

# Modelos de regressão quantílica com fração de cura: uma aplicação aos dados de COVID-19 grave na população materna

Ornella Scardua Ferreira

Universidade Federal do Espírito Santo - UFES  
Departamento de Estatística  
Orientadora: Profa. Dra. Agatha Sacramento Rodrigues  
Coorientador: Prof. Dr. Patrick Borges

3 de fevereiro de 2023

## Resumo

Neste artigo, abordamos modelos paramétricos de regressão quantílica para dados de sobrevivência com possibilidade de cura em que as distribuições são convenientemente reparametrizadas em termos do  $q$ -ésimo quantil ligado a covariáveis por intermédio de uma função logarítmica. Desenvolvemos o modelo de regressão quantílica de mistura padrão com distribuição Gompertz generalizada e, por meio de um estudo de simulação de Monte Carlo, comparamos com o modelo que considera a distribuição Gompertz generalizada em uma versão defeituosa (por isso, modelo defeituoso). Mostramos que os parâmetros do modelo defeituoso não ficam sobrecarregados ao estimar, simultaneamente, a fração de cura e os parâmetros do tempo de vida dos indivíduos que estão sujeitos à falha, pois seus resultados de simulação foram melhores em comparação ao modelo de mistura padrão quando os dados foram gerados pelo modelo defeituoso. Além disso, o modelo defeituoso e de mistura são aplicados aos dados públicos de gestantes e puérperas de 10 a 55 anos internadas com Síndrome Respiratória Aguda Grave por COVID-19 que estão disponíveis no portal openDataSUS, do Ministério da Saúde. Portanto, este estudo ainda possibilitou analisar os efeitos de variáveis de caracterização, sintoma e comorbidade na fração de cura e em diferentes quantis dos tempos de sobrevivência dessas mulheres que estavam vivendo a gestação ou puerpério durante a pandemia de COVID-19 no Brasil.

**Palavras-chaves:** Análise de Sobrevivência; COVID-19; Dados públicos; Distribuição Gompertz generalizada; Fração de cura; Gestantes e puérperas; Modelo defeituoso; Modelo de mistura padrão; Regressão quantílica.

## Abstract

In this article, we discuss quantile regression models for data survival with a cure rate in which the distributions are conveniently reparameterized in terms of the " $q$ th" quantile and linked to covariates linked to a logarithmic function. We developed the standard mixture quantile regression model with generalized Gompertz distribution for the susceptible individuals. Through a Monte Carlo simulation study, we compared it with the model that considers the generalized Gompertz distribution in a defective version (hence, defective model). We evaluated that the parameters of the flawed model are manageable when simultaneously estimating the cure fraction and the parameters of the lifetime of the individuals that are subject to failure, presenting better simulation results

than the standard mixture model when the defective model generated the data. In addition, the flawed and mixture models referred to public data on pregnant and puerperal women aged 10 to 55 years hospitalized with Severe Acute Respiratory Syndrome due to COVID-19, which are available on the openDataSUS portal of the Ministry of Health. Therefore, this study also made it possible to analyze the effects of characterization, symptom, and comorbidity variables on the cure fraction and on different quantiles of survival times of these women who were experiencing pregnancy or puerperium during the COVID-19 pandemic in Brazil.

**Key-words:** COVID-19; Cure rate; Defective model; Generalized Gompertz distribution; Quantile regression; Standard mixture model; Survival Analysis; Pregnant and postpartum women; Public data.

## 1 Introdução

Descoberta em dezembro de 2019 na cidade de Wuhan, na China, a COVID-19, causada pelo coronavírus SARS-CoV-2, é uma infecção aguda potencialmente grave com alto poder de propagação entre seres humanos e de distribuição global, sendo declarada como uma Emergência de Saúde Pública de Importância Internacional (ESPII) em 30 de janeiro de 2020 (WHO, 2020a) e caracterizada como uma pandemia em 11 de março do mesmo ano pela Organização Mundial da Saúde (OMS) (WHO, 2020b). Até 12 de novembro de 2022, foram quase 650 milhões de casos de infecções e mais de 6,6 milhões de mortes por essa doença ao redor do mundo (DONG; DU; GARDNER, 2020).

A COVID-19 é transmitida de pessoa infectada para pessoa sadia através de gotículas de secreção eliminadas pela boca ou nariz ou aerossóis (micropartículas de secreção respiratória). Os sintomas mais comuns, segundo a OPAS (2022) informa, são febre, fadiga e tosse seca, enquanto os menos improváveis são perda de paladar ou olfato, dor de garganta, vômito e diarreia. Apesar da maioria dos infectados apresentarem sinais leves a moderados da doença ou mesmo nenhum sintoma (os assintomáticos), alguns desenvolvem sintomas graves que podem evoluir para óbito.

