ .
- (S<sub>2</sub>) Defina  $K$  como sendo uma variável aleatória discreta, com  $K \geq 0$ . A sua respectiva função de probabilidade condicional, dado um valor de  $M$ , é representado por  $f_{K|M=m}$  tal que  $\mathbb{P}(0 \leq K \leq m | M = m) = 1$  para  $0 \leq M$ . Caso exista a presença de causas competitivas, a variável  $K$  caracteriza o número mínimo de fatores que devem ser ativados para a ocorrência do evento de interesse.
- (S<sub>3</sub>) Considere as variáveis aleatórias contínuas  $Z_1, Z_2, \dots$  independentes, com uma mesma função de sobrevivência  $S_0(\cdot)$ , independentes de  $M$  e  $K$ . Então  $Z = (Z_1, Z_2, \dots)$  correspondem aos tempos até o acontecimento ou ativação das causas latentes.
- (S<sub>4</sub>) O tempo até o acontecimento do fenômeno de interesse é representado por uma variável aleatória estendida, denominada por  $T$  a qual é definida por:

$$T = \begin{cases} \infty, & \text{se } M = 0 \\ Z_{(K,M)}, & \text{se } M \geq 1, \end{cases}$$

em que  $Z_{(1,M)} \leq \dots \leq Z_{(K,M)} \leq \dots \leq Z_{(M,M)}$  representam as estatísticas de ordem associadas as variáveis aleatórias  $Z_i$  com  $1 \leq i \leq M$ . No caso em que  $M = 0$ , o indivíduo é classificado como curado, portanto não será submetido ao nenhum dos fatores latentes e  $T = \infty$ . Por outro lado, se  $M \geq 1$ , então os modelos de cura fundamentados em esquemas de ativação latente assumem que um tempo de falha observado  $T$  (quando o evento de interesse do indivíduo ocorre, isto é, quando ele falha) é originado por pelo menos um dos FR's (COONER *et al.*, 2007; COONER; BANERJEE; MCBEAN, 2006).

A função de sobrevivência de  $T$  será representado por  $S_p(\cdot)$  e esta pode ser definida pelo seguinte teorema.

**Teorema 1.** Ante as suposições (S<sub>1</sub>) a (S<sub>4</sub>), a função de sobrevivência é caracterizada pela seguinte expressão:

$$S_p(t) = p_1 + (1 - p_1)S^*(t), \quad t > 0,$$

em que

$$S^*(t) = \frac{1}{1 - p_1} \sum_{m=1}^{\infty} \sum_{k=1}^m \sum_{j=m-k+1}^m \binom{m}{j} S_0^j(t) [1 - S_0(t)]^{m-j} f_{K|M=m}(k) f_M(m), \quad (2.15)$$

em que a série, dada pela Equação (2.15), converge uniformemente. Adicionalmente,  $S^*(\cdot)$  tem as seguintes propriedades satisfeitas:

( $P_1$ )  $S^*(\cdot)$  é uma função contínua.

( $P_2$ )  $S^*(\cdot)$  é uma função decrescente.

( $P_3$ )  $\lim_{t \rightarrow 0^+} S^*(t) = 1$ .

( $P_4$ )  $\lim_{t \rightarrow \infty} S^*(t) = 0$ .

Essas propriedades são referentes a função de sobrevivência relacionada a uma variável aleatória contínua positiva. As propriedades ( $P_1$ ) – ( $P_3$ ) também valem para  $S_p(\cdot)$ , já a ( $P_4$ ) não vale pois  $\lim_{t \rightarrow \infty} S(t) = p_1$ . Para verificar detalhes da prova desse teorema ver [Delgado \(2014\)](#). Pelo teorema citado anteriormente, podemos observar que:

1. Se  $p_1 > 0$ , então o modelo com FR's latentes será um modelo de mistura. A fração de cura deste modelo é representada por  $p_1 = \mathbb{P}(M = 0)$  e a sua respectiva função de sobrevivência será dada pela equação (2.15).
2. Um modelo de mistura, dada a fração de cura  $p_1$  e função de sobrevivência dos que estão suscetíveis  $S$ , é associado a um modelo com causas de risco latentes, em que este número de causas segue uma distribuição de Bernoulli. A função de sobrevivência dos tempos de ativação é dada por  $S_0(t) = S(t)$  e  $\mathbb{P}(M = 0) = p_1$ .
3. Quando  $p_1 = 0$  teremos um modelo sem fração de cura.
4. O modelo de tempo de promoção é equivalente ao modelo de fatores de risco latentes com  $K = 1$ ,  $M$  seguindo uma distribuição Poisson e os tempos de ativação das causas seguindo uma distribuição exponencial.

Então, o modelo de fração de cura com causas ou riscos latentes proposto por [Cooner et al. \(2007\)](#) descreve a origem do tempo até o acontecimento do evento que está sendo estudado,  $T$ ; generaliza tanto o modelo de mistura quanto o de promoção. Dito isto, assim como [Roman et al. \(2013\)](#), nós consideramos as três especificações, são elas:

1.  $K = 1$  implica que a ativação do primeiro fator latentes leva à ocorrência do evento, ou seja,  $Z_{(1,M)} = \min(T_1, T_2, \dots, T_M)$ .
2.  $K = q$  implica na ativação do  $q$ -ésimo fator de risco, ou seja,  $W = Z_{(q,M)}$ .
3.  $K = M$ , o tempo observado é  $Y = Z_{(M,M)} = \max(T_1, T_2, \dots, T_M)$  e isto implica a ativação de todos os fatores de risco.

As três especificações (ou mecanismos de ativações) citadas acima serão representadas pelas, respectivas, nomenclaturas:  $FR_{min}$ ,  $FR_{rand}$  e  $FR_{max}$ . Nas próximas subseções, apresentaremos os detalhes sobre esses três cenários.

### 2.3.1 Ativação pelo tempo mínimo

Esta abordagem foi originalmente proposta por [Yakovlev \*et al.\* \(1993\)](#), [Yakovlev \(1996\)](#), [Tsodikov, Yakovlev e Asselain \(1996\)](#) e neste caso, denomina-se como esquema de primeira ativação (ou mínimos de tempos). Sob esta perspectiva, [Nie \*et al.\* \(2017\)](#) fez uma abordagem explicativa com relação aos conceitos e a metodologia do modelo de causas (riscos) competitivos aplicado a dados de sobrevivência em epidemiologia clínica. [Nascimento \*et al.\* \(2019\)](#) estudaram o efeito de covariáveis nas causas de morte dos pacientes com anemia falciforme. Para representar esse cenário, iremos considerar a nomenclatura  $FR_{min}$ .

Nesse cenário, assume-se que uma única ativação leva a ocorrência do evento de interesse, ou seja, a primeira ativação dentro todos os fatores latentes é responsável pela ocorrência do evento de interesse. A variável aleatória que representa esse tempo é dada por

$$T = \begin{cases} \infty, & \text{se } M = 0, \\ Z_{(1,M)} = \min\{Z_1, \dots, Z_M\}, & \text{se } M \geq 1. \end{cases}$$

De (2.15), obtemos as funções de sobrevivência e de densidade dos suscetíveis ao evento de interesse, dada, respectivamente, por:

$$S_F^*(T) = \frac{1}{1-p_1} \sum_{m=1}^{\infty} S_0^m(t) f_M(m) \quad \text{e} \quad f_F^*(t) = \frac{f_0(t)}{1-p_1} \sum_{m=1}^{\infty} m S_0^{m-1}(t) f_M(m). \quad (2.16)$$

As equações (2.16) podem ser reescritas, respetivamente, pela seguinte forma:

$$S_F^*(T) = \frac{g_M(S_0(t)) - p_1}{1-p_1} \quad \text{e} \quad f_F^*(t) = \frac{g'_M(S_0(t))}{1-p_1}. \quad (2.17)$$

Utilizando o resultado obtido em (2.17), a função de sobrevivência correspondente aos indivíduos suscetíveis e não suscetíveis,  $S(t) = p_1 + (1-p_1)S_F^*(t)$  é tal que

$$S_p(t) = \sum_{m=0}^{\infty} S_0^m(t) f_M(m). \quad (2.18)$$

Na Tabela 2, temos alguns casos particulares das funções de sobrevivência estendida levando em consideração (2.18). Nós atribuímos algumas distribuições específicas para a quantidade de causas latentes  $M$ , com  $\mathbb{P}(M=0) > 0$ .

Tabela 2 – Proporção de curados,  $p_1$ ; Funções de sobrevivência da parcela suscetíveis ao evento  $S_F^*(t)$  e função de sobrevivência da população suscetíveis e não suscetíveis  $S_p(t)$  considerando o esquema de ativação mínimo.

Distribuição	$p_1 = \mathbb{P}(M = 0)$	$S_F^*(t)$	$S_p(t)$
Bernoulli( $\theta$ )	$\theta$	$S_0(t)$	$\theta + (1 - \theta)S_0(t)$
Poisson( $\theta$ )	$\exp(-\theta)$	$\frac{1 - \exp(-\theta S_0(t))}{1 - \exp(-\theta)}$	$\exp\{-\theta(1 - S_0(t))\}$
Geométrica( $1 - \theta$ )	$1 - \theta$	$\frac{(1 - \theta)S_0(t)}{1 - \theta S_0(t)}$	$\frac{1 - \theta}{1 - \theta S_0(t)}$

### 2.3.2 Ativação pelo tempo máximo

A estrutura de causas latentes obtidos pelo esquema de ativação mínimo, isto é, de primeira ativação, pode ser inadequada para certas situações, por exemplo, o desenvolvimento de doenças cardiovasculares (DCV). Neste exemplo, as DCV podem ocorrer quando uma sequência de fatores de risco não observáveis se manifesta como genética, stress, emoções.

Neste cenário, uma alternativa é a quantidade de FR's ser todos ativados para que o evento de interesse ocorra. A título de exemplo, considere uma situação para diferentes tipos de câncer, onde exista um mecanismo que envolve a mutação da metástase. Durante um período curto de tempo desse processo, a resposta imune do paciente também é ativada para iniciar  $M$  respostas contra essa mutação. Quando não existe mutação, não haverá resposta imune, ou seja,  $M = 0$ . Cada resposta imune é um fator latente capaz de resistir à ocorrência do evento de interesse, que pode ser por exemplo, uma doença ou à morte até seu tempo de promoção (destruição),  $Y_k$  com  $k = 1, \dots, M$ . Para representar esse cenário, iremos considerar a nomenclatura  $FR_{max}$ . O evento de interesse ocorrerá depois da ativação de todos os  $M$  fatores, sendo assim de acordo com [Delgado \(2014\)](#), o tempo de falha observado é dado por:

$$T = \begin{cases} \infty, & \text{se } M = 0, \\ Z_{(M,M)} = \max\{Z_1, \dots, Z_M\}, & \text{se } M \geq 1. \end{cases}$$

Vale ressaltar que essa abordagem seria análogo a duração de um sistema em paralelo com  $M$  componentes, caso  $M \geq 1$ . Através da equação (2.15), obtemos a função de sobrevivência das pessoas suscetíveis que é dada por:

$$S_{max}^*(t) = 1 - \frac{1}{1 - p_1} \sum_{m=1}^{\infty} F_0^m(t) f_M(m). \quad (2.19)$$

Reescrevendo a Equação (2.19), obtemos o seguinte resultado:

$$S_{max}^*(t) = \frac{1 - g_M(F_0(t))}{1 - p_1}, \quad (2.20)$$

em que  $g_M$  é calculado pela função geradora de probabilidade da variável  $M$ . Portanto, a função de sobrevivência associada à população,  $S_{p_{max}}(t) = p_1 + (1 - p_1)S_{max}^*(t)$ , dá em:

$$S_{p_{max}}(t) = 1 + p_1 - \sum_{m=0}^{\infty} F_0^m(t) f_M(m). \quad (2.21)$$

Na Tabela 3, temos alguns casos particulares das funções de sobrevivência estendida levando em consideração a (2.21) em que atribuímos algumas distribuições específicas para a quantidade de causas latentes  $M$ , com  $\mathbb{P}(M = 0) > 0$ .

Tabela 3 – Proporção de curados,  $p_1$ ; Funções de sobrevivência da parcela suscetíveis ao evento  $S_{max}^*$  e função de sobrevivência da população suscetíveis e não suscetíveis  $S_{p_{max}}(t)$  considerando o esquema de ativação máximo.

Distribuição	$p_1 = \mathbb{P}(M = 0)$	$S_{max}^*(t)$	$S_{p_{max}}(t)$
Bernoulli( $\theta$ )	$\theta$	$S_0(t)$	$\theta + (1 - \theta)S_0(t)$
Poisson( $\theta$ )	$\exp(-\theta)$	$\frac{1}{1 - \exp(-\theta)} [1 - \exp(-\theta S_0(t))]$	$1 + \exp(-\theta) [1 - \exp(\theta F_0(t))]$
Geométrica( $1 - \theta$ )	$1 - \theta$	$\frac{1 - F_0(t)}{1 - \theta F_0(t)}$	$1 - \theta + \theta \frac{1 - F_0(t)}{1 - \theta F_0(t)}$

### 2.3.3 Ativação aleatória

Diferente dos mecanismos anteriores, o esquema de ativação aleatória ocorre quando o tempo observado é associado à ocorrência do  $k$ -ésimo fator de risco, ou seja, temos que  $1 < k < M$  e para este cenário, denotaremos por  $FR_{rand}$ . Este cenário pode ser empregado para modelar, em situações as quais a ativação "aleatória" de um dos fatores de risco pode ser suficiente para a ocorrência do evento de interesse. Na literatura, vários estudos já foram realizados considerando esse esquema de ativação, como pode ser visto em: [Cooner et al. \(2007\)](#), [Cancho, Castro e Dey \(2013\)](#), [Bourguignon, Silva e Cordeiro \(2014\)](#), [Cancho et al. \(2013\)](#) entre outros.

Nesse cenário, a distribuição condicional, referente ao número de causas latentes que serão ativadas para a ocorrência do fenômeno de interesse, tem uma distribuição uniforme ou seja, a  $k$ -ésima ativação produzida pelo acontecimento do evento de interesse, e sua respectiva probabilidade é  $1/m$ . Diante disto, temos que:

$$f_{k|M}(k) = \frac{1}{m}, \quad k = 1, \dots, m; m \in 1, 2, \dots$$

De acordo com [Delgado \(2014\)](#), temos o seguinte teorema

**Teorema 2.** No mecanismo aleatório com distribuição uniforme, a função de sobrevivência do modelo de causas latentes representa um modelo de mistura padrão cuja função de sobrevivência dos indivíduos expostos ao evento de interesse é a correspondente aos tempos de ativação, isto é,  $S_0(t)$ . A fração de cura é representada por  $p_1 = \mathbb{P}(M = 0)$ , então temos que:

$$S_p(\cdot) = p_1 + (1 - p_1)S_0(\cdot).$$

Sendo assim, podemos observar que a distribuição a qual atribuímos para a variável aleatória  $M$  inclui apenas na determinação da fração de curados  $p_1$ . A prova deste teorema pode ser encontrado em [Delgado \(2014\)](#).

## 2.4 Conclusão

Neste capítulo apresentamos as principais propriedades do modelo de fração de cura usual e do modelo de fração de cura zero inflacionado. Além disso apresentamos as principais propriedades de cada esquema de ativação, isto é, mínimo, máximo e aleatório. Essa estrutura permite acomodar diferentes possibilidades para as causas competitivas.

Com o objetivo de analisar dados de sobrevivência com fração de cura, fração de zero inflacionado e causas competitivas, no Capítulo 3 abordaremos o modelo de fração de cura zero inflacionado sob diferentes esquemas de ativação que é baseado em [Jr, Moreira e Louzada \(2017\)](#) para estimar a proporção de curados e zeros ajustados. A distribuição Weibull e Poisson é considerada para os indivíduos em risco e causas competitivas, respectivamente.



---

# MODELOS DE FRAÇÃO DE CURA ZERO INFLACIONADO CONSIDERANDO DIFERENTES ESQUEMAS DE ATIVAÇÃO

---

---

Em alguns estudos em análises de dados de sobrevivência, a incidência do fenômeno de interesse pode estar relacionada a diferentes causas de falha, podendo ser a primeira, a última ou intermediária. No entanto a quantidade destas são desconhecidas, ou seja, presume-se que houve a ocorrência de  $k$  ( $0 < k \leq M$ ) fatores de risco para desencadear o fenômeno de interesse. Nesse contexto, propomos um modelo de taxa de cura de promoção ajustada para zero (ZICR) em um cenário de risco competitivo para dados de sobrevivência. Consideramos que a quantidade  $M$  de causas de falhas possui uma distribuição de probabilidade de Poisson e que o tempo até a ocorrência de cada uma das causas segue uma distribuição de base Weibull.

## 3.1 Introdução

Neste capítulo, nós apresentamos uma extensão do modelo de tempo de promoção estudado em [Yakovlev \(1996\)](#), [Chen, Ibrahim e Sinha \(1999\)](#) e [Jr, Moreira e Louzada \(2017\)](#) incorporando excesso de zeros na modelagem sob diferentes esquemas de ativação (mínimo, máximo e aleatório). Além disso, apresentamos uma rápida descrição do modelo de mistura com fração de cura estudado por [Yakovlev \*et al.\* \(1993\)](#), [Tsodikov, Yakovlev e Asselain \(1996\)](#) e [Chen, Ibrahim e Sinha \(1999\)](#). Eles propuseram modelos de taxa de cura com base no perfil de evolução de um tumor.

Como descrito pelos autores, em seus estudos, o modelo assume que, para um indivíduo da população, existem  $M$  células tumorais competindo para a ocorrência da metástase e  $M$  segue uma distribuição de Poisson com média  $\theta$ . Para a  $i$ -ésima célula carcinogênica, tem o tempo aleatório  $Z_i$ , classificado como tempo de incubação, para ocasionar uma massa de câncer

perceptível. Além disso, a variável aleatória  $M$  é utilizada para representar o número de células afetadas por bactérias ou infecção viral. No artigo dos autores mencionados anteriormente, foi assumido que: as variáveis aleatórias (v.a)  $Z_i, i = 1, 2, \dots$ , são i.i.d com função de distribuição comum  $F(t) = 1 - S(t)$ . As variáveis  $Z_i$  são independentes de  $M$ .

Para incorporar, no modelo, os pacientes não suscetíveis a enfermidade estudada (recidiva do câncer), isto é, os pacientes cujo o número inicial de células cancerígenas seja igual a 0 e, teoricamente, com tempo de sobrevivência ao infinito, admitimos que  $P(Z_0 = \infty) = 1$ .

Por fim, o tempo para os indivíduos suscetíveis ao evento de interesse é definido por uma das variáveis aleatórias  $T = \min\{Z_i, 0 \leq i \leq N\}$ ,  $T = \max\{Z_i, 0 \leq i \leq N\}$  e  $T = T_{(q)}$  em que estas v.a's representam respectivamente, os cenários mínimo, máximo e aleatório. Adicionalmente, nós consideramos a elaboração do modelo proposto por [Chen, Ibrahim e Sinha \(1999\)](#) (ver Seção 2.1) e utilizado por [Jr, Moreira e Louzada \(2017\)](#) para a elaboração do modelo para cada esquema de ativação em que são descritos nas seções a seguir.

### 3.2 Modelos de fração de cura no cenário $FR_{max}$

O tempo para os indivíduos suscetíveis ao evento de interesse é definido pela variável aleatória  $T = \max\{Z_i, 0 \leq i \leq M\}$ , e conseqüentemente, a função de sobrevivência populacional de  $T$ , incluindo os suscetíveis e não suscetíveis, é dada por:

$$\begin{aligned}
 S_p(t) &= \mathbb{P}(T > t | M \geq 0) \\
 &= \mathbb{P}(M = 0) + \mathbb{P}(Z_1 > t, \dots, Z_M > t, M \geq 1) \\
 &= \mathbb{P}(M = 0) + \sum_{m=1}^{\infty} \{1 - [F_0(t)]^m\} \frac{\theta^m}{m!} \exp(-\theta) \\
 &= 1 - \exp\{-\theta(1 - F_0(t))\} + \exp\{-\theta\}.
 \end{aligned} \tag{3.1}$$

De (3.1), obtemos a respectiva função densidade dada por

$$f_p(t) = -\frac{d}{dt}S_p(t) = \theta f_0(t) \exp(-\theta(1 - F_0(t))). \tag{3.2}$$

Assim como as funções de sobrevivência descritas na Tabela 1, nós verificamos que as Equações (3.1) e (3.2) também são funções impróprias. Dessa forma, nós calcularemos a função de sobrevivência dos indivíduos expostos ao evento de interesse, chamada de  $S^*(t)$ . Então, a expressão para  $S^*(t)$  é dada por:

$$S_p^*(t) = P(T > t | M \geq 1) = \frac{1 - \exp(-\theta(1 - F_0(t)))}{1 - \exp(-\theta)}. \tag{3.3}$$

Note que a função de sobrevivência, dada por (3.3), é uma função própria pois as condições de que  $S_p^*(0) = 1$  e  $S_p^*(\infty) = 0$  e, assim, obtemos a função de densidade de probabilidade de

quem está suscetível à recorrência do evento analisado:

$$f_p^*(t) = -\frac{d}{d(t)}S^*(t) = \left( \frac{\exp(-\theta[1 - F_0(t)])}{1 - \exp(-\theta)} \right) \theta f_0(t). \quad (3.4)$$

Com isso, nós obtemos a relação matemática entre os modelos estudado por [Berkson e Gage \(1952\)](#) e [Chen, Ibrahim e Sinha \(1999\)](#), em que esta é dada, respectivamente, pelas Equações (2.1) e (3.1), isto é:

$$S_p(t) = \exp(-\theta) + (1 - \exp(-\theta))S_p^*(t), \quad t \geq 0, \quad (3.5)$$

$$f_p(t) = (1 - \exp(-\theta))f_p^*(t), \quad t \geq 0,$$

em que,  $S^*(t)$  e  $f^*(t)$  são funções próprias dadas por (3.3) e (3.4), respectivamente. Deste modo, o modelo estudado por [Chen, Ibrahim e Sinha \(1999\)](#) pode ser reescrito como modelo de taxa de cura, em que a taxa é dada por  $p = \exp(-\theta)$ .

### 3.3 Modelos de fração de cura no cenário $FR_{rand}$

O tempo para os indivíduos suscetíveis ao evento de interesse é definido pela variável aleatória  $T_{(k)}$  em que  $1 < k < M$ . A função de sobrevivência de  $T$ , incluindo os suscetíveis e não suscetíveis, é dada por:

$$\begin{aligned} S_p(t) &= \mathbb{P}(T > t | M \geq 0) \\ &= \mathbb{P}(M = 0) + \mathbb{P}(Z_1 > t, \dots, Z_M > t, M \geq 1) \\ &= \mathbb{P}(M = 0) + \sum_{m=1}^{\infty} S(t) \frac{\theta^m}{m!} \exp(-\theta) \\ &= \exp\{-\theta\} + (1 - \exp\{-\theta\})(1 - F_0(t)). \end{aligned} \quad (3.6)$$

De (3.6), obtemos a função densidade representada por

$$f_p(t) = -\frac{d}{dt}S_p(t) = f_0(t)(\exp\{-\theta\} - 1). \quad (3.7)$$

As equações dada pelas funções de sobrevivência ( $S_p(\cdot)$ ) e densidade ( $f_p(\cdot)$ ) são funções impróprias.

Considere a seguinte função de sobrevivência populacional dada pela expressão:

$$S_p^*(t) = P(T > t | M \geq 1) = \frac{(1 - \exp(-\theta))(1 - F_0(t))}{1 - \exp(-\theta)} = 1 - F_0(t). \quad (3.8)$$

A função de sobrevivência, dada por (3.8) é uma função própria pois as condições de que  $S_p^*(0) = 1$  e  $S_p^*(\infty) = 0$  são satisfeita e assim, obtemos a função de densidade de probabilidade

dos indivíduos suscetível à recorrência do evento de interesse:

$$f_p^*(t) = -\frac{d}{d(t)}S^*(t) = f_0(t).$$

### 3.4 Modelos de fração de cura no cenário $FR_{min}$

Nesse cenário, o tempo para os indivíduos suscetíveis ao evento de interesse é como sendo  $T = \min\{Z_i, 0 \leq i \leq M\}$ , e conseqüentemente, a função de sobrevivência populacional de  $T$ , incluindo os suscetíveis e não suscetíveis, é dado por:

$$\begin{aligned} S_p(t) &= \mathbb{P}(T > t | M \geq 0) \\ &= \mathbb{P}(M = 0) + \mathbb{P}(Z_1 > t, \dots, Z_M > t, M \geq 1) \\ &= \mathbb{P}(M = 0) + \sum_{m=1}^{\infty} S(t)^m \frac{\theta^m}{m!} \exp(-\theta) \\ &= \exp\{-\theta + \theta S(t)\} = \exp\{-\theta F(t)\}. \end{aligned} \quad (3.9)$$

De (3.9), obtemos função densidade populacional representada por

$$f_p(t) = -\frac{d}{dt}S_p(t) = \theta f_0(t) \exp(-\theta F_0(t)). \quad (3.10)$$

Note que as equações representadas pelas expressões (3.9) e (3.10) são funções impróprias. Portanto, assim como nos esquemas de ativação máximo e aleatório, considere a função de sobrevivência dos não curados dada pela expressão:

$$S_p^*(t) = P(T > t | M \geq 1) = \frac{\exp(-\theta F_0(t)) - \exp(-\theta)}{1 - \exp(-\theta)}. \quad (3.11)$$

A função de sobrevivência, dada por (3.11), é uma função própria pois as condições de que  $S_p^*(0) = 1$  e  $S_p^*(\infty) = 0$  são satisfeitas.

A função de densidade de probabilidade de quem está suscetível à recorrência do evento de interesse é dada por:

$$f_p^*(t) = -\frac{d}{d(t)}S^*(t) = \left( \frac{\exp(-\theta F_0(t))}{1 - \exp(-\theta)} \right) \theta f_0(t). \quad (3.12)$$

Por fim, nós chegamos à relação matemática entre o modelo de taxa de cura, conforme apresentado por Berkson e Gage (1952)) por meio da expressão (2.1) e (3.9), isto é:

$$S_p(t) = \exp(-\theta) + (1 - \exp(-\theta))S_p^*(t), \quad t \geq 0, \quad (3.13)$$

$$f_p(t) = (1 - \exp(-\theta))f_p^*(t), \quad t \geq 0,$$

em que, agora,  $S_p^*(t)$  e  $f_p^*(t)$  são funções próprias dadas por (3.11) e (3.12).

### 3.5 Modelo de fração de cura zero inflacionado na presença de covariável

Por meio dessa abordagem, que incorpora o processo dos esquemas de ativação referente a todas as causas de risco no modelo inflacionado no zero, é possível obter uma visão completa de todos os fatores de risco envolvidos nos pacientes. Isso significa lidar com a probabilidade de uma parcela dos enfermos que apresentaram o evento de interesse logo no início do estudo, resultando em tempo de sobrevivência igual a zero, e outra fração de indivíduos que não experimentaram o evento de interesse e terminaram sem apresentar o mesmo, o que chamamos de taxa de cura. Nossa proposta é uma modificação na equação da função de sobrevivência, conforme definido em (2.9).

A seguir, consideramos o modelo de taxa de cura zero inflacionado, conforme definido na expressão (3.5). Portanto, propomos uma função de sobrevivência (imprópria) da seguinte forma

$$S_p(t) = p_1 + (1 - p_0 - p_1)S_p^*(t), \quad t \geq 0,$$

em que  $S_p^*(t)$ , para cada esquema de ativação, é dada, respectivamente, por (3.3) ou (3.8) ou (3.11) e os parâmetros  $p_0$  e  $p_1$  são determinados, respectivamente, pelas expressões:

$$p_0 = \exp(-\nu) \quad \text{e} \quad p_1 = \exp(-\theta) \quad \nu, \theta > 0.$$

Para assegurarmos que  $p_0$ ,  $p_1$  e  $(1 - p_0 - p_1)$  estejam no intervalo (0,1), de acordo com Pereira, Botter e Sandoval (2013), Jr, Lemeshow e Sturdivant (2013, p. 199), Souza *et al.* (2022) e Jr, Moreira e Louzada (2017), nós consideramos para o link entre dois vetores de covariáveis aos parâmetros relativo à inflação zero e à taxa de cura, respectivamente, como sendo:

$$p_{0i} = \exp(-\nu_i) \quad \text{e} \quad p_{1i} = \exp(-\theta_i),$$

em que  $\nu_i = -\log\left(\frac{\exp\{x_{1i}^T \beta_1\}}{1 + \exp\{x_{1i}^T \beta_1\} + \exp\{x_{2i}^T \beta_2\}}\right)$ ,  $\theta_i = -\log\left(\frac{\exp\{x_{2i}^T \beta_2\}}{1 + \exp\{x_{1i}^T \beta_1\} + \exp\{x_{2i}^T \beta_2\}}\right)$ ,  $\beta_1$  e  $\beta_2$  são vetores de coeficientes de regressão a serem estimados, onde o primeiro relaciona a influência das covariáveis no excesso de zeros e o segundo relaciona a influência das covariáveis na fração de cura.

Para definir a especificação do modelo, isto é, para estabelecer a forma paramétrica de  $S_p^*(t)$ , consideraremos que  $f_0(t)$  e  $F_0(t)$  sejam, respectivamente, a função densidade e de probabilidade cumulativa da distribuição Weibull. Esta distribuição possui dois parâmetros:  $\alpha_1 > 0$  e  $\alpha_2 > 0$ , eles são os parâmetros de forma e escala, respectivamente. Sendo assim, nós conectaremos os parâmetros da Weibull como sendo:  $\alpha_{1i} = \exp\{x_{3i}^T \beta_3\}$  e  $\alpha_{2i} = \exp\{x_{4i}^T \beta_4\}$ . Essa escolha se dá pelo fato de  $g_1(\cdot)$  e  $g_2(\cdot)$  serem funções estritamente monótonas e duas vezes diferenciáveis. Conforme mencionado na subseção 2.2.1, esses são as funções de ligações mais apropriados para os parâmetros não negativos.

Assim, apresentamos a estrutura do modelo de taxa de cura inflacionada no zero sob os diferentes esquemas de ativação.

$$S_p(t) = \exp(-\theta) + (1 - \exp(-\nu) - \exp(-\theta))S_p^*(t),$$

onde  $S_p^*(t)$  e  $f_p^*(t)$  para cada esquema de ativação é dada, pela Tabela 4.

Tabela 4 – Função de sobrevivência  $S_p^*(t)$  e função densidade  $f_p^*(t)$  para os esquemas de ativação.

Esquema de ativação	$S_p^*(t)$	$f_p^*(t)$
Mínimo	$\frac{\exp(-\theta F_0(t)) - \exp(-\theta)}{1 - \exp(-\theta)}$	$\left(\frac{\exp(-\theta F_0(t))}{1 - \exp(-\theta)}\right) \theta f_0(t)$
Máximo	$\frac{1 - \exp(-\theta(1 - F_0(t)))}{1 - \exp(-\theta)}$	$\left(\frac{\exp[-\theta(1 - F_0(t))]}{1 - \exp(-\theta)}\right) \theta f_0(t)$
Aleatório	$1 - F_0(t)$	$f_0(t)$
$F_0(t) = 1 - \exp\left\{\left(\frac{t}{\theta}\right)^\alpha\right\}$ e $f_0(t) = \frac{\alpha}{\theta} \frac{t}{\theta}^{\alpha-1} \exp\left\{\left(\frac{t}{\theta}\right)^\alpha\right\}$		

### 3.5.1 Inferência

Nessa seção, nós apresentamos o método que utilizamos para fazer a inferência sobre os parâmetros. Para isto, a abordagem utilizada foi a de máxima verossimilhança. Com relação à contribuição de cada indivíduo para a função de probabilidade, notamos que existem diferentes subgrupos entre os pacientes, são eles: (i) indivíduos que já apresentaram o evento de interesse no tempo inicial, isto é, tempo zero; (ii) os não suscetíveis ao evento; ou (iii) suscetíveis ao evento. Desta forma, através da expressão (3.14), temos a função de verossimilhança de cada situação descrita acima para tempo  $t_i$ , isto é:

$$\begin{cases} p_{0i}, & \text{se } t_i = 0, \\ (1 - p_{0i} - p_{1i})f_p^*(t_i), & \text{se } t_i \text{ é totalmente observado,} \\ p_{1i} + (1 - p_{0i} - p_{1i})S_p^*(t_i), & \text{se } t_i \text{ é censurado.} \end{cases} \quad (3.14)$$

Considerando os dados da forma  $\mathbb{D} = (t_i, \delta_i, x_i = \{x_{1i}, x_{2i}, x_{3i}, x_{4i}\})$  em que  $\delta_i = 1$  significa que não houve censura enquanto que  $\delta_i = 0$  houve censura e  $\mathbf{x}_i$  é o vetor de covariáveis relacionado com a observação  $i$ . Por outro lado, o vetor de covariáveis podem ser considerados iguais, ou seja,  $x_1 = x_2 = x_3 = x_4$ . Seja  $\boldsymbol{\omega} = (\alpha_1, \alpha_2)$  o vetor de parâmetros da distribuição Weibull, e, por fim, temos  $(\beta_\nu, \beta_\theta)$  os parâmetros da regressão associados a proporção de inflação de zeros e da taxa de cura, respectivamente.

A função de verossimilhança para o modelo de sobrevivência de taxa de cura inflacionado no zero sob os esquemas de ativação, dado um vetor de parâmetros,  $\zeta = (\boldsymbol{\omega}, \beta_\nu^\top, \beta_\theta^\top)^\top$  será estimado via Estimador de Máxima Verossimilhança (EMV) sendo baseada em  $n$  observações. A estrutura da função de verossimilhança será a mesma para cada esquema de ativação. Sendo

assim, a função de verossimilhança para o modelo de fração de cura zero inflacionado sob diferentes esquemas de ativação é dada por:

$$\mathcal{L}(\zeta, \mathbb{D}) \propto \prod_{t_i=0} \{p_{0i}\} \prod_{t_i>0} \left\{ [(1 - p_{0i} - p_{1i})f_p^*(t_i)]^{\delta_i} [p_{1i} + (1 - p_{0i} - p_{1i})S_p^*(t_i)]^{1-\delta_i} \right\}. \quad (3.15)$$

A estimativa de  $\hat{\zeta} = (\hat{\omega}, \hat{\beta}_v, \hat{\beta}_\theta)$ , pelo método de máxima verossimilhança, pode ser obtida resolvendo o sistema não linear da equação  $U(\zeta) = 0$ . Porém como não tem forma fechada, nós utilizamos um método iterativo, chamada de quasi-Newton (maiores detalhes ver Shanno (1970)), através do software RStudio Team (2020).

Assim como Migon, Gamerman e Louzada (2014) e Ospina e Ferrari (2012), para um tamanho amostral grande, uma inferência para os parâmetros é dada pela matriz de segundas derivadas da log-verossimilhança usando a matriz de informação observada, isto é,  $\mathbf{I}(\vartheta) = \{\partial^2 \ell(\zeta) / \partial \zeta \partial \zeta^T\}^{-1}$ , calculando em  $\zeta = \hat{\zeta}$ . Portanto a aproximação do intervalo de confiança com  $(1 - \alpha)100\%$  para os parâmetros  $\alpha_1, \alpha_2, \beta_v$  e  $\beta_\theta$  são dados, respectivamente, por:

$$\hat{\alpha}_1 \pm \xi_{\alpha/2} \sqrt{\text{Var}(\hat{\alpha}_1)}, \quad \hat{\alpha}_2 \pm \xi_{\alpha/2} \sqrt{\text{Var}(\hat{\alpha}_2)}, \quad \hat{\beta}_v \pm \xi_{\alpha/2} \sqrt{\text{Var}(\hat{\beta}_v)}, \quad \hat{\beta}_\theta \pm \xi_{\alpha/2} \sqrt{\text{Var}(\hat{\beta}_\theta)},$$

em que  $\xi_{\alpha/2}$  é percentil superior da distribuição normal padrão.

## 3.6 Estudos de simulação

### 3.6.1 Propriedades assintóticas

Nesta seção, analisamos os desempenhos dos EMV's dos modelos propostos considerando diferentes tamanhos de amostra. Com intuito de avaliar os efeitos das covariáveis, consideramos uma variável binária ( $X$ ), isto é, assume 0 ou 1 em que esta é oriunda da distribuição Bernoulli com parâmetro 0.5. Esta variável pode ser considerada, por exemplo, uma situação em que existam indivíduos pertencentes ao grupo de controle ou não, sendo atribuídos valor de covariável de 0 e 1, respectivamente.

Suponha que o tempo de ocorrência de um evento de interesse tenha a função de distribuição cumulativa imprópria  $F(t)$  dada por (2.10), ou seja,

$$F(t) = p_0 + (1 - p_0 - p_1)F_0(t).$$

Nosso objetivo é gerar amostras aleatórias de tamanho  $n$  simulando tempos de sobrevivência de indivíduos, onde cada amostra compreende uma proporção  $p_0$  de tempos inflacionados de zero, uma fração de  $p_1$  e com uma proporção  $(1 - p_0 - p_1)$  de tempos de falha desenhados de uma distribuição Weibull com parâmetros  $\alpha$  e  $\theta$ .

O seguinte algoritmo foi usado para gerar dados do modelo inflacionados a zero sob o esquema de ativação máximo:

1. Determine os valores desejados para os parâmetros  $\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}$ ;
2. Simule  $x_i$  seguindo uma Bernoulli com parâmetro igual a 0.5 e calcule  $p_{0i}, p_{1i}, \alpha_{1i}, \alpha_{2i}$ ;
3. Simule  $u_i$  seguindo uma distribuição Uniforme no intervalo (0,1);
4. Se  $u_i \leq p_{0i}$ , faça  $s_i = 0$ ;
5. Se  $u_i > 1 - p_{1i}$ , faça  $s_i = \infty$ ;
6. Se  $p_{0i} < u_i \leq 1 - p_{1i}$ , simule  $v_i$  seguindo uma distribuição uniforme no intervalo  $(p_{0i}, 1 - p_{1i})$  e tome  $s_i$  como raiz de  $F(s_i) - v_i = 0$ ;
7. Gere  $w_i$  de uma Uniforme no intervalo  $(0, \max(s_i))$ , mas considerando apenas valores finitos de  $s_i$ ;
8. Obtenha  $t_i = \min(s_i, w_i)$ , se  $t_i < w_i$ , então  $\delta_i = 1$ , caso contrário,  $\delta_i = 0$ ;
9. Repita  $n$  vezes os passos de 2 até 8. Note que a censura segue uma distribuição uniforme com alcance limitado, pois assim é possível manter as taxas de censura razoáveis (ROCHA *et al.*, 2017, p. 12).