Uma das populações em que as manifestações clínicas da COVID-19 se mostraram bastante críticas foi a das mulheres na gestação e no puerpério (período consecutivo ao parto). Para se ter uma ideia, a razão de morte materna aumentou cerca de 30% no Brasil somente no primeiro ano pandêmico, período no qual a vacina ainda não estava disponível à população brasileira geral. No ano seguinte, em 2021, foram quase 107 óbitos maternos para cada 100 mil nascimentos vivos, provocando um aumento de 94% na taxa de mortalidade materna em comparação a 2019 (UNFPA Brasil, 2022) e fazendo com que o Brasil figurasse na primeira posição entre os países com os maiores números de mortes maternas por COVID-19 em todo o mundo (DINIZ; BRITO; RONDON, 2022).

O alto número de mortes de gestantes e puérperas por COVID-19 e consequentes elevados níveis de mortalidade materna parte de um problema já antigo nos sistemas de saúde brasileiros: as fragilidades na rede de atenção à saúde, em especial à saúde materna. A má qualidade dos serviços hospitalares prestados à população materna aflixe tanto as mulheres com baixo risco obstétrico, que podem não receber os cuidados pré-natais necessários para o acompanhamento natural da gestação, quanto as mulheres com alto risco obstétrico, para as quais os cuidados maternos e perinatais podem ser comprometidos devido à falta ou inadequação da infraestrutura hospitalar (BITTENCOURT et al., 2016).

Além disso, mulheres grávidas e no pós-parto são mais vulneráveis às infecções respiratórias virais graves em razão das mudanças fisiológicas e imunológicas que ocorrem durante e após o período da gravidez (HUGHES, 2020). Portanto, estar grávida pode ser considerado como fator de risco para o desenvolvimento de complicações graves após a infecção por SARS-CoV-2, que podem rapidamente evoluir para óbito a depender da idade gestacional ou se no puerpério, da idade, da comorbidade declarada e do sintoma manifestado pela paciente. Em pesquisa sobre mortalidade materna por COVID-19 na América Latina<sup>1</sup>, Maza-Arnedo et al. (2022) revelaram que, dentre as 447 mortes identificadas, a

<sup>1</sup> Nesse estudo, foram considerados os países latinos Honduras ( $n = 126$ ), Paraguai ( $n = 86$ ), Colômbia ( $n = 84$ ), Equador ( $n = 55$ ), Peru ( $n = 34$ ), República Dominicana ( $n = 30$ ), Bolívia ( $n = 21$ ) e Costa Rica ( $n = 11$ ).

maioria se deu no puerpério e que a média de idade foi de 31 anos, com mais de 25% ocasionadas em mulheres com idade materna avançada (35 anos ou mais). Já a partir de dados clínicos, esse estudo foi ainda capaz de mostrar que aproximadamente metade (49,4%) das gestantes e puérperas mortas eram obesas e que 62% delas apresentaram saturação de oxigênio abaixo de 95%.

Diante dessa situação, não só identificar as características que levam ao óbito, mas também investigar como elas impactam no tempo de sobrevida de gestantes e puérperas afetadas pelo vírus SARS-CoV 2 é essencial para médicos detectarem precocemente a gravidade da doença. Assim, intervenções preventivas poderiam ser feitas e atrasos nos cuidados a pacientes obstétricas, evitados.

Em estudos de sobrevivência, os modelos mais usuais para esses fins são os modelos de regressão de Cox (COX, 1972) e de taxa de falha acelerada (PRENTICE, 1978), que assumem que todo indivíduo irá falhar<sup>2</sup> com probabilidade igual a 1 ao longo do período de observação. Entretanto, devemos considerar que uma parcela desses indivíduos pode responder favoravelmente ao tratamento, tornando-se curada ou imune ao evento de interesse (não recidiva de uma doença, adimplência junto a um banco, não reincidência em um crime de trânsito etc). Por isso, a proporção de curados<sup>3</sup> (ou imunes) tem sido extensivamente estudada e aplicada à análise de sobrevivência tradicional, sobretudo em estudos clínicos (TSODIKOV; YAKOVLEV; ASSELAINE, 1996).

Modelos que levam em conta uma fração de cura, conhecidos como modelos de fração de cura ou modelos de sobrevivência de longo prazo, são constantemente usados para explicar simultaneamente pacientes imunes e suscetíveis ao evento de interesse. Existem abordagens que consideram a fração de cura no cenário de fatores latentes competitivos (BORGES et al., 2016) e fatores latentes competitivos e fragilidade (SILVA, 2011). Uma abordagem comum na presença de fração de cura é o modelo de mistura padrão abordado por Boag (1949) e estendido por Berkson e Gage (1952), que, como o próprio nome sugere, “mistura” a distribuição do tempo de falha postulada para os suscetíveis com a distribuição degenerada<sup>4</sup> do tempo de sobrevivência dos insuscetíveis, resultando em um modelo para as duas sub-populações. Dessa forma, obtém-se os efeitos das covariáveis sobre a proporção de cura e os efeitos das covariáveis que impactam no tempo de sobrevivência dos indivíduos que estão sujeitos a apresentarem a falha.