Para a simulação, consideramos que o tempo de falha segue uma distribuição Weibull com parâmetros  $\alpha_1$  e  $\alpha_2$ . Com isso, a estrutura é dada por oito parâmetros ( $\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \beta_{30}, \beta_{31}, \beta_{40}, \beta_{41}$ ). Podemos obter as estimativas dos parâmetros Weibull, taxa de cura e inflação zero através das expressões subsequentes:

$$\begin{aligned}
 v &= -\log \left( \frac{\exp\{\beta_{10} + x_i\beta_{11}\}}{1 + \exp\{\beta_{10} + x_i\beta_{11}\} + \exp\{\beta_{20} + x_i\beta_{21}\}} \right), \\
 \theta_{1i} &= -\log \left( \frac{\exp\{\beta_{20} + x_i\beta_{21}\}}{1 + \exp\{\beta_{10} + x_i\beta_{11}\} + \exp\{\beta_{20} + x_i\beta_{21}\}} \right), \\
 \alpha_{1i} &= \exp\{\beta_{30} + x_i\beta_{31}\}, \\
 \alpha_{2i} &= \exp\{\beta_{40} + x_i\beta_{41}\}.
 \end{aligned} \tag{3.16}$$

Considerando os parâmetros estabelecidos no modelo de regressão definido acima, nós realizamos um estudo de simulação considerando 3 diferentes cenários para os parâmetros do modelo, são eles, dado pela Tabela 5:

Tabela 5 – Valor dos parâmetros sob diferentes cenários

Parâmetros	Cenários		
	1	2	3
$\beta_{10}$	-2.00	-0.50	-0.50
$\beta_{11}$	0.5	1.50	0.50
$\beta_{20}$	-1.00	-0.75	-1.00
$\beta_{21}$	0.75	3.00	1.00
$\beta_{30}$	0.50	-20	-0.75
$\beta_{31}$	0.50	1.50	1.00
$\beta_{40}$	1.50	-1.25	1.25
$\beta_{41}$	2.00	1.00	1.00

A covariável considerada no processo de simulação é binária, isto é, assume 0 ou 1. Sendo assim no cenário 1 temos que  $p_0$  assume 9 ou 11%, respectivamente. Enquanto que  $p_1$  assume 24.47 ou 38.90%.

No cenário 2,  $p_0$  assume, respectivamente, 18.63% ou 21.19%, enquanto  $p_1$  assume 30.72 ou 57.61%. E por último, no cenário 3, temos que  $p_0$  assume 30.72 ou 33.33% enquanto que  $p_1$  assume 18.63 ou 33.33%.

Inicialmente, quando comparamos os três cenários, o cenário 1 (—■—) é considerado com baixa incidência de pacientes que apresentaram o evento de interesse antes do início do estudo. Enquanto que os cenários 2 (—●—) e 3 (—▲—) são considerados com média e alta incidência, respectivamente.

As Figuras 3, 4, 5 apresentam as curvas de sobrevivência estimadas por meio do estimador de Kaplan-Meier que ilustram a diferença na taxa de inflação de zeros e fração de cura entre os três cenários definidos para os parâmetros. Esses cenários foram criados para simular uma variedade de situações possíveis em um estudo real, proporcionando uma análise visual útil para compreender o impacto dos diferentes cenários na sobrevivência do paciente..

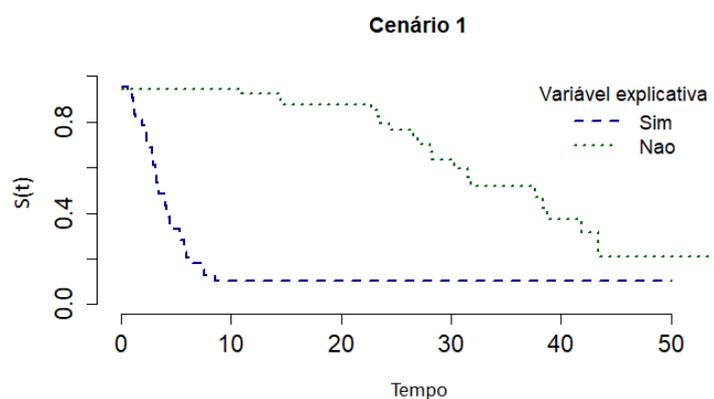


Figura 3 – Curvas de sobrevivência de KM dos dados simulados baseado nos valores de parâmetros do cenário 1.

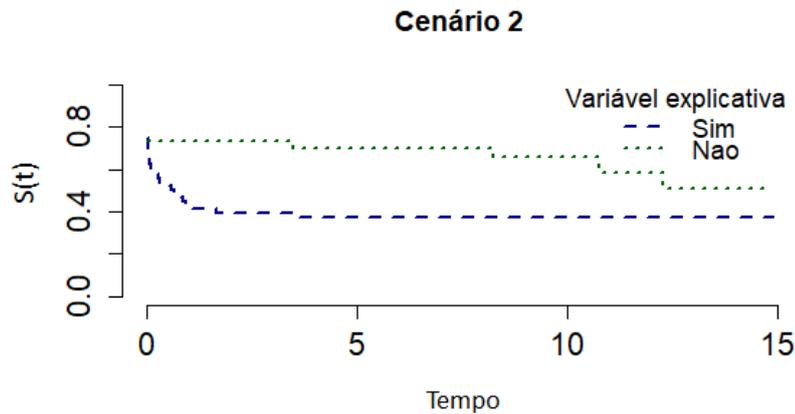


Figura 4 – Curvas de sobrevivência de KM dos dados simulados baseado nos valores de parâmetros do cenário 2.

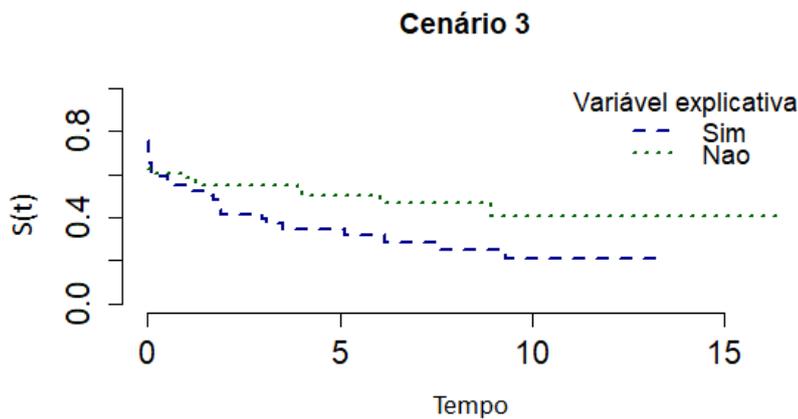


Figura 5 – Curvas de sobrevivência de KM dos dados simulados baseado nos valores de parâmetros do cenário 3.

No decorrer dessa seção, nós apresentamos os resultados da simulação obtidos nos três esquemas de ativação para os diferentes cenários descritos acima, onde ao tamanho da amostra ( $n$ ) são atribuídos os seguintes valores: 100, 250, 500, 750 e 1.000; além disso, foram consideradas simulações de Monte Carlo com 1.000 repetições. Para avaliar o desempenho das EMV, foram considerados diferentes valores para os parâmetros ( $\beta_{30}, \beta_{31}, \beta_{40}$  e  $\beta_{41}$ ) que estão associados com distribuição de tempo Weibull, além de considerar diferentes frações de dados ajustados por zero ( $\beta_{10}$  e  $\beta_{11}$ ) e taxa de doença ( $\beta_{20}$  e  $\beta_{21}$ ) relacionado a dados censurados.

Obtivemos estimativas de máxima verossimilhança dos parâmetros e seus, respectivos, erros padrão para cada cenário. Essas estimativas foram usadas para calcular o viés, a raiz do erro quadrático médio (EQMR) e a probabilidade de cobertura (PC) para cada tamanho de amostra e cenário estabelecido. Para cada processo de simulação, calculamos a raiz do erro quadrático médio, isto é,  $EQMR(\zeta) = \sqrt{Viés(\zeta)^2 + Var(\zeta)}$  onde o viés é dada por  $Viés(\zeta) = \hat{\zeta} - \zeta$ .

Além disso, a probabilidade de cobertura (ou probabilidade de cobertura assintótica) indica a quantidade de vezes em que o valor verdadeiro do parâmetro está dentro do intervalo de assintótico estimado de 95 % de confiança. No entanto, mesmo para estratégias precisas, é improvável que a cobertura empírica resultante seja igual à probabilidade de cobertura verdadeira igual a 95% devido ao erro de Monte Carlo. Então este é um erro de simulação e foi avaliado analiticamente. Se a probabilidade de cobertura real para uma estratégia para 0,95, então esperaríamos que 95% das coberturas empíricas simuladas estivessem dentro de  $1,96 \times EP(\text{cober})$  de 0,95, em que *cober* é a cobertura nominal esperada, 0,95. O EP foi  $\sqrt{0,95(1 - 0,95)/1000} = 0,00689$ , então o intervalo será aproximadamente 0,014, e podemos esperar resultados, quando a verdadeira probabilidade de cobertura para 0,95 é entre 0,936 e 0,964.

Nos gráficos de probabilidade de cobertura, as linhas cinza pontilhadas em torno de 0,95 representam o intervalo esperado para a cobertura para grandes tamanhos de amostra. Os limites superior e inferior do intervalo são 0,936 e 0,964.

Por meio do método delta com uma aproximação de Taylor de primeira ordem, obtemos as estimativas dos erros padrão para  $p_0$  e  $p_1$ , que estão associados com a proporção de zero ajustado e parâmetros de taxa de cura. Além disso, os EMV's foram obtidos usando a função `optim` em R e o algoritmo de maximização BFGS. Isso é feito usando o software livre [RStudio Team \(2020\)](#).

As Figuras 6 - 8 mostram os resultados do estudo de simulação considerando os esquemas de ativação máximo, aleatório e mínimo, respectivamente. De acordo com os resultados, podemos concluir o seguinte:

- Em geral, em todos os esquemas de ativação, a estimativa de máxima verossimilhança está próxima dos zeros nos três cenários de simulação analisados. Em outras palavras, está próximo dos valores originais que foram usados para definir os parâmetros de cada cenário inicialmente. Porém através da Figura 6, nós vemos que no cenário 1 (—■—), o parâmetro  $\beta_{21}$  precisa, pelo menos  $n = 500$ , para atingir a convergência enquanto que, na Figura 8, o mesmo acontece nos cenários 1 (—■—) e 2 (—●—) para os parâmetros  $\beta_{11}$  e  $\beta_{21}$ .
- Observe que os vieses diminuem para zero à medida que aumentamos o tamanho da amostra; Além disso, vemos que, em geral, a probabilidade de cobertura se aproxima do valor esperado de 95 %. No entanto, olhando para a Figura 7, percebemos que o parâmetro  $\beta_{31}$  converge mais lentamente para um PC nominal do que o parâmetro  $\beta_{40}$ . Também percebemos que o cenário 3 (—▲—) requer um tamanho de amostra de 300 para convergir para sua PC nominal.
- Em geral, em todos os esquemas de ativação, para os cenários que apresentam maior presença de indivíduos curados e zeros, ou seja, cenários 2 (—●—) e 3 (—▲—) que representam valores moderados e altos, respectivamente; as estimativas obtidas através do EMV, os valores de viés e PC da estimativa de regressão dos parâmetros associados a

$p_0 = \exp(-v)$  e  $p_1 = \exp(-\theta)$ , têm melhor desempenho em relação a cenário 1 (baixa presença de indivíduos curados,  $\text{---}\blacksquare\text{---}$ ), já que esses dois últimos cenários possuem maior quantidade de zeros e dados censurados;

- Por outro lado, o cenário 1 ( $\text{---}\blacksquare\text{---}$ ) que tem menor presença de zeros e indivíduos não curados, o EMV e os valores de viés e PC dos parâmetros de regressão estimados associados a  $\alpha_1$  e  $\alpha_2$ , tem melhor desempenho entre os três cenários, pois cenário 1 tem o maior tempo de observação.

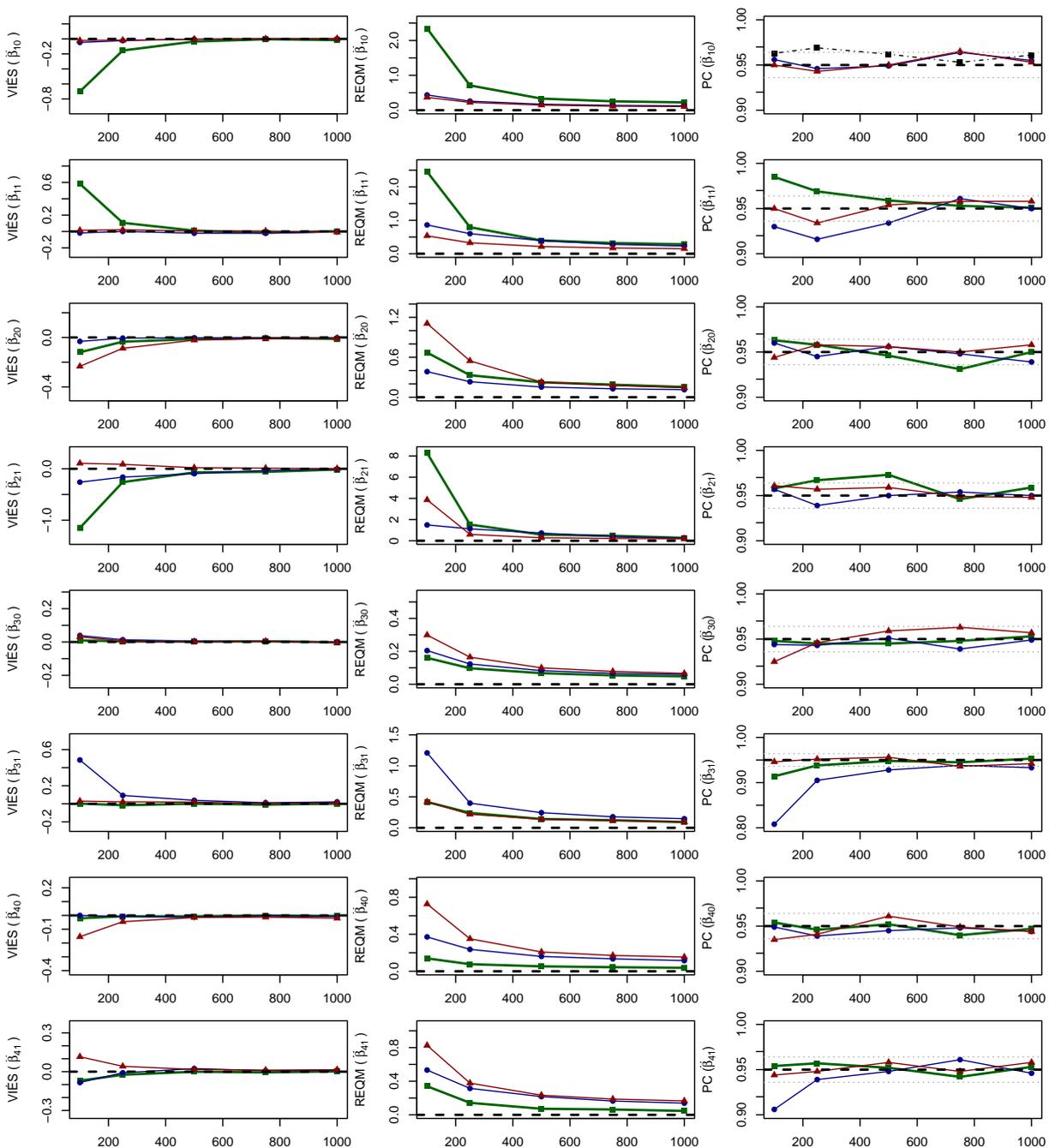


Figura 6 – Viés, REQM e probabilidade de cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação máximo

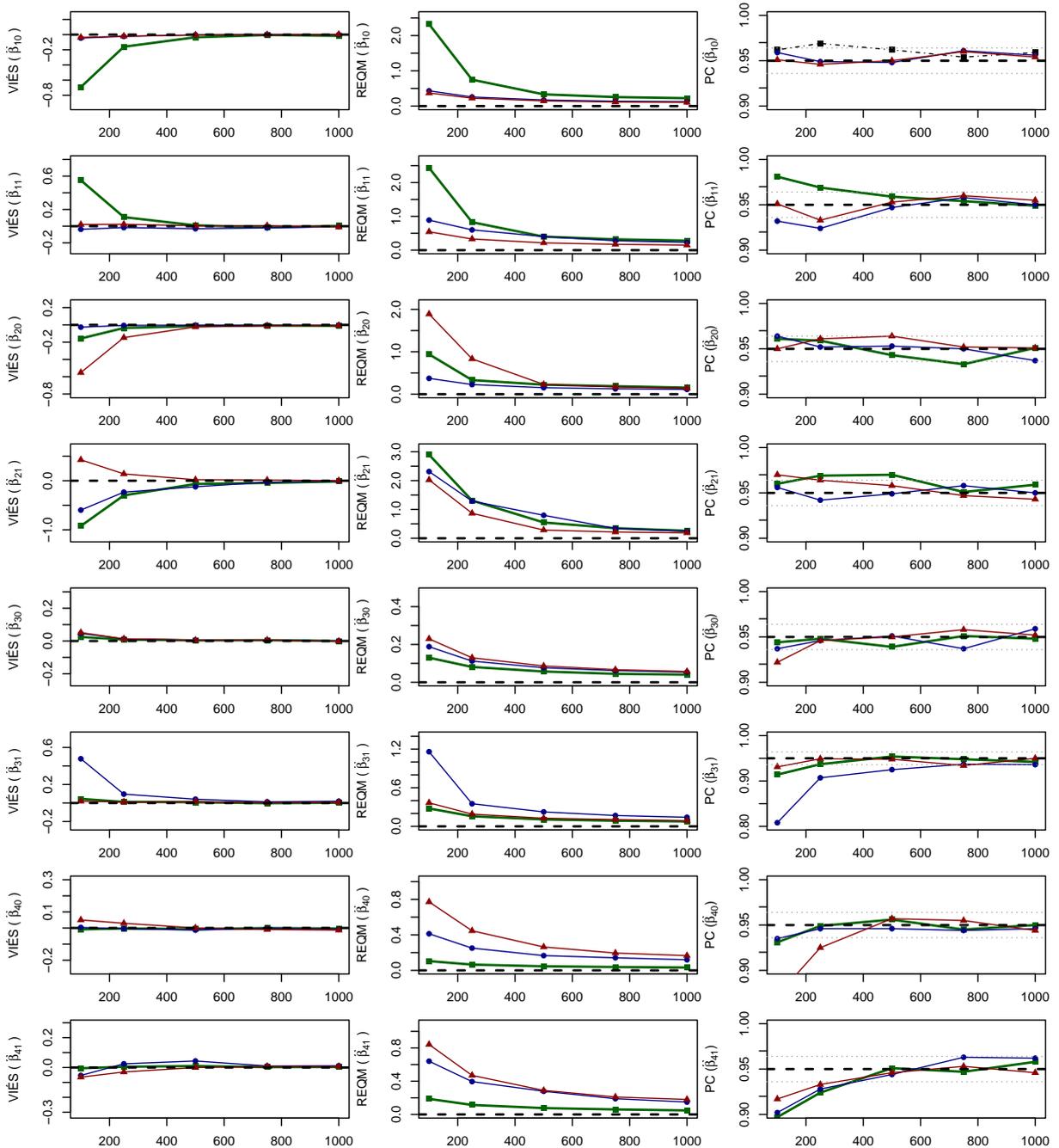


Figura 7 – Viés, REQM e probabilidade de cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação aleatório

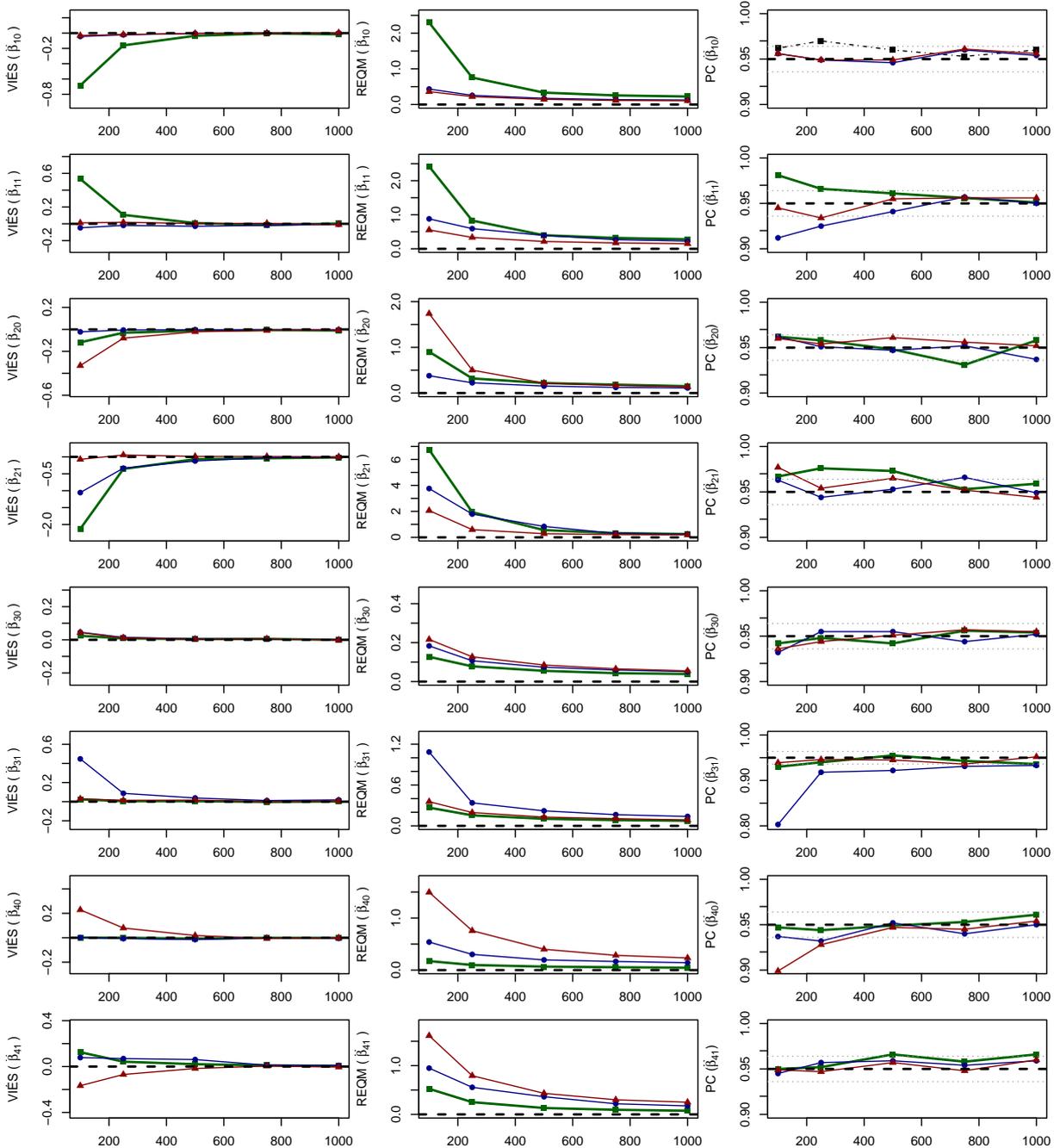


Figura 8 – Viés, REQM e probabilidade de cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação mínimo

### 3.6.2 Especificação incorreta do modelo

Estudos de simulação de especificação incorreta de modelos são frequentemente conduzidos para avaliar o desempenho de modelos estatísticos sob especificação incorreta. Nesses estudos, os critérios *Akaike Information Criterion* (AIC) e *Bayesian Information Criterion* (BIC) são comumente usados para comparar e selecionar modelos com base em sua qualidade de ajuste e complexidade.

Os critérios AIC e BIC fornecem medidas quantitativas da compensação entre o ajuste do modelo e a complexidade. Eles penalizam modelos com complexidade excessiva, desencorajando o *overfitting*, enquanto favorecem modelos que fornecem um bom ajuste aos dados. No contexto de estudos de simulação de especificação incorreta de modelos, AIC e BIC podem ser usados para avaliar o impacto da especificação incorreta na seleção do modelo e avaliar a confiabilidade desses critérios na identificação do verdadeiro modelo subjacente. A motivação por trás desses estudos de simulação de especificação incorreta de modelos é entender como os modelos propostos nesse capítulo se comportam quando são aplicados a dados que violam suas suposições subjacentes.

Dessa forma, nesta seção, nós apresentamos estudos de simulação de Monte Carlo para verificar a robustez da estatística dos critério de seleção sob especificação incorreta dos modelos propostos na seção 3.2 - 3.4. Os estudos de simulação foram realizados considerando os mesmos cenários da subseção anterior em que para cada cenário foi gerado o respectivo modelo e estimado considerando os diferentes esquemas de ativação e os resultados são comparados com o verdadeiro modelo subjacente para avaliar o impacto da má especificação.

Os dados fornecidos apresentam os resultados de um estudo de especificação incorreta, em que foram gerados diferentes cenários e analisados os percentuais de casos em que o modelo com o menor AIC foi selecionado corretamente. Em cada cenário, foram considerados diferentes tamanhos de amostra ( $n$ ) níveis de proporção de zeros e curados. A análise dos resultados quando consideramos o modelo ZICR com o esquema de ativação mínimo é a seguinte:

- Para o cenário 1, independentemente do tamanho da amostra, o método de seleção "mínimo" obteve os melhores resultados em termos de proporção de vezes em que o modelo com o menor AIC foi selecionado corretamente. À medida que o tamanho da amostra aumenta, a proporção de casos em que o modelo com o menor AIC foi selecionado corretamente também aumenta, enquanto as proporções para os métodos "máximo" e "aleatório" diminuem.
- No cenário 2, os resultados são semelhantes ao cenário 1, com o método "mínimo" novamente apresentando os melhores resultados em termos de proporção de seleção correta do modelo com menor AIC. À medida que o tamanho da amostra aumenta, a proporção de casos em que o modelo com o menor AIC foi selecionado corretamente aumenta, enquanto as proporções para os métodos "máximo" e "aleatório" diminuem.
- No cenário 3, novamente o método "mínimo" mostra os melhores resultados em todas as situações. À medida que o tamanho da amostra aumenta, a proporção de casos em que o modelo com o menor AIC foi selecionado corretamente também aumenta, enquanto as proporções para os métodos "máximo" e "aleatório" diminuem.

Quanto aos resultados obtidos considerando o modelo ZICR sob os esquemas de ativação máximo e aleatório, podemos concluir que

- O cenário 2, com média proporção de curados e zeros ajustados obteve um desempenho mediano quanto a discriminação do verdadeiro modelo considerando o critério AIC.
- Em todos os cenários, nós observamos que o modelo sob esquema de ativação máximo, teve um maior taxa de erro acusando o modelo ZICR sob o esquema de ativação aleatório como verdadeiro.
- Sob o modelo ZICR considerando o esquema de ativação aleatório, nós observamos que este foi o que apresentou maior taxa de má especificação considerando o critério de seleção AIC. Isso dá indícios que modelos complexos são vulneráveis quando usamos AIC como método de seleção.
- No cenário 3, novamente o método de seleção via AIC mostrou bons resultados em todas as situações. À medida que o tamanho da amostra aumenta, a proporção de casos em que o modelo com o menor AIC foi selecionado corretamente também aumenta, enquanto as proporções para os métodos "máximo" e "aleatório" diminuem.

Em resumo, os resultados mostram consistentemente que o método de seleção do modelo com o menor AIC é eficaz em selecionar corretamente o modelo adequado quando comparado aos modelos "máximo" e "aleatório". Além disso, à medida que o tamanho da amostra aumenta, a capacidade de selecionar corretamente o modelo com o menor AIC também aumenta. Isso sugere que um tamanho de amostra maior proporciona uma melhor discriminação entre os modelos e aumenta a probabilidade de selecionar o modelo correto com base no AIC.

Tabela 6 – % de caso com o menor AIC e esquema de ativação mínimo como modelo verdadeiro

Cenário 1			
n	Mínimo	Máximo	Aleatório
100	52%	30%	18%
250	68%	22%	10%
500	75%	19%	6%
750	83%	15%	2%
1000	90%	9%	1%
Cenário 2			
n	Mínimo	Máximo	Aleatório
100	51%	23%	26%
250	57%	22%	21%
500	62%	22%	15%
750	64%	25%	11%
1000	70%	20%	10%
Cenário 3			
n	Mínimo	Máximo	Aleatório
100	53%	23%	24%
250	61%	21%	17%
500	66%	24%	10%
750	70%	23%	7%
1000	75%	20%	5%

Tabela 7 – % de casos com menor AIC e esquema de ativação máximo como modelo verdadeiro

Cenário 1			
n	Mínimo	Máximo	Aleatório
100	23%	43%	33%
250	18%	54%	28%
500	7%	66%	27%
750	4%	75%	21%
1000	2%	81%	17%
Cenário 2			
n	Mínimo	Máximo	Aleatório
100	36%	44%	20%
250	31%	51%	18%
500	18%	56%	26%
750	16%	62%	22%
1000	12%	68%	20%
Cenário 3			
n	Mínimo	Máximo	Aleatório
100	23%	43%	34%
250	18%	54%	28%
500	7%	68%	25%
750	5%	75%	20%
1000	3%	82%	15%

Tabela 8 – % de casos com menor AIC e esquema de ativação aleatório como modelo verdadeiro

Cenário 1			
n	Mínimo	Máximo	Aleatório
100	30%	25%	45%
250	26%	24%	50%
500	18%	21%	61%
750	13%	18%	69%
1000	14%	16%	70%
Cenário 2			
n	Mínimo	Máximo	Aleatório
100	42%	37%	21%
250	41%	35%	24%
500	35%	32%	32%
750	34%	32%	33%
1000	30%	31%	39%
Cenário 3			
n	Mínimo	Máximo	Aleatório
100	42%	31%	26%
250	41%	32%	27%
500	30%	27%	43%
750	26%	27%	47%
1000	25%	26%	48%

## 3.7 Aplicações para dados reais

### 3.7.1 Dados sobre insulina

Inicialmente, o conjunto de dados, denominado conjunto de dados da insulina, usado nesta aplicação foi estudado por [Calsavara et al. \(2019a\)](#) que o aplicaram para o modelo de regressão defeituoso para dados de sobrevivência. Este conjunto de dados referem-se a um estudo retrospectivo sobre o uso de insulina em mulheres no período gestacional com diagnóstico de diabetes gestacional antes das 24 semanas de gestação acompanhadas durante o pré-natal no Hospital das Clínicas da Universidade de São Paulo, Brasil, a partir de 2012 a 2015. O estudo investiga o uso da insulina durante a gestação. O tempo de vida analisado é o período entre a primeira consulta médica e o início do uso da insulina. No entanto, algumas gestantes tiveram o evento estudado nos primeiros sete dias após a primeira consulta, sendo considerado um tempo de sobrevivência zero para esse grupo. Por outro lado, algumas mulheres não precisaram de insulina durante a gravidez, o que é considerado como uma cura. Esta aplicação apresenta uma situação em que o evento de interesse, que é o uso de insulina por mulheres com diabetes gestacional,

A diabetes mellitus gestacional (DMG) é uma condição complexa influenciada por vários fatores de ativação latentes. As causas da DMG podem ser multifatoriais, envolvendo fatores genéticos e ambientais. Sob um esquema de ativação latente, fatores como resistência à insulina, alterações hormonais e predisposição genética podem contribuir para a ativação ou aparecimento da DMG durante a gravidez. Essas causas competetivas podem interagir e levar ao comprometimento do metabolismo da glicose, afetando a capacidade do corpo de regular os níveis de açúcar no sangue. Compreender os esquemas de ativação latente envolvidos na DMG pode ajudar os profissionais de saúde a identificar indivíduos em risco, implementar medidas preventivas e fornecer estratégias de manejo adequadas para garantir o bem-estar da mãe e do bebê.

Nesse contexto, podemos considerar um dos três esquemas de ativação. Se, por exemplo, o uso de insulina for indicado pela ocorrência de uma das causas, teremos o esquema mínimo de ativação. Se a indicação for feita quando todas as causas ocorrerem, teremos um esquema de ativação máxima e, caso contrário, o esquema de ativação aleatória. Por meio de nossa abordagem (incorporando o processo de esquemas de ativação referentes a todas as causas de risco no modelo inflacionado a zero), é possível ter uma visão completa de todos os fatores de risco envolvidos no processo.

Na Figura 9, nós obtemos a curva de sobrevivência estimada via o método de Kaplan-Meier para dados de insulina e visualizamos que essa aplicação tem uma incidência de zero ajustado, isto é, fração de mulheres as quais precisaram da aplicação de insulina antes da 1 semana. Adicionalmente, essa aplicação possui uma alta taxa de cura, ou seja, gestantes que não precisaram aplicar insulina para administrar a diabetes gestacional.

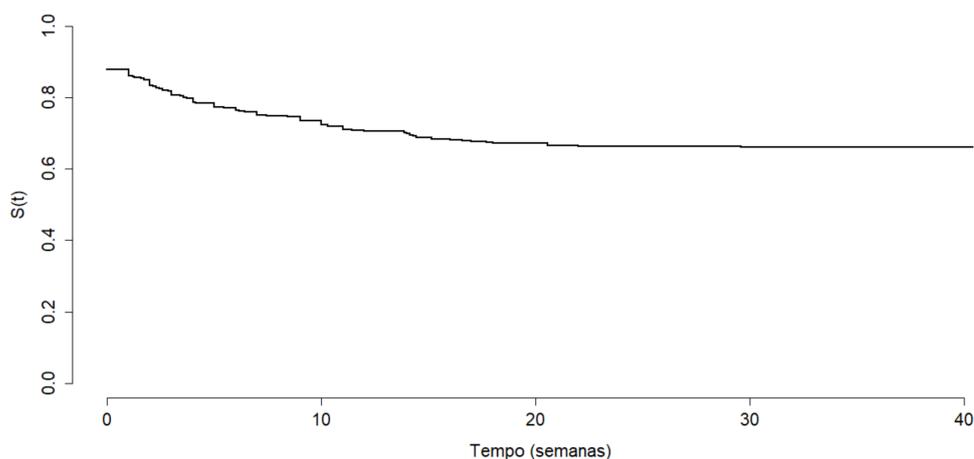


Figura 9 – Curva de sobrevida estimada obtida pelo método de Kaplan-Meier para dados de insulina

Como nosso evento de interesse é o tempo entre a primeira consulta médica e o início do uso de insulina pela gestante, nosso objetivo é verificar o efeito das variáveis: história familiar de diabetes, história anterior de GDM, história de tabagismo, macrossomia fetal anterior, hipertensão crônica anterior, IMC pré-gravidez, Idade. Na Tabela 9, nós apresentamos as frequências

observadas de cada variável mencionada anteriormente e a média e desvio padrão (D.P) par as variáveis contínuas.

Tabela 9 – Frequência observada das variáveis presente no banco de dados da insulina

Variável	N (%)
Histórico familiar de diabetes	
Não	143 (36,7)
Sim	247 (63,3)
Histórico anterior de GDM	
Primigesta	94 (23,1)
Não	249 (63,8)
Sim	47 (12,1)
Histórico tabagismo	
Não	356 (91,3)
Sim	34 (8,7)
Macrossomia fetal anterior	
Não	356 (91,3)
Sim	34 (8,7)
Hipertensão Crônica Anterior	
Não	282 (72,3)
Sim	108 (27,7)
Idade maternal (em anos)	média = 32,73; D.P = 6,15)

Portanto, o objetivo é analisar se essas variáveis influenciam tanto o uso de insulina nos primeiros sete dias quanto na fração de pacientes que não precisaram da aplicação da insulina. Vinculamos os parâmetros  $p_0$  e  $p_1$  às covariáveis por meio das seguintes equações, respectivamente:

$$p_0 = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta}_1)}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_1) + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_2)}, \quad \text{e} \quad p_1 = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta}_2)}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_1) + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_2)},$$

em que  $\mathbf{x}_i^\top = (x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i})$  e está associado ao vetor de covariáveis, para  $i = 1, \dots, 390$ . Nós também temos que  $\boldsymbol{\beta}_1^\top = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{16})$  são os coeficientes de regressão relacionados às gestantes com necessidade de uso de insulina na primeira consulta médica e  $\boldsymbol{\beta}_2^\top = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \beta_{25}, \beta_{26})$  são os coeficientes de regressão relacionados à distribuição Weibull.

Em seguida, ajustamos 18 modelos de forma independentes de acordo com os modelos descritos nas Equações (3.3), (3.8) e (3.11). O modelo, sob cada mecanismo, foi ajustado de forma independente considerando o efeito das variáveis mencionadas anteriormente. Além disso, para todos os modelos ajustados, a distribuição Weibull foi considerada para o tempo. Os resultados das estimativas dos modelos ajustados são apresentados nas Tabelas 10, 11 e 12. De acordo com os critérios de seleção AIC e BIC calculados para os modelos ajustados, portanto podemos concluir que o cenário considerando o esquema de ativação aleatória apresentou o melhor ajuste.

De acordo com os resultados, podemos obter a inferência para a influência dos fatores de risco para explicar a taxa de cura, ou seja, observamos que as variáveis diabetes familiar, hipertensão prévia e IMC pré-gestacional são importantes para isso. Por outro lado, individualmente, nenhum dos fatores de risco tem efeito significativo para descrever o uso de insulina nos primeiros sete dias, como pode ser visto nos parâmetros  $\beta_{11}$ . Ainda de acordo com a Tabela 10, 11 e 12, temos as estimativas e o intervalo de confiança (com 95%) das proporções de zeros ajustados ( $p_0$ ) e indivíduos curados ( $p_1$ ) sobre o quê as estimativas dos erros padrão para  $p_0$  e  $p_1$  foram obtidas usando o método delta com uma aproximação de Taylor de primeira ordem. Os resultados mostram que para todos os esquemas de ativação, ambas as proporções são estatisticamente significativas para o modelo. A Figura 10 mostra a função de sobrevivência estimada de acordo com os três esquemas de ativação para cada variável.

Ajustamos um modelo completo considerando todos os fatores de risco citados anteriormente. Os resultados deste modelo são dados na Tabela 13. Observe que a variável macrossomia prévia é significativa apenas para explicar o uso de insulina nos primeiros sete dias, enquanto as variáveis Histórico de diabetes e tabagismo são importantes apenas para explicar a distribuição basal.

A Tabela 14 mostra os resultados do modelo final ajustado considerando a variável macrossomia prévia como parâmetro de ligação para  $p_0$  e as variáveis história de diabetes e tabagismo como ligação para  $p_1$ . As variáveis estudadas no modelo são significativas para descrever o uso de insulina na primeira semana e a taxa de cura. Como a macrossomia fetal prévia tem efeito positivo ( $\beta_{11} > 0$ ), então a probabilidade de mulheres grávidas usarem insulina nos primeiros sete anos aumenta. Por outro lado, ainda de acordo com os resultados, o tabagismo não é importante na distribuição ao longo da vida. Por fim, a história familiar de diabetes tem efeito negativo ( $\beta_{22} < 0$ ) na distribuição ao longo da vida, ou seja, para o uso de insulina durante a gestação.