Outra abordagem é considerar modelos baseados em distribuições defeituosas (BALKA; DESMOND; MCNICHOLAS, 2009). As distribuições defeituosas são obtidas ao mudar o espaço paramétrico de um dos parâmetros da distribuição de tal forma que o limite da função de sobrevivência convirja para um valor limitado no intervalo (0, 1) quando o tempo vai para infinito. Com isso, as distribuições defeituosas têm a vantagem de permitir uma fração de cura sem adicionar nenhum parâmetro extra na modelagem, diferentemente dos modelos de mistura padrão (SCUDILIO et al., 2019). Portanto, modelos defeituosos atendem ao princípio da parcimônia, dada a desnecessidade do parâmetro da proporção de curados, mas ao mesmo tempo podem sobrecarregar seus parâmetros, já que eles concentrarem informações tanto do grupo dos suscetíveis quanto do grupo dos indivíduos que não estão sob o risco da falha.

Todos os modelos de sobrevida mencionados acima são essencialmente modelos que se concentram em estimar os parâmetros da taxa de risco, escala e fração de cura, mas que não abordam a estimativa da função quantílica para qualquer indivíduo com valores de covariáveis escolhidos arbitrariamente (GUPTA et al., 2016). No entanto, os médicos podem estar interessados em avaliar o estágio inicial do acompanhamento, por exemplo, e a regressão quantílica pode modelar diretamente o quantil inferior de interesse para fornecer os efeitos das covariáveis específicos para esse quantil com interpretações simples na escala de tempo estudada (PENG, 2021). Sob essa abordagem, é possível saber como os coeficientes de regressão de uma dada covariável mudam para diferentes quantis de tempo de sobrevivência.

<sup>2</sup> Nesse contexto, “falhar” significa apresentar o evento de interesse, enquanto “falha” é o próprio evento de interesse.

<sup>3</sup> Também conhecidos como sobreviventes de longa duração, são indivíduos que não apresentarão o evento de interesse, independente do tempo que são observados. Diferente das censuras, que consistem em indivíduos que não experimentaram o evento durante o tempo de observação.

<sup>4</sup> Indivíduos curados ou imunes nunca irão falhar, ou seja, o tempo de vida é sempre igual a infinito. Portanto, a distribuição da variável tempo de vida é degenerada no infinito.

Na literatura, existem muitos trabalhos que abordam a regressão quantílica para dados de sobrevivência: [Ying, Jung e Wei \(1995\)](#), [Honore, Khan e Powell \(2002\)](#), [Portnoy \(2003\)](#), [Yin, Zeng e Li \(2008\)](#) e [Peng e Huang \(2008\)](#) são alguns exemplos. Sem contar [Noufaily e Jones \(2013\)](#), que introduziram uma abordagem totalmente paramétrica na função de regressão quantílica para evitar o problema do cruzamento de quantis (já que, nesse caso, os quantis estimados obrigatoriamente seguem ordenação crescente no intervalo  $(0,1)$ ); e [Xue, Xie e Strickler \(2018\)](#), que incluíram um componente não paramétrico na forma funcional da regressão e/ou na distribuição associada ao tempo de falha.

A regressão quantílica é apropriada para modelos heterocedásticos ([SANTOS, 2012](#)), caso dos modelos com fração de cura cuja heterogeneidade advém da mistura das sub-populações dos suscetíveis e insuscetíveis. Nessa conjuntura, esses modelos são conhecidos como modelos de regressão quantílica com fração de cura, e podemos citar [Wu e Yin \(2017\)](#), [Gupta et al. \(2016\)](#), [Narisetty e Koenker \(2022\)](#), [Rodrigues, Borges e Santos \(2021\)](#) e [Velten \(2021\)](#) como algumas abordagens. As três primeiras propostas incluíram um componente não paramétrico na forma funcional da equação de regressão ou na distribuição do tempo de vida do modelo ou em ambas. Já os trabalhos de [Rodrigues, Borges e Santos \(2021\)](#) e [Velten \(2021\)](#) propõem um modelo de regressão quantílica puramente paramétrico para dados de sobrevivência na presença de sobreviventes de longo prazo ao considerar, respectivamente, a distribuição generalizada de Gompertz em sua versão defeituosa e a distribuição Weibull exponenciada para os suscetíveis em um modelo de mistura padrão, ambas convenientemente reparametrizadas em termos do  $q$ -ésimo quantil ligado a covariáveis por meio de uma função de ligação logarítmica.

Tudo isso posto e instigados pelo trabalho de [Rodrigues, Borges e Santos \(2021\)](#), vamos comparar o modelo proposto por esses autores com um modelo de regressão quantílica de mistura padrão que considera como distribuição da variável tempo de falha a Gompertz generalizada. Desenvolveremos uma análise bayesiana para estimar os parâmetros do modelo defeituoso e de mistura e faremos uso da simulação de Monte Carlo para avaliar seus respectivos desempenhos, tendo em vista que o principal objetivo deste artigo é verificar se os parâmetros do modelo defeituoso ficam sobrecarregados ao estimar a fração de cura e os parâmetros do tempo de vida dos suscetíveis simultaneamente. Além do mais, ambos os modelos serão usados para investigar quais são os reais impactos de variáveis de caracterização, sintoma e comorbidade no tempo de sobrevivência de gestantes e puérperas internadas com Síndrome Respiratória Aguda Grave (SRAG) por consequência da doença de COVID-19.

Este artigo está organizado da seguinte forma: a formulação dos modelos de mistura padrão e defeituoso e a inferência dos parâmetros sob o paradigma bayesiano são descritos na Seção 3. Na Seção 4, realizamos um estudo de simulação de Monte Carlo para demonstrar o desempenho dos modelos propostos. Na Seção 5, é apresentada a aplicação aos dados de gestantes e puérperas internadas com SRAG por COVID-19 e a seleção do melhor modelo entre os modelos candidatos. Encerramos com as considerações finais e propostas de trabalhos futuros na Seção 6.

## 2 Objetivos

1. Desenvolver o modelo de regressão quantílica de mistura padrão com distribuição Gompertz generalizada para o tempo de vida dos suscetíveis.
2. Comparar o modelo de regressão do item 1 com o modelo de regressão quantílica defeituoso com distribuição Gompertz generalizada de [Rodrigues, Borges e Santos \(2021\)](#) por meio de estudos de simulação. Avaliar, portanto, se os parâmetros do modelo defeituoso ficam sobrecarregados ao estimar, simultaneamente, a fração de cura e os parâmetros do tempo de vida dos suscetíveis.
3. Aplicar os modelos estudados aos dados de gestantes e puérperas internadas com síndrome respiratória aguda grave por COVID-19. Os dados e os objetivos da aplicação são explicados na Seção 5.

### 3 Materiais e métodos

Nesta seção, apresentamos a distribuição Gompertz generalizada na Subseção 3.1, o modelo de mistura padrão Gompertz generalizada com reparametrização quantílica na Subseção 3.2 e o modelo defeituoso Gompertz generalizada com reparametrização quantílica na Subseção 3.3. Na Subseção 3.4, são apresentadas as funções de verossimilhança dos modelos de mistura padrão e defeituoso para os quais os parâmetros serão estimados via inferência bayesiana.

#### 3.1 Distribuição Gompertz generalizada

A distribuição Gompertz generalizada (GG) introduzida por [El-Gohary, Alshamrani e Al-Otaibi \(2013\)](#) considera que o tempo de vida  $T$ , condicional aos parâmetros  $\lambda$ ,  $\alpha$  e  $\theta$ , tem funções densidade e de sobrevivência dadas, respectivamente, por

$$f(t | \lambda, \alpha, \theta) = \lambda \theta \exp \left\{ \alpha t - \frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1) \right\} \left( 1 - \exp \left\{ -\frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1) \right\} \right)^{\theta-1}, \quad t > 0 \quad (1)$$

e

$$S(t | \lambda, \alpha, \theta) = 1 - \left( 1 - \exp \left\{ -\frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1) \right\} \right)^{\theta}, \quad (2)$$

em que  $\lambda > 0$  e  $\alpha > 0$  são os parâmetros de escala e  $\theta > 0$  é o parâmetro de forma. A distribuição GG inclui as seguintes distribuições como casos especiais: (i) distribuição exponencial generalizada quando  $\alpha$  vai para zero; (ii) distribuição Gompertz quando  $\theta = 1$ ; e (iii) distribuição exponencial quando  $\alpha$  vai para zero e  $\theta = 1$ .

A função de risco da distribuição GG é dada por

$$h(t | \lambda, \alpha, \theta) = \frac{\lambda \theta \exp\{\alpha t - \frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1)\} \left( 1 - \exp\{-\frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1)\} \right)^{\theta-1}}{1 - \left( 1 - \exp\{-\frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1)\} \right)^{\theta}}.$$

De acordo com [El-Gohary, Alshamrani e Al-Otaibi \(2013\)](#), a forma da função de risco da distribuição GG é: (i) crescente se  $\alpha > 0$  e  $\theta = 1$ ; (ii) constante se  $\alpha = 0$  e  $\theta = 1$ ; (iii) crescente quando  $\theta > 1$ ; (iv) decrescente quando  $\alpha = 0$  e  $\theta < 1$ ; e (v) banheira se  $\alpha > 0$  e  $\theta < 1$ .