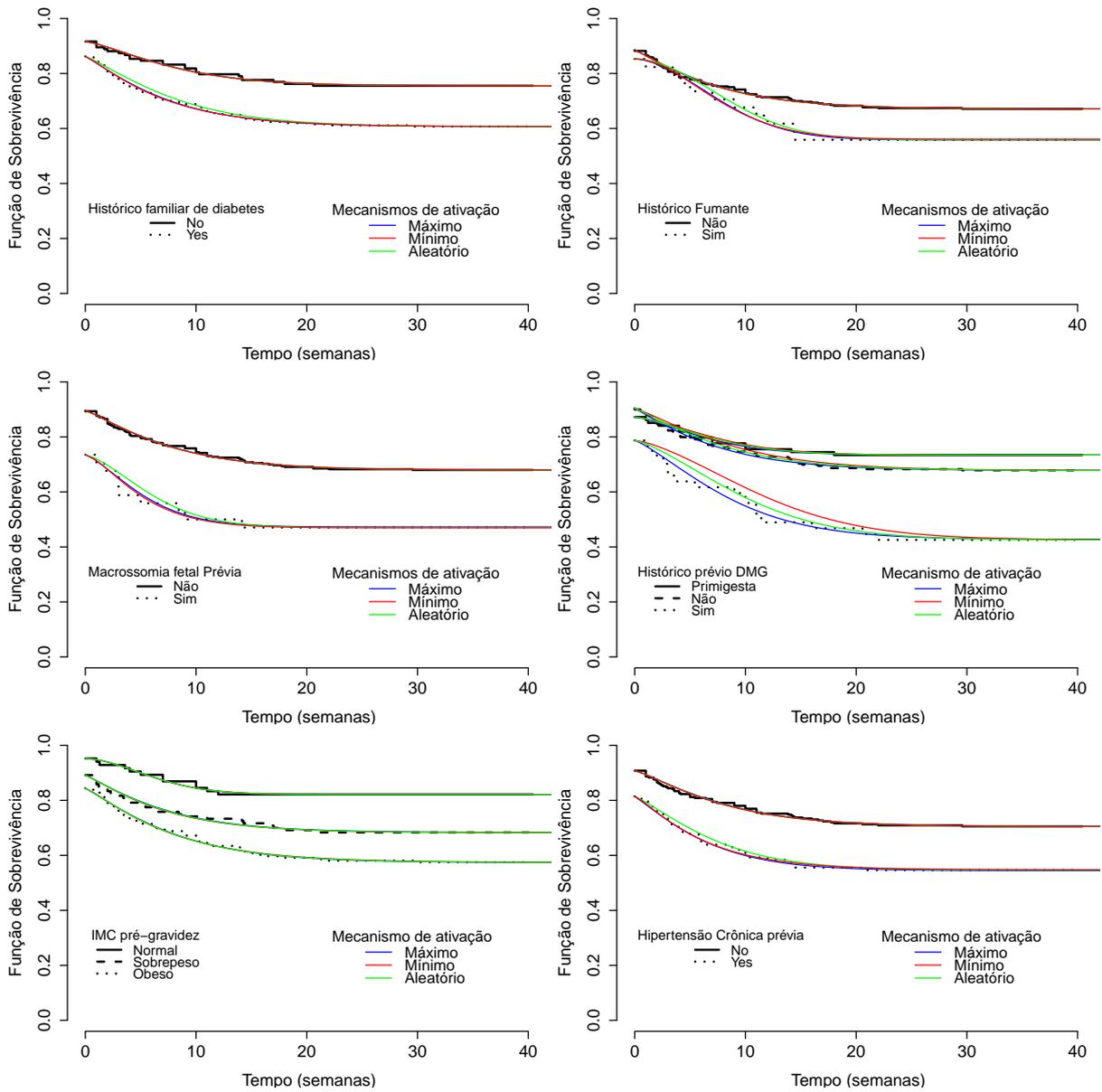


Figura 10 – Curva de sobrevida estimada obtida por Kaplan-Meier para dados de insulina considerando as variáveis história prévia de GDM, história de tabagismo, macrossomia fetal prévia, hipertensão crônica anterior, IMC pré-gravidez e história familiar de diabetes, e estimativas das respectivas funções de sobrevivência de acordo com ajuste do modelo ZICR sob diferentes de mecanismos de ativação latente

Tabela 10 – EMV, EP, IC com 95%, AIC e BIC das variáveis Família diabetes Macrosonomia prévia obtidos por ZICR sob diferentes mecanismos de ativação considerando os dados de insulina

Parâmetros	Mínimo					Máximo					Aleatório					
	EMV	Inferior	Superior	EP	EMV	Inferior	Superior	EP	EMV	Inferior	Superior	EP	EMV	Inferior	Superior	EP
$\beta_{10}$ (intercepto)	-0.650	-1.348	0.048	0.356	-0.651	-1.349	0.047	0.356	-0.653	-1.351	0.045	0.356	-0.653	-1.351	0.045	0.356
$\beta_{11}$ (família diabetes)	0.031	-0.782	0.844	0.415	0.032	-0.781	0.846	0.415	0.034	-0.779	0.848	0.415	0.034	-0.779	0.848	0.415
$\beta_{20}$ (intercepto)	1.547	1.097	1.998	0.230	1.546	1.096	1.996	0.230	1.545	1.095	1.995	0.230	1.545	1.095	1.995	0.230
$\beta_{21}$ (família diabetes)	-0.682	-1.220	-0.144	0.275	-0.681	-1.219	-0.143	0.274	-0.680	-1.218	-0.142	0.274	-0.680	-1.218	-0.142	0.274
AIC		1188.0048				1187.9497				1187.9162				1187.9162		
BIC		1219.7339				1219.6788				1219.6454				1219.6454		
$p_0$ (No)	0.084	0.039	0.129	0.023	0.084	0.039	0.129	0.023	0.084	0.039	0.129	0.023	0.084	0.039	0.129	0.023
$p_0$ (Yes)	0.138	0.095	0.181	0.022	0.138	0.095	0.181	0.022	0.138	0.095	0.181	0.022	0.138	0.095	0.181	0.022
$p_1$ (No)	0.755	0.685	0.826	0.036	0.755	0.685	0.826	0.036	0.755	0.685	0.826	0.036	0.755	0.685	0.826	0.036
$p_1$ (Yes)	0.607	0.546	0.668	0.031	0.607	0.546	0.668	0.031	0.607	0.546	0.668	0.031	0.607	0.546	0.668	0.031
$\beta_{10}$ (intercepto)	-0.736	-1.128	-0.343	0.200	-0.735	-1.127	-0.343	0.200	-0.735	-1.127	-0.343	0.200	-0.735	-1.127	-0.343	0.200
$\beta_{11}$ (macrosonomia prévia)	0.736	-0.268	1.739	0.512	0.735	-0.268	1.739	0.512	0.735	-0.268	1.739	0.512	0.735	-0.268	1.739	0.512
$\beta_{20}$ (intercepto)	1.143	0.886	1.400	0.131	1.142	0.885	1.399	0.131	1.142	0.885	1.399	0.131	1.142	0.885	1.399	0.131
$\beta_{21}$ (macrosonomia prévia)	-0.572	-1.428	0.284	0.437	-0.559	-1.414	0.297	0.436	-0.567	-1.423	0.290	0.437	-0.567	-1.423	0.290	0.437
AIC		1188.5019				1188.6249				1188.4775				1188.4775		
BIC		1220.2311				1220.3540				1220.2067				1220.2067		
$p_0$ (No)	0.104	0.072	0.135	0.016	0.104	0.073	0.135	0.016	0.104	0.073	0.135	0.016	0.104	0.073	0.135	0.016
$p_0$ (Yes)	0.265	0.116	0.414	0.076	0.264	0.117	0.411	0.075	0.265	0.116	0.414	0.076	0.265	0.116	0.414	0.076
$p_1$ (No)	0.679	0.630	0.728	0.025	0.679	0.630	0.728	0.025	0.679	0.630	0.728	0.025	0.679	0.630	0.728	0.025
$p_1$ (Yes)	0.469	0.301	0.638	0.086	0.473	0.304	0.641	0.086	0.471	0.302	0.639	0.086	0.471	0.302	0.639	0.086

Tabela 11 – EMV, EP, IC com 95%, AIC e BIC das variáveis Hipertensão prévia e fumante obtidos por ZICR sob diferentes mecanismos de ativação considerando os dados de insulina

Parâmetro	Mínimo			Máximo			Aleatório			
	EMV	Inferior	Superior	EMV	Inferior	Superior	EMV	Inferior	Superior	EP
$\beta_{10}$ (intercepto)	-0.787	-1.251	-0.324	0.237	0.237	0.237	-0.788	-1.251	-0.324	0.237
$\beta_{11}$ (hipertensão prévia)	0.415	-0.320	1.150	0.375	0.413	-0.321	0.415	-0.319	1.150	0.375
$\beta_{20}$ (intercepto)	1.246	0.952	1.541	0.150	1.251	0.956	1.247	0.953	1.542	0.150
$\beta_{21}$ (hipertensão prévia)	-0.530	-1.063	0.003	0.272	-0.550	-1.083	-0.538	-1.072	-0.005	0.272
AIC		1186.9748			1186.6558			1186.4715		
BIC		1218.7040			1218.3850			1218.2006		
$p_0$ (No)	0.092	0.059	0.126	0.017	0.092	0.017	0.092	0.059	0.125	0.017
$p_0$ (Yes)	0.184	0.112	0.257	0.037	0.186	0.037	0.185	0.113	0.258	0.037
$p_1$ (No)	0.705	0.652	0.758	0.027	0.706	0.027	0.705	0.652	0.758	0.027
$p_1$ (Yes)	0.548	0.454	0.642	0.048	0.544	0.048	0.546	0.452	0.640	0.048
$\beta_{10}$ (intercepto)	-0.620	-0.999	-0.240	0.194	-0.619	-0.999	-0.620	-1.000	-0.240	0.194
$\beta_{11}$ (fumante)	-0.074	-1.212	1.065	0.581	-0.074	-1.213	-0.071	-1.210	1.067	0.581
$\beta_{20}$ (intercepto)	1.143	0.884	1.401	0.132	1.144	0.885	1.143	0.884	1.401	0.132
$\beta_{21}$ (fumante)	-0.491	-1.298	0.315	0.411	-0.504	-1.311	-0.499	-1.308	0.309	0.412
AIC		1194.0061			1193.8797			1193.8464		
BIC		1225.7352			1225.6088			1225.5756		
$p_0$ (No)	0.115	0.082	0.148	0.017	0.115	0.082	0.115	0.082	0.148	0.017
$p_0$ (Yes)	0.146	0.029	0.264	0.060	0.147	0.028	0.147	0.028	0.267	0.061
$p_1$ (No)	0.671	0.622	0.720	0.025	0.671	0.622	0.671	0.622	0.720	0.025
$p_1$ (Yes)	0.561	0.395	0.728	0.085	0.558	0.392	0.559	0.392	0.726	0.085

Tabela 12 – EMV, EP, IC com 95% e AIC, BIC das variáveis primigesta, IMC pré-gravídes obtidos por ZICR sob diferentes mecanismos de ativação considerando os dados de insulina

Parâmetro	Mínimo					Máximo					Aleatório					
	EMV	Inferior	Superior	EP	EMV	Inferior	Superior	EP	EMV	Inferior	Superior	EP	EMV	Inferior	Superior	EP
$\beta_{10}$ (intercepto)	0.335	-0.059	0.730	0.201	0.166	-0.252	0.583	0.213	0.268	-0.130	0.665	0.203	0.048	-0.130	0.665	0.203
$\beta_{11}$ (primigesta)	0.048	-0.532	0.629	0.296	0.151	-0.451	0.753	0.307	0.084	-0.501	0.669	0.298	0.821	-0.501	0.669	0.298
$\beta_{11}$ (Não)	-0.189	-0.640	0.262	0.230	-0.100	-0.573	0.374	0.241	-0.158	-0.612	0.297	0.232	0.108	-0.612	0.297	0.232
$\beta_{20}$ (intercepto)	2.404	2.001	2.806	0.205	2.030	1.601	2.459	0.219	2.222	1.834	2.611	0.198	0.683	1.834	2.611	0.198
$\beta_{21}$ (primigesta)	-0.326	-0.892	0.239	0.288	-0.079	-0.676	0.518	0.305	-0.207	-0.767	0.353	0.286	0.683	-0.767	0.353	0.286
$\beta_{21}$ (Não)	-0.274	-0.749	0.201	0.242	-0.104	-0.605	0.396	0.255	-0.193	-0.656	0.270	0.236	0.156	-0.656	0.270	0.236
AIC																
BIC																
$p_0$ (Primigesta)	0.128	0.061	0.194	0.034	0.128	0.061	0.194	0.034	0.128	0.061	0.194	0.034	0.128	0.061	0.194	0.034
$p_1$ (Primigesta)	0.734	0.697	0.771	0.019	0.734	0.697	0.772	0.019	0.734	0.697	0.771	0.019	0.734	0.697	0.771	0.019
$p_0$ (Não)	0.096	-0.021	0.214	0.060	0.096	-0.019	0.212	0.059	0.096	-0.021	0.214	0.060	0.096	-0.021	0.214	0.060
$p_1$ (Não)	0.678	0.588	0.769	0.046	0.678	0.588	0.768	0.046	0.678	0.588	0.768	0.046	0.678	0.588	0.768	0.046
$p_0$ (Sim)	0.212	0.154	0.271	0.030	0.212	0.153	0.271	0.030	0.213	0.154	0.272	0.030	0.213	0.154	0.272	0.030
$p_1$ (Sim)	0.426	0.285	0.567	0.072	0.427	0.286	0.568	0.072	0.425	0.284	0.566	0.072	0.425	0.284	0.566	0.072
$\beta_{10}$ (intercepto)	0.185	-0.050	0.419	0.119	0.072	-0.168	0.312	0.122	0.135	-0.100	0.371	0.120	0.427	-0.100	0.371	0.120
$\beta_{21}$ (Normal)	0.394	-0.156	0.944	0.281	0.474	-0.084	1.031	0.285	0.427	-0.125	0.980	0.282	0.427	-0.125	0.980	0.282
$\beta_{11}$ (Sobrepeso)	-0.019	-0.416	0.377	0.203	0.010	-0.398	0.417	0.208	-0.009	-0.409	0.392	0.204	2.109	-0.409	0.392	0.204
$\beta_{20}$ (intercepto)	2.249	1.984	2.515	0.135	1.963	1.691	2.234	0.138	2.222	1.848	2.369	0.133	0.683	1.848	2.369	0.133
$\beta_{21}$ (Normal)	-0.230	-0.671	0.210	0.225	-0.005	-0.453	0.442	0.228	-0.121	-0.559	0.317	0.224	0.683	-0.559	0.317	0.224
$\beta_{21}$ (Sobrepeso)	-0.141	-0.594	0.311	0.231	-0.045	-0.515	0.425	0.240	-0.094	-0.546	0.359	0.231	0.683	-0.546	0.359	0.231
AIC																
BIC																
$p_0$ (Normal)	0.048	0.002	0.093	0.023	0.048	0.003	0.093	0.023	0.048	0.003	0.093	0.023	0.048	0.003	0.093	0.023
$p_1$ (Normal)	0.822	0.767	0.877	0.028	0.821	0.766	0.876	0.028	0.821	0.767	0.876	0.028	0.821	0.767	0.876	0.028
$p_0$ (Sobrepeso)	0.109	0.056	0.161	0.027	0.108	0.055	0.161	0.027	0.108	0.055	0.161	0.027	0.108	0.055	0.161	0.027
$p_1$ (Sobrepeso)	0.683	0.600	0.765	0.042	0.684	0.602	0.766	0.042	0.683	0.601	0.765	0.042	0.683	0.601	0.765	0.042
$p_0$ (Obesa)	0.156	0.071	0.240	0.043	0.156	0.074	0.238	0.042	0.156	0.072	0.240	0.043	0.156	0.072	0.240	0.043
$p_1$ (Obesa)	0.575	0.504	0.646	0.036	0.574	0.504	0.645	0.036	0.574	0.504	0.645	0.036	0.574	0.504	0.645	0.036

Tabela 13 – EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR sob diferentes mecanismos de ativação considerando conjunto de dados de insulina

Parâmetro	Mínimo			Máximo			Aleatório			
	EMV	Inferior	Superior	EMV	Inferior	Superior	EMV	Inferior	Superior	EP
$\beta_{10}$ (intercepto)	-1.405	-2.542	-0.268	-0.961	-2.087	0.166	-0.987	-2.103	0.129	0.569
$\beta_{11}$ (diabetes prévia)	0.414	-0.421	1.250	0.036	-0.794	0.865	0.058	-0.768	0.884	0.421
$\beta_{12}$ (Fumante)	0.197	-0.980	1.375	-0.268	-1.410	0.873	-0.245	-1.381	0.892	0.580
$\beta_{13_1}$ (primigesta)	1.081	-0.110	2.272	0.343	-0.899	1.585	0.245	-0.977	1.467	0.623
$\beta_{13_2}$ (Não)	-0.221	-1.188	0.747	-0.405	-1.389	0.579	-0.474	-1.457	0.510	0.502
$\beta_{14_1}$ (normal)	-0.810	-2.103	0.483	-0.077	-1.381	1.228	-0.009	-1.317	1.298	0.667
$\beta_{14_2}$ (Sobrepeso)	-0.126	-0.975	0.722	0.161	-0.698	1.020	0.170	-0.682	1.022	0.435
$\beta_{15}$ (Macrossomia fetal anterior)	1.032	-0.033	2.097	1.047	-0.026	2.119	1.073	0.005	2.140	0.545
$\beta_{16}$ (Hipertensão crônica)	0.494	-0.300	1.288	0.208	-0.584	0.999	0.258	-0.532	1.048	0.403
$\beta_{20}$ (Intercepto)	-0.350	-1.456	0.756	0.456	-0.438	1.351	0.382	-0.505	1.268	0.452
$\beta_{21}$ (diabetes prévia)	-0.197	-0.824	0.429	-0.799	-1.442	-0.156	-0.791	-1.429	-0.153	0.325
$\beta_{22}$ (Fumante)	-0.341	-1.209	0.528	-1.298	-2.589	-0.007	-1.300	-2.578	-0.022	0.652
$\beta_{23_1}$ (primigesta)	1.990	0.864	3.117	0.838	-0.353	2.030	0.718	-0.439	1.876	0.591
$\beta_{23_2}$ (Não)	1.195	0.251	2.139	0.747	-0.092	1.586	0.682	-0.166	1.530	0.432
$\beta_{24_1}$ (normal)	0.368	-0.495	1.230	1.396	0.417	2.376	1.538	0.532	2.544	0.513
$\beta_{24_2}$ (Sobrepeso)	0.167	-0.458	0.792	0.635	-0.076	1.347	0.685	-0.026	1.396	0.363
$\beta_{25}$ (Macrossomia fetal anterior)	0.172	-0.757	1.101	0.175	0.001	0.349	0.275	-0.712	1.261	0.503
$\beta_{26}$ (Hipertensão crônica)	-0.139	0.386	-0.664	-0.658	-0.983	-0.333	-0.605	-1.334	0.124	0.372
AIC		1181.7420			1191.4410			1189.2750		
BIC		1324.5230			1334.2220			1332.0570		

Tabela 14 – EMV, EP, IC com 95%, AIC e BIC obtidos pelo modelo ZICR sob diferentes mecanismos de ativação ajustados para o conjunto de dados insulina

Parâmetros	Mínimo			
	EMV	Inferior	Superior	EP
$\beta_{10}$	-0.770	-1.157	-0.383	0.197
$\beta_{11}$ (Prior macrosomia)	1.118	0.279	1.958	0.428
$\beta_{20}$	1.582	1.175	1.989	0.207
$\beta_{21}$ (Smoking)	-0.447	-1.168	0.275	0.368
$\beta_{22}$ (History diabetes)	-0.675	-1.136	-0.214	0.235
AIC	1205.4			
BIC	1260.9			
Parâmetros	Máximo			
	EMV	Inferior	Superior	EP
$\beta_{10}$	-0.769	-1.156	-0.382	0.197
$\beta_{11}$ (Prior macrosomia)	1.111	0.272	1.950	0.428
$\beta_{20}$	1.586	1.178	1.995	0.209
$\beta_{21}$ (Smoking)	-0.493	-1.216	0.230	0.369
$\beta_{22}$ (History diabetes)	-0.674	-1.135	-0.212	0.235
AIC	1187.257			
BIC	1238.8171			
Parâmetros	Aleatório			
	EMV	Inferior	Superior	EP
$\beta_{10}$	-0.769	-1.156	-0.383	0.197
$\beta_{11}$ (Prior macrosomia)	1.115	0.275	1.954	0.428
$\beta_{20}$	1.584	1.176	1.991	0.208
$\beta_{21}$ (Smoking)	-0.473	-1.196	0.250	0.369
$\beta_{22}$ (History diabetes)	-0.674	-1.135	-0.213	0.235
AIC	1187.149			
BIC	1238.709			

### 3.7.2 Hemorragia intracraniana em pacientes com câncer

Este conjunto de dados foi inicialmente estudado por Yamaki *et al.* (2021) o qual refere-se a um estudo retrospectivo de todos os procedimentos cirúrgicos consecutivos para o tratamento de Hemorragia intracraniana (em inglês ICH) em pacientes com câncer admitidos a partir de 2008 no Instituto do Câncer do Estado de São Paulo – Universidade de São Paulo, Brasil (ICESP). Neste estudo foram incluídos pacientes com hemorragias intraparenquimatosas (HPI), hematomas subdurais crônicos (CSDH), hematomas subdurais agudos (HSDA) e hematomas epidurais submetidos a procedimentos neurocirúrgicos de emergência.

Além disso, os autores excluíram complicações hemorrágicas no pós-operatório imediato de tumores cerebrais. O objetivo, desse estudo, é analisar quais fatores estão associados à sobrevida e/ou óbito do paciente após a cirurgia. Em 38 casos, os pacientes vieram a óbito no mesmo dia de admissão antes da cirurgia, então esta ocorrência será considerada como tendo um

tempo de sobrevivência igual a zero. Por outro lado, outra parcela dos pacientes não faleceu após a cirurgia, o que consideramos como cura.

A Hemorragia intracraniana apresenta situações as quais o evento de interesse tem riscos competitivos para a ocorrência do mesmo. Segundo [Caceres e Goldstein \(2012\)](#), os fatores de risco mais importantes para a ocorrência deste evento de interesse incluem hipertensão (HTN) e angiopatia amiloide cerebral. Eles destacam ainda que ICH pode estar relacionada aos efeitos não observáveis da ingestão diária de álcool de forma excessiva, metástase, efeitos do colesterol alto, genética, anticoagulação, entre outras causas. Diante desse cenário, através do modelo proposto, nós consideramos os esquemas de ativação máximo, mínimo e aleatório para entender o mecanismo de ativação do evento de interesse.

Como nosso evento de interesse é o óbito em que analisamos o tempo entre a data após a cirurgia e a morte do paciente, nosso objetivo é verificar o efeito das variáveis descritas na Tabela 15 a qual temos as frequências observadas de cada variável mencionada anteriormente.

Tabela 15 – Frequência observada das variáveis presente no banco de dados dos pacientes com câncer e hemorragia intracraniana

Variável	N (%)
Gênero	
Masculino	56 (60,2)
Feminino	37 (39,8)
Primary Site	
0 - Others	80 (85,1)
1- Hemato	14 (14,9)
Primary site	
0 - Primary CNS	14 (14,9)
1 - Others	80 (85,1)
Metástase	
Não	45 (47,9)
Sim	49 (52,1)
CNS Metastasis (0-No/1- Yes)	
Não	69 (73,4)
Sim	25 (26,6)
Não CNS nem meta	
Não	39 (38,9)
Sim	55 (61,1)
Idade (em anos)	média = 62.71; D.P = 4.15)

Portanto, nós analisaremos se essas variáveis influenciam as chances do paciente tanto no fato de vir a óbito no mesmo dia da admissão antes da cirurgia quanto na fração de pacientes curados após a cirurgia. Os parâmetros  $p_0$  e  $p_1$  foram conectados às covariáveis por meio das

seguintes equações, respectivamente:

$$p_0 = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta}_1)}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_1) + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_2)}, \quad e \quad p_1 = \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta}_2)}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_1) + \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_2)},$$

em que  $\mathbf{x}_i^\top = (x_{1i}, x_{2i}, x_{3i}, x_{4i}, x_{5i}, x_{6i}, x_{7i})$  e está associado ao vetor de covariáveis, para  $i = 1, \dots, 93$ . Nós também temos que  $\boldsymbol{\beta}_1^\top = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}, \beta_{16})$  são os coeficientes de regressão relacionados aos pacientes que faleceram antes de realizar a cirurgia e  $\boldsymbol{\beta}_2^\top = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \beta_{25}, \beta_{26})$  são os coeficientes de regressão relacionados à distribuição ao longo da vida.

Inicialmente ajustamos a estimativa de Kaplan-Meier da função de sobrevivência dos dados. No gráfico 11, observamos que os dados são apropriados para a metodologia proposta uma vez que temos a fração de zero e de cura. Diferentemente da aplicação anterior, sobre pacientes em gestação com diabetes mellitus gestacional, nós temos uma alta taxa de zero ajustado, ou seja, muitos pacientes morreram na mesma data de internação e baixa fração de cura, ou seja, após a cirurgia baixo índice de curados após 5 anos.

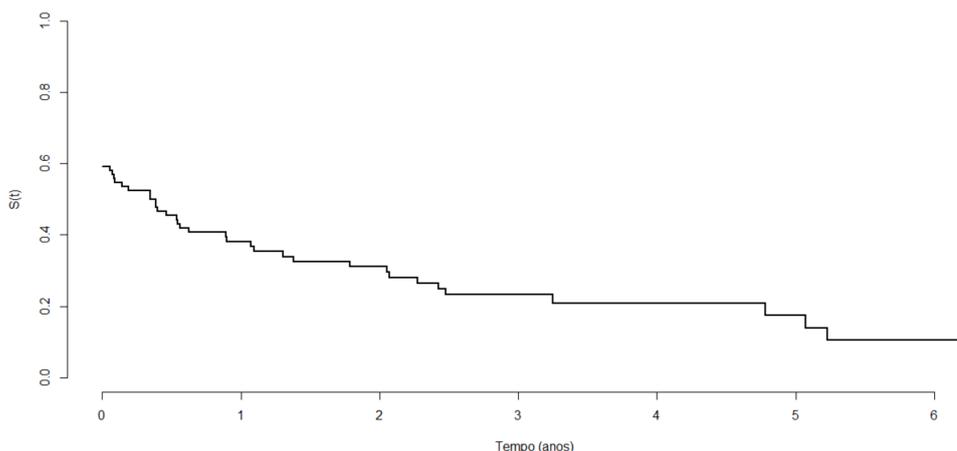


Figura 11 – Curva de sobrevivência estimada obtida pelo método de Kaplan-Meier para dados de hemorragia intracraniana

Assim como na subseção anterior, nós ajustamos 18 modelos de forma independentes baseados nos modelos expressados nas Equações (3.3), (3.8) e (3.11). O modelo, sob cada esquema de ativação, foi ajustado considerando o efeito das variáveis descritas previamente. A distribuição Weibull foi considerada para o tempo para todos os modelos ajustados. Os resultados das estimativas dos modelos ajustados são mostrados nas Tabelas 16, 17 e 18.

Baseado nos critérios de seleção AIC e BIC calculados para cada modelo ajustado, percebemos em geral que o esquema de ativação máximo obteve menor valor e portanto podemos concluir que este cenário apresentou o melhor ajuste. A Figura 12 mostra a função de sobrevivência estimada baseado em cada esquema de ativação para cada variável ajustada de formas independentes. Observamos que em geral os esquemas de ativação mínimo e máximo obtiveram boas estimativas quando comparado com a curva de Kaplan Meier. Por outro lado, as

estimativas considerando o esquema de ativação aleatório, em algumas variáveis, destoou um pouco. Adicionalmente, visualizamos, por exemplo, na variável CNS Metástase que o modelo proposto se ajustou bem nas situações em que existia fração de cura e outra parcela não tinha fração de cura,

De acordo com os resultados, podemos obter a inferência para a influência dos fatores de risco para explicar a taxa de cura, ou seja, observamos que nenhuma das variáveis foram significativas para explicar a fração de zero ajustado, isto é, nenhuma das 7 variáveis analisadas estão associadas ao evento de ter morrido antes da cirurgia. Os resultados obtidos são razoáveis uma vez que o evento de interesse é de alta gravidade e provavelmente não tem variáveis suficientes que explicam o tempo até o evento.

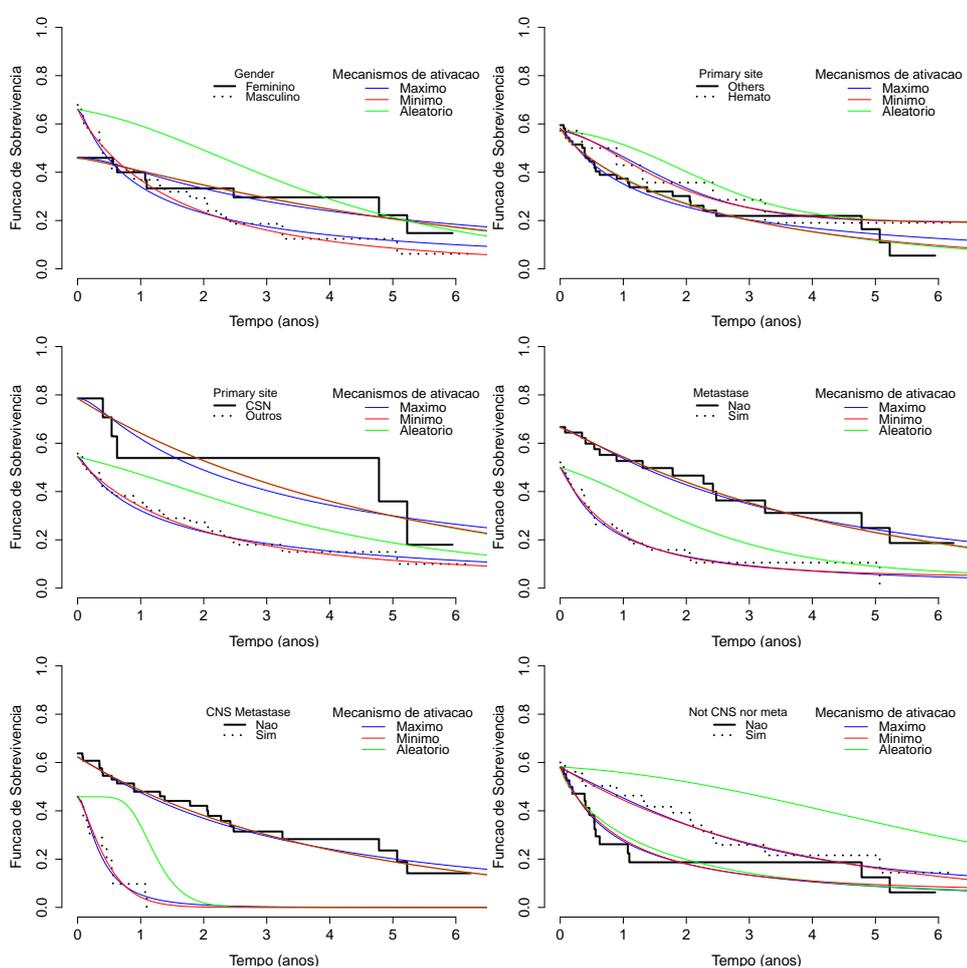


Figura 12 – Curva de sobrevivida estimada obtida por Kaplan-Meier para dados de Hemorragia intracraniana considerando as variáveis e as respectivas estimativas da função de sobrevivência de acordo com ajuste do modelo ZICR sob diferentes de mecanismos de ativações latente

Tabela 16 – EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR das variáveis Gênero, e Metástase sob diferentes mecanismos de ativação considerando os dados de HI.

Parâmetros	Mínimo						Máximo						Aleatório					
	EMV	LI	LS	DP	EMV	LI	LS	DP	EMV	LI	LS	DP	EMV	LI	LS	DP		
$\beta_{10}$ (Intercepto)	0.1640	-0.4900	0.8180	0.3340	0.1630	-0.4840	0.8090	0.3300	0.1640	-0.5030	0.8320	0.3410						
$\beta_{11}$ (Gênero)	-0.8220	-1.6920	0.0480	0.4440	-0.8290	-1.6800	0.0220	0.4340	-0.7950	-1.7000	0.1090	0.4620						
$\beta_{20}$ (Intercepto)	-6.5560	-74.7840	61.6710	34.8100	<0.00001	-328.2580	232.6950	143.1000	-6.2480	-91.9670	79.4710	43.7340						
$\beta_{21}$ (Gênero)	1.8260	-68.6130	72.2660	35.9390	-8.2730	-956.0350	939.4900	483.5520	2.9270	-83.1140	88.9680	43.8980						
AIC			261.6428				260.1182				261.7472							
BIC			281.9036				280.3790				282.0080							
$p_{1 y=0}$	0.5406	0.3799	0.7013	0.0820	0.5406	0.3798	0.7013	0.0820	0.5405	0.3798	0.7013	0.0820						
$p_{1 y=1}$	0.3392	0.2158	0.4627	0.0630	0.3393	0.2158	0.4628	0.0630	0.3393	0.2158	0.4628	0.0630						
$p_{1 y=0}$	0.0007	-0.0444	0.0457	0.0230	0.0000	0.0000	0.0000	<0.001	0.0009	-0.0756	0.0773	0.0390						
$p_{1 y=1}$	0.0058	-0.0942	0.1057	0.0510	0.0000	0.0000	0.0000	<0.001	0.0230	-0.1416	0.1877	0.0840						
$\beta_{10}$ (Intercepto)	-0.6930	-1.3130	-0.0730	0.3160	-0.6920	-1.3120	-0.0730	0.3160	-0.6920	-1.3230	-0.0610	0.3220						
$\beta_{11}$ (Metástase)	0.7810	-0.0900	1.6530	0.4450	0.6920	-0.1470	1.5310	0.4280	0.7800	-0.1040	1.6650	0.4510						
$\beta_{20}$ (Intercepto)	-11.5180	-139.9750	116.9380	65.5390	-8.6560	-37.8410	20.5300	14.8910	-6.7120	<0.00001	91.5990	50.1590						
$\beta_{21}$ (Metástase)	9.1510	-119.3330	137.6350	65.5530	<0.00001	-163.3690	107.9210	69.2070	4.3260	-94.0310	102.6840	50.1820						
AIC			261.2698				259.9545				262.1183							
BIC			281.5306				280.2153				282.3791							
$p_{1 y=0}$	0.3333	0.1962	0.4705	0.0700	0.3335	0.1963	0.4707	0.0700	0.3333	0.1961	0.4705	0.0700						
$p_{1 y=1}$	0.4997	0.3585	0.6408	0.0720	0.4999	0.3588	0.6411	0.0720	0.5001	0.3589	0.6412	0.0720						
$p_{1 y=0}$	0.000007	0.000007	0.000007	<0.001	0.0001	-0.0038	0.0040	0.0020	0.0008	-0.0795	0.0812	0.0410						
$p_{1 y=1}$	0.0429	-0.0630	0.1487	0.0540	0.0000	0.0000	0.0000	<0.001	0.0421	-0.0735	0.1577	0.0590						

Tabela 17 – EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR das variáveis Not CNS nor Meta e Others Primary Site sob diferentes mecanismos de ativação considerando os dados de HI.

Parâmetros	Mínimo				Máximo				Aleatório			
	EMV	LI	LS	DP	EMV	LI	LS	DP	EMV	LI	LS	DP
$\beta_{10}$ (Intercepto)	-0.2030	-0.9060	0.4990	0.3590	-0.3180	-0.9620	0.3260	0.3290	-0.2310	-0.9510	0.4890	0.3670
$\beta_{11}$ (Not CNS nor Meta)	-0.1270	-1.0110	0.7570	0.4510	0.1470	-0.8220	1.1170	0.4950	-0.0460	-1.3010	1.2100	0.6410
$\beta_{20}$ (Intercepto)	-2.0880	-4.5570	0.3810	1.2600	<0.00001	-275.3680	160.7710	111.2600	-2.3910	-6.2230	1.4400	1.9550
$\beta_{21}$ (Not CNS nor Meta)	-7.0270	-131.3810	117.3260	63.4460	55.5670	-162.5270	273.6610	111.2720	-0.5030	-17.6860	16.6800	8.7670
AIC		268.9871				266.6630				270.0002		
BIC		289.2479				286.9238				290.2610		
$p_{0 y=0}$	0.4206	0.2638	0.5774	0.0800	0.4211	0.2643	0.5779	0.0800	0.4211	0.2643	0.5779	0.0800
$p_{0 y=1}$	0.4182	0.2869	0.5495	0.0670	0.4174	0.2880	0.5467	0.0660	0.4182	0.2869	0.5495	0.0670
$p_{1 y=0}$	0.0639	-0.0772	0.2050	0.0720	0.0000	0.0000	0.0000	<0.001	0.0485	-0.1220	0.2191	0.0870
$p_{1 y=1}$	0.0001	-0.0078	0.0079	0.0040	0.0876	-0.1555	0.3306	0.1240	0.0305	-0.4536	0.5146	0.2470
$\beta_{10}$ (Intercepto)	-0.3300	-0.7790	0.1180	0.2290	-0.3320	-0.7790	0.1150	0.2280	-0.3320	-0.7790	0.1150	0.2280
$\beta_{11}$ (Others Primary Site)	0.4420	-0.8350	1.7200	0.6520	0.4490	-0.8200	1.7190	0.6480	0.4510	-0.8180	1.7200	0.6470
$\beta_{20}$ (Intercepto)	-7.2180	-58.4000	43.9630	26.1130	<0.00001	-162.6310	103.8350	67.9760	-11.5520	<0.00001	521.5720	272.0020
$\beta_{21}$ (Others Primary Site)	6.5200	-44.6910	57.7300	26.1280	28.7260	-104.5170	161.9690	67.9810	10.8620	<0.00001	543.9890	272.0040
AIC		271.9602				270.9596				271.8818		
BIC		292.2210				291.2204				292.1426		
$p_{0 y=0}$	0.4180	0.3102	0.5258	0.0550	0.4177	0.3099	0.5255	0.0550	0.4177	0.3099	0.5255	0.0550
$p_{0 y=1}$	0.4276	0.1689	0.6863	0.1320	0.4267	0.1680	0.6854	0.1320	0.4286	0.1699	0.6873	0.1320
$p_{1 y=0}$	0.0004	-0.0211	0.0220	0.0110	0.0000	0.0000	0.0000	<0.00001	0.0000	-0.0039	0.0039	0.0020
$p_{1 y=1}$	0.1901	-0.0471	0.4273	0.1210	0.1938	-0.0316	0.4192	0.1150	0.1909	-0.0306	0.4124	0.1130

Tabela 18 – EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR das variáveis CSNS Metastasi, CSNS Primary Site e Idade sob diferentes mecanismos de ativação considerando os dados de HI.

Parâmetros	Mínimo						Máximo						Aleatório					
	EMV	LI	LS	DP	EMV	LI	LS	DP	EMV	LI	LS	DP	EMV	LI	LS	DP		
$\beta_{10}$ (Intercepto)	-0.5030	-0.9900	-0.0160	0.2480	-0.4990	-0.9860	-0.0130	0.2480	-0.5030	-0.9910	-0.0150	0.2490	-0.5030	-0.9910	-0.0150	0.2490		
$\beta_{11}$ (CNS Metastasis)	0.6700	-0.2690	1.6090	0.4790	0.6600	-0.2780	1.5990	0.4790	0.6700	-0.2690	1.6100	0.4790	0.6700	-0.2690	1.6100	0.4790		
$\beta_{20}$ (Intercepto)	-26.2600	-419.9520	367.4320	<0.00001	-7.7960	-25.0370	9.4440	8.7960	-8.8310	<0.00001	174.2450	93.4060	-8.8310	<0.00001	174.2450	93.4060		
$\beta_{21}$ (CNS Metastasis)	-16.0640	-583.4470	551.3190	<0.00001	-1.4730	-63.5360	60.5910	31.6650	-4.8380	<0.00001	610.8690	314.1360	-4.8380	<0.00001	610.8690	314.1360		
AIC	251.2723																	
BIC	271.5331																	
$p_0   y=0$	0.3768	0.2631	0.4905	0.0580	0.3776	0.2639	0.4912	0.0580	0.3768	0.2631	0.4905	0.0580	0.3768	0.2631	0.4905	0.0580		
$p_0   y=1$	0.5417	0.3418	0.7416	0.1020	0.5401	0.3402	0.7401	0.1020	0.5417	0.3418	0.7416	0.1020	0.5417	0.3418	0.7416	0.1020		
$p_1   y=0$	0.0000	0.0000	0.0000	<0.001	0.0003	-0.0037	0.0042	0.0020	0.0001	-0.0175	0.0177	0.0090	0.0001	-0.0175	0.0177	0.0090		
$p_1   y=1$	0.0000	0.0000	0.0000	<0.001	0.0000	-0.0019	0.0020	0.0010	0.0000	0.0000	0.0000	<0.001	0.0000	0.0000	0.0000	<0.001		
$\beta_{10}$ (Intercepto)	-1.2990	-2.5760	-0.0230	0.6510	-1.2990	-2.5760	-0.0230	0.6510	-1.2990	-2.5770	-0.0220	0.6520	-1.2990	-2.5770	-0.0220	0.6520		
$\beta_{11}$ (CNS Primary Site)	1.1960	-0.2070	2.5990	0.7160	1.1210	-0.2300	2.4730	0.6890	1.2500	-0.1450	2.6450	0.7120	1.2500	-0.1450	2.6450	0.7120		
$\beta_{20}$ (Intercepto)	-18.2430	-242.2440	205.7590	<0.00001	<0.00001	-438.0310	336.2930	197.5320	-8.4190	<0.00001	148.6210	80.1220	-8.4190	<0.00001	148.6210	80.1220		
$\beta_{21}$ (CNS Primary Site)	15.6900	-208.3720	239.7520	<0.00001	7.0740	-558.6200	572.7670	288.6190	6.4320	<0.00001	163.4970	80.1350	6.4320	<0.00001	163.4970	80.1350		
AIC	268.7753																	
BIC	289.0361																	
$p_{0 y=0}$	0.2143	-0.0013	0.4299	0.1100	0.2143	-0.0013	0.4299	0.1100	0.2142	-0.0014	0.4298	0.1100	0.2142	-0.0014	0.4298	0.1100		
$p_{0 y=1}$	0.4556	0.3459	0.5654	0.0560	0.4557	0.3459	0.5654	0.0560	0.4557	0.3459	0.5654	0.0560	0.4557	0.3459	0.5654	0.0560		
$p_{1 y=0}$	0.0000	0.0000	0.0000	<0.001	0.0000	0.0000	0.0000	<0.001	0.0002	-0.0273	0.0276	0.0140	0.0002	-0.0273	0.0276	0.0140		
$p_{1 y=1}$	0.0393	-0.1508	0.2294	0.0970	0.0000	0.0000	0.0000	<0.001	0.0657	-0.1009	0.2323	0.0850	0.0657	-0.1009	0.2323	0.0850		
$\beta_{10}$ (Intercepto)	-0.2770	-0.7000	0.1460	0.2160	-0.2780	-0.6990	0.1430	0.2150	-0.2560	-0.8200	0.3090	0.2880	-0.2560	-0.8200	0.3090	0.2880		
$\beta_{11}$ (Idade)	-0.3250	-0.7950	0.1450	0.2400	-0.3190	-0.7850	0.1460	0.2380	-0.1750	-0.6250	0.2760	0.2300	-0.1750	-0.6250	0.2760	0.2300		
$\beta_{20}$ (Intercepto)	-8.6930	-21.5430	4.1580	6.5570	<0.00001	-136.9600	38.1870	44.6800	-2.6050	-8.2520	3.0430	2.8810	-2.6050	-8.2520	3.0430	2.8810		
$\beta_{21}$ (Idade)	-3.8060	-9.0330	1.4210	2.6670	<0.00001	-52.7010	13.8860	16.9860	0.1740	-2.7460	3.0950	1.4900	0.1740	-2.7460	3.0950	1.4900		
AIC	269.8195																	
BIC	290.0803																	
	286.9198																	
	287.9953																	
	267.7345																	
	289.1294																	
	272.8075																	
	271.4817																	
	252.5467																	
	251.2210																	
	272.8075																	
	271.4817																	
	252.5467																	
	251.2210																	
	266.6590																	
	292.3194																	

## 3.8 Conclusão

Neste capítulo, foi realizado um estudo de simulação para avaliar o desempenho do modelo de fração de cura inflacionado em zero com três esquemas de ativação diferentes. Utilizamos a distribuição Weibull para modelar o tempo de vida e consideramos diferentes proporções de inflação de zero e cura. Os resultados mostraram um bom desempenho assintótico do modelo.

O modelo proposto para diferentes esquemas de ativações foram aplicados em dois conjuntos de dados: o primeiro foi sobre a necessidade de utilização da insulina em 390 mulheres gestantes que foram diagnosticadas com diabetes gestacional antes das 24 semanas de gravidez, em que os modelos estudados tiveram uma bom ajuste com relação aos dados aplicados. Com isso, considerando o modelo sob esquema de ativação aleatório, percebemos que a fração de curados estimada para o grupo de mulheres grávidas sem presença de diabético na família foi de aproximadamente 75% enquanto que para as que tinha diabético na família de de 61%. Com relação de gestantes que apresentaram o evento de interesse (uso de insulina) antes de começar o estudo, isto é, a proporção de inflação de zeros, os valores foram, respectivamente, 8% e 14% para os grupos sem e com diabetes na família.

A segunda aplicação foi referente a um estudo sobre hemorragia intracraniana em que analisamos fatores de riscos que poderiam estar associados a fração de cura e aos pacientes que morreram na mesma data de admissão. Nós concluímos que, considerando a inferência clássica, as variáveis estudadas não apresentaram, individualmente, significância estatística e portanto não foi ajustado um modelo final considerando diferentes covariáveis..



---

# ABORDAGEM BAYESIANA DO MODELO ZICR SOB DIFERENTES ESQUEMAS DE ATIVAÇÃO

---

Neste Capítulo os modelos de fração de cura zero inflacionados sob os esquemas de ativação mínimo, máximo e aleatório propostos no Capítulo 3 são analisados sob a perspectiva Bayesiana para estimar os parâmetros de interesse. Para a aplicabilidade da metodologia proposta nós consideramos o conjunto de dados reais sobre mulheres gestantes com diabetes gestacional apresentado no subseção 3.7 do Capítulo 3.

## 4.1 Introdução

Análise Bayesiana é uma estrutura atraente em problemas práticos e tem crescido em popularidade nos últimos anos. Aqui, assumimos que o leitor possui um conhecimento básico sobre procedimentos Bayesianos. Para uma visão geral das técnicas Bayesianas, o leitor pode consultar [Migon, Gamerman e Louzada \(2014\)](#).

Nesta abordagem os parâmetros são tratados como variáveis aleatórias, a distribuição associada a tais variáveis é conhecida como distribuição a *priori*. A distribuição a *priori* é uma parte fundamental da inferência Bayesiana e existem diferentes tipos de distribuição a *priori* disponíveis na literatura. A informação prévia pode ser proveniente de estudos anteriores, conhecimento de especialistas, ou mesmo de dados históricos. A partir da combinação da informação a *priori* com a informação dos dados, a abordagem Bayesiana permite obter uma distribuição a *posteriori* dos parâmetros do modelo.

Nos modelos de fração de cura zero inflacionado, a abordagem Bayesiana permite obter a distribuição *posteriori* dos parâmetros que descrevem a proporção de pacientes com tempo iguais a zero, fração de curados e as funções de sobrevivência e densidade de probabilidade

para os pacientes suscetíveis. Além disso, essa abordagem permite realizar inferências sobre as diferenças entre grupos ou tratamentos, e incorporar covariáveis que possam estar relacionadas com a proporção de curados ou com a sobrevivência dos pacientes.

Neste capítulo, apresentamos os resultados da análise Bayesiana do modelo ZICR sob diferentes esquemas de ativação em que os parâmetros de interesse foram estimados utilizando a técnica de Monte Carlo Markov Chains (MCMC), que é uma ferramenta computacional para simular a distribuição *a posteriori* dos parâmetros. Adicionalmente, nós realizamos estudos de simulação com diferentes cenários para os parâmetros e tamanho de amostra, a fim de avaliar a performance de precisão e acurácia das estimativas.

Por fim, o conjunto de dados sobre diabetes gestacional utilizado no Capítulo 3 foi analisado, ilustrando a aplicação prática da metodologia proposta.

## 4.2 Distribuições *a priori* e *a posteriori*

Na abordagem bayesiana, os parâmetros são tratados como variáveis aleatórias e a inferência do vetor de parâmetros é realizada por meio da distribuição *a posteriori*, obtida a partir da distribuição *a priori* e da função de verossimilhança dos dados observados.

Considerando o modelo ZICR do Capítulo 3, temos que a distribuição *a priori*, representada por  $\pi(p_0, p_1, \boldsymbol{\omega})$ , para os parâmetros  $p_0$ ,  $p_1$  e  $\boldsymbol{\omega}$  estão relacionados, respectivamente, a fração de zeros, cura e o vetor de parâmetros da distribuição Weibull a qual representa o tempo de sobrevivência. Considerando  $p_0$  e  $p_1$  independentes de  $\boldsymbol{\omega}$ , nós podemos obter a distribuição *a priori* conjunta dada por

$$\pi(p_0, p_1, \boldsymbol{\omega}) = \pi(p_1|p_0)\pi(p_0)\pi(\boldsymbol{\omega}).$$

A escolha de uma distribuição *a priori* para  $\pi(p_0, p_1|\boldsymbol{\omega}, \mathbb{D})$  não é trivial. Em situações em que não há muitas informações prévias disponíveis, pode ser interessante especificar uma distribuição *a priori* não informativa, isto é, que coloca a maior ênfase na distribuição *a posteriori* resultante dos dados observados. O exemplo mais conhecido de distribuição *a priori* não informativa é a de Jeffreys (Jeffreys (1998)) em que esta é obtida utilizando a matriz de informação de Fisher a qual em muitos casos não possui uma forma fechada mesmo considerando distribuições de probabilidade que são algebricamente tratáveis como por exemplo, Log-normal, Gama, Weibull, entre outras. Nesse sentido, um exemplo de *a priori* para o parâmetro  $p_0$  seria a de Jeffrey, isto é,  $p_0 \propto \frac{1}{\sqrt{p_0(1-p_0)}}$  e uma distribuição *a priori* para  $p_1$  é proveniente de uma distribuição Uniforme  $(0, 1 - p_0)$  uma vez que  $0 < p_1 < 1 - p_0$  (mais detalhes ver Júnior (2016)).

A distribuição *a posteriori* conjunta de  $p_0, p_1$  e  $\boldsymbol{\omega}$ , dado as informações provenientes dos dados é representada por

$$\pi(p_0, p_1, \boldsymbol{\omega}|\mathbb{D}) \propto \pi(p_0, p_1, \boldsymbol{\omega}) \mathcal{L}_{ZICR}(\boldsymbol{\zeta}, \mathbb{D}). \quad (4.1)$$

No modelo ZICR, a função de verossimilhança  $\mathcal{L}_{ZICR}(\zeta, \mathbb{D})$  é dada por (3.15). Substituindo a função de verossimilhança (3.15) em (4.1), nós obtemos a expressão da distribuição a posteriori conjunta, dada por:

$$\begin{aligned} \pi(p_0, p_1, \omega | \mathbb{D}) &\propto \pi(p_0, p_1, \omega) \prod_{t_i=0} \{p_0\} \prod_{t_i>0} [(1-p_0-p_1)f_p^*(t_i)]^{\delta_i} \times \\ &\times [p_1 + (1-p_0-p_1)S_p^*(t_i)]^{1-\delta_i}, \end{aligned} \quad (4.2)$$

em que as expressões de  $f_p^*(t)$  e  $S_p^*(t)$  são descritas na Tabela 4 para os modelos ZICR sob os esquemas de ativação mínimo, máximo e aleatório.

Nós consideramos a mesma estrutura proposta pelo Júnior (2016) e descrito na subseção 2.2.1, isto é, seja o modelo de regressão ZICR definido através dos componentes sistemáticos representado pela equação (2.12) em que as funções de ligações  $H$  e  $g_j$  são definidas através de uma regressão logística multinomial representada pela expressão (2.13).

A distribuição Weibull foi considerada para o tempo, então as funções de ligações  $g_1$  e  $g_2$  utilizadas para os parâmetros da Weibull foram  $g_1(\omega_{1i}) = \log(\omega_{1i})$  e  $g_2(\omega_{2i}) = \log(\omega_{2i})$ . Sendo assim,  $\omega_{1i} = \exp\{x_{3i}^\top \beta_3\}$  e  $\omega_{2i} = \exp\{x_{4i}^\top \beta_4\}$ . Por último, nós estabelecemos a relação entre os parâmetros  $p_0$  e  $p_1$  através da expressão (2.14). Dessa forma considerando a estrutura descrita anteriormente entre os parâmetros do modelo e covariáveis, nós temos que a distribuição a posteriori (4.2) pode ser rescrita como:

$$\begin{aligned} \pi(\beta | \mathbb{D}, x) &\propto \pi(\beta) \prod_{t_i=0} \left[ \frac{e^{x_{1i}^\top \beta_1}}{1 + e^{x_{1i}^\top \beta_1} + e^{x_{2i}^\top \beta_2}} \right] \prod_{t_i>0} \left[ \frac{f_p^*(t_i | \eta(\beta), x)}{1 + e^{x_{1i}^\top \beta_1} + e^{x_{2i}^\top \beta_2}} \right]^{\delta_i} \times \\ &\times \prod_{t_i>0} \left[ \frac{e^{x_{1i}^\top \beta_1} + S_p^*(t_i | \eta(\beta), x)}{1 + e^{x_{1i}^\top \beta_1} + e^{x_{2i}^\top \beta_2}} \right]^{1-\delta_i}, \end{aligned} \quad (4.3)$$

em que as expressões de  $f_p^*(\cdot)$  e  $S_p^*(\cdot)$  são descritas na Tabela 4 para os modelos ZICR sob os esquemas de ativação mínimo, máximo e aleatório.

A distribuição a priori conjunta considerada para os parâmetros  $\beta_j$   $j = 1, \dots, k+2$  é dada por

$$\pi(\beta) \propto \pi(\beta_1) \pi(\beta_2) \dots \pi(\beta_{k+2}), \quad (4.4)$$

em que  $\beta_j \sim N(0, b_j)$ ,  $j = 1, \dots, k+2$  e  $b_j$  são valores grandes que controlam a variância para produzir *priori*'s suaves.

As densidades marginais a posteriori de cada um dos parâmetros são obtidas integrando a equação (4.3) com relação a cada um dos parâmetros, porém como a distribuição a posteriori, não pertence a uma família de distribuições conhecidas e, portanto, não possui uma forma de expressão fechada. As estimativas dos parâmetros do modelo ZICR sob os esquemas de ativações

considerando a abordagem bayesiana é simulada através dos algoritmos MCMC. Nesta tese, nós desenvolvemos e implementamos os códigos no software livre [RStudio Team \(2020\)](#) usando o pacote [R2OpenBUGS](#) (ver [Sturtz, Ligges e Gelman \(2005\)](#)) para gerar cadeias de distribuições marginais a posteriori via métodos MCMC. Esta é uma abordagem computacional simples em que só precisamos especificar a verossimilhança e as distribuições a priori dos parâmetros.

### 4.3 Estudos de Simulação

Nessa seção apresentamos o estudo de simulação com o objetivo de avaliar a performance quanto ao poder de recuperação dos parâmetros, isto é, analisar as estimativas dos parâmetros para os modelos ZICR sob os três esquemas de ativação: primeira, última e aleatória considerando a perspectiva bayesiana. Para simular os dados utilizados no estudo Bayesiano, os procedimentos foram os mesmos utilizados na seção 3.6 do Capítulo 3 em que consideramos a abordagem clássica. Além disso, nós consideramos os mesmos cenários descritos na Tabela 5, o qual pode ser visto novamente a seguir

Tabela 19 – Valor dos parâmetros sob diferentes cenários

Parâmetros	Cenários		
	1	2	3
$\beta_{10}$	-2.00	-0.50	-0.50
$\beta_{11}$	0.5	1.50	0.50
$\beta_{20}$	-1.00	-0.75	-1.00
$\beta_{21}$	0.75	3.00	1.00
$\beta_{30}$	0.50	-20	-0.75
$\beta_{31}$	0.50	1.50	1.00
$\beta_{40}$	1.50	-1.25	1.25
$\beta_{41}$	2.00	1.00	1.00

Nós simulamos uma cadeia de tamanho 45000 para cada parâmetro, em que descartamos as primeiras 5000 iterações para eliminar o efeito dos valores iniciais e as amostras restantes foram obtidas considerando um espaçamento de tamanho 10, resultando em uma amostra final com  $n$  igual 4000 sobre a qual a inferência a posteriori é baseada. As Figuras 13, 14 e 15 mostram a performance obtida pelo processo de simulação através das estimativas médias, viés, o EQM e a intervalo de credibilidade para os parâmetros dos modelo ZICR sob os três esquemas de ativação. Sendo assim, analisando os resultados obtidos, temos que:

- No geral, observamos que em todos os cenários o intervalo de credibilidade está próxima da cobertura nominal esperada, isto é, 95%. Considerando o modelo ZICR sob os cenários máximos e aleatório, o parâmetro  $\beta_{41}$  precisou de um tamanho amostral maior para a probabilidade de cobertura convergir ao nível nominal. Enquanto que o parâmetro  $\beta_{31}$  obteve

uma convergência mais lenta em todos cenários, isto é, com baixa e alta fração de curados e zeros.

- De um modo geral, ao analisar o viés e o REQM nós percebemos que o cenário 1 (—■—), para os parâmetros associados a fração de zeros e cura ( $\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}$ ) obteve uma convergência mais lenta quando comparado aos cenários 2 e 3. Por outro lado, a medida que o tamanho da amostra aumenta, nós notamos a diminuição do viés e REQM.

- O Cenário 3 (—▲—), em geral, obteve o menor viés para os parâmetros associados as frações de cura e zeros e ocorre por este cenário, dentre os três, ter a maior proporção de zeros inflacionados e taxa de curados.

- Os resultados da simulação demonstram que os modelos ZICR, utilizando diferentes esquemas de ativação e abordagem bayesiana, apresentaram um desempenho satisfatório. A abordagem bayesiana permitiu a incorporação de informações a priori e a obtenção de intervalos de credibilidade para os parâmetros estimados.

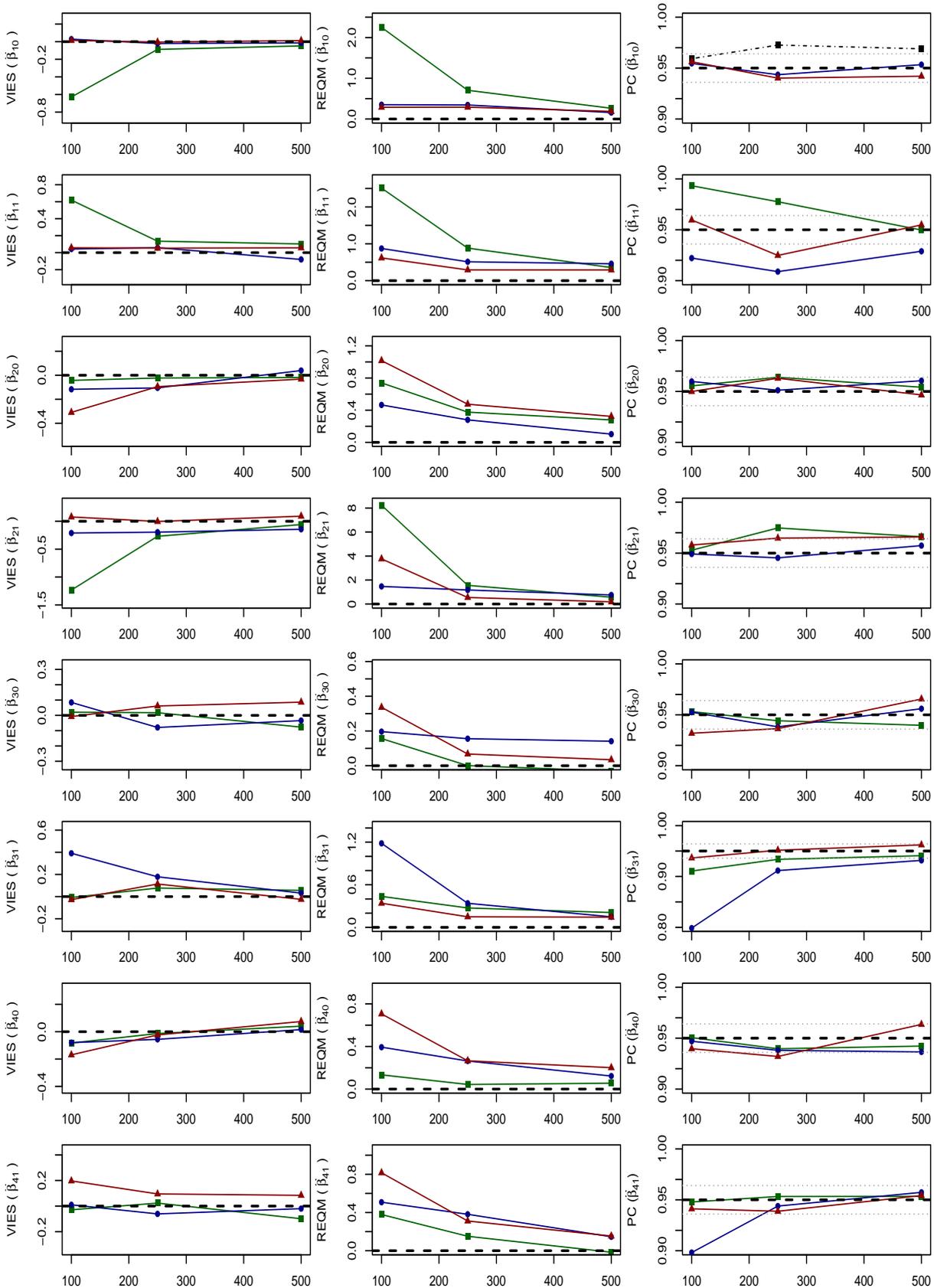


Figura 13 – Abordagem Bayesiana para estimação dos parâmetros para o Viés, REQM e cobertura empírica para intervalo de credibilidade com 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação máximo

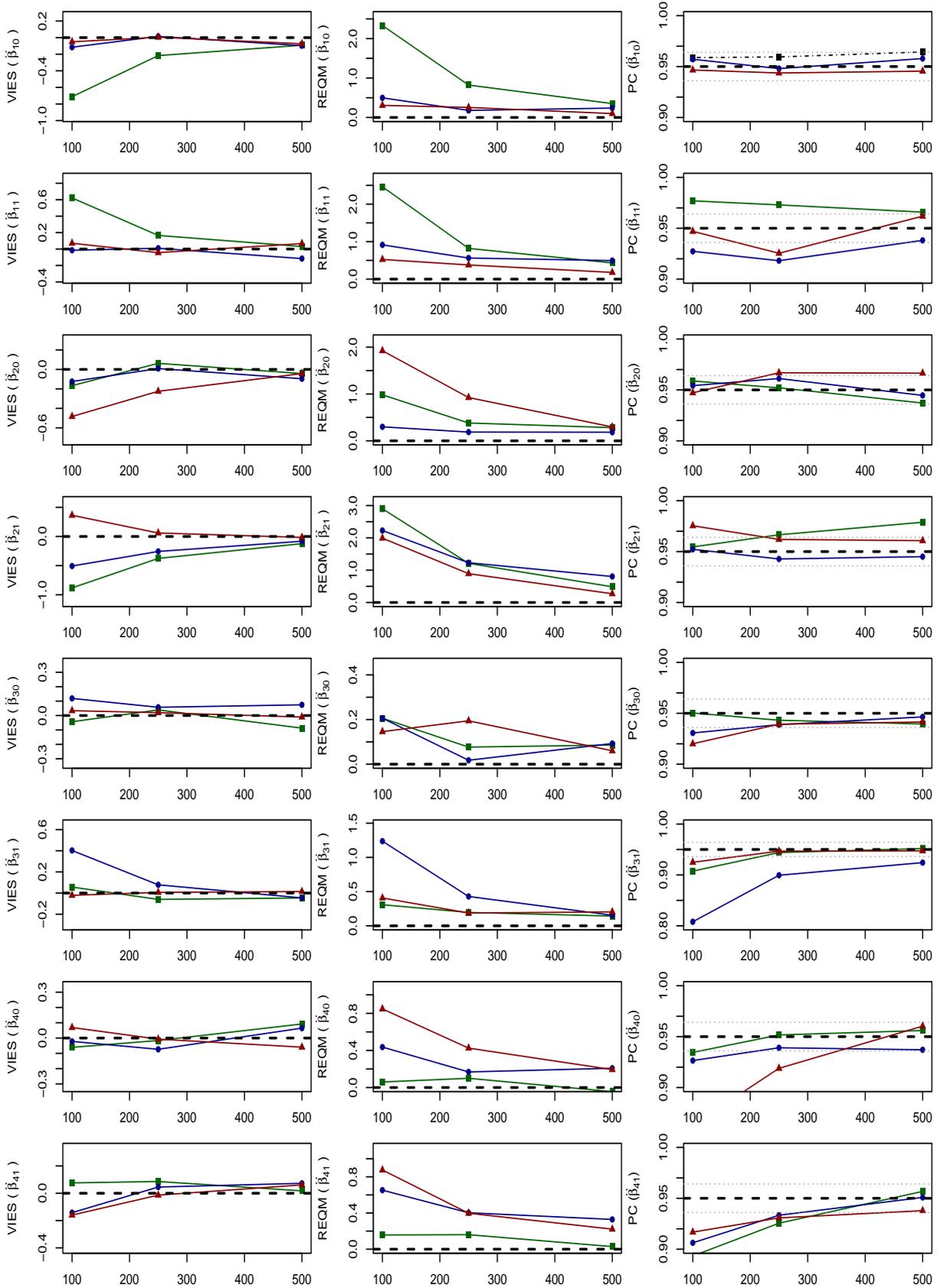


Figura 14 – Abordagem Bayesiana para estimação dos parâmetros - Viés, REQM e cobertura empírica para intervalos de credibilidade com 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação aleatório

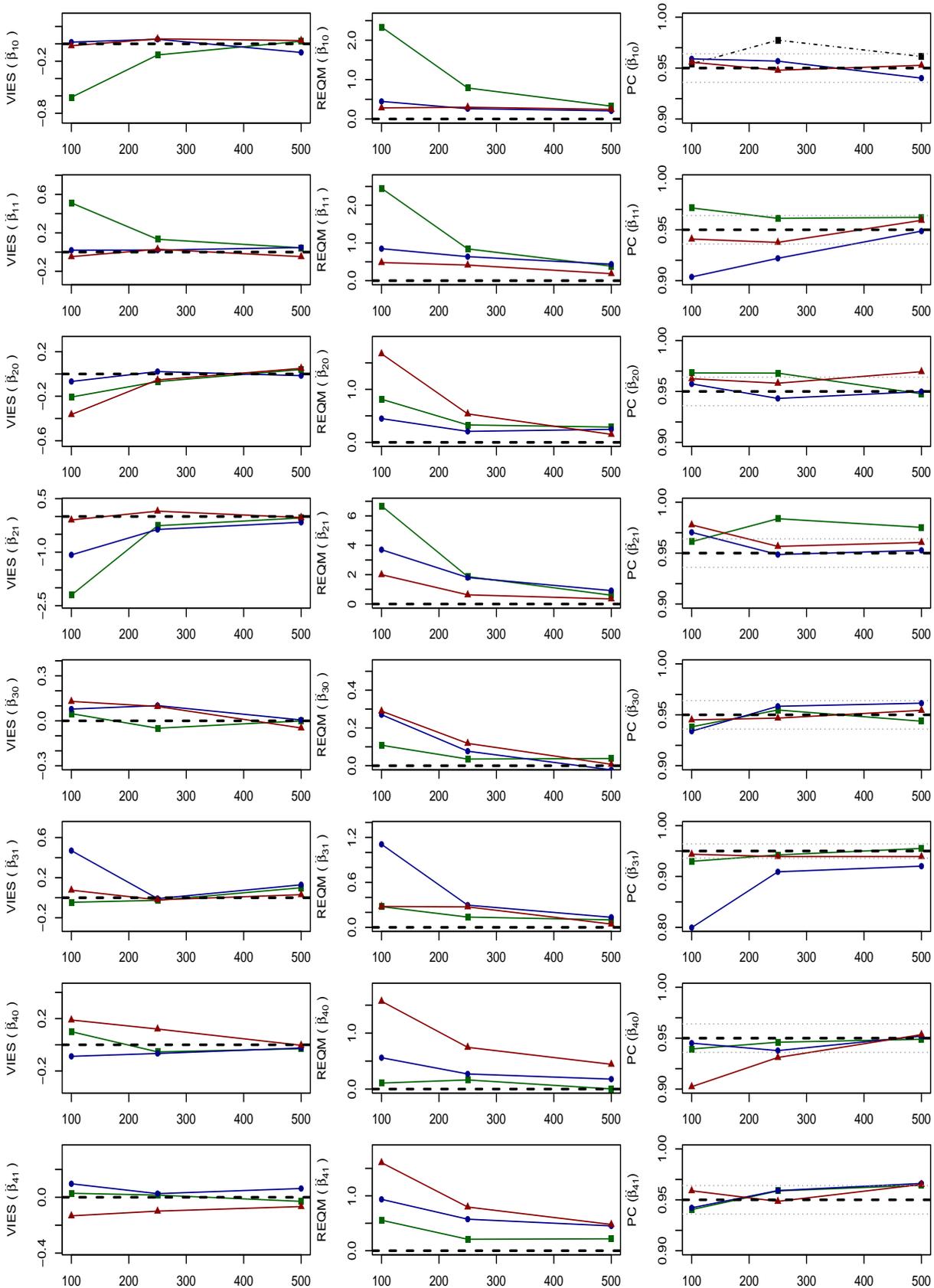


Figura 15 – Abordagem Bayesiana para estimação dos parâmetros - Viés, REQM e cobertura empírica para intervalos de credibilidade com 95 % versus tamanho da amostra para dados simulados sob os três cenários de parâmetros do modelo de taxa de cura de promoção ajustado por zero sob esquema de ativação mínimo

## 4.4 Aplicação de dados sobre insulina

Nesta seção nós consideramos o conjunto de dados sobre o uso da insulina descrito na seção 3.7 (Subseção 3.7.1) para avaliarmos a aplicabilidade do modelo proposto com abordagem bayesiana. Sendo assim, inferência será feita sob o ponto de vista Bayesiano em que consideramos o modelo ZICR sob os três esquemas de ativações obtidos na Tabela 14 da seção 3.7.

Para realizar a análise, foi utilizada uma simulação de cadeia com um total de 15.000 observações para cada parâmetro. Para garantir que os resultados não fossem influenciados pelos valores iniciais, foram descartadas as primeiras 5.000 iterações. Além disso, para evitar problemas de autocorrelação, nós consideramos um espaçamento de tamanho 10, isto é, nós pegamos as observações de 10 em 10. Com essa abordagem, obtivemos uma amostra final de tamanho 1.000 para cada parâmetro, a partir da qual foram realizadas as análises de inferência a posteriori.

Em um primeiro momento, nós ajustamos individualmente um modelo ZICR para cada esquema de ativação com abordagem Bayesiana para cada uma das variáveis descritas na Tabela 14, isto é, as variáveis histórico fumante, histórico de diabetes e macrosomia à priori.

As Tabelas 20, 21 e 22 mostram os resultados das estimativas para cada variável considerando o modelo ZICR sob os três esquemas de ativação. Utilizando os resultados das estimativas, construímos o gráfico de KM para cada variável o que pode ser visualizado pelas Figuras 16, 17 e 18. Os resultados obtidos mostram que as estimativas nos três cenários foram bem parecidas, e o valor do Deviance e DIC foram bem próximos. Além disso, nós observamos que as estimativas estão próximas a curva de sobrevivência estimada via KM (ver Figuras 18, 17 e 16). Assim como no Capítulo 3, a curva de sobrevivência estimada dos três cenários visualmente estão bem próximas umas das outras.

As Figuras 33, 35, 37, 39, 41, 43, 45, 47 e 49 (ver Apêndice A.2) mostram o gráfico chamado de *traplot* em que este é utilizado para avaliar a convergência e a mistura de uma simulação Markov Chain Monte Carlo. Nós observamos que os traços, para cada parâmetro do modelo (representados por  $\beta_i$ ) estão se movendo livremente pelo espaço de parâmetros e não estão presas, portanto isto indica que a simulação MCMC está convergindo.

Com relação a convergência das estimativas dos parâmetros para as variáveis macrosomia prévia, histórico tabagismo e histórico diabetes, nós obtemos os gráficos médios, também conhecidos como gráficos de traço de média, representados pelas Figuras 32, 34, 36, 38, 40, 42, 44, 46, 48 (ver Apêndice A.2). Esses gráficos mostram a média dos valores de cada parâmetro ao longo do tempo, em uma única figura, permitindo uma comparação visual dos diferentes parâmetros. Dessa forma, nós observamos que as estimativas convergiram para os respectivos valores de cada parâmetros.

Por fim, podemos concluir que as estimativas obtidas pelas abordagens clássica e bayesiana apresentam poucas diferenças significativas para esse conjunto de dados. As estimativas

obtidas para as frações de indivíduos com tempo iguais a zero e de indivíduos curados também foram similares nas duas abordagens. Isso sugere que as abordagens clássica e Bayesiana forneceram estimativas consistentes para esses parâmetros importantes. Por último, com relação a escolha do esquema de ativação que melhor descreve a explicatividade das causas competitivas que influenciam o evento de interesse (aplicação de insulina), observamos que o DIC e Deviance foram muito próximos então a escolha se dar por um conjunto de fatores: ajuste visual, relevância médica e medias de ajuste obtidas e portanto o esquema de ativação aleatório seria o escolhido.

Tabela 20 – Resumos *a posteriori* dos parâmetros do modelo ZICR sob os mecanismos de ativação para a variável histórico diabetes

Parâmetros	Mínimo			
	Média	L.I	L.S	D.P
$\beta_{10}$	-0.683	-1.368	-0.0319	0.341
$\beta_{11}$	0.054	-0.706	0.859	0.395
$\beta_{20}$	1.520	1.087	1.975	0.229
$\beta_{21}$	-0.6530	-1.175	-0.127	0.271
Deviance	1180.3003			
DIC	1188			
Parâmetros	Máximo			
	Média	L.I	L.S	D.P
$\beta_{10}$	-0.682	-1.379	-0.027	0.344
$\beta_{11}$	0.058	-0.699	0.8505	0.398
$\beta_{20}$	1.517	1.090	1.966	0.225
$\beta_{21}$	-0.646	-1.182	-0.121	0.270
Deviance	1180.1514			
DIC	1188			
Parâmetros	Aleatório			
	Média	L.I	L.S	D.P
$\beta_{10}$	-0.683	-1.354	-0.0424	0.336
$\beta_{11}$	0.0612	-0.701	0.839	0.391
$\beta_{20}$	1.519	1.092	1.962	0.222
$\beta_{21}$	-0.648	-1.175	-0.1359	0.264
Deviance	1180.191			
DIC	1189			

Tabela 21 – Resumos a *posteriori* dos parâmetros do modelo ZICR sob os mecanismos de ativação para a variável macrosomia a priori

Parâmetros	Mínimo			
	Média	L.I	L.S	D.P
$\beta_{10}$	-0.736	-1.124	-0.361	0.196
$\beta_{11}$	0.684	-0.301	1.65	0.496
$\beta_{20}$	1.132	0.885	1.392	0.129
$\beta_{21}$	-0.561	-1.427	0.295	0.438
Deviance	1180.8424			
DIC	1189			
Parâmetros	Máximo			
	Média	L.I	L.S	D.P
$\beta_{10}$	-0.741	-1.137	-0.354	0.197
$\beta_{11}$	0.693	-0.297	1.673	0.485
$\beta_{20}$	1.138	0.886	1.403	0.131
$\beta_{21}$	-0.564	-1.435	0.323	0.458
Deviance	1181.2453			
DIC	1189			
Parâmetros	Aleatório			
	Média	L.I	L.S	D.P
$\beta_{10}$	-0.736	-1.124	-0.361	0.195
$\beta_{11}$	0.684	-0.301	1.650	0.496
$\beta_{20}$	1.132	0.885	1.392	0.129
$\beta_{21}$	-0.561	-1.427	0.295	0.433
Deviance	1180.842			
DIC	1189			

Tabela 22 – Resumos a *posteriori* dos parâmetros do modelo ZICR sob os mecanismos de ativação para a variável histórico fumante

Parâmetros	Mínimo			
	Média	L.I	L.S	desvio padrão
$\beta_{10}$	-0.633	-1.008	-0.267	0.189
$\beta_{11}$	-0.094	-1.242	0.975	0.563
$\beta_{20}$	1.137	0.885	1.391	0.130
$\beta_{21}$	-0.485	-1.316	0.341	0.429
Deviance	1186.374			
DIC	1194			
Parâmetros	Máximo			
	Média	L.I	L.S	desvio padrão
$\beta_{10}$	-0.636	-1.025	-0.271	0.189
$\beta_{11}$	-0.091	-1.228	0.963	0.550
$\beta_{20}$	1.135	0.882	1.395	0.131
$\beta_{21}$	-0.483	-1.277	0.336	0.409
Deviance	1186.249			
DIC	1195			
Parâmetros	Aleatório			
	Média	L.I	L.S	desvio padrão
$\beta_{10}$	-0.629	-1.006	-0.266	0.189
$\beta_{11}$	-0.083	-1.211	0.965	0.552
$\beta_{20}$	1.137	0.884	1.397	0.131
$\beta_{21}$	-0.482	-1.335	0.359	0.443
Deviance	1186.266			
DIC	1194			

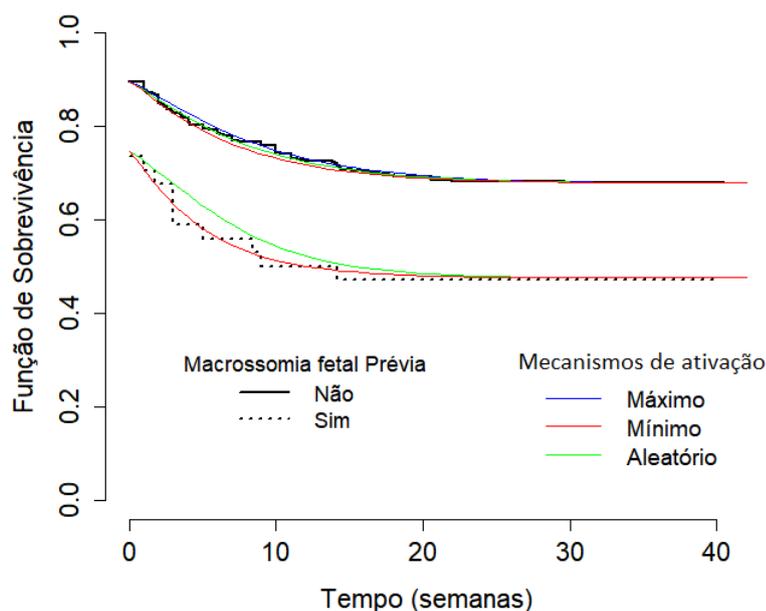


Figura 16 – Curva de Sobrevivência estimada pelo modelo ZICR sob os três esquemas de ativação com abordagem bayesiana para a variável Microssomia anterior.