#### 3.2 Modelo de mistura padrão

O modelo de mistura padrão procede atribuindo uma variável latente Bernoulli  $Z$  com probabilidade de sucesso  $\mathbb{P}(Z = 1) = 1 - p_0$  e probabilidade de fracasso  $\mathbb{P}(Z = 0) = p_0$  associada a cada observação.

Assim, a função de sobrevivência é dada por

$$\begin{aligned} S(t) &= P(T \geq t) = \sum_{z=0}^1 P(T \geq t, Z = z) = P(T \geq t, Z = 0) + P(T \geq t, Z = 1) \\ &= P(T \geq t | Z = 0)P(Z = 0) + P(T \geq t | Z = 1)P(Z = 1) \\ &= p_0 + (1 - p_0)S_1(t), \end{aligned}$$

em que  $S_1(\cdot)$  denota a função de sobrevivência dos indivíduos suscetíveis ao evento de interesse.

A função densidade é dada por

$$f(t) = (1 - p_0)f_1(t),$$

sendo  $f_1(\cdot)$  a função densidade dos indivíduos suscetíveis ao evento de interesse.

A função quantílica da distribuição GG é dada por

$$\mu_q = \mu(q | \lambda, \alpha, \theta) = \frac{1}{\alpha} \log \left( 1 - \frac{\alpha}{\lambda} \log \left( 1 - q^{1/\theta} \right) \right), \quad 0 < q < 1, \quad (3)$$

em que o  $q$ -ésimo quantil  $\mu_q$  é obtido a partir da função de distribuição acumulada, isto é,

$$F(t | \lambda, \alpha, \theta) = q = 1 - S(t | \lambda, \alpha, \theta) = \left( 1 - \exp \left\{ -\frac{\lambda}{\alpha} (e^{\alpha t} - 1) \right\} \right)^\theta,$$

$S(t | \lambda, \alpha, \theta)$  dada em (2).

No que segue, vamos reparametrizar a função densidade em (1) em termos do  $q$ -ésimo quantil  $\mu_q$ , de tal forma que  $\theta$  pode ser escrito como

$$\theta = \frac{\log(q)}{\log \left\{ 1 - \exp \left[ -\frac{\lambda}{\alpha} (e^{\alpha \mu_q} - 1) \right] \right\}},$$

em que esse resultado é obtido ao isolar  $\theta$  na Equação (3).

Agora, construímos o modelo de regressão impondo que o quantil  $\mu_q$  de  $T$  satisfaça a seguinte relação funcional:

$$\mu_q(\boldsymbol{\beta}, \mathbf{x}) = \mu_q = \exp(\mathbf{x}^\top \boldsymbol{\beta}_q), \quad (4)$$

em que  $\mathbf{x}^\top = (1, x_1, \dots, x_p)$  é o vetor de covariáveis e  $\boldsymbol{\beta}_q = (\beta_{q0}, \beta_{q1}, \dots, \beta_{qp})^\top$  é o vetor de parâmetros da regressão a serem estimados.

Também assumimos que a fração de cura  $p_0$  pode ser relacionada a um vetor de  $r$  covariáveis,  $\mathbf{z}^\top = (1, z_1, \dots, z_r)$ , por meio da função logito

$$p_0(\boldsymbol{\gamma}, \mathbf{z}) = p_0 = \exp\{\mathbf{z}^\top \boldsymbol{\gamma}\} / (1 + \exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}), \quad (5)$$

em que  $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_r)^\top$  é o vetor de parâmetros da regressão que também devem ser estimados. Vale lembrar que os vetores das covariáveis  $\mathbf{x}$  e  $\mathbf{z}$ , como muitas vezes acontece na prática, podem ser os mesmos, ou seja,  $\mathbf{x} = \mathbf{z}$ .

Assim,  $\boldsymbol{\vartheta}_q = (\boldsymbol{\beta}_q^\top, \boldsymbol{\gamma}^\top, \lambda, \alpha)^\top$  é o vetor de parâmetros.

Ao considerar a reparametrização anterior e as covariáveis na função quantílica e na fração de cura, tem-se que a função densidade do modelo de mistura padrão é

$$\begin{aligned} f(t | \boldsymbol{\vartheta}_q, \mathbf{x}, \mathbf{z}) &= \left( 1 - \frac{\exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}}{1 + \exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}} \right) \times \lambda \left[ \frac{\log(q)}{\log \left\{ 1 - \exp \left[ -\frac{\lambda}{\alpha} (e^{\alpha \exp(\mathbf{x}^\top \boldsymbol{\beta}_q)} - 1) \right] \right\}} \right] \\ &\times \exp \left\{ \alpha t - \frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1) \right\} \\ &\times \left( 1 - \exp \left\{ -\frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1) \right\} \right)^{\frac{\log(q)}{\log \left\{ 1 - \exp \left[ -\frac{\lambda}{\alpha} (e^{\alpha \exp(\mathbf{x}^\top \boldsymbol{\beta}_q)} - 1) \right] \right\}} - 1}, \quad t > 0, \end{aligned} \quad (6)$$