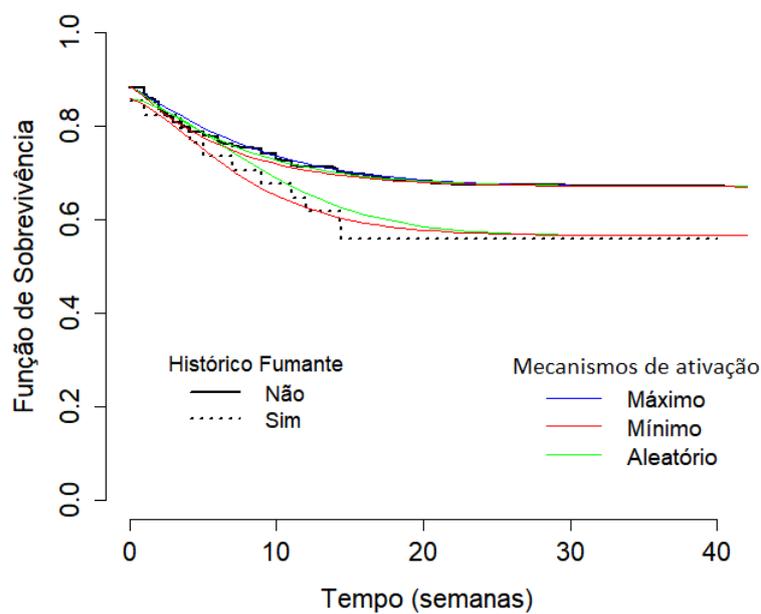


Figura 17 – Curva de Sobrevivência estimada pelo modelo ZICR sob os três esquemas de ativação com abordagem bayesiana para a variável Histórico tabagismo.

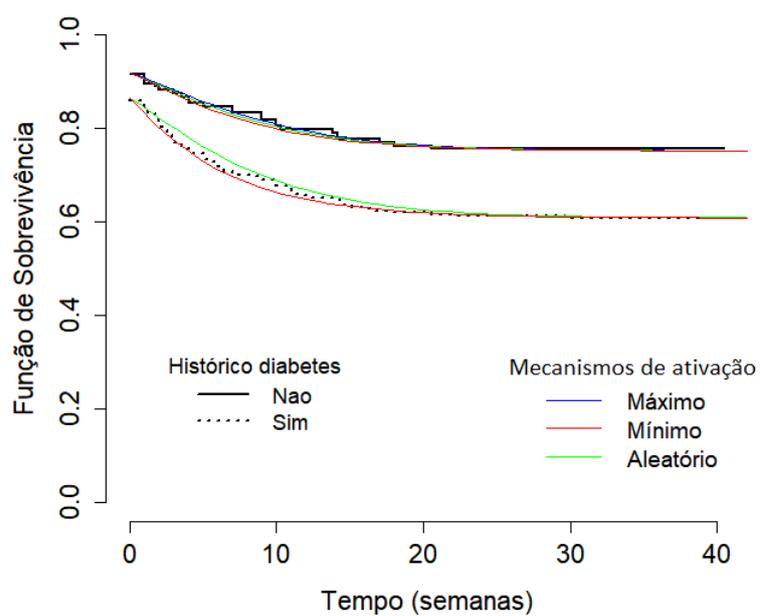


Figura 18 – Curva de Sobrevivência estimada pelo modelo ZICR sob os três esquemas de ativação com abordagem bayesiana para a variável Histórico Diabetes na família.

## 4.5 Conclusão

Neste capítulo apresentamos uma abordagem Bayesiana dos modelos de fração de cura zero inflacionados sob os esquemas de ativação máximo, mínimo e aleatório. Os resultados mostram que a abordagem Bayesiana pode ser uma alternativa promissora para a estimação dos parâmetros dos modelos de fração de cura zero inflacionados, especialmente em situações onde há pouca informação disponível ou incerteza sobre a distribuição dos parâmetros. Além disso, ambas as abordagens podem ser usadas de forma confiável para analisar os dados e obter estimativas precisas dos parâmetros, e a escolha da abordagem pode depender de outros fatores, como o tempo disponível e a necessidade de informações adicionais sobre a incerteza das estimativas.

---

# MODELOS DE FRAÇÃO DE CURA ZERO INFLACIONADO SOB DIFERENTES ESQUEMAS DE ATIVAÇÃO COM FRAGILIDADE

---

---

Neste capítulo propomos uma extensão para o modelo de fração de cura zero inflacionado sob diferentes esquemas de ativação conforme apresentado no Capítulo 3, em que foi incorporado um efeito aleatório na modelagem da sobrevida, chamado de fragilidade. A inclusão da variável de fragilidade, nos permite estimar a heterogeneidade devido à presença de fatores não observáveis. As estimativas dos parâmetros do modelo são baseadas no método de máxima verossimilhança e os intervalos de confiança são obtidos por métodos assintóticos. Nós realizamos um estudo de simulação de Monte Carlo para avaliar as propriedades dos estimadores dos parâmetros do modelo. Finalmente, para ilustrar a nova metodologia, apresentamos um exemplo de aplicação do modelo proposto em dois conjuntos de dados relacionados a sobrevida. Nós utilizamos as duas aplicações apresentadas no Capítulo 3 na seção 3.7 considerando o modelo proposto.

## 5.1 Introdução

Uma abordagem comumente usada na análise de sobrevivência é o modelo de fração de cura proposto por Tsodikov, Yakovlev e Asselain (1996) e Chen, Ibrahim e Sinha (1999), que é usado quando alguns indivíduos da população não são suscetíveis ao evento de interesse. No entanto, há situações em que essa suposição de cura pode não ser adequada ou apropriada.

Por exemplo, em estudos de câncer, pode haver indivíduos que nunca experimentam o evento de interesse, o que pode levar a estimativas imprecisas ou enviesadas da taxa de cura.

Além disso, pode haver indivíduos que nunca experimentam o evento de interesse, mas que não estão realmente curados e ainda estão sob risco de experimentar o evento no futuro.

Os modelos de fragilidade têm como característica principal a inclusão de um efeito aleatório, ou seja, uma variável aleatória não observável é incorporada a essa modelagem para representar essas características não observáveis entre os sujeitos. Esse efeito aleatório, denominado fragilidade, pode ser inserido na função de risco de linha de base com o objetivo de controlar a heterogeneidade não observável dos sujeitos em estudo. Essa inclusão pode ser de forma aditiva ou multiplicativa e ao longo dos últimos anos, vários autores consideram uma abordagem em que a fragilidade tem um efeito multiplicativo sobre a função de risco, isto é, esses modelos com fragilidade multiplicativa são uma generalização do modelo de riscos proporcionais proposto por Cox (1972). Por exemplo, Andersen *et al.* (1993), Hougaard (1995) realizaram uma revisão da literatura sobre modelos de fragilidade multiplicativa enquanto Yu e Peng (2008), Calsavara, Tomazella e Fogo (2013), Leão *et al.* (2018), entre outros, consideraram os modelos de fragilidade com fração de cura.

## 5.2 Modelo de fragilidade

O termo fragilidade refere-se ao efeito aleatório incorporado na função de risco, sendo que este pode atuar de maneira aditiva ou multiplicativa na função de risco individual. As distribuições mais utilizadas na literatura para este termo têm sido Weibull, exponencial, log-normal e gama em que esta última é frequentemente utilizada devido ao fácil tratamento algébrico.

A variável de fragilidade incorporada nessa modelagem não se restringe apenas a mensurar a heterogeneidade dos indivíduos mas também é possível avaliar a ausência de alguma covariável que não tenha sido incluído no modelo. Isto é, se por exemplo, uma covariável que deveria ser incluída mas que por algum motivo não foi incorporada, então isso fará com que o valor de  $v_i$  aumente e conseqüentemente influenciará as inferências sobre os parâmetros incluídos no modelo.

Seja  $V$  uma variável aleatória (v.a) que representa o termo de fragilidade em que esta v.a é não negativa e não observável a qual é incorporada multiplicativamente na função de risco de base. O modelo de fragilidade multiplicativo é uma extensão do modelo proposto por Cox (1972). A função de risco do  $i$ -ésimo indivíduo em um determinado momento  $t$  condicionada a fragilidade é representada por:

$$h_i(t) = h(t|v_i) = v_i h_0(t) \exp(\mathbf{x}_i^T \boldsymbol{\beta}), \quad i = 1, \dots, n, \quad (5.1)$$

em que  $v_1, v_2, \dots, v_n$  representam a fragilidade para o  $i$ -ésimo indivíduo,  $h_0(t)$  é a função de risco de base comum a todos a população pertencente ao estudo,  $\mathbf{x}$  é o vetor de covariáveis e  $\boldsymbol{\beta}$  é o vetor de coeficientes associados as respectivas covariáveis.

O modelo representado pela Equação (5.1) apresenta uma estrutura de risco proporcional condicionada ao efeito aleatório, simbolizado pela variável  $V$ . Como a variável de fragilidade age de forma multiplicativa na função de risco de cada indivíduo temos que o risco aumenta se  $v_i > 1$ , diminui para  $v_i < 1$  e quando  $v_i = 1$  temos o modelo de riscos proporcionais de Cox (1972). Dessa forma, quanto maior for o valor de  $v_i$ , os indivíduos estarão mais suscetíveis para a ocorrência do evento de interesse.

Diante da diversidade de distribuições de probabilidade que podem ser assumidas para modelar a variável de fragilidade, diversos autores tem proposto estudos com diferentes distribuições. A distribuição Gama é mais utilizada para a modelagem da fragilidade devido ao fácil tratamento algébrico, ver por exemplo (VAUPEL; MANTON; STALLARD, 1979). Tomazella (2003) realizou um estudo comparativo entre as distribuições gama, lognormal, gaussiana inversa para aos modelos de regressão em processos de Poisson.

De acordo com Elbers e Ridder (1982), para que o modelo seja identificável, é necessário que a distribuição da variável de fragilidade tenha uma média finita. Nesse sentido, para manter a identificabilidade do modelo, é conveniente atribuir alguma distribuição com média 1. Dessa forma, nesta tese, nós assumimos a distribuição gama para a fragilidade, isto é,  $V \sim \text{Gama}(1/\alpha, 1/\alpha)$  em que  $\mathbb{E}(V) = 1$  e  $\text{Var}(V) = \alpha$ . Assim  $\alpha > 0$  mede a heterogeneidade não observável entre os indivíduos. Quando o valor de  $\alpha$  é pequeno, isso indica que a heterogeneidade observada na população é pequena, ou seja, que os indivíduos são bastante semelhantes entre si. Neste caso, a maior parte da variabilidade nos dados pode ser explicada pelos efeitos fixos (referem-se aos parâmetros que descrevem o efeito das variáveis covariáveis) do modelo e a contribuição dos efeitos aleatórios (representados pela variável de fragilidade) é relativamente pequena. Por outro lado, quando o valor de  $\alpha$  é grande, isso indica que a heterogeneidade observada é alta, ou seja, que existem diferenças significativas entre os indivíduos. Neste caso, a variabilidade nos dados não pode ser explicada apenas pelos efeitos fixos do modelo, e a contribuição dos efeitos aleatórios é relativamente grande. Isso sugere que existe uma grande variabilidade entre os indivíduos, que não pode ser explicada apenas pelas variáveis incluídas no modelo.

Considerando o modelo (5.1) sem a presença de covariáveis, temos que a função de risco de cada indivíduo condicional a variável de fragilidade é dada por

$$h(t|v_i) = v_i h_0(t), \quad i = 1, \dots, n, \quad (5.2)$$

em que  $h_0(t)$  é a função de risco basal (ou função de risco na ausência de fragilidade) e  $v_i$  é o valor da fragilidade para o  $i$ -ésimo indivíduo. A função de sobrevivência condicional, ou seja, a probabilidade de que um indivíduo sobreviva até um determinado tempo  $t$  condicional ao valor de sua fragilidade  $v_i$ , pode ser obtida a partir da função de risco dada em (5.2).

A função de sobrevivência condicional a variável de fragilidade,  $S(t|v_i)$ , é dada por:

$$S(t|v_i) = S_0(t)^{v_i}, \quad (5.3)$$

em que  $S_0(t)$  representa a função de sobrevivência basal a qual é comum a todos os indivíduos da população.

Para obtermos a função de verossimilhança, nós precisamos obter a função de sobrevivência não condicional a qual é obtida integrando-se sobre todas as possíveis realizações do efeito aleatório (termo de fragilidade) ponderadas pela distribuição de probabilidade dessa variável, representado pela Equação (5.4), dada por

$$S(t) = \mathbb{E}_V[S(t|v_i)] = \int_0^\infty S(t|v_i) f_V(v_i) dv_i, \quad (5.4)$$

em que  $f_V(v_i)$  é a função densidade de probabilidade da fragilidade  $v_i$ . A Equação (5.4) mostra que a função de sobrevivência marginal é a média ponderada de todas as possíveis funções de sobrevivência condicional  $S(t|v_i)$ , onde a ponderação é realizada pela distribuição de probabilidade da fragilidade  $v_i$ . Note que a função de sobrevivência marginal  $S(t)$  representa a probabilidade de sobrevivência para um indivíduo selecionado aleatoriamente da população, sem considerar seu valor específico de fragilidade.

A transformada de Laplace tem muitas propriedades úteis que permitem simplificar cálculos e manipulações matemáticas. Em particular, no contexto da análise de sobrevivência com efeitos aleatórios, a transformada de Laplace é uma ferramenta útil para se obter a função de sobrevivência marginal a partir da função de risco e da distribuição de probabilidade do efeito aleatório representado pela Equação (5.4).

**Definição 1.** A transformada de Laplace para qualquer função  $f: [0, \infty] \rightarrow \mathbb{R}$  é definida por

$$L(s) = \mathbb{E}(e^{-st}) = \int_0^\infty e^{-st} f(t) dt, \quad (5.5)$$

para qualquer  $s \geq 0$  em que a integral dada pela Equação (5.5) converge.

A partir da expressão (5.5) e aplicando a expressão dada pela Equação (5.4), podemos obter a função de sobrevivência marginal,  $S(t)$ , em que consideramos  $f(t)$  como sendo  $f_V(v_i)$  e  $s$  é a função de risco de base acumulada  $H_0(t)$ . Logo,

$$S(t) = \mathbb{E}_V[S(t|v_i)] = \int_0^\infty e^{-H_0(t)v_i} f_V(v_i) dv_i = L_V[H_0(t)],$$

em que  $f_V(\cdot)$  é a função de densidade da variável  $V$  condicional,  $H_0(\cdot)$  é a função de risco da linha de base cumulativa e  $L_V(\cdot)$  representa a transformada de Laplace. Então, a transformada de Laplace para distribuição gama( $1/\alpha, 1/\alpha$ ) é dada por:

$$L_V(s) = (1 + s\alpha)^{-1/\alpha}. \quad (5.6)$$

Substituindo  $s = H_0(t)$  na Equação (5.6), nós obtemos a função de sobrevivência própria sob a fragilidade gama, que é dada pela Equação (5.7).

$$S(t) = (1 + H_0(t)\alpha)^{-1/\alpha}. \quad (5.7)$$

A partir da Equação (5.7), podemos obter a função densidade no modelo de fragilidade gama. Essa função é dada por

$$f(t) = (1 + \alpha H_0(t))^{-(\alpha+1)/\alpha} h_0(t), \quad (5.8)$$

em que  $h_0(t)$  é a função de risco basal e  $H_0(t)$  é a função de risco acumulada da distribuição basal.

### 5.3 Modelo de fração de cura zero ajustado com fragilidade gama

Souza (2020) abordou um modelo de fração de cura zeros ajustados com fragilidade, chamado de ZICR-F, como uma alternativa para modelar dados de sobrevivência com fração de cura e fração de zeros. Nesse trabalho, as distribuições gama e log-normal foram consideradas para as variáveis de fragilidade e tempo de sobrevida, respectivamente. Para o conjunto de dados reais analisados por esses autores foi observada uma heterogeneidade significativa. O modelo zero inflacionado com fragilidade pode ser uma alternativa interessante para analisar dados de sobrevivência com heterogeneidade não observada.

A seguir, consideramos o modelo de taxa de cura, conforme definido na expressão (3.5). Para modelar os dados de sobrevivência com longa duração, fração de zeros e incorporando a fragilidade, a função de sobrevivência populacional do modelo de fração de cura zero inflacionado com fragilidade (ZICR-F) é dada por (ver Souza (2020)),

$$S_p(t|v) = p_1 + (1 - p_0 - p_1)S_0(t|v). \quad (5.9)$$

em que  $p_0$  representa a proporção de tempos de sobrevivência iguais a zero,  $p_1$  representa a proporção de indivíduos curados na população e  $S(t|v)$  é função de sobrevivência condicional a variável de fragilidade, representada pela Equação (5.3).

A função de sobrevivência populacional não condicional,  $S_0(t|v)$ , para o modelo descrito pela expressão (5.9) pode ser obtida usando a transformada de Laplace. Dessa forma, considerando a transformada de Laplace, a função de sobrevivência populacional não condicional do modelo ZICR-F é dada por

$$S_p(t) = p_1 + (1 - p_0 - p_1)L_v[H_0(t)], \quad (5.10)$$

em que a v.a de fragilidade nesse modelo tem ponto de massa em zero com probabilidade  $(1 - p_0 - p_1)$ .

Assumindo que a distribuição da fragilidade  $V$  segue uma distribuição gama com parâmetros  $1/\alpha$  e  $1/\alpha$ . A função de sobrevivência populacional não condicional do modelo ZICR-F é obtida substituindo (5.6) em (5.10) em que esta é dada por:

$$S_p(t) = p_1 + (1 - p_0 - p_1)(1 + H_0(t)\alpha)^{-1/\alpha}. \quad (5.11)$$

E, por consequência de (5.11), nós obtemos a função densidade não condicional dada por

$$f_p(t) = (1 - p_0 - p_1)(1 + \alpha H_0(t))^{-(\alpha+1)/\alpha} h_0(t),$$

Essas expressões permitem que a heterogeneidade seja modelada por meio da fragilidade, permitindo a estimação de parâmetros relacionados à fragilidade e aos efeitos das covariáveis no tempo de sobrevivência. Além disso, observe que o modelo dado pela equação 5.11 tem as seguintes propriedades:

- Se  $p_0 = 0$ ,  $p_1 = 1$  obtemos o modelo de fragilidade gama (5.7);
- Se  $p_0 = 0$  e  $\alpha \rightarrow \infty$ , tem-se o modelo de mistura padrão proposto por Berkson e Gage (1952);

A partir dessas propriedades, nós podemos observar que o modelo de fração de cura zero inflacionado com fragilidade abrange alguns sub-modelos. Dessa forma podemos fazer testes de hipóteses com o objetivo de verificar se existe evidências de zero inflacionados na população, bem como curados ou se a heterogeneidade não observada está presente.

### 5.3.1 Modelo ZICR-F log-normal com Fragilidade gama

Nessa subseção, nós apresentamos o modelo de fração de cura zero inflacionado com fragilidade, assumindo a distribuição log-normal para modelar o tempo de vida de indivíduos sob risco. Assim as funções de risco acumulada e risco de base do modelo log normal com média e variância representadas por  $\mu$  e  $\sigma$ , respectivamente são dadas por:

$$H_0(t) = -\ln(1 - \Phi(\ln(t)/\sigma)) \quad e \quad h_0(t) = \frac{\frac{1}{\sqrt{2\pi t}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2\right]}{1 - \Phi\left[\frac{\ln(t)-\mu}{\sigma}\right]}. \quad (5.12)$$

Substituindo a função de risco acumulada  $H_0(t)$  de (5.12) em (5.11), nós obtemos a função de sobrevivência populacional, de todos os indivíduos, do modelo ZICR-F log-normal com fragilidade gama a qual é representada por:

$$S_p(t) = p_1 + (1 - p_0 - p_1)(1 - \ln(1 - \Phi(\ln(t)/\sigma))H_0(t)\alpha)^{-1/\alpha}, \quad (5.13)$$

com função densidade populacional imprópria dada por

$$f_p(t) = (1 - p_0 - p_1) (1 - \ln(1 - \Phi(\ln(t)/\sigma)) H_0(t) \alpha)^{-1/\alpha} \frac{\frac{1}{\sqrt{2\pi t \sigma}} \exp\left[-\frac{1}{2} \left(\frac{\ln(t) - \mu}{\sigma}\right)^2\right]}{1 - \Phi\left[\frac{\ln(t) - \mu}{\sigma}\right]}.$$

Devido à relação entre  $\alpha$  e a variância da distribuição gama, ela pode ser interpretada como diretamente proporcional à heterogeneidade dos dados. A Figura 19 mostra o comportamento da função de risco e de sobrevivência para diferentes valores de  $\alpha$  considerando Log-Normal como a função de linha de base.

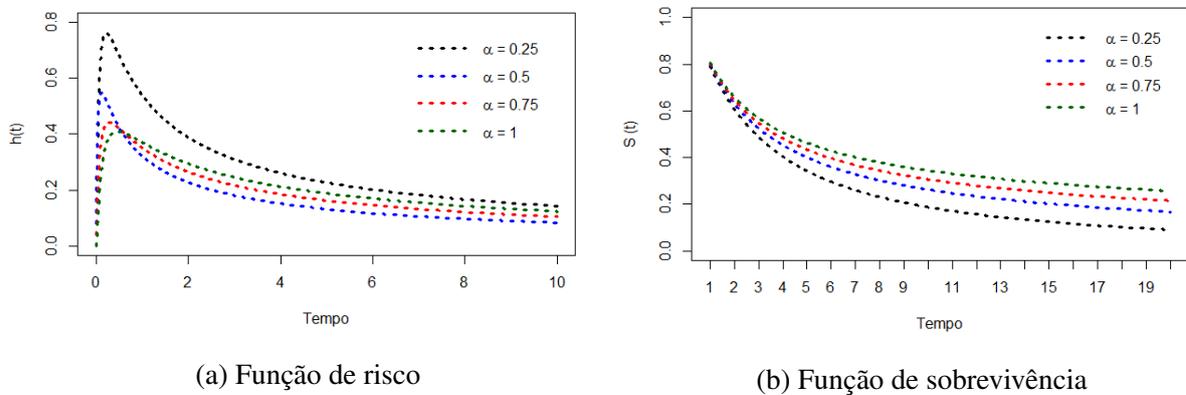


Figura 19 – Exemplo de funções não condicional de Sobrevivência (b) e risco (a) para o modelo gamma ZICR-F para diferentes valores de fragilidade ( $\alpha$ ) considerando a distribuição Log-normal para o tempo.

## 5.4 Especificação do Modelo

Nesta seção, nós propomos os modelos zero inflacionado com fração de cura na presença de fragilidade gama para esquemas de ativação mínimo, máximo e aleatório. Para isso, nós consideramos a mesma estrutura usada no Capítulo 3 em que  $M$  representa o número de causas competitivas para a ocorrência do evento de interesse para um indivíduo da população. Assim como descrito no Capítulo 3 considere  $M$  como sendo uma variável não observável a qual segue uma distribuição Poisson com parâmetro  $\theta$  e a variável  $Z_j$ , com  $j = 1, \dots, M$ . Por fim, as variáveis  $Z_j$  são i.i.d. variáveis aleatórias com função de distribuição  $F(t) = 1 - S(t)$  em que  $\mathbb{P}(Z_0 = \infty)$  além de  $Z_j$  serem independentes de  $M$ .

O tempo de falha ao fenômeno de interesse dos indivíduos é uma variável aleatória que representa o tempo até a ocorrência do evento de interesse, como a falha de um componente, a morte de um indivíduo, recidiva de um câncer, entre outros. Nesse contexto, existem diferentes formas de definir o tempo de falha  $T$  com base nas variáveis aleatórias  $Z_i$ . Uma delas é definir  $T$  como o mínimo, ou seja,  $T = \min Z_i, 0 \leq i \leq M$ . Outra forma é definir  $T$  como o máximo, ou

seja,  $T = \max Z_i, 0 \leq i \leq M$ . A terceira e última forma a qual nós consideramos é  $T$  como sendo  $T = T_{(q)}$ , em que  $q$  é um número entre 0 e  $M$  e representa o esquema aleatório.

Com o objetivo de incorporar a fragilidade no modelo ZICR sob os diferentes esquemas de ativação, nós consideramos uma estrutura similar a qual foi considerada no Capítulo 3 e portanto, inicialmente, nós apresentamos os principais resultados obtidos. Dessa forma, a Tabela 23 mostra os resultados das expressões 3.1, 3.6 e 3.9 em que representam a função de sobrevivência de todos os indivíduos da população para os esquemas de ativação máximo, aleatório e mínimo, respectivamente.

Tabela 23 – Função de sobrevivência populacional de T e densidade para cada cenário do modelo ZICR

Esquema de ativação	$S_p(t)$	$f_p(t)$
Mínimo	$\exp\{-\theta F_0(t)\}$	$\theta f_0(t) \exp\{-\theta F_0(t)\}$
Máximo	$1 - \exp\{-\theta[1 - F_0(t)]\} + \exp\{-\theta\}$	$\theta f_0(t) \exp\{-\theta[1 - F_0(t)]\}$
Aleatório	$\exp\{-\theta\} + [1 - \exp\{-\theta\}][1 - F_0(t)]$	$f_0(t)(\exp\{-\theta\} - 1)$

Através da relação  $F(t) = 1 - S(t)$  em que  $S(t)$  é dado pela Equação (5.7) e substituindo nas expressões obtidas em 3.1, 3.6 e 3.9 (resumidas na Tabela 23), nós obtemos a função de sobrevivência populacional com fragilidade gama para cada esquema de ativação, descrita na Tabela 24.

Tabela 24 – Função de sobrevivência populacional com fragilidade (ZICR-F) gama para cada esquema de ativação

Esquema de ativação	$S_p(t)$
Mínimo	$\exp\{-\theta(1 - (1 + H_0(t)\alpha)^{-1/\alpha})\}$
Máximo	$1 - \exp\{-\theta(1 + H_0(t)\alpha)^{-1/\alpha}\} + \exp\{-\theta\}$
Aleatório	$\exp\{-\theta\} + [1 - \exp\{-\theta\}](1 + H_0(t)\alpha)^{-1/\alpha}$

A função densidade populacional para o modelo proposto sob o esquema de ativação mínimo pode ser obtida através da seguinte manipulação algébrica em que, inicialmente, nós consideramos a mesma expressão obtida em (3.10):

$$f_p(t) = -\frac{d}{dt}S_p(t) = \theta f_0(t) \exp(-\theta F_0(t)). \tag{5.14}$$

Substituindo  $f_0(t)$  pela expressão obtida em (5.8) e usando a relação  $F(t) = 1 - S(t)$  em que  $S(t)$  é dada por (5.7), a função densidade do modelo de fração de cura com fragilidade gama sob o esquema de ativação mínimo é dada por :

$$f_p(t) = \theta(1 + \alpha H_0(t))^{-(\alpha+1)/\alpha} h_0(t) \exp(-\theta(1 - (1 + H_0(t)\alpha)^{-1/\alpha})). \tag{5.15}$$

Analogamente, nós realizamos as mesmas manipulações algébricas para obter as funções densidades populacionais dos esquemas de ativação máximo e aleatório em que nesses casos substituímos  $f_0(t)$  (5.8) nas equações (3.2) e (3.7), respectivamente. Dessa forma, a Tabela 25 mostra os resultados obtidos das funções densidades populacionais do modelo de fração de cura zero ajustado com fragilidade gama,  $f_p(t)$ , para cada esquema de ativação.

Tabela 25 – Função de densidade com fragilidade gama para cada esquema de ativação

Esquema de ativação	$f_p(t)$
Mínimo	$\theta(1 + \alpha H_0(t))^{-(\alpha+1)/\alpha} h_0(t) \exp\{-\theta(1 - (1 + H_0(t)\alpha)^{-1/\alpha})\}$
Máximo	$\theta(1 + \alpha H_0(t))^{-(\alpha+1)/\alpha} h_0(t) \exp\{-\theta(1 + H_0(t)\alpha)^{-1/\alpha}\}$
Aleatório	$(1 + \alpha H_0(t))^{-(\alpha+1)/\alpha} h_0(t) (\exp\{-\theta\} - 1)$

Note que as expressões para  $S_p(t)$  e  $f_p(t)$  do modelo ZICR-F gama para cada esquema de ativação, dadas nas Tabelas 24 e 25, são funções impróprias. Assim como foi feito nas seções 3.2, 3.3 e 3.4, nós calcularemos a função de sobrevivência dos não curados ( $S^*(t)$ ) através da elaboração do modelo em que se tem apenas indivíduos expostos ao evento de interesse, isto é, quando as causas competitivas encontra-se em todos os indivíduos ( $M \geq 1$ ). Assim, utilizando o mesmo artifício de Chen, Ibrahim e Sinha (1999), dada pela equação (2.8), o cálculo da função de sobrevivência  $S_p^*(t)$  para o modelo ZICR-F gamma para cada esquema de ativação é dado por

$$S_p^*(t) = P(T > t | M \geq 1) = \frac{\mathbb{P}(Z_1 > t, \dots, Z_M > t, M \geq 1)}{1 - P(M = 0)}. \quad (5.16)$$

Sendo assim, para obtermos as expressões da Equação (5.16) para o modelo proposto com fragilidade gama pra cada esquema de ativação, considere inicialmente os resultados obtidos nas Equações (3.3), (3.8) e (3.11) para  $S^*(t)$  do modelo ZICR para os três esquemas de ativação, os que são mostrados na Tabela 26.

Tabela 26 – Função de Sobrevivência para os indivíduos não curados ( $S^*(t)$ ) do modelo ZICR

Activation scheme	$S_p^*(t)$	$f_p^*(t)$
Minimum	$\frac{\exp(-\theta F_0(t)) - \exp(-\theta)}{1 - \exp(-\theta)}$	$\left(\frac{\exp(-\theta F_0(t))}{1 - \exp(-\theta)}\right) \theta f_0(t)$
Maximum	$\frac{1 - \exp(-\theta(1 - F_0(t)))}{1 - \exp(-\theta)}$	$\left(\frac{\exp(-\theta[1 - F_0(t)])}{1 - \exp(-\theta)}\right) \theta f_0(t)$
Random	$1 - F_0(t)$	$f_0(t)$

Seja  $F(t) = 1 - S(t)$  em que  $S(t)$  é dada pela Equação (5.7) e substituindo na Equação obtida em 3.11, nós obtemos a função de sobrevivência dos não curados para o modelo ZICR-F gamma sob o esquema de ativação mínimo em que esta é dada por:

$$\begin{aligned}
 S_p^*(t) = P(T > t | M \geq 1) &= \frac{\exp(-\theta F_0(t)) - \exp(-\theta)}{1 - \exp(-\theta)} \\
 &= \frac{\exp(-\theta(1 - S(t))) - \exp(-\theta)}{1 - \exp(-\theta)} \\
 &= \frac{\exp(-\theta(1 - (1 + H_0(t)\alpha)^{-1/\alpha})) - \exp(-\theta)}{1 - \exp(-\theta)}. \quad (5.17)
 \end{aligned}$$

Por consequência, a função de densidade de probabilidade associada ao modelo em (5.17) é dada por

$$f_p^*(t) = -\frac{d}{d(t)}S^*(t) = \left(\frac{\exp(-\theta F_0(t))}{1 - \exp(-\theta)}\right) \theta f_0(t) \quad (5.18)$$

$$= \left(\frac{\exp(-\theta(1 - (1 + H_0(t)\alpha)^{-1/\alpha}))}{1 - \exp(-\theta)}\right) \theta f_0, \quad (5.19)$$

em que  $H_0$  é a função de risco acumulada da distribuição basal para o tempo de sobrevivência. De forma similar, nós obtemos os resultados de  $S^*(t)$  e  $f^*(t)$  para os esquemas de ativação máximo e aleatório. As Tabelas 27 e 28 mostram, respectivamente, as expressões para  $S_p^*(t)$  e  $f_p^*(t)$  de cada esquema de ativação para o modelo ZICR-F gamm.

Tabela 27 – Expressões para  $S_p^*(t)$  para o modelo ZICR-F considerando cada esquema de ativação

Esquema de ativação	$S_p^*(t)$
Mínimo	$\frac{\exp(-\theta(1 - (1 + H_0(t)\alpha)^{-1/\alpha})) - \exp(-\theta)}{1 - \exp(-\theta)}$
Máximo	$\frac{1 - \exp(-\theta((1 + H_0(t)\alpha)^{-1/\alpha}))}{1 - \exp(-\theta)}$
Aleatório	$(1 + H_0(t)\alpha)^{-1/\alpha}$

Tabela 28 – Expressões para  $f_p^*(t)$  do modelo ZICR-F considerando cada esquema de ativação

Esquema de ativação	$f_p^*(t)$
Mínimo	$\left(\frac{\exp(-\theta(1 - (1 + H_0(t)\alpha)^{-1/\alpha}))}{1 - \exp(-\theta)}\right) \theta f_0$
Máximo	$\left(\frac{\exp(-\theta[(1 + H_0(t)\alpha)^{-1/\alpha}])}{1 - \exp(-\theta)}\right) \theta f_0$
Aleatório	$f_0(t)$

Note que as funções de sobrevivência e densidade são funções próprias, uma vez que as condições  $S_p^*(0) = 1$  e  $S_p^*(\infty) = 0$  estão satisfeitas. Em que  $f_0$ ,  $H_0(t)$  e  $h_0(t)$  são as funções densidade, de risco cumulativo e função de risco da distribuição basal, respectivamente. Dessa

forma, nós propomos uma modificação na equação da função de sobrevivência, dado por (5.9). A seguir, consideramos o modelo de taxa de cura zero inflacionado com fragilidade gama em que propomos uma nova função de sobrevivência da seguinte forma

$$S_{pop}(t) = p_1 + (1 - p_1 - p_0)S_p^*(t), \quad (5.20)$$

em que  $S_p^*(t)$ , para cada esquema de ativação (mínimo, máximo e aleatório), é dado na Tabela 27 e os parâmetros  $p_0$  e  $p_1$  são determinados, respectivamente, pelas expressões:  $p_0 = \exp(-\nu)$  e  $p_1 = \exp(-\theta)$  com  $\nu, \theta > 0$ .

Para assegurarmos que  $p_0$ ,  $p_1$  e  $(1 - p_0 - p_1)$  estejam no intervalo  $(0,1)$ , conforme já descrito no Capítulo 3, nós consideramos para o link entre dois vetores de covariáveis aos parâmetros relativo à inflação zero e à taxa de cura, respectivamente, como sendo:

$$p_{0i} = \exp(-\nu_i) \quad \text{e} \quad p_{1i} = \exp(-\theta_i),$$

em que  $\nu_i = -\log\left(\frac{\exp\{x_{1i}^T\beta_1\}}{1 + \exp\{x_{1i}^T\beta_1\} + \exp\{x_{2i}^T\beta_2\}}\right)$ ,  $\theta_i = -\log\left(\frac{\exp\{x_{2i}^T\beta_2\}}{1 + \exp\{x_{1i}^T\beta_1\} + \exp\{x_{2i}^T\beta_2\}}\right)$ ,  $\beta_1$  e  $\beta_2$  são vetores de coeficientes de regressão a serem estimados, onde o primeiro relaciona a influência das covariáveis no excesso de zeros e o segundo relaciona a influência das covariáveis na fração de cura.

Para definir a especificação do modelo, isto é, para estabelecer a forma paramétrica de  $S_p^*$ , consideraremos que  $f_0(t)$  e  $F_0(t)$  sejam, respectivamente, a função densidade e de probabilidade cumulativa da distribuição log-normal. Esta distribuição possui dois parâmetros:  $\mu$  e  $\sigma > 0$ , eles são os parâmetros de forma e escala, respectivamente. Sendo assim, nós conectaremos os parâmetros da log-normal como sendo:  $\mu_{1i} = \exp\{x_{3i}^T\beta_3\}$  e  $\sigma_{2i} = \exp\{x_{4i}^T\beta_4\}$ .

A construção da função de verossimilhança segue a mesma lógica apresentada no capítulo 3. Dessa forma temos que a função de verossimilhança é composta de três partes: (i) dos indivíduos com evento no início do tempo (hora zero); (ii) dos indivíduos não suscetíveis ao evento de interesse, e (iii) dos indivíduos suscetíveis ao evento de interesse. A expressão (5.21) apresenta a contribuição da verossimilhança de cada parcela

$$\begin{cases} p_{0i}, & \text{se } t_i = 0, \\ (1 - p_{0i} - p_{1i})f_p^*(t_i), & \text{se } t_i \text{ é totalmente observado,} \\ p_{1i} + (1 - p_{0i} - p_{1i})S_p^*(t_i), & \text{se } t_i \text{ é censurado.} \end{cases} \quad (5.21)$$

Considerando os dados da forma  $\mathbb{D} = (t_i, \delta_i, x_i = \{x_{1i}, x_{2i}, x_{3i}, x_{4i}\})$  em que  $\delta_i = 1$  significa que não houve censura enquanto que  $\delta_i = 0$  houve censura e  $x_i$  é o vetor de covariáveis relacionado com a observação  $i$ . Por outro lado, o vetor de covariáveis podem ser considerados iguais, ou seja,  $x_1 = x_2 = x_3 = x_4$ . Seja  $(\mu, \sigma)$  o vetor de parâmetros da distribuição Log-normal, e, por fim, temos  $(\beta_\nu, \beta_\theta)$  os parâmetros da regressão associados a proporção de inflação de zeros e da taxa de cura, respectivamente.