e a função de sobrevivência,

$$\begin{aligned} S(t | \boldsymbol{\vartheta}_q, \mathbf{x}, \mathbf{z}) &= \frac{\exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}}{1 + \exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}} + \left( 1 - \frac{\exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}}{1 + \exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}} \right) \\ &\times \left[ 1 - \left( 1 - \exp \left\{ \frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1) \right\} \right)^{\frac{\log(q)}{\log \left\{ 1 - \exp \left[ -\frac{\lambda}{\alpha} (e^{\alpha \exp(\mathbf{x}^\top \boldsymbol{\beta}_q)} - 1) \right] \right\}} - 1} \right], \quad t > 0. \end{aligned} \quad (7)$$

### 3.3 Modelo defeituoso

Esse modelo foi proposto por [Rodrigues, Borges e Santos \(2021\)](#) e descrevemo-lo a seguir.

A distribuição GG se torna uma distribuição defeituosa se  $\alpha < 0$ . Nesse cenário, a fração de cura correspondente é:

$$\lim_{t \rightarrow \infty} S(t | \lambda, \alpha, \theta) = 1 - \left[ 1 - \exp \left( \frac{\lambda}{\alpha} \right) \right]^\theta = p_0(\lambda, \alpha, \theta) \in [0, 1]. \quad (8)$$

Conforme destacado em [Scudilio et al. \(2019\)](#), as distribuições defeituosas têm a vantagem de permitir uma fração de cura sem adicionar parâmetros extras na modelagem. Como queremos reparametrizar em função do quantil, uma desvantagem é que a função quantílica não é bem definida no modelo GG defeituoso, pois a distribuição é defeituosa.

Para contornar essa restrição, vamos reescrever a distribuição defeituosa como um modelo de mistura padrão, ou seja, vamos separar as populações suscetíveis e imunes sem incluir um parâmetro adicional na modelagem. Assim, além de garantir a vantagem da parcimônia, também poderemos descrever os quantis da população suscetível.

Definimos uma variável indicadora  $Y$ , que assume valor 0 se o indivíduo é imune e 1 se o indivíduo é suscetível. Seja  $\mathbb{P}[Y = 0] = p_0(\lambda, \alpha, \theta)$  e  $\mathbb{P}[Y = 1] = 1 - p_0(\lambda, \alpha, \theta)$ , em que  $p_0(\lambda, \alpha, \theta)$  é dado pela Equação (8). A função de sobrevivência defeituosa é dada por

$$S(t \mid \lambda, \alpha, \theta) = p_0(\lambda, \alpha, \theta) + [1 - p_0(\lambda, \alpha, \theta)]S_2(t \mid \lambda, \alpha, \theta), \quad (9)$$

em que  $S_2(\cdot)$  denota a função de sobrevivência dos suscetíveis.

Da Equação (9), obtemos uma expressão para  $S_2$  como sendo

$$S_2(t \mid \lambda, \alpha, \theta) = \frac{S(t \mid \lambda, \alpha, \theta) - p_0(\lambda, \alpha, \theta)}{1 - p_0(\lambda, \alpha, \theta)}. \quad (10)$$

Ao considerar a relação em (10), notamos que  $S_2$  é uma função de sobrevivência própria com uma função quantílica dada por

$$\mu_q^1 = \mu^1(q \mid \lambda, \alpha, \theta) = \frac{\log\left(\lambda - \alpha \log\left(1 - q^{\frac{1}{\theta}}\left(1 - \exp\left\{\frac{\lambda}{\alpha}\right\}\right)\right)\right) - \log(\lambda)}{\alpha}, \quad 0 < q < 1. \quad (11)$$

Reparametrizamos a função densidade em (1) em termos do  $q$ -ésimo quantil  $\mu_q^1$ , tal que  $\theta$  pode ser escrito como

$$\theta = -\frac{\log(q)}{\log\left(1 - \exp\left\{\frac{\lambda}{\alpha}\right\}\right) - \log\left(1 - \exp\left\{-\frac{\lambda}{\alpha}\left(e^{\alpha\mu_q^1} - 1\right)\right\}\right)}, \quad (12)$$

em que esse resultado é obtido ao isolar  $\theta$  na Equação (11).

É importante destacar que, ao escrevermos a função de sobrevivência para indivíduos suscetíveis (Equação 10) em função da função de sobrevivência da distribuição Gompertz generalizada defeituosa e da fração de cura  $p_0(\lambda, \alpha, \theta)$ , temos que o espaço paramétrico para  $\alpha$  na distribuição da Gompertz defeituosa generalizada reparametrizada é  $(-\infty, 0)$ .