A função de verossimilhança para o modelo de sobrevivência de taxa de cura inflacionado no zero com fragilidade sob os esquemas de ativação, dado um vetor de parâmetros,  $\zeta = (\mu, \sigma, \alpha, \beta_v^\top, \beta_\theta^\top)^\top$  será estimado via EMV e é baseada em  $n$  observações. Por fim, a função de verossimilhança, sob censura não informativa é dada por:

$$L(\zeta, \mathbb{D}) \propto \prod_{t_i=0} \{p_{0i}\} \prod_{t_i>0} \left\{ [(1 - p_{0i} - p_{1i})f_p^*(t_i)]^{\delta_i} [p_{1i} + (1 - p_{0i} - p_{1i})S_p^*(t_i)]^{1-\delta_i} \right\}.$$

## 5.5 Estudo de simulação

Nessa seção, analisamos a performance dos EMV's dos modelos ZICR-F sob diferentes esquemas de ativação da mesma forma que foi feita na seção 3.6 do capítulo 3. A estimação através do método de máxima verossimilhança é realizada com a utilização do método de maximização BFGS da rotina R `optim()`. Para avaliar o desempenho dos estimadores, nós realizamos estudos de simulações para analisar as probabilidades de cobertura dos intervalos de confiança de 95% para os EMV's. As linhas cinza pontilhadas em torno de 0,95 representam os limites superior e inferior do intervalo são 0,936 e 0,964. Além disso, no estudo de simulação Monte Carlo também calculamos os resultados de desvios e EQM para os parâmetros estimados, para analisar o desempenho a medida que tamanho da amostra aumenta. Para o cálculo dos intervalos de confiança, nós consideramos a mesma metodologia assintótica descrita na subseção 3.5.1.

Com intuito de avaliar os efeitos das covariáveis, consideramos uma variável binária ( $X$ ) a qual assume 0 ou 1 e é proveniente da distribuição Bernoulli com parâmetro  $p$  igual à 0,5. Esta variável pode ser considerada, por exemplo, uma situação em que existam indivíduos pertencentes ao grupo de controle ou não, sendo atribuídos valor de covariável de 0 e 1, respectivamente.

O estudo de simulação é baseado em 1000 réplicas amostrais. Então, nós realizamos simulações de Monte Carlo considerando o tamanho da amostra  $n = 100, 250, 500, 750$  e 1000. Três estudos de simulação são realizados para o modelo proposto. A descrição do método de geração da amostra, ou seja, todos os detalhes da distribuição simulada do tempo de sobrevivência, e os resultados obtidos em relação ao método de estimação proposto são os mesmos descritos na seção 3.6 do Capítulo 3. Os cenários de parâmetros considerados nesse estudo de simulação são iguais aos da Tabela 5 do Capítulo 3 só que acrescentando diferentes valores para a fragilidade em que  $\alpha$  assume os valores 0.25 e 0.50 conforme Tabela 29.

Tabela 29 – Valor dos parâmetros sob diferentes cenário

Parâmetros	Cenários		
	1	2	3
$\beta_{10}$	-2.00	-0.50	-0.50
$\beta_{11}$	0.5	1.50	0.50
$\beta_{20}$	-1.00	-0.75	-1.00
$\beta_{21}$	0.75	3.00	1.00
$\beta_{30}$	0.50	-20	-0.75
$\beta_{31}$	0.50	1.50	1.00
$\beta_{40}$	1.50	-1.25	1.25
$\beta_{41}$	2.00	1.00	1.00
$\alpha$	(0,25;0,50)	(0,25;0,50)	(0.25,0.50)

As Figuras (20, 26, 27, 22, 28, 22, 21, 30 e 31) a seguir mostram os resultados obtidos através da simulação para os três cenários simulados de parâmetros, em que para cada cenário foi considerado diferentes valores de fragilidade ( $\alpha = 0,25$  e  $0,50$ ) em que em cada cenário a linha verde corresponde o modelo proposto com fragilidade igual à  $0,25$  enquanto que a linha azul corresponde ao valor de fragilidade igual a  $0,50$ . O tamanho da amostra assumiu valores iguais a  $n = 100, 250, 500, 750$  e  $1000$ . Sendo assim, a estrutura é dada por 9 parâmetros ( $\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \beta_{30}, \beta_{31}, \beta_{40}, \beta_{41}, \alpha$ ). Podemos obter as estimativas dos parâmetros da lognormal, taxa de cura e inflação zero através das expressões subsequentes:

$$v = -\log \left( \frac{\exp\{\beta_{10} + x_i\beta_{11}\}}{1 + \exp\{\beta_{10} + x_i\beta_{11}\} + \exp\{\beta_{20} + x_i\beta_{21}\}} \right),$$

$$\theta_{1i} = -\log \left( \frac{\exp\{\beta_{20} + x_i\beta_{21}\}}{1 + \exp\{\beta_{10} + x_i\beta_{11}\} + \exp\{\beta_{20} + x_i\beta_{21}\}} \right),$$

$$\mu_{1i} = \exp\{\beta_{30} + x_i\beta_{31}\}$$

$$\sigma_{2i} = \exp\{\beta_{40} + x_i\beta_{41}\}.$$

Os valores dos parâmetros são escolhidos para avaliarmos a performance da estimativa de máxima verossimilhança sob diferentes parâmetros de forma e escala ( $\beta_{30}, \beta_{31}, \beta_{40}, \beta_{41}$ , associados à distribuição de tempo a qual assumimos a distribuição log-normal), e também sob uma composição de diferentes proporções de zero ajustado ( $\beta_{10}$  e  $\beta_{11}$ ), taxas de curados ( $\beta_{20}$  e  $\beta_{21}$ ) e índice de fragilidade. Com resultados parecidos ao obtidos no capítulo anterior 3, podemos observar nas Figuras 20 até 31 que:

- Em geral, os vieses e os erros quadráticos médios diminuem à medida que o tamanho da amostra aumenta; observamos também que, em geral, a probabilidade de abrangência, ou seja, a proporção do tempo que o intervalo contém o valor verdadeiro de interesse, está próximo de 95%, como esperado;
- O parâmetro de fragilidade, representado por  $\alpha$ , possui um viés alto em tamanho de amostras pequenas nos cenários 1, 2 e 3, isto é, com baixo, médio e alto índice de fração de cura e zeros, respectivamente. Porém observamos que esse viés diminui à medida que o tamanho da amostra aumenta.
- Independente da quantidade de proporção de zeros e curados, o cenário em que consideramos o valor da fragilidade igual a 0,25 (linha verde) necessitou de um tamanho amostral maior para o viés diminuir e portanto teve um pior desempenho quando comparado ao cenário da linha azul ( $\alpha = 0,50$ ).
- O cenário com fragilidade igual a 0,5 (linha azul) em geral possuem um viés menor quando comparado ao cenário com fragilidade igual a 0,25 (linha verde) além de convergir mais rapidamente.
- De um modo geral, os parâmetros associados a distribuição log-normal, isto é,  $\beta_{10}$ ,  $\beta_{11}$ ,  $\beta_{20}$  e  $\beta_{21}$  também possuem um viés pequeno mas que converge a medida que o tamanho da amostra aumenta.

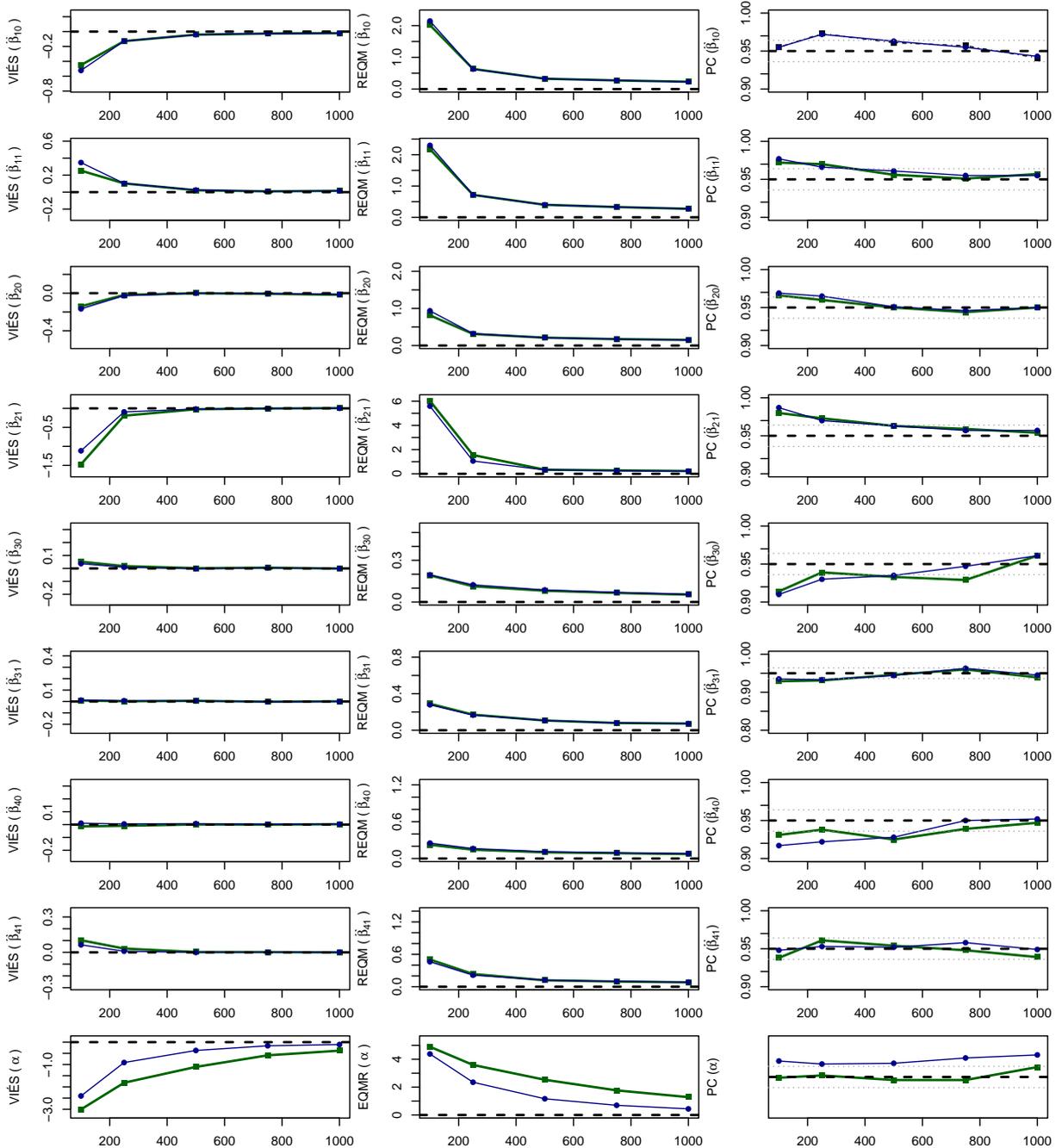


Figura 20 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 1 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo.

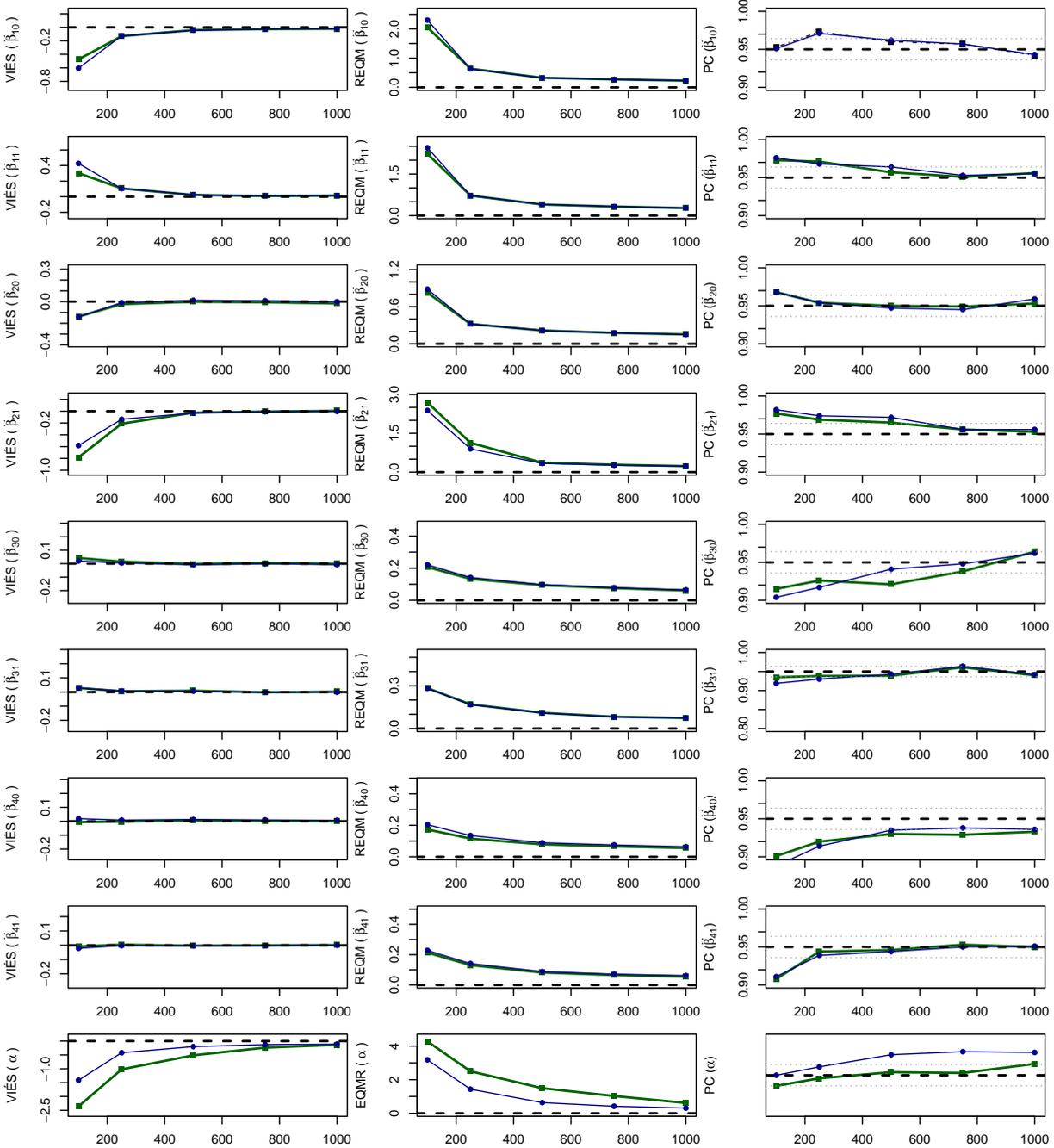


Figura 21 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 1 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação aleatório.

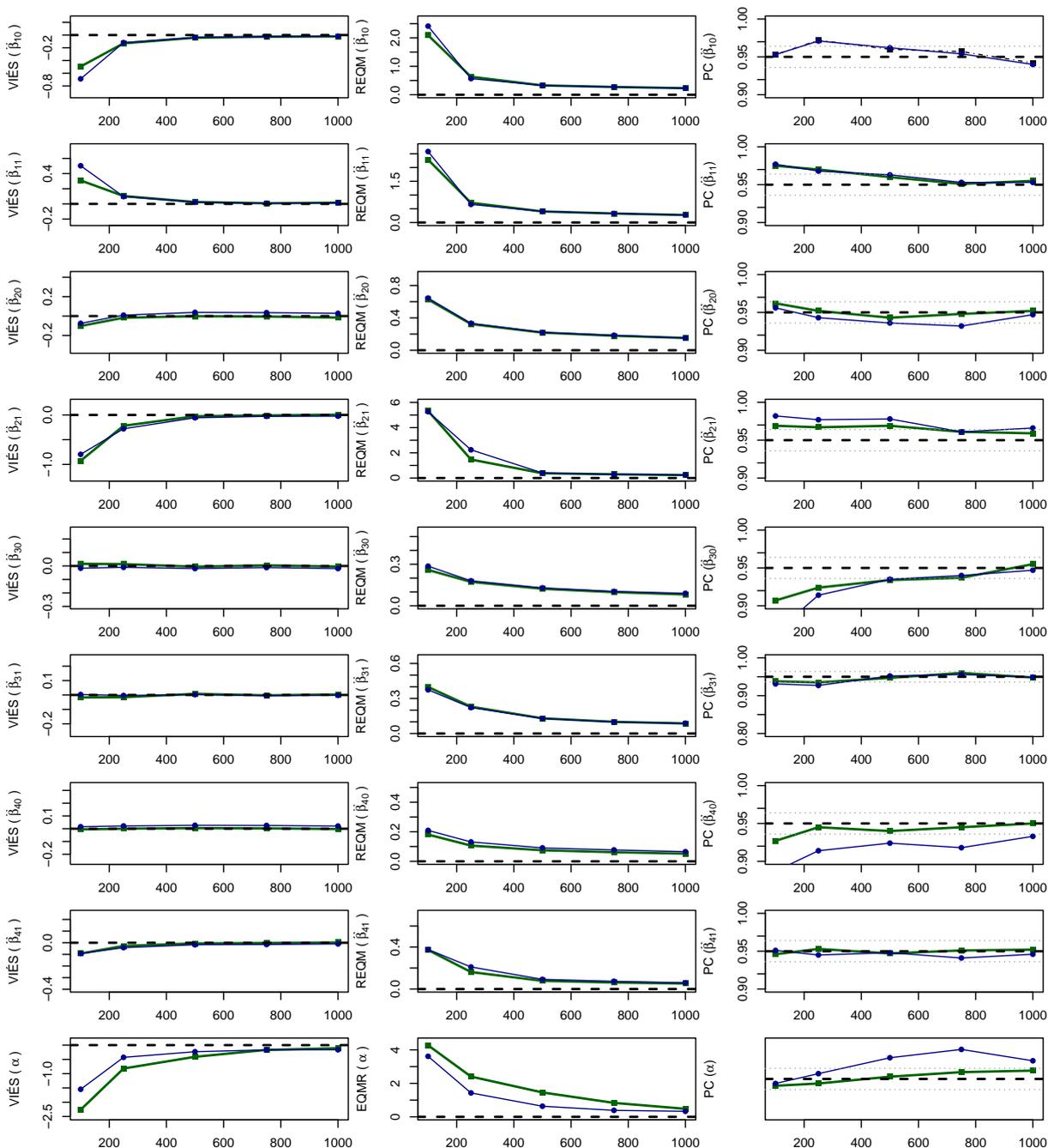


Figura 22 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 1 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação máximo.

## 5.6 Aplicação em dados reais

Nesta seção, apresentamos a aplicação do modelo de ZICR-F gama sob diferentes esquemas de ativação introduzido na Seção 5.4 nos mesmos conjunto de dados reais analisados no capítulo 3 isto é, Insulina Data Set e Hemorragia intracraniana em pacientes com câncer.

### 5.6.1 Insulina Data set

Nesta subseção, apresentamos a aplicação do modelo de ZICR-F sob diferentes esquemas de ativação utilizando o modelo final obtido no capítulo 3. Portanto, ajustamos o modelo com as mesmas covariáveis selecionadas mostradas na Tabela 14.

Na Figura 23, nós obtemos as estimativas através do modelo proposto ZICR-F para cada esquema de ativação considerando as variáveis explicativas descritas na Tabela 9. Os resultados mostram que as curvas estimada para cada esquema de ativação praticamente se sobrepõe, coincidindo com os resultados obtidos na seção anterior em que, nesta aplicação, visualmente não obtemos diferença entre os três esquemas de ativação. Por fim, verificamos que o modelo proposto obteve bons resultados para todas as variáveis .

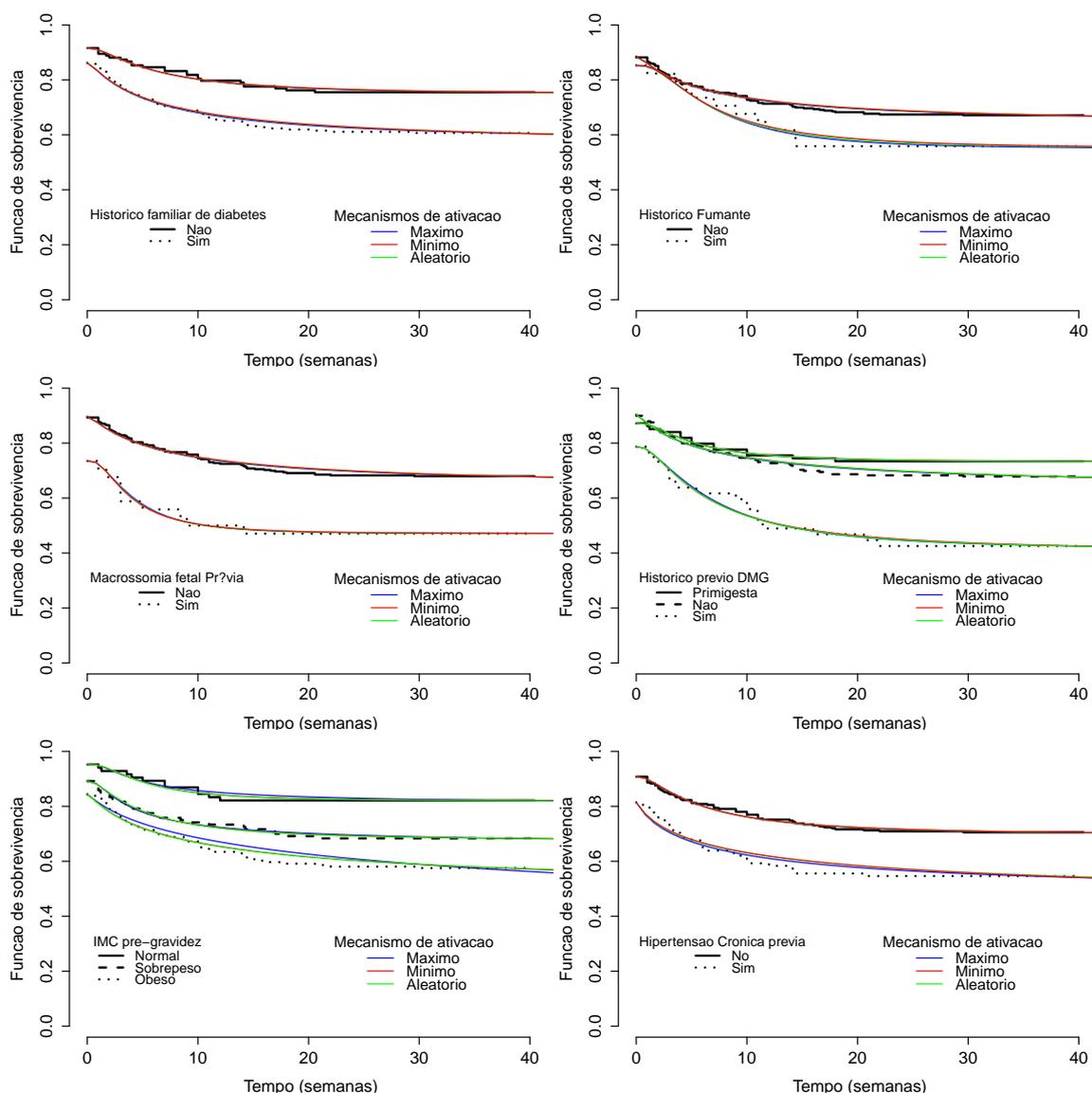


Figura 23 – Curva de sobrevida estimada obtida por Kaplan-Meier para dados de Insulina e as respectivas estimativas da função de sobrevivência de acordo com ajuste do modelo ZICR-F sob diferentes de mecanismos de ativações latente.

A Tabela 30 mostra os resultados do modelo final ZICR-F considerando os três cenários de ativação, e verificamos que nesse caso o esquema de ativação máximo apresentou o menor AIC e BIC. Por outro lado, os valores do AIC e BIC foram superiores aos encontrados no modelo sem fragilidade (capítulo 3 - Tabela 14). Nós verificamos também que nessa aplicação os esquemas de ativação tiveram valores bem próximos, isto significa que nesse caso os fatores competitivos que influenciam tanto no fato da mulher em período gestacional ter aplicação da insulina logo nos primeiros 7 dias ou na fração de cura, podem ocorrer apenas por 1 mas também por todas as causas competitivas. Considerando o critério AIC, o método de esquema de ativação máximo foi escolhido e baseado nessa escolha ajustamos os resíduos quantílicos randomizados para analisar a adequabilidade do modelo. O gráfico QQ plot dos resíduos quantílicos randomizados normalizados está representado na Figura 24 a qual verificamos que os resíduos estão alinhados na linha vermelha e portanto mostra uma boa adequação do modelo final considerando o esquema de ativação máximo.

Tabela 30 – EMV, EP, IC com 95%, AIC e BIC obtidos pelo modelo ZICR-F sob diferentes mecanismos de ativação ajustados para o conjunto de dados insulina

Parâmetros	Mínimo			
	EMV	Inferior	Superior	EP
$\beta_{10}$	-0.923	-1.344	-0.503	0.215
$\beta_{11}$ (Prior macrosomia)	1.108	0.268	1.947	0.428
$\beta_{20}$	1.407	0.901	1.914	0.258
$\beta_{21}$ (Smoking)	-0.311	-1.182	0.344	0.389
$\beta_{22}$ (History diabetes)	-0.758	-1.289	-0.227	0.271
$\alpha$		0.002		
AIC		1205.4		
BIC		1260.9		
Parâmetros	Máximo			
	EMV	Inferior	Superior	EP
$\beta_{10}$	-0.876	-1.281	-0.472	0.207
$\beta_{11}$ (Prior macrosomia)	1.107	0.268	1.946	0.428
$\beta_{20}$	1.462	1.002	1.992	0.235
$\beta_{21}$ (Smoking)	-0.419	-1.216	0.230	0.369
$\beta_{22}$ (History diabetes)	-0.726	-1.228	-0.224	0.256
$\alpha$		0.0001		
AIC		1205.2		
BIC		1260.7		
Parâmetros	Aleatório			
	EMV	Inferior	Superior	EP
$\beta_{10}$	-0.899	-1.312	-0.486	0.211
$\beta_{11}$ (Prior macrosomia)	1.110	0.275	1.949	0.428
$\beta_{20}$	1.433	1.176	1.991	0.247
$\beta_{21}$ (Smoking)	-0.358	-1.130	0.417	0.393
$\beta_{22}$ (History diabetes)	-0.736	-1.251	-0.221	0.263
$\alpha$		0.0005		
AIC		1205.7		
BIC		1261.1		

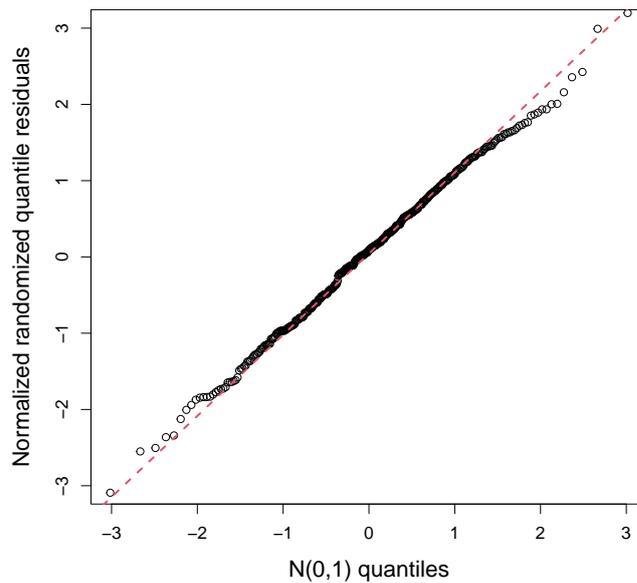


Figura 24 – Gráfico QQ plot dos resíduos quantílicos randomizados do modelo final sob esquema de ativação final

### 5.6.2 Hemorragia intracraniana

No segundo momento, ajustamos 2 modelos de forma independente de acordo com os modelos descritos na Tabela 27. O modelo, sob cada mecanismo, foi ajustado de forma independente considerando o efeito das variáveis mencionadas anteriormente. Além disso, para todos os modelos ajustados, a distribuição log-normal foi considerada para o tempo relacionada ao evento de interesse. Os resultados das estimativas dos modelos ajustados são apresentados na Figura 25 e na Tabela 31. A Figura 25 mostra que o modelo proposto sob o esquema de ativação máximo, em alguns casos, não obteve boas estimativas quando confrontado as estimativas de Kaplan Meier. Por outro lado, os esquemas Mínimo e aleatório performaram bem. Além disso, de acordo com os critérios de seleção AIC e BIC calculados para os modelos ajustados, percebemos que os valores desses critérios para o cenário máximo ficou acima dos demais.

Por fim, de um modo geral, concluímos que o cenário considerando o esquema de ativação aleatória apresentou o melhor ajuste. Na Tabela 31, analisamos que o termo da fragilidade ( $\alpha$ ) não foi significativo em nenhum dos cenários considerados. Ao comparar os resultados obtidos pelo modelo sem fragilidade do (capítulo 3), nós observamos que os valores obtidos pelos critérios AIC e BIC foram um pouco maior portanto o modelo ZICR-F nesse banco de dados com o objetivo de verificar a existência de heterogeneidade não observada foi não significativa.

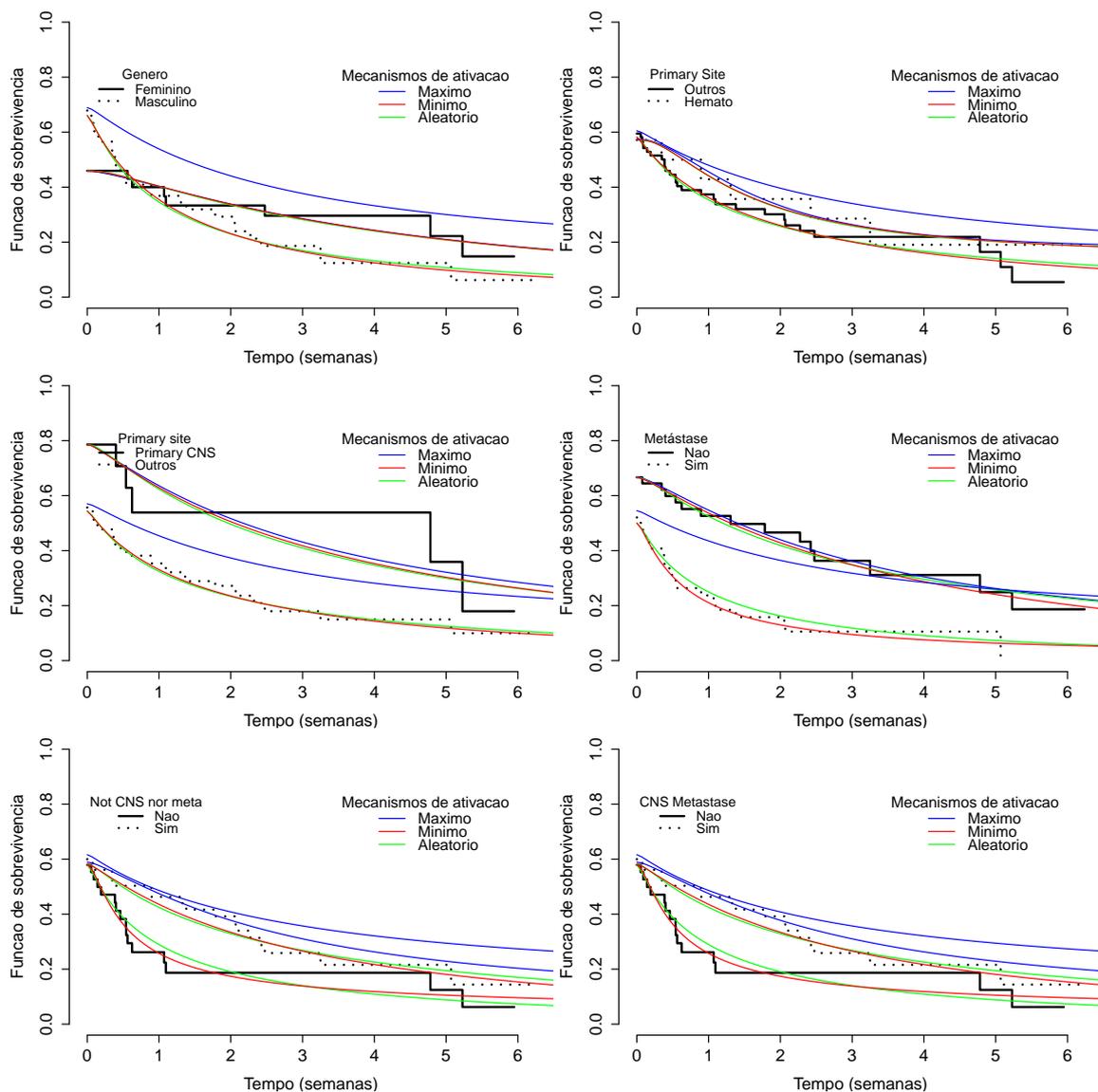


Figura 25 – Curva de sobrevivencia estimada obtida por Kaplan-Meier para dados de Hemorragia intracraniana considerando as variáveis e as respectivas estimativas da função de sobrevivência de acordo com ajuste do modelo ZICR-F sob diferentes de mecanismos de ativações latente

Tabela 31 – EMV, EP, IC com 95% e AIC, BIC obtidos por ZICR das variáveis CSNS Metastasi, CSNS Primary Site e Idade sob diferentes mecanismos de ativação considerando os dados de HI.

Parâmetro	Mínimo					Máximo					Aleatório						
	EMV	LI	LS	EP	IC	EMV	LI	LS	EP	IC	EMV	LI	LS	EP	IC		
$\beta_{1,1}$ intercepto	0.167776	-0.557	0.893	0.37	0.2120744	-0.47	0.894	0.348	0.1628058	-0.486	0.811	0.811	0.331	0.1628058	-0.486	0.811	0.331
$\beta_{1,2}$ Gênero	-0.83419	-1.746	0.078	0.465	-0.7690888	-1.662	0.124	0.456	-0.8292955	-1.682	0.023	0.023	0.435	-0.8292955	-1.682	0.023	0.435
$\beta_{2,1}$ intercepto	-5.21348	-66.382	55.955	31.209	-2.9591057	-5.988	0.07	1.546	-8.6790382	-311.235	293.877	293.877	154.365	-8.6790382	-311.235	293.877	154.365
$\beta_{2,2}$ Gênero	-5.33811	-86.514	75.838	41.416	1.6524404	-1.5	4.805	1.609	-4.8245857	-982.188	972.538	972.538	498.655	-4.8245857	-982.188	972.538	498.655
$\alpha$	0.412	-0.51741	1.34	0.474	0.00006	-	-	-	0.0001369	-	-	-	-	0.0001369	-	-	-
AIC	262.42																
BIC	285.22																
$p_0$ (NO)	0.54050	0.08200	0.37978	0.70122	0.54029	0.08200	0.37957	0.70101	0.54057	0.08200	0.37985	0.70129	0.54057	0.08200	0.37985	0.70129	0.54057
$p_0$ (YES)	0.33929	0.06300	0.21582	0.46277	0.31075	0.05900	0.19512	0.42639	0.33928	0.06300	0.21581	0.46276	0.33928	0.06300	0.21581	0.46276	0.33928
$p_1$ (NO)	0.00249	0.07700	-0.14843	0.15340	0.02267	0.03300	-0.04201	0.08735	0.00008	0.01200	-0.02344	0.02260	0.00008	0.01200	-0.02344	0.02260	0.00008
$p_1$ (YES)	0.00002	0.00000	0.00002	0.00002	0.14684	0.05300	0.04296	0.25072	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\beta_{1,1}$ intercepto	-0.69523	-1.315	-0.075	0.316	-0.6033095	-1.268	0.061	0.339	-0.693047	-1.313	-0.073	0.316	-0.693047	-1.313	-0.073	0.316	-0.693047
$\beta_{1,2}$ Metastase	0.706974	-0.167	1.581	0.446	0.6955209	-0.195	1.586	0.454	0.6929719	-0.146	1.532	0.428	0.6929719	-0.146	1.532	0.428	0.6929719
$\beta_{2,1}$ intercepto	-27.5829	-215.07	159.904	95.657	-2.2686471	-4.325	-0.212	1.049	-9.6933128	-201.024	181.638	181.638	97.618	-9.6933128	-201.024	181.638	97.618
$\beta_{2,2}$ Metastase	23.36665	-158.948	205.681	93.017	1.1182071	-1.145	3.382	1.155	-2.3613256	-421.203	416.481	416.481	213.695	-2.3613256	-421.203	416.481	213.695
$\alpha$	0.412	-7.53	8.35	4.05	0.000005	-	-	-	0.0000007	-	-	-	-	0.0000007	-	-	-
AIC	262.00																
BIC	284.79																
$p_0$ (NO)	0.333	0.070	0.196	0.470	0.331	0.070	0.194	0.469	0.333	0.070	0.196	0.471	0.333	0.070	0.196	0.471	0.333
$p_0$ (YES)	0.499	0.072	0.358	0.640	0.454	0.068	0.321	0.588	0.500	0.072	0.359	0.641	0.499	0.072	0.359	0.641	0.499
$p_1$ (NO)	7.00E-13	0	7.00E-13	7.00E-13	0.063	0.059	-0.05	0.18	0.0000	0.0040	0.0000	0.0079	0.0000	0.0040	0.0000	0.0079	0.0000
$p_1$ (YES)	0.00728	0.063	-0.1162	0.130758	0.13	0.05	0.03	0.23	0.00000	0.00100	0.00000	0.00196	0.00000	0.00100	0.00000	0.00196	0.00000
$\beta_{1,1}$ intercepto	-0.28	-0.71	0.14	0.22	-0.19	-0.64	0.26	0.23	-0.28	-0.70	0.14	0.22	-0.28	-0.70	0.14	0.22	-0.28
$\beta_{1,2}$ Idade	-0.31	-0.78	0.16	0.24	-0.19	-0.66	0.27	0.24	-0.32	-0.79	0.14	0.24	-0.32	-0.79	0.14	0.24	-0.32
$\beta_{2,1}$ intercepto	-7.95	-20.08	4.17	6.19	-1.39	-2.32	-0.46	0.48	-20.21	-75.23	34.81	28.07	-20.21	-75.23	34.81	28.07	-20.21
$\beta_{2,2}$ Idade	-3.36	-8.25	1.54	2.50	0.11	-0.94	1.16	0.53	-8.29	-29.48	12.90	10.81	-8.29	-29.48	12.90	10.81	-8.29
$\alpha$	0.0011	-0.145	0.149	0.07	0.0000002	-	-	-	0.0033784	-	-	-	0.0033784	-	-	-	0.0033784
AIC	270.93																
BIC	293.72																
AIC	283.77																
BIC	306.57																
AIC	269.22																
BIC	292.02																

## 5.7 Considerações finais

Nós introduzimos uma metodologia baseada em dados de sobrevivência inflacionados de zero com fragilidade o qual estende o modelo estudado em [Jr, Moreira e Louzada \(2017\)](#). Nesse sentido, uma vantagem de nossa abordagem é acomodar as situações que os pacientes expostos ao evento de interesse não têm risco homogêneo e existem várias causas competitivas para a ocorrência do evento de interesse o que não é possível no modelo de mistura padrão zero inflacionado. Para ilustrar a metodologia aqui apresentada, nós reaplicamos os dados de sobrevivência, a fim de avaliar a existência de alguma heterogeneidade não observada nos dados. Neste cenário, as informações são exploradas por meio da modelagem conjunta do tempo de sobrevivência zero, juntamente com os tempos de sobrevivência dos indivíduos curados.