Se substituirmos (12) em (1), (2) e (8), temos as funções densidade e de sobrevivência reparametrizadas e a fração de cura dadas, respectivamente, por

$$\begin{aligned} f(t \mid \lambda, \alpha, \mu_q^1) &= \lambda \left[ -\frac{\log(q)}{\log\left(1 - \exp\left\{\frac{\lambda}{\alpha}\right\}\right) - \log\left(1 - \exp\left\{-\frac{\lambda}{\alpha}\left(e^{\alpha\mu_q^1} - 1\right)\right\}\right)} \right] \\ &\quad \times \exp\left\{\alpha t - \frac{\lambda}{\alpha}(\exp\{\alpha t\} - 1)\right\} \\ &\quad \times \left(1 - \exp\left\{-\frac{\lambda}{\alpha}(\exp\{\alpha t\} - 1)\right\}\right)^{-\frac{\log(q)}{\log\left(1 - \exp\left\{\frac{\lambda}{\alpha}\right\}\right) - \log\left(1 - \exp\left\{-\frac{\lambda}{\alpha}\left(e^{\alpha\mu_q^1} - 1\right)\right\}\right)} - 1}, \end{aligned} \quad (13)$$

$$S(t \mid \lambda, \alpha, \mu_q^1) = 1 - \left(1 - \exp\left\{-\frac{\lambda}{\alpha}(\exp\{\alpha t\} - 1)\right\}\right)^{-\frac{\log(q)}{\log\left(1 - \exp\left\{\frac{\lambda}{\alpha}\right\}\right) - \log\left(1 - \exp\left\{-\frac{\lambda}{\alpha}\left(e^{\alpha\mu_q^1} - 1\right)\right\}\right)}} \quad (14)$$



e

$$p_0(\lambda, \alpha, \mu_q^1) = 1 - \left(1 - \exp\left\{\frac{\lambda}{\alpha}\right\}\right)^{-\frac{\log(q)}{\log\left(1 - \exp\left\{\frac{\lambda}{\alpha}\right\}\right) - \log\left(1 - \exp\left\{-\frac{\lambda}{\alpha}\left(e^{\alpha\mu_q^1} - 1\right)\right\}\right)}}. \quad (15)$$

Agora, construímos o modelo de regressão impondo que  $\mu_q^1$  de  $T$  satisfaça a seguinte relação funcional:

$$\mu_q^1(\beta_q, \mathbf{x}) = \mu_q^1 = \exp(\mathbf{x}^\top \beta_q), \quad (16)$$

em que  $\mathbf{x}^\top = (1, x_1, \dots, x_p)$  são os vetores de covariáveis e  $\beta_q = (\beta_{q0}, \beta_{q1}, \dots, \beta_{qp})^\top$  é o vetor desconhecido dos parâmetros de regressão a serem estimados.

Os parâmetros  $\alpha$  e  $\lambda$  também podem ser ligados a covariáveis, uma possibilidade para tornar as estimativas do modelo mais verossímeis. Assim, para  $\alpha$ , podemos ter

$$\log(-\alpha) = \mathbf{w}^\top \boldsymbol{\gamma}, \quad (17)$$

em que  $\mathbf{w}^\top = (1, w_1, \dots, w_r)$  são os vetores de covariáveis e  $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_r)^\top$  é o vetor desconhecido dos parâmetros de regressão.

E para  $\lambda$ ,

$$\lambda = \exp(\mathbf{z}^\top \mathbf{v}),$$

em que  $\mathbf{z}^\top = (1, z_1, \dots, z_s)$  são os vetores de covariáveis e  $\mathbf{v} = (v_0, v_1, \dots, v_s)^\top$  é o vetor desconhecido dos parâmetros de regressão.

Assim, o vetor de parâmetros é  $\boldsymbol{\vartheta}_q^1 = (\beta_q^\top, \boldsymbol{\gamma}^\top, \mathbf{v}^\top)^\top$ .

### 3.4 Inferência

Considere que o tempo de vida  $T$  pode não ser observado, isto é, ele pode ser limitado por um tempo de falha censurado à direita e seja  $C$  o tempo de censura. Em uma amostra de tamanho  $n$ , observamos o  $i$ -ésimo tempo de vida  $t_i = \min(T_i, C_i)$  e o  $i$ -ésimo indicador de falha  $\delta_i = I(T_i \leq C_i)$ , onde  $\delta_i = 1$  se  $T_i$  for observado e  $\delta_i = 0$  caso contrário, para  $i = 1, \dots, n$ . Assumimos que cada tempo de censura  $C_i$  é independente do tempo de vida  $T_i$  e consideramos a suposição de censura não informativa, ou seja, a distribuição da censura não envolve parâmetros da distribuição de  $T$ .