Os resultados mostraram que o novo modelo teve um desempenho satisfatório. Apesar do novo modelo de fração cura zero inflacionado com fragilidade sob diferentes esquemas de ativação apresentar resultados ligeiramente piores em termos de AIC quando comparado ao modelo de taxa de cura inflado de zero apresentado no capítulo 3, é importante observar que o desempenho do modelo proposto pode ser observado através da aplicação em outros de bancos de dados cujos possuam características de fração de zeros ajustados e fração de cura.



---

## CONCLUSÕES

---

### 6.1 Observações finais

Nesta tese de doutorado, nós propomos modelos para analisar dados na área de saúde. Primeiro apresentamos um modelo de fração de cura inflacionado no zero sob diferentes esquemas de ativação, ou seja, introduzimos uma metodologia baseada em dados de sobrevivência inflacionados no zero que estendem o modelo estudado por [Jr, Moreira e Louzada \(2017\)](#) em que eles modificaram o modelo de taxa de cura padrão proposto por [Berkson e Gage \(1952\)](#) para uma configuração de risco de crédito.

A nossa abordagem permitiu estimar o esquema de ativação que melhor descreve a ocorrência do evento de interesse através dos esquemas latentes de ativação, os quais são chamados de: mínimo, máximo e aleatório. Estes podem ser incorporados em situações as quais são necessárias que todas ou alguma das causas de risco sejam ativadas para que o evento de interesse ocorra. Adicionalmente, conseguimos estimar as proporções de três subpopulações de pacientes que podem compor um estudo clínico: pacientes que experimentam o evento de interesse logo no início do estudo, outra parcela que acontece ao longo do estudo e os que não ocorre o evento de interesse. Para isso, muda-se a função de sobrevivência imprópria para contabilizar o excesso de zeros, que representa a taxa de pacientes que tem a ocorrência do evento de interesse logo no início do estudo.

Porém o modelo estudado inicialmente supõe que todos os pacientes expostos ao evento de interesse têm risco homogêneo, porém, a existência de uma heterogeneidade pode ser medida através da inclusão de uma variável aleatória chamada de fragilidade. Então a segunda parte da tese trata-se de uma extensão do modelo de mistura padrão zero ajustado com fragilidade estudado [Souza \(2020\)](#) o qual incorporamos causas competitivas para a ocorrência do evento de interesse e também uma estrutura latente de ativação isto é, diferentes esquemas de ativação para os ricos competitivos.

A apresentação desse novo modelo possibilitou uma alternativa na busca de modelos que considerem fração de curado, abrindo a possibilidade para novas generalizações, por exemplo, na mesma formulação que o modelo framework foi apresentado em [Cooner, Banerjee e McBean \(2006\)](#) e [Cooner et al. \(2007\)](#), incorporando excesso de zeros na modelagem e fragilidade. Sendo assim, contribuímos com a literatura estatística no que diz respeito à análise de inflação zero, que não havia sido incorporada à análise de sobrevivência que visa lidar com o risco de pacientes experimentarem o evento de interesse logo no início de estudos sob diferentes esquemas de ativação considerando a estrutura latente de ativação. Por fim, preenchemos uma lacuna faltante dentro da literatura a qual não existia modelo de fração cura zero inflacionados com riscos competitivos.

## 6.2 Propostas futuras

Para dar continuidade nos estudos dos modelos de fração de cura zero inflacionados (ou ajustados) sob diferentes esquemas de ativação, consideramos a distribuição Weibull para o tempo de vida e a distribuição Poisson para as causas competitivas, mas diferentes funções de densidade de linha de base podem ser consideradas para as causas competitivas. Assim, uma análise mais aprofundada pode ser realizada a fim de propor novos modelos de misturas inflacionadas de zero, com diferentes funções de densidade para acomodar o tempo até os dados do evento.

Na modelagem considerando modelos fração de cura zero inflacionados (ou ajustados) com fragilidades sob diferentes esquemas de ativação percebemos que a fragilidade possui um viés alto em pequenas amostras, assim desenvolver uma correção por exemplo, via *Bootstrap* podem ser consideradas para trabalhos futuros. Por exemplo [Vos et al. \(2015\)](#) propôs correção de viés e de fatores populacionais sobre riscos cumulativos ao longo da vida no coorte clínica retrospectiva de portadores com riscos de câncer de mama da mutação BRCA1/2.

Por fim, desenvolver uma abordagem Bayesiana para o modelos proposto no Capítulo 5. Por exemplo, [Souza et al. \(2022\)](#), propôs uma abordagem bayesiana ao modelo de fração de cura zero inflacionado em que os autores aplicaram a teoria desenvolvida no conjunto de dados relacionados a mulheres diagnosticadas com câncer cervical invasivo no Brasil.

## REFERÊNCIAS

---

---

AALEN, O. Nonparametric inference in connection with multiple decrement models. **Scandinavian Journal of Statistics**, JSTOR, p. 15–27, 1976. Citado na página 23.

\_\_\_\_\_. Nonparametric estimation of partial transition probabilities in multiple decrement models. **The Annals of Statistics**, JSTOR, p. 534–545, 1978. Citado na página 23.

ADAMIDIS, K.; LOUKAS, S. A lifetime distribution with decreasing failure rate. **Statistics & Probability Letters**, Elsevier, v. 39, n. 1, p. 35–42, 1998. Citado na página 24.

ALTSHULER, B. Theory for the measurement of competing risks in animal experiments. **Mathematical Biosciences**, Elsevier, v. 6, p. 1–11, 1970. Citado na página 23.

ANDERSEN, P. K.; BORGAN, Ø.; GILL, R. D.; KEIDING, N. Nonparametric estimation. In: **Statistical models based on counting processes**. [S.l.]: Springer, 1993. p. 176–331. Citado na página 94.

BENDER, R.; AUGUSTIN, T.; BLETTNER, M. Generating survival times to simulate cox proportional hazards models. **Statistics in medicine**, Wiley Online Library, v. 24, n. 11, p. 1713–1723, 2005. Citado na página 21.

BENÍTEZ-PAREJO, N.; ÁGUILA, M. R. del; PÉREZ-VICENTE, S. Survival analysis and cox regression. **Allergologia et immunopathologia**, Elsevier, v. 39, n. 6, p. 362–373, 2011. Citado na página 21.

BERKSON, J.; GAGE, R. P. Survival curve for cancer patients following treatment. **Journal of the American Statistical Association**, Taylor & Francis, v. 47, n. 259, p. 501–515, 1952. Citado nas páginas 22, 27, 28, 31, 32, 45, 46, 98 e 117.

BEYERSMANN, J.; LATOUCHE, A.; BUCHHOLZ, A.; SCHUMACHER, M. Simulating competing risks data in survival analysis. **Statistics in medicine**, Wiley Online Library, v. 28, n. 6, p. 956–971, 2009. Citado na página 23.

BOAG, J. W. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, v. 11, n. 1, p. 15–53, 1949. Citado nas páginas 22, 27, 28 e 31.

BOURGUIGNON, M.; SILVA, R. B.; CORDEIRO, G. M. A new class of fatigue life distributions. **Journal of Statistical computation and simulation**, Taylor & Francis, v. 84, n. 12, p. 2619–2635, 2014. Citado na página 40.

CACERES, J. A.; GOLDSTEIN, J. N. Intracranial hemorrhage. **Emergency medicine clinics of North America**, Elsevier, v. 30, n. 3, p. 771–794, 2012. Citado na página 71.

CALSAVARA, V. F.; RODRIGUES, A. S.; ROCHA, R.; TOMAZELLA, V.; LOUZADA, F. Defective regression models for cure rate modeling with interval-censored data. **Biometrical Journal**, Wiley Online Library, v. 61, n. 4, p. 841–859, 2019. Citado na página 61.

- CALSAVARA, V. F.; RODRIGUES, A. S.; ROCHA, R.; LOUZADA, F.; TOMAZELLA, V.; SOUZA, A. C.; COSTA, R. A.; FRANCISCO, R. P. Zero-adjusted defective regression models for modeling lifetime data. **Journal of Applied Statistics**, Taylor & Francis, v. 46, n. 13, p. 2434–2459, 2019. Citado na página 24.
- CALSAVARA, V. F.; TOMAZELLA, V. L.; FOGO, J. C. The effect of frailty term in the standard mixture model. **Chilean Journal of Statistics**, v. 4, n. 2, p. 95–109, 2013. Citado na página 94.
- CANCHO, V. G.; BANDYOPADHYAY, D.; LOUZADA, F.; YIQI, B. The destructive negative binomial cure rate model with a latent activation scheme. **Statistical methodology**, Elsevier, v. 13, p. 48–68, 2013. Citado nas páginas 25 e 40.
- CANCHO, V. G.; BARRIGA, G.; LEÃO, J.; SAULO, H. Survival model induced by discrete frailty for modeling of lifetime data with long-term survivors and change-point. **Communications in Statistics-Theory and Methods**, Taylor & Francis, p. 1–12, 2019. Citado na página 25.
- CANCHO, V. G.; CASTRO, M. de; DEY, D. K. Long-term survival models with latent activation under a flexible family of distributions. **Brazilian Journal of Probability and Statistics**, JSTOR, p. 585–600, 2013. Citado na página 40.
- CANCHO, V. G.; LOUZADA-NETO, F.; BARRIGA, G. D. The poisson-exponential lifetime distribution. **Computational Statistics & Data Analysis**, Elsevier, v. 55, n. 1, p. 677–686, 2011. Citado na página 24.
- CHEN, M.-H.; IBRAHIM, J. G.; SINHA, D. A new bayesian model for survival data with a surviving fraction. **Journal of the American Statistical Association**, Taylor & Francis, v. 94, n. 447, p. 909–919, 1999. Citado nas páginas 9, 11, 22, 23, 28, 31, 43, 44, 45, 93 e 101.
- CLAYTON, D. G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. **Biometrika**, Oxford University Press, v. 65, n. 1, p. 141–151, 1978. Citado na página 25.
- COONER, F.; BANERJEE, S.; CARLIN, B. P.; SINHA, D. Flexible cure rate modeling under latent activation schemes. **Journal of the American Statistical Association**, Taylor & Francis, v. 102, n. 478, p. 560–572, 2007. Citado nas páginas 25, 36, 37, 40 e 118.
- COONER, F.; BANERJEE, S.; MCBEAN, A. M. Modelling geographically referenced survival data with a cure fraction. **Statistical methods in medical research**, Sage Publications Sage CA: Thousand Oaks, CA, v. 15, n. 4, p. 307–324, 2006. Citado nas páginas 24, 25, 36 e 118.
- COX, D. R. Regression models and life-tables. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 34, n. 2, p. 187–202, 1972. Citado nas páginas 21, 94 e 95.
- DELGADO, J. J. F. **Modelos flexíveis para dados de tempos de vida em um cenário de riscos competitivos e mecanismos de ativação latentes**. Tese (Doutorado) — Universidade de São Paulo, 2014. Citado nas páginas 23, 37, 39, 40 e 41.
- ELBERS, C.; RIDDER, G. True and spurious duration dependence: The identifiability of the proportional hazard model. **The Review of Economic Studies**, Wiley-Blackwell, v. 49, n. 3, p. 403–409, 1982. Citado na página 95.

- EWELL, M.; IBRAHIM, J. G. The large sample distribution of the weighted log rank statistic under general local alternatives. **Lifetime Data Analysis**, Springer, v. 3, n. 1, p. 5–12, 1997. Citado na página 22.
- FAN, J.; LI, R. Variable selection for cox's proportional hazards model and frailty model. **Annals of Statistics**, JSTOR, p. 74–99, 2002. Citado na página 21.
- FISHER, L. D.; LIN, D. Y. Time-dependent covariates in the cox proportional-hazards regression model. **Annual review of public health**, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 20, n. 1, p. 145–157, 1999. Citado na página 21.
- GIOVANNETTI, G.; RICCHIUTI, G.; VELUCCHI, M. Size, innovation and internationalization: a survival analysis of italian firms. **Applied Economics**, Taylor & Francis, v. 43, n. 12, p. 1511–1520, 2011. Citado na página 21.
- GOLDMAN, A. I. Survivorship analysis when cure is a possibility: a monte carlo study. **Statistics in Medicine**, Wiley Online Library, v. 3, n. 2, p. 153–163, 1984. Citado na página 22.
- GONZALES, J. F. B.; TOMAZELLA, V. L. D.; TACONELLI, J. P. Estimac ao paramétrica do modelo de mistura com fragilidade gama na presença de covariáveis. **Rev. Bras. Biom**, v. 31, n. 2, p. 233–247, 2013. Citado na página 25.
- GRAY, R. J.; TSIATIS, A. A. A linear rank test for use when the main interest is in differences in cure rates. **Biometrics**, JSTOR, p. 899–904, 1989. Citado na página 22.
- GREENHOUSE, J. B.; WOLFE, R. A. A competing risks derivation of a mixture model for the analysis of survival data. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 13, n. 25, p. 3133–3154, 1984. Citado na página 22.
- HOEM, J. M. Point estimation of forces of transition in demographic models. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 33, n. 2, p. 275–289, 1971. Citado na página 23.
- HOUGAARD, P. Frailty models for survival data. **Lifetime data analysis**, Springer, v. 1, n. 3, p. 255–273, 1995. Citado na página 94.
- IBRAHIM, J. G.; CHEN, M.-H.; SINHA, D.; IBRAHIM, J.; CHEN, M. **Bayesian survival analysis**. [S.l.]: Springer, 2001. v. 2. Citado na página 22.
- JEFFREYS, H. **The theory of probability**. [S.l.]: OuP Oxford, 1998. Citado na página 80.
- JR, D. W. H.; LEMESHOW, S.; STURDIVANT, R. X. **Applied logistic regression**. [S.l.]: John Wiley & Sons, 2013. v. 398. Citado nas páginas 35 e 47.
- JR, M. Ribeiro de O.; MOREIRA, F.; LOUZADA, F. The zero-inflated promotion cure rate model applied to financial data on time-to-default. **Cogent Economics & Finance**, Taylor & Francis, v. 5, n. 1, p. 1395950, 2017. Citado nas páginas 9, 11, 23, 26, 32, 41, 43, 44, 47, 115 e 117.
- JÚNIOR, M. R. d. O. Models for inflated data applied to credit risk analysis. Universidade Federal de São Carlos, 2016. Citado nas páginas 34, 80 e 81.
- KAMINSKI, D.; GEISLER, C. Survival analysis of faculty retention in science and engineering by gender. **science**, American Association for the Advancement of Science, v. 335, n. 6070, p. 864–866, 2012. Citado na página 21.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American statistical association**, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958. Citado na página 21.

KUK, A. Y.; CHEN, C.-H. A mixture model combining logistic regression with proportional hazards regression. **Biometrika**, Oxford University Press, v. 79, n. 3, p. 531–541, 1992. Citado na página 22.

KUŞ, C. A new lifetime distribution. **Computational Statistics & Data Analysis**, Elsevier, v. 51, n. 9, p. 4497–4509, 2007. Citado na página 24.

KYRIAKOPOULOS, C. E.; CHEN, Y.-H.; CARDUCCI, M. A.; LIU, G.; JARRARD, D. F.; HAHN, N. M.; SHEVRIN, D. H.; DREICER, R.; HUSSAIN, M.; EISENBERGER, M. *et al.* Chemohormonal therapy in metastatic hormone-sensitive prostate cancer: long-term survival analysis of the randomized phase iii e3805 chaarted trial. **Journal of Clinical Oncology**, American Society of Clinical Oncology, v. 36, n. 11, p. 1080, 2018. Citado na página 21.

LANE, W. R.; LOONEY, S. W.; WANSLEY, J. W. An application of the cox proportional hazards model to bank failure. **Journal of Banking & Finance**, Elsevier, v. 10, n. 4, p. 511–531, 1986. Citado na página 21.

LEÃO, J.; BOURGUIGNON, M.; GALLARDO, D. I.; ROCHA, R.; TOMAZELLA, V. A new cure rate model with flexible competing causes with applications to melanoma and transplantation data. **Statistics in Medicine**, Wiley Online Library, v. 39, n. 24, p. 3272–3284, 2020. Citado na página 23.

LEÃO, J.; LEIVA, V.; SAULO, H.; TOMAZELLA, V. Incorporation of frailties into a cure rate regression model and its diagnostics and application to melanoma data. **Statistics in Medicine**, Wiley Online Library, v. 37, n. 29, p. 4421–4440, 2018. Citado nas páginas 25 e 94.

LOUZADA, F.; MOREIRA, F. F.; OLIVEIRA, M. R. de. A zero-inflated non default rate regression model for credit scoring data. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 47, n. 12, p. 3002–3021, 2018. Citado na página 24.

LOUZADA-NETO, F. Polyhazard models for lifetime data. **Biometrics**, Wiley Online Library, v. 55, n. 4, p. 1281–1285, 1999. Citado na página 24.

LU, K.; TSIATIS, A. A. Multiple imputation methods for estimating regression coefficients in the competing risks model with missing cause of failure. **Biometrics**, Wiley Online Library, v. 57, n. 4, p. 1191–1197, 2001. Citado na página 24.

\_\_\_\_\_. Comparison between two partial likelihood approaches for the competing risks model with missing cause of failure. **Lifetime Data Analysis**, Springer, v. 11, n. 1, p. 29–40, 2005. Citado na página 24.

MALLER, R. A.; ZHOU, X. **Survival analysis with long-term survivors**. [S.l.]: John Wiley & Sons, 1996. Citado na página 22.

MARTIN, T. G.; WINTLE, B. A.; RHODES, J. R.; KUHNERT, P. M.; FIELD, S. A.; LOW-CHOY, S. J.; TYRE, A. J.; POSSINGHAM, H. P. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. **Ecology letters**, Wiley Online Library, v. 8, n. 11, p. 1235–1246, 2005. Citado nas páginas 23 e 32.

MASUD, A.; TU, W.; YU, Z. Variable selection for mixture and promotion time cure rate models. **Statistical methods in medical research**, SAGE Publications Sage UK: London, England, v. 27, n. 7, p. 2185–2199, 2018. Citado na página 22.

MCCULLAGH, P.; NELDER, J. A. Generalized linear models, volume 37 of. **Monographs on statistics and applied probability**, 1989. Citado na página 35.

MIGON, H. S.; GAMERMAN, D.; LOUZADA, F. **Statistical inference: an integrated approach**. [S.l.]: CRC press, 2014. Citado nas páginas 49 e 79.

MILANI, E. A.; CALSAVARA, V. F.; RODRIGUES, J. S.; TOMAZELLA, C. D.; BONINI, C. R.; ANTUNES, C. C.; SANTOS, D. G. d.; ZANOTTI, M. A.; TOMAZELLA, V. L. D. Análise da sobrevida de mulheres diagnosticadas com câncer de mama triplo negativo utilizando modelo de longa duração. Universidade Federal de Goiás, 2021. Citado na página 22.

MINAYO, M. C. d. S. Saúde-doença: uma concepção popular da etiologia. **Cadernos de saúde Pública**, SciELO Brasil, v. 4, n. 4, p. 363–381, 1988. Citado na página 24.

MORITA, L. H.; TOMAZELLA, V. L.; BALAKRISHNAN, N.; RAMOS, P. L.; FERREIRA, P. H.; LOUZADA, F. Inverse gaussian process model with frailty term in reliability analysis. **Quality and Reliability Engineering International**, Wiley Online Library, 2020. Citado na página 25.

NASCIMENTO, E. M. do; LOBO, C. L. de C.; PEREIRA, B. de B.; BALLAS, S. K. Survival probability in patients with sickle cell anemia using the competitive risk statistical model. **Mediterranean journal of hematology and infectious diseases**, Catholic University in Rome, v. 11, n. 1, 2019. Citado nas páginas 23 e 38.

NELSON, W. Hazard plotting for incomplete failure data. **Journal of Quality Technology**, Taylor & Francis, v. 1, n. 1, p. 27–52, 1969. Citado na página 23.

NIE, Z.; OU, Y.; QU, Y.; YUAN, H.; LIU, X. A new perspective of survival data on clinical epidemiology: introduction of competitive risk model. **Zhonghua liu Xing Bing xue za zhi= Zhonghua Liuxingbingxue Zazhi**, v. 38, n. 8, p. 1127–1131, 2017. Citado na página 38.

NUNES, A.; SARMENTO, E. de M. Business demography dynamics in portugal: a non-parametric survival analysis. In: **The shift to the entrepreneurial society**. [S.l.]: Edward Elgar Publishing, 2012. Citado na página 21.

OAKES, D. A model for association in bivariate survival data. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 44, n. 3, p. 414–422, 1982. Citado na página 25.

ONCOLOGY, A. S. of C. *et al.* Cancer. net. 2008. Citado na página 24.

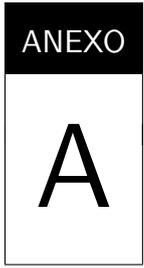
OSPINA, R.; FERRARI, S. L. A general class of zero-or-one inflated beta regression models. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 6, p. 1609–1623, 2012. Citado na página 49.

PEREIRA, G. H.; BOTTER, D. A.; SANDOVAL, M. C. A regression model for special proportions. **Statistical Modelling**, Sage Publications Sage India: New Delhi, India, v. 13, n. 2, p. 125–151, 2013. Citado nas páginas 34, 35 e 47.

- PETERSON, A. V. **Nonparametric estimation in the competing risks problem**. [S.l.]: Department of Statistics, Stanford University, 1975. Citado na página 23.
- ROCHA, R.; NADARAJAH, S.; TOMAZELLA, V.; LOUZADA, F.; EUDES, A. New defective models based on the kumaraswamy family of distributions with application to cancer data sets. **Statistical methods in medical research**, SAGE Publications Sage UK: London, England, v. 26, n. 4, p. 1737–1755, 2017. Citado na página 50.
- RODRIGUES, J.; CANCHO, V. G.; CASTRO, M. de; LOUZADA-NETO, F. On the unification of long-term survival models. **Statistics & Probability Letters**, Elsevier, v. 79, n. 6, p. 753–759, 2009. Citado na página 23.
- ROMAN, M. *et al.* Modelos para dados de sobrevivência na presença de diferentes esquemas de ativação baseados na distribuição geométrica. Universidade Federal de São Carlos, 2013. Citado nas páginas 9, 11, 25, 35 e 37.
- RONDEAU, V.; SCHAFFNER, E.; CORBIERE, F.; GONZALEZ, J. R.; MATHOULIN-PÉLISSIER, S. Cure frailty models for survival data: Application to recurrences for breast cancer and to hospital readmissions for colorectal cancer. **Statistical methods in medical research**, Sage Publications Sage UK: London, England, v. 22, n. 3, p. 243–260, 2013. Citado na página 25.
- RStudio Team. **RStudio: Integrated Development Environment for R**. Boston, MA, 2020. Disponível em: <<http://www.rstudio.com/>>. Citado nas páginas 49, 53 e 82.
- SHANKER, R.; HAGOS, F. On poisson–sujatha distribution and its applications to model count data from biological sciences. **Biometrics & Biostatistics International Journal**, v. 3, n. 4, p. 1–7, 2016. Citado na página 32.
- SHANNO, D. F. Conditioning of quasi-newton methods for function minimization. **Mathematics of computation**, v. 24, n. 111, p. 647–656, 1970. Citado na página 49.
- SOUZA, H. C. C. de; LOUZADA, F.; RAMOS, P. L.; JÚNIOR, M. R. de O.; PERDONÁ, G. d. S. C. A bayesian approach for the zero-inflated cure model: an application in a brazilian invasive cervical cancer database. **Journal of Applied Statistics**, Taylor & Francis, v. 49, n. 12, p. 3178–3194, 2022. Citado nas páginas 47 e 118.
- SOUZA, H. C. Cavenague de. **Modelos de sobrevida para dados inflacionados de zeros aplicados a progressão do trabalho de parto**. Tese (Doutorado) — Universidade de São Paulo, 2020. Citado nas páginas 25, 97 e 117.
- SOUZA, H. C. Cavenague de; LOUZADA, F.; OLIVEIRA, M. R. de; FAWOLE, B.; AKINTAN, A.; OYENEYIN, L.; SANNI, W.; PERDONÁ, G. d. S. C. The log-normal zero-inflated cure regression model for labor time in an african obstetric population. **Journal of Applied Statistics**, Taylor & Francis, v. 49, n. 9, p. 2416–2429, 2022. Citado nas páginas 24 e 25.
- STROUSE, J. J.; FEARS, T. R.; TUCKER, M. A.; WAYNE, A. S. Pediatric melanoma: risk factor and survival analysis of the surveillance, epidemiology and end results database. **J Clin Oncol**, v. 23, n. 21, p. 4735–4741, 2005. Citado na página 21.
- STURTZ, S.; LIGGES, U.; GELMAN, A. E. R2winbugs: a package for running winbugs from r. 2005. Citado na página 82.

- TOMAZELLA, V. L. D. **Modelagem de dados de eventos recorrentes via processo de Poisson com termo de fragilidade**. Tese (Doutorado) — Universidade de São Paulo, 2003. Citado na página 95.
- TSODIKOV, A. D.; YAKOVLEV, A. Y.; ASSELAIN, B. **Stochastic models of tumor latency and their biostatistical applications**. [S.l.]: World Scientific, 1996. v. 1. Citado nas páginas 22, 38, 43 e 93.
- VAUPEL, J. W.; MANTON, K. G.; STALLARD, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. **Demography**, Springer, v. 16, n. 3, p. 439–454, 1979. Citado nas páginas 25 e 95.
- VIANA, P. V. de S.; PAIVA, N. S.; VILLELA, D. A. M.; BASTOS, L. S.; BIERRENBACH, A. L. de S.; BASTA, P. C. Factors associated with death in patients with tuberculosis in brazil: survival analysis with competitive risks. 2019. Citado na página 23.
- VOS, J. R.; HSU, L.; BROHET, R. M.; MOURITS, M. J.; VRIES, J. de; MALONE, K. E.; OOSTERWIJK, J. C.; BOCK, G. H. de. Bias correction methods explain much of the variation seen in breast cancer risks of brca1/2 mutation carriers. **Journal of Clinical Oncology**, American Society of Clinical Oncology, v. 33, n. 23, p. 2553, 2015. Citado na página 118.
- YAKOVLEV, A.; ASSELAIN, B.; BARDOU, V.; FOURQUET, A.; HOANG, T.; ROCHEFEDIERE, A.; TSODIKOV, A. A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. **Biometrie et Analyse de Donnees Spatio-Temporelles**, v. 12, p. 66–82, 01 1993. Citado na página 43.
- YAKOVLEV, A. Y. Threshold models of tumor recurrence. **Mathematical and computer modelling**, Elsevier, v. 23, n. 6, p. 153–164, 1996. Citado nas páginas 9, 11, 22, 38 e 43.
- YAKOVLEV, A. Y.; ASSELAIN, B.; BARDOU, V.; FOURQUET, A.; HOANG, T.; ROCHEFEDIERE, A.; TSODIKOV, A. A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. **Biometrie et analyse de donnees spatio-temporelles**, Société Française de Biométrie, ENSA Rennes France, v. 12, p. 66–82, 1993. Citado nas páginas 22 e 38.
- YAMAKI, V. N.; TELLES, J. P. M.; PAIVA, W. S.; TEIXEIRA, M. J.; NEVILLE, I. S. Surgical treatment of spontaneous intracranial hemorrhage in patients with cancer: analysis of prognostic factors. **Journal of Clinical Neuroscience**, Elsevier, v. 83, p. 140–145, 2021. Citado na página 70.
- YIN, G.; IBRAHIM, J. G. Cure rate models: a unified approach. **Canadian Journal of Statistics**, Wiley Online Library, v. 33, n. 4, p. 559–570, 2005. Citado na página 22.
- YIQI, B.; CANCHO, V. G.; NETO, F. L. *et al.* The weibull-negative-binomialmodel under latent failure causes with randomized activation schemes. São Carlos, SP, Brasil., 2011. Citado na página 25.
- YU, B.; PENG, Y. Mixture cure models for multivariate survival data. **Computational Statistics & Data Analysis**, Elsevier, v. 52, n. 3, p. 1524–1532, 2008. Citado na página 94.





# FIGURAS



### A.1 Figuras - Capítulo 3

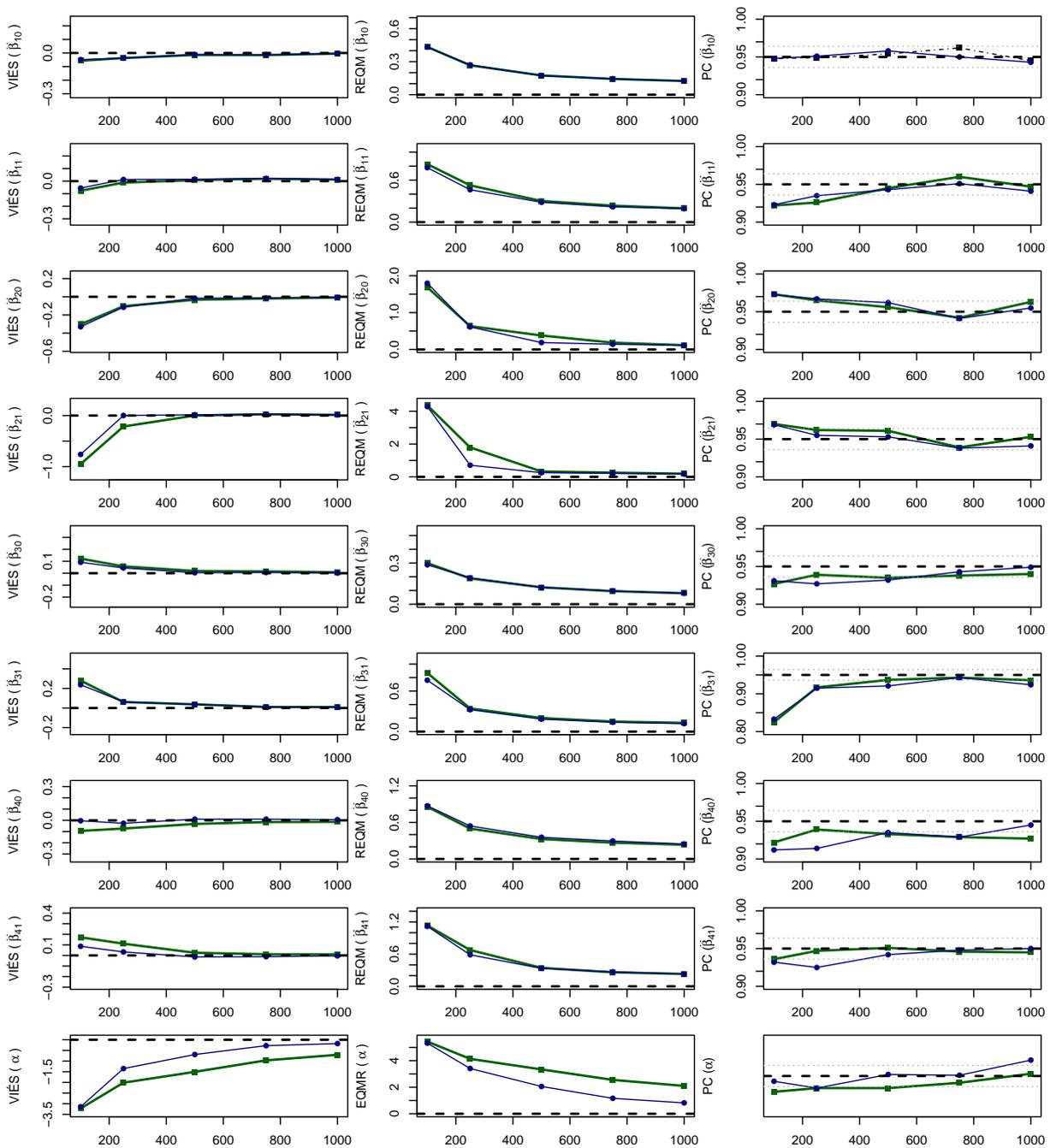


Figura 26 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 2 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo

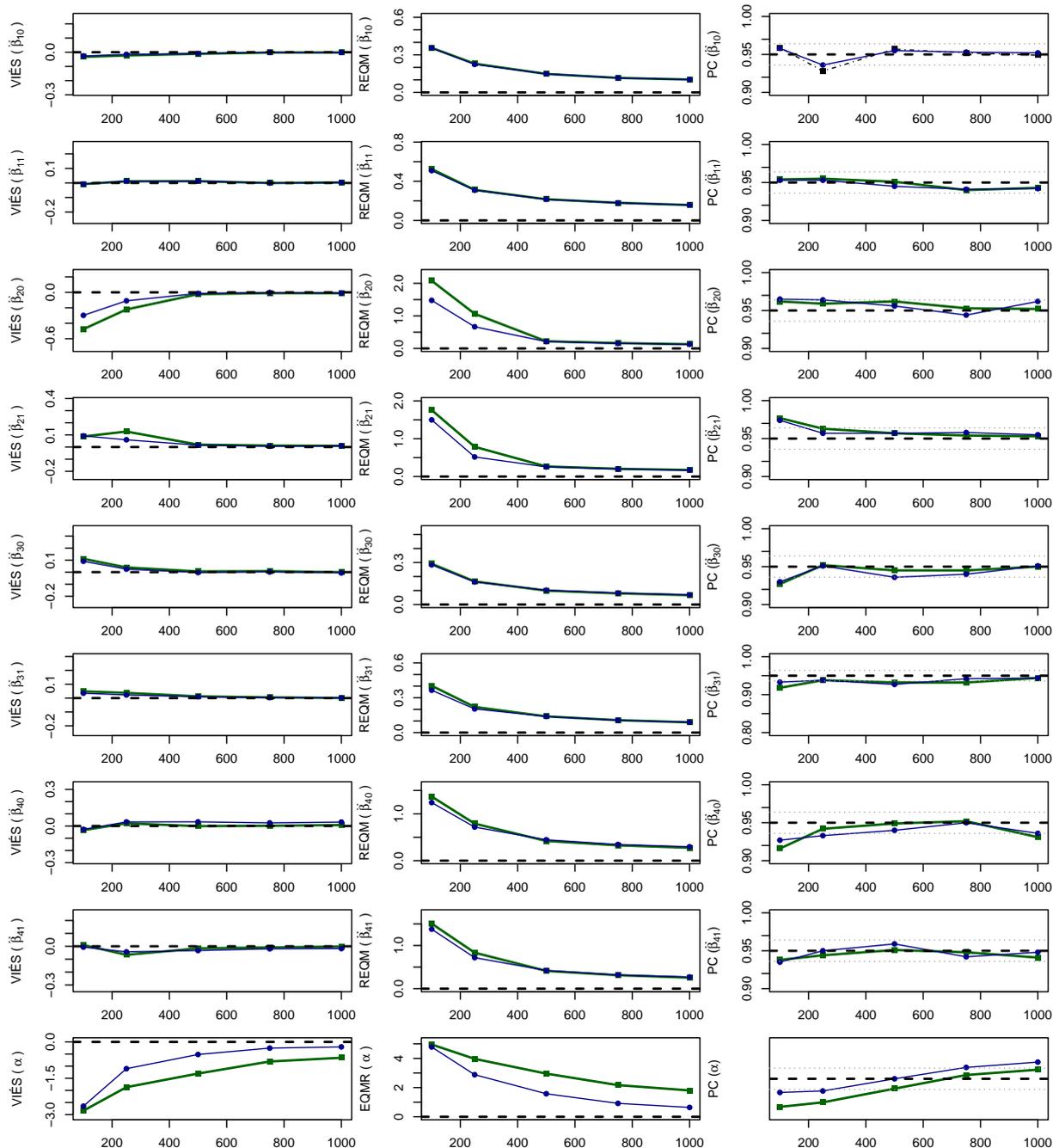


Figura 27 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 3 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo

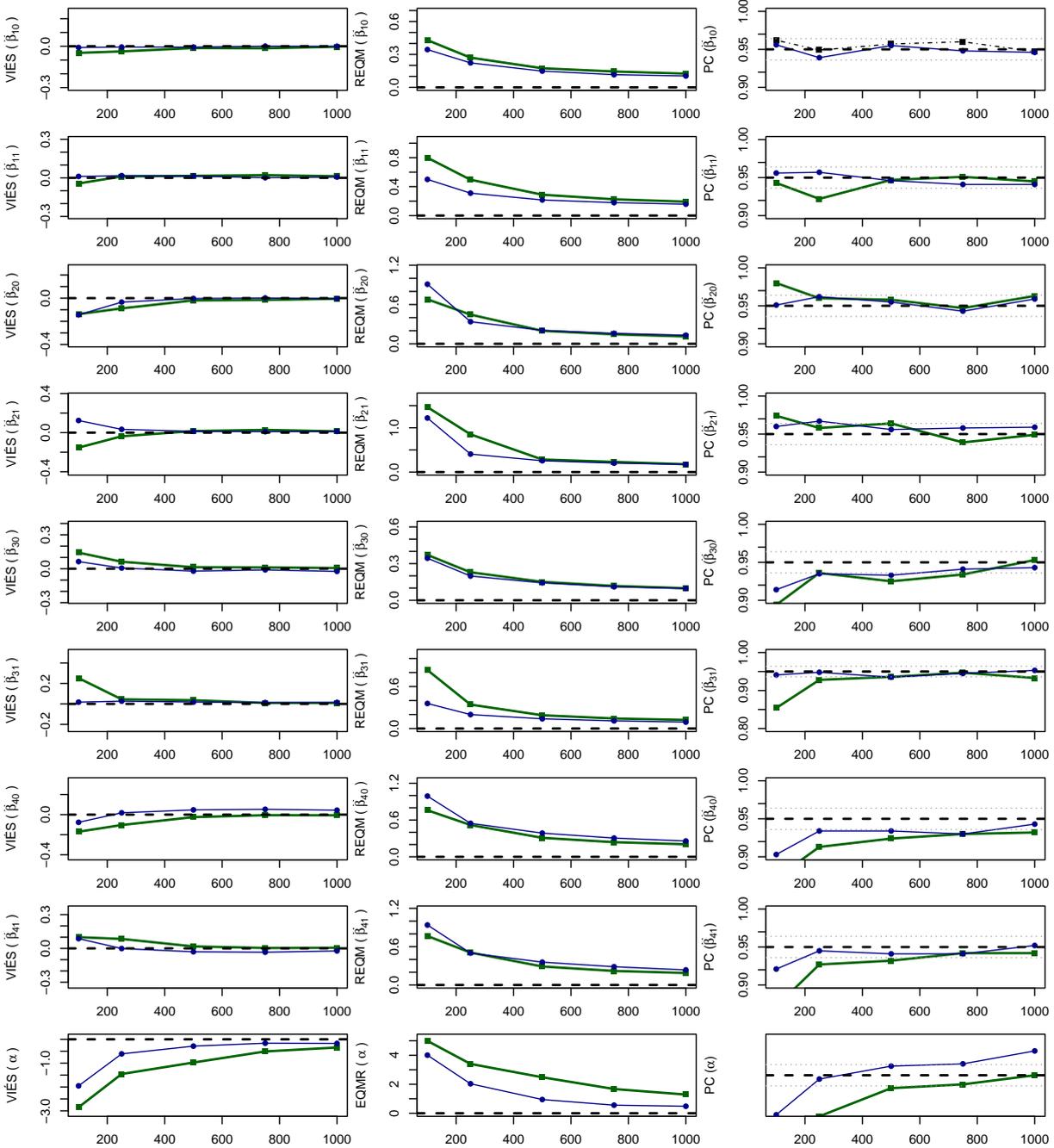


Figura 28 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 2 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo

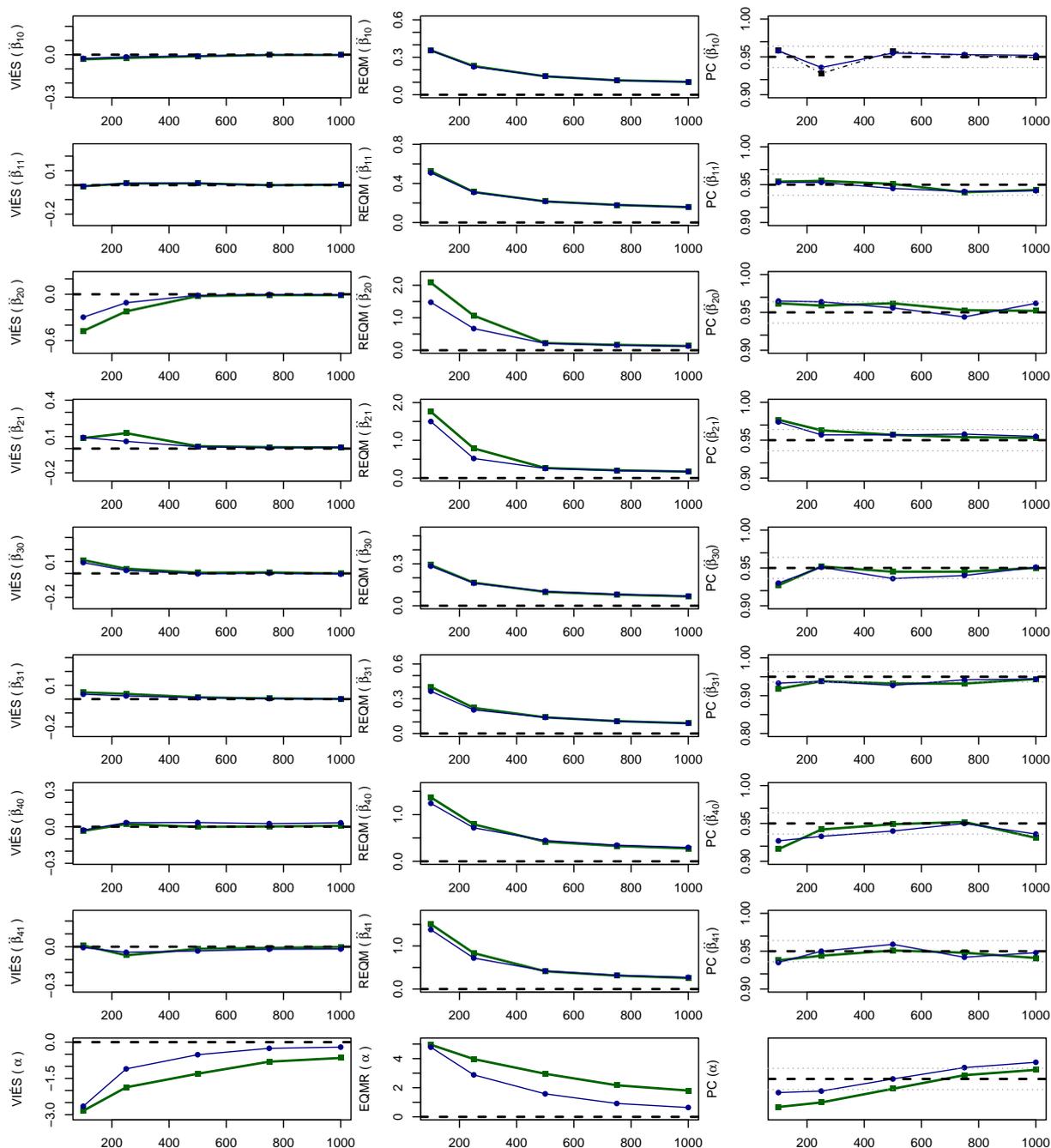


Figura 29 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 3 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo

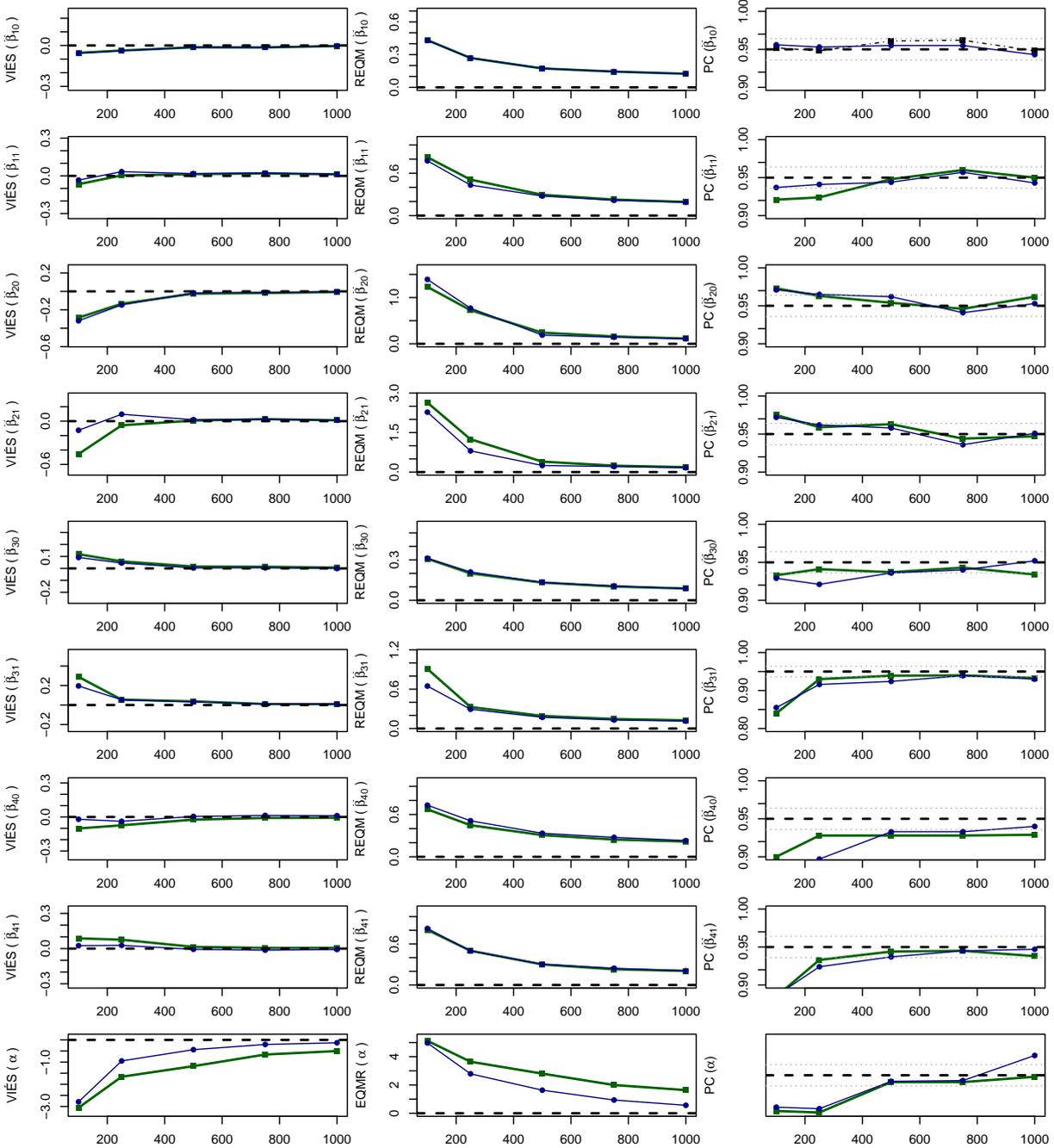


Figura 30 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 2 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo

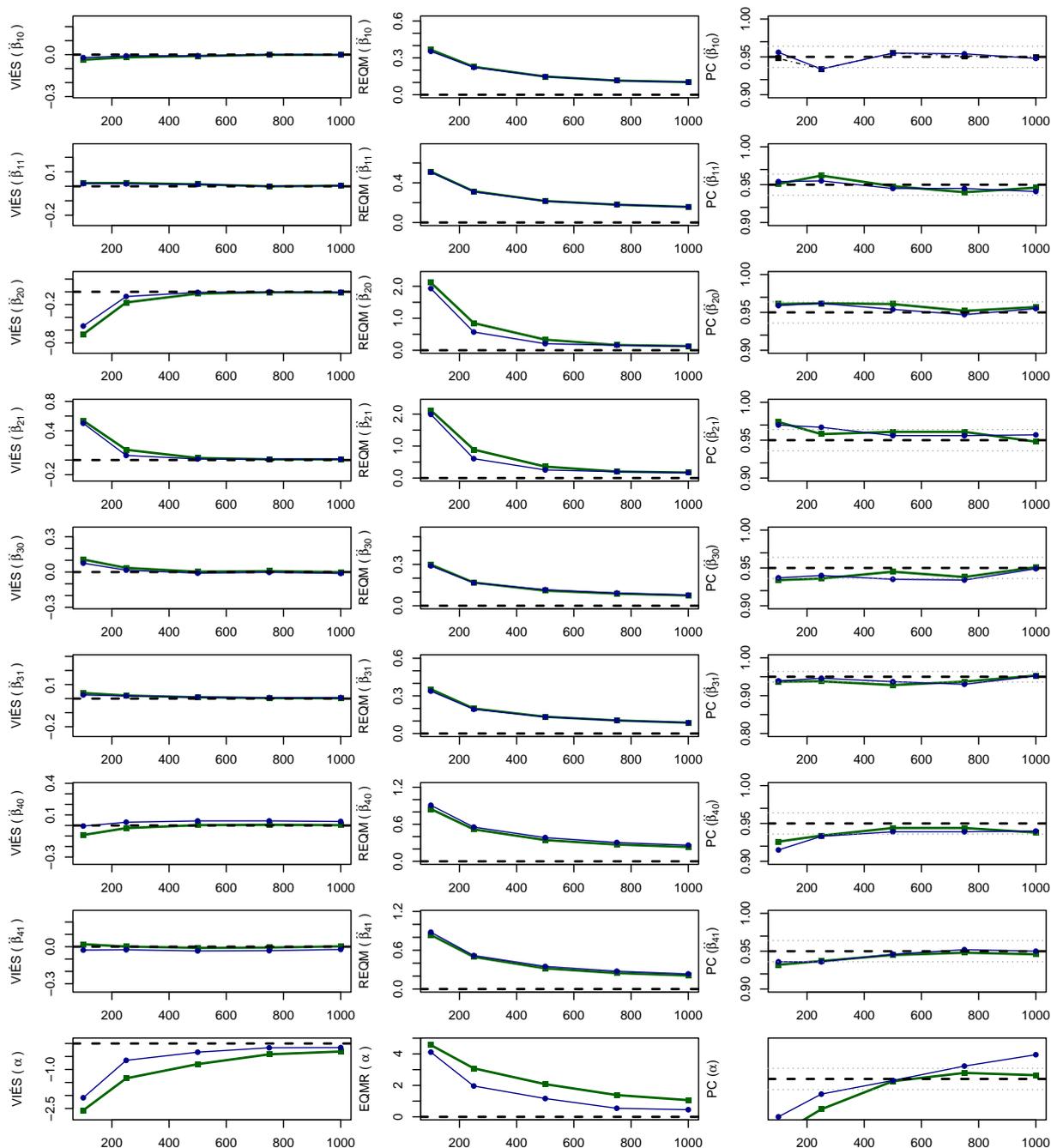


Figura 31 – Viés, REQM e cobertura empírica para intervalos de confiança de 95 % versus tamanho da amostra para dados simulados sob o cenário 3 de parâmetros do modelo zero ajustado de fração de cura com fragilidade sob esquema de ativação mínimo

## A.2 Figuras - Aplicação: Capítulo 4

### A.2.1 Variável Histórico Fumante

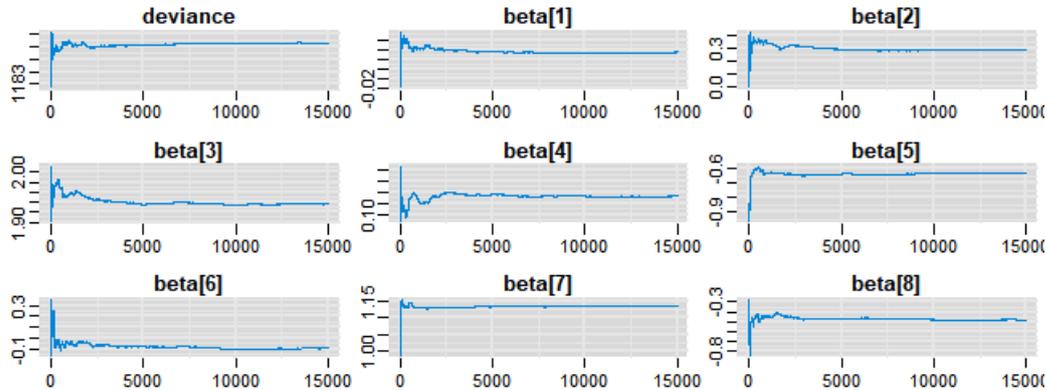


Figura 32 – Gráfico traço de média para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo

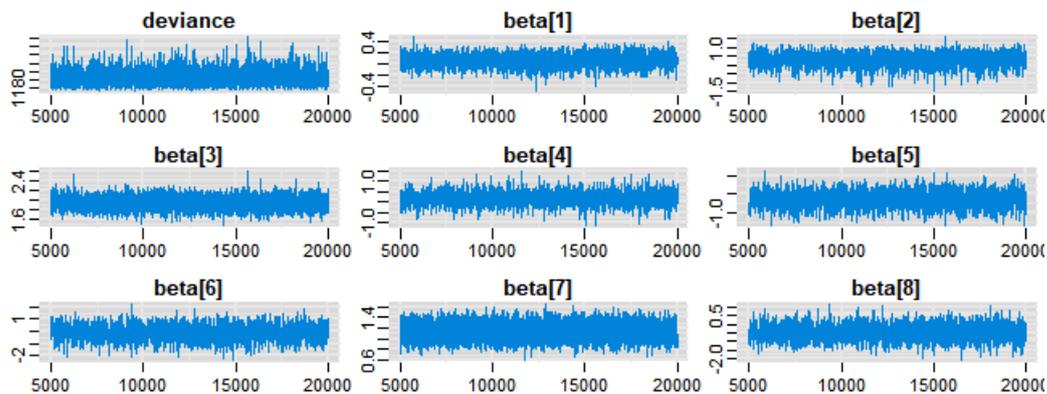


Figura 33 – Gráfico Histórico das Cadeias para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo

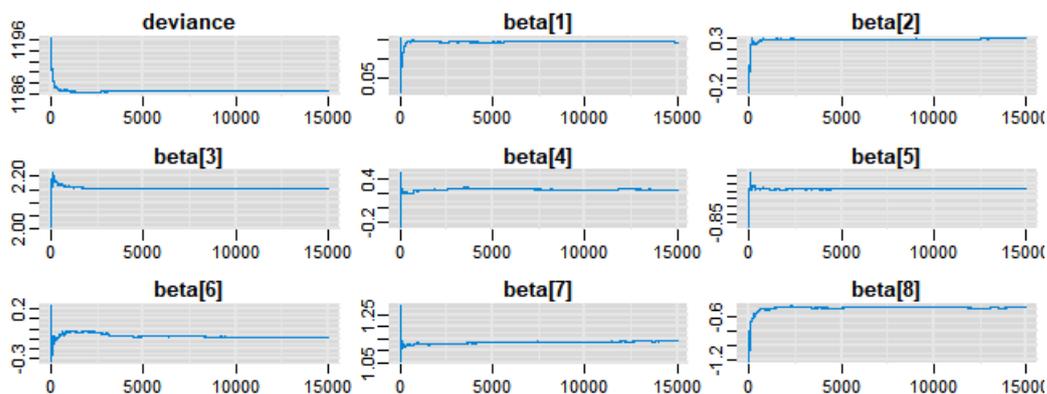


Figura 34 – Gráfico traço de média para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação mínimo

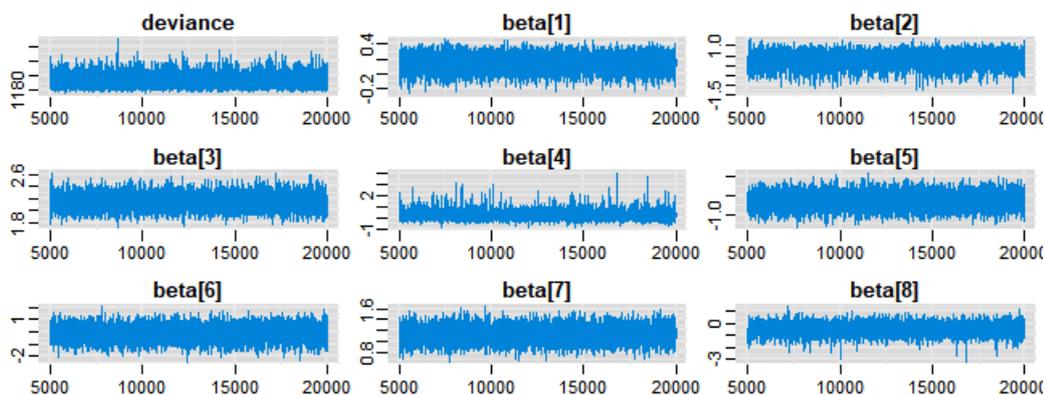


Figura 35 – Gráfico Histórico das Cadeias para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação mínimo

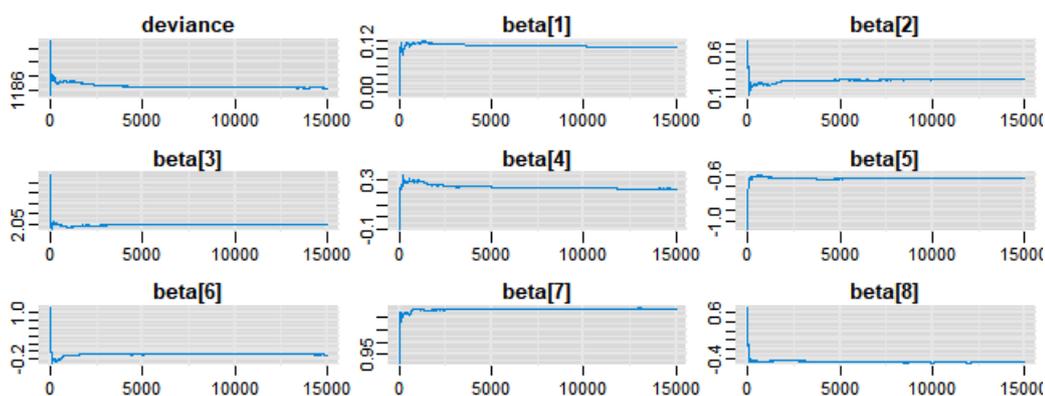


Figura 36 – Gráfico traço de média para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório

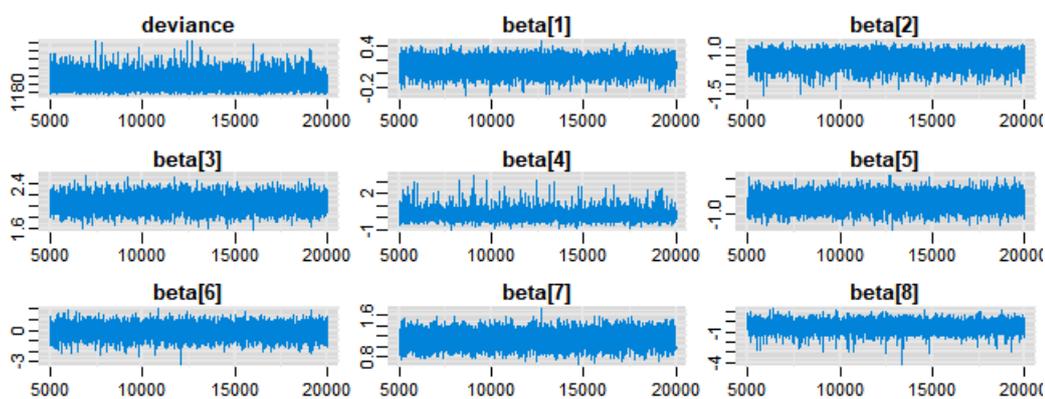


Figura 37 – Gráfico Histórico das Cadeias para a variável histórico fumante considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório

### A.2.2 Variável Histórico Familiar Diabetes

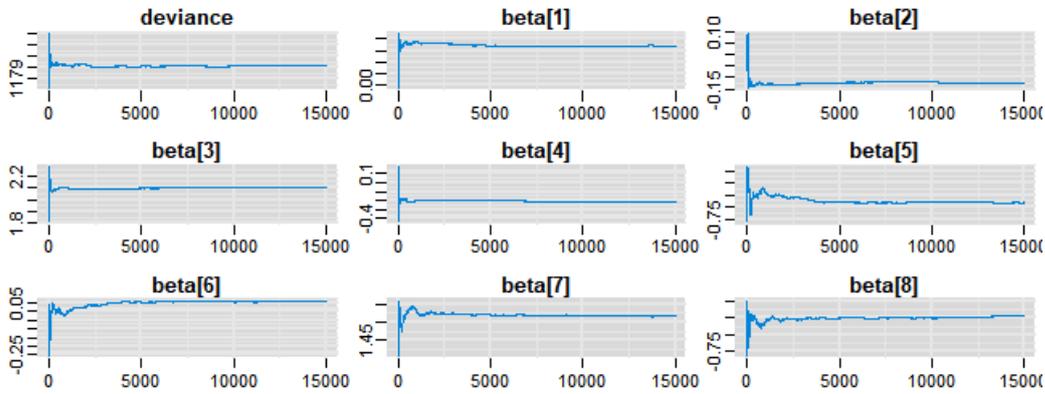


Figura 38 – Gráfico traço de média para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo

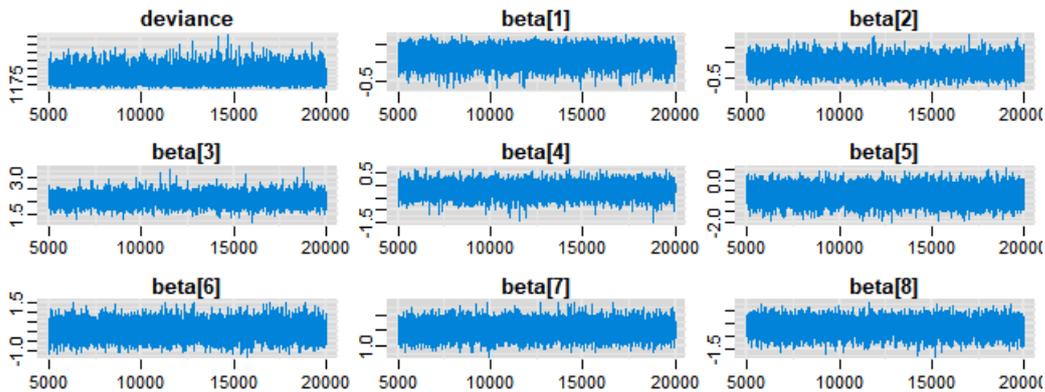


Figura 39 – Gráfico Histórico das Cadeias para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo

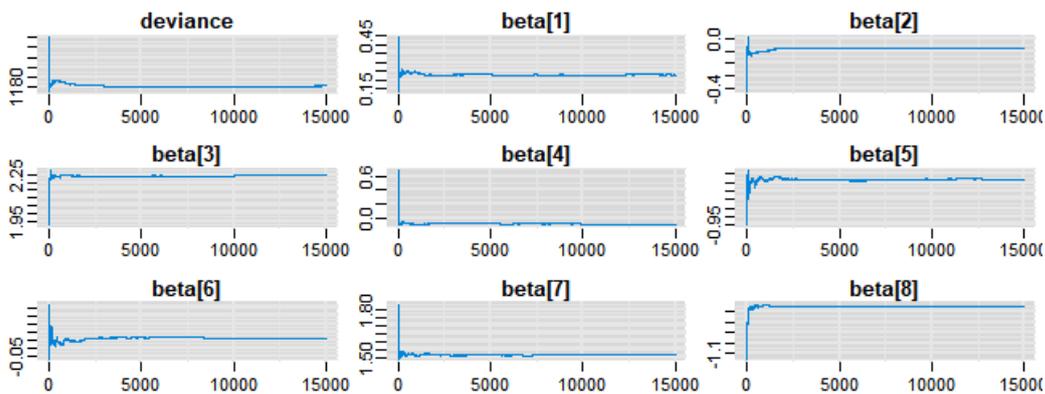


Figura 40 – Gráfico traço de média para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação mínimo

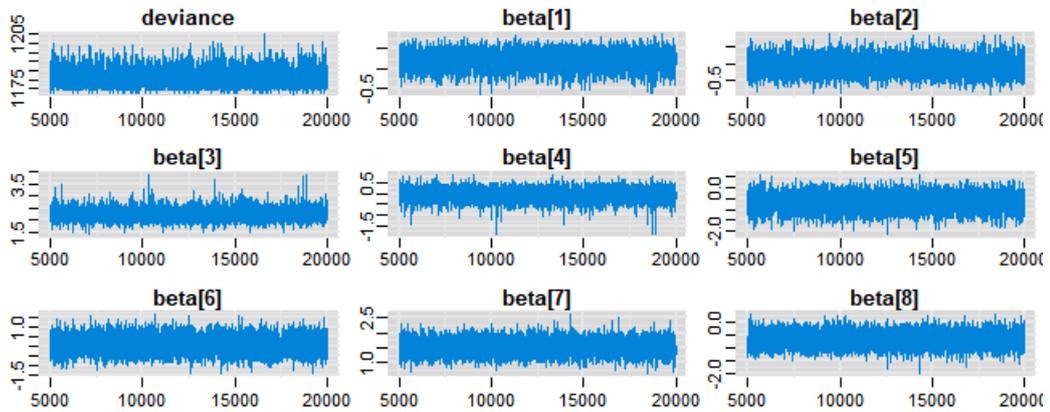


Figura 41 – Gráfico Histórico das Cadeias para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação mínimo

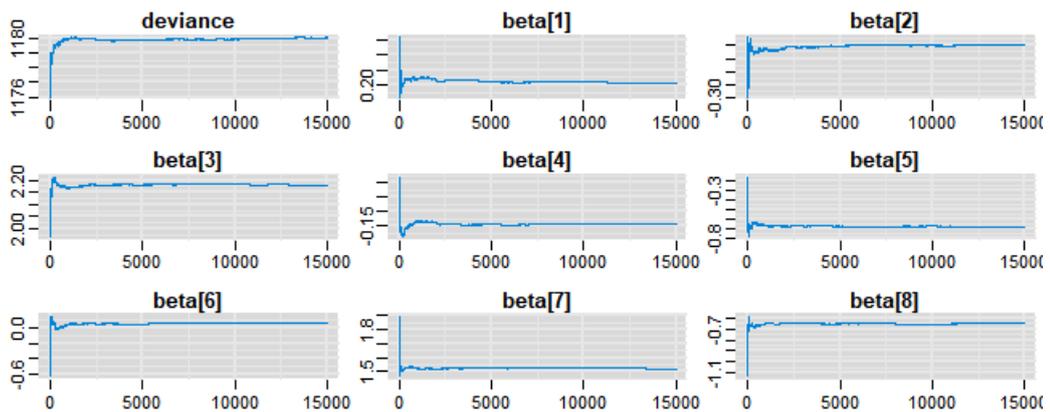


Figura 42 – Gráfico traço de média para a variável histórico diabetes modelo ZICR com abordagem bayesiana e esquema de ativação aleatório

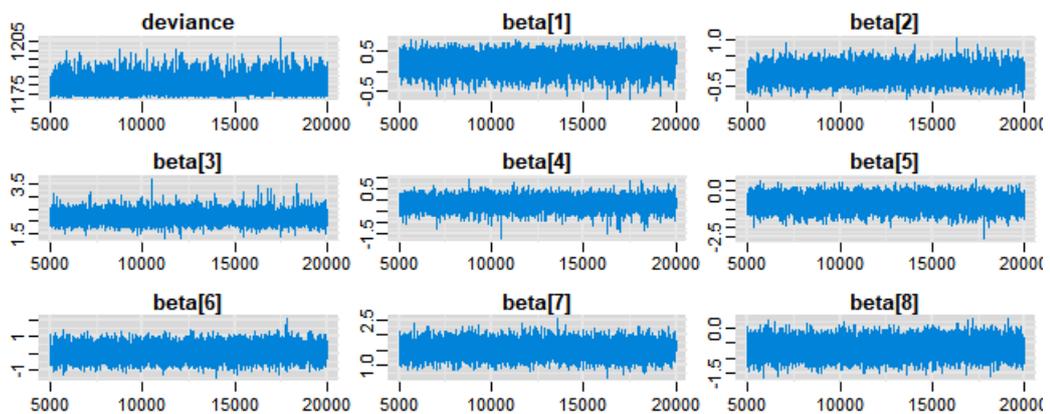


Figura 43 – Gráfico Histórico das Cadeias para a variável histórico diabetes considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório

### A.2.3 Variável Macrosomia fetal

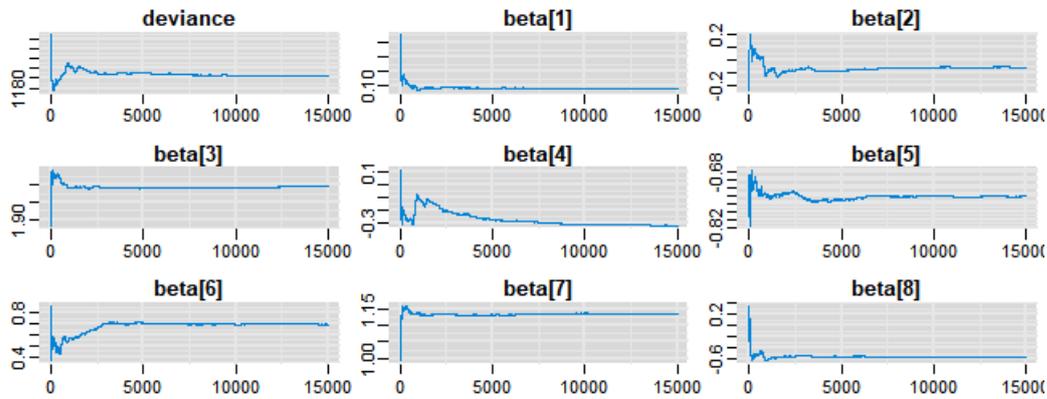


Figura 44 – Gráfico traço de média para a variável Macrosomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo

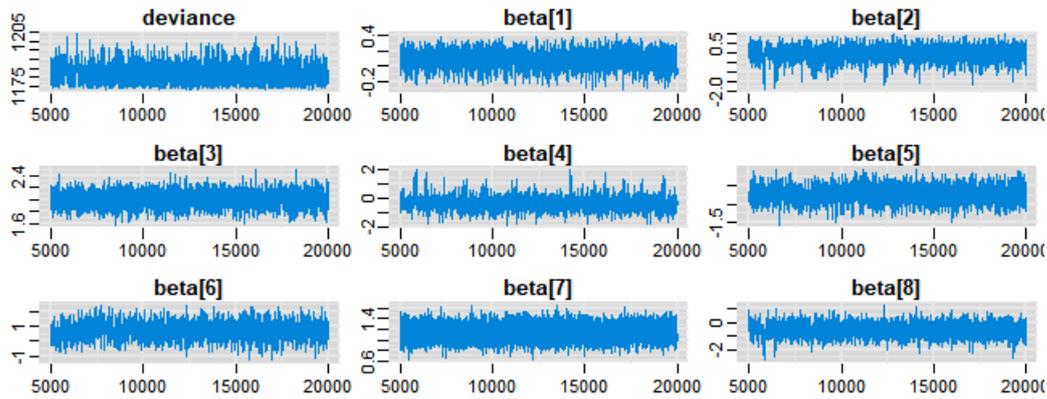


Figura 45 – Gráfico Histórico das Cadeias para Macrosomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação máximo

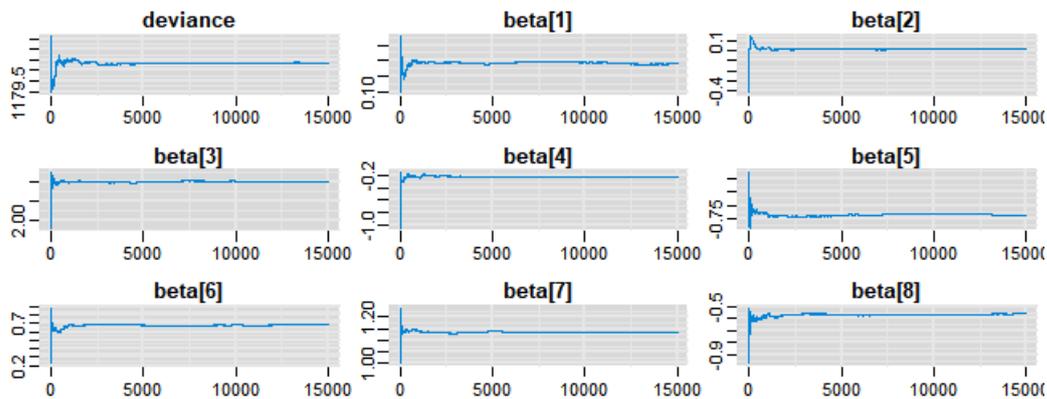


Figura 46 – Gráfico traço de média para a variável Macrosomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação mínimo

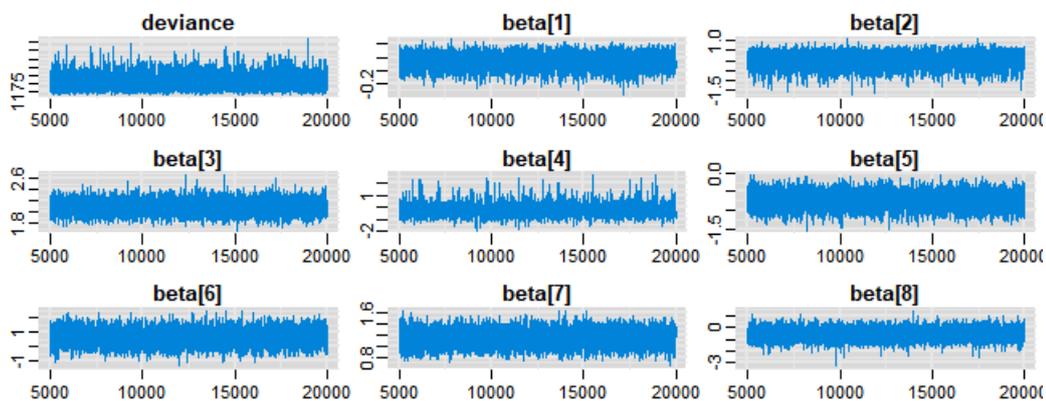


Figura 47 – Gráfico Histórico das Cadeias para Macrosomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação mínimo

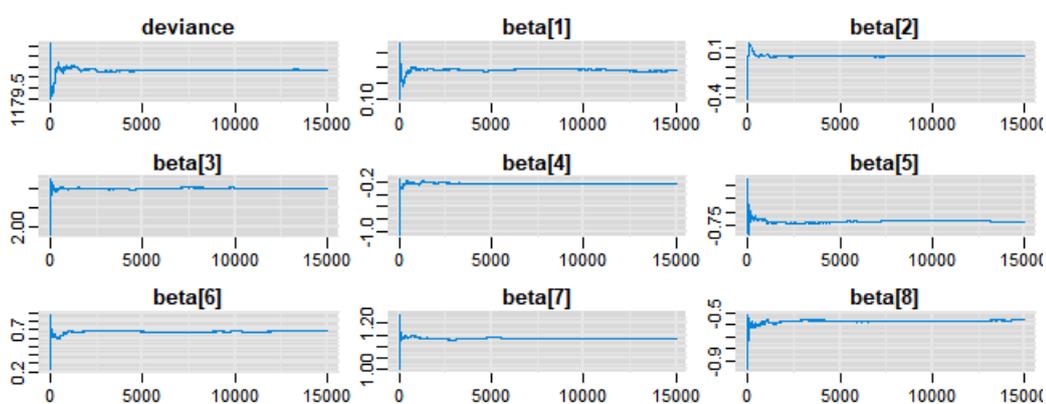


Figura 48 – Gráfico traço de média para a variável Macrosomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório

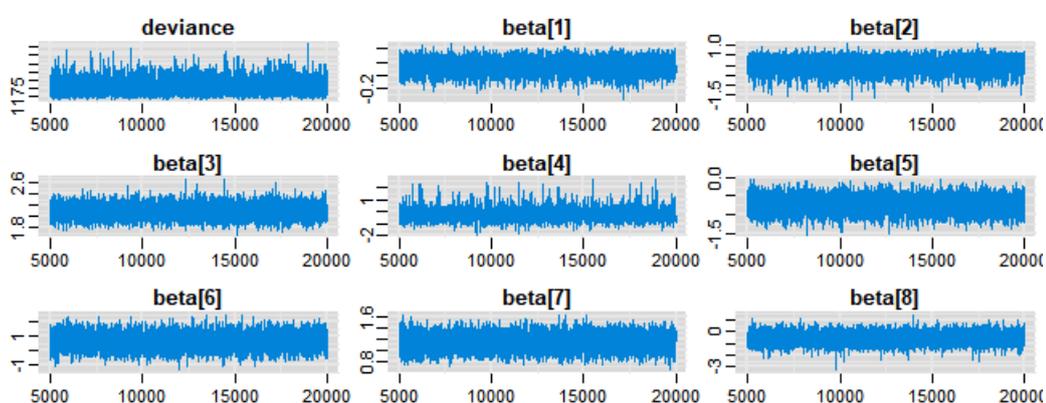


Figura 49 – Gráfico Histórico das Cadeias para Macrosomia fetal considerando o modelo ZICR com abordagem bayesiana e esquema de ativação aleatório