Ao considerar o modelo de mistura padrão, a função de verossimilhança de  $\boldsymbol{\vartheta}_q$  pode ser escrita como

$$\begin{aligned} L(\boldsymbol{\vartheta}_q; \mathbf{D}) &\propto \prod_{i=1}^n [f(t_i | \boldsymbol{\vartheta}_q, \mathbf{x}_i, \mathbf{z}_i)]^{\delta_i} [S(t_i | \boldsymbol{\vartheta}_q, \mathbf{x}_i, \mathbf{z}_i)]^{1-\delta_i} \\ &\propto \prod_{i=1}^n \left\{ \left(1 - \frac{\exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}}{1 + \exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}}\right) \times \lambda \left[ \frac{\log(q)}{\log\left\{1 - \exp\left[-\frac{\lambda}{\alpha}\left(e^{\alpha \exp(\mathbf{x}_i^\top \beta_q)} - 1\right)\right]\right\}} \right] \right. \\ &\quad \times \exp\left\{\alpha t - \frac{\lambda}{\alpha}(\exp\{\alpha t\} - 1)\right\} \\ &\quad \times \left(1 - \exp\left\{-\frac{\lambda}{\alpha}(\exp\{\alpha t\} - 1)\right\}\right)^{\frac{\log(q)}{\log\left\{1 - \exp\left[-\frac{\lambda}{\alpha}\left(e^{\alpha \exp(\mathbf{x}_i^\top \beta_q)} - 1\right)\right]\right\}} - 1} \left. \right\}^{\delta_i} \\ &\quad \times \left\{ \frac{\exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}}{1 + \exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}} + \left(1 - \frac{\exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}}{1 + \exp\{\mathbf{z}^\top \boldsymbol{\gamma}\}}\right) \right. \\ &\quad \times \left[ 1 - \left(1 - \exp\left\{\frac{\lambda}{\alpha}(\exp\{\alpha t\} - 1)\right\}\right)^{\frac{\log(q)}{\log\left\{1 - \exp\left[-\frac{\lambda}{\alpha}\left(e^{\alpha \exp(\mathbf{x}_i^\top \beta_q)} - 1\right)\right]\right\}} - 1} \right] \left. \right\}^{1-\delta_i}, \quad (18) \end{aligned}$$



em que  $f(t_i | \boldsymbol{\vartheta}_q, \mathbf{x}_i, \mathbf{z}_i)$  é dado por (6),  $S(t_i | \boldsymbol{\vartheta}_q, \mathbf{x}_i, \mathbf{z}_i)$  é dado por (7) e  $\mathbf{D} = (\mathbf{t}, \mathbf{x}, \mathbf{z}, \boldsymbol{\delta})$ .

Enquanto a função de verossimilhança de  $\boldsymbol{\vartheta}_q^1$  ao considerar o modelo defeituoso é dada por

$$\begin{aligned}
L(\boldsymbol{\vartheta}_q^1; \mathbf{D}) &\propto \prod_{i=1}^n \left[ f(t_i | \boldsymbol{\vartheta}_q^1, \mathbf{x}_i, \mathbf{z}_i) \right]^{\delta_i} \left[ S(t_i | \boldsymbol{\vartheta}_q^1, \mathbf{x}_i, \mathbf{z}_i) \right]^{1-\delta_i} \\
&\propto \prod_{i=1}^n \left\{ \lambda \left[ -\frac{\log(q)}{\log(1 - \exp\{\frac{\lambda}{\alpha}\}) - \log(1 - \exp\{-\frac{\lambda}{\alpha}(e^{\alpha \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_q) - 1)})\}} \right] \right. \\
&\quad \times \exp\left\{ \alpha t - \frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1) \right\} \\
&\quad \times \left( 1 - \exp\left\{ -\frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1) \right\} \right)^{-\frac{\log(q)}{\log(1 - \exp\{\frac{\lambda}{\alpha}\}) - \log(1 - \exp\{-\frac{\lambda}{\alpha}(e^{\alpha \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_q) - 1)})\}} - 1} \Bigg\}^{\delta_i} \\
&\quad \left\{ 1 - \left( 1 - \exp\left\{ -\frac{\lambda}{\alpha} (\exp\{\alpha t\} - 1) \right\} \right)^{-\frac{\log(q)}{\log(1 - \exp\{\frac{\lambda}{\alpha}\}) - \log(1 - \exp\{-\frac{\lambda}{\alpha}(e^{\alpha \exp(\mathbf{x}_i^\top \boldsymbol{\beta}_q) - 1)})\}} \right) \right\}^{1-\delta_i} \quad (19)
\end{aligned}$$

em que  $f(t_i | \boldsymbol{\vartheta}_q^1, \mathbf{x}_i, \mathbf{z}_i)$  é dado em (13),  $S(t_i | \boldsymbol{\vartheta}_q^1, \mathbf{x}_i, \mathbf{z}_i)$  é dado em (14) e  $\mathbf{D} = (\mathbf{t}, \boldsymbol{\gamma}, \mathbf{v}, \boldsymbol{\delta})$ .

O processo inferencial será feito sob o enfoque bayesiano, e a distribuição a posteriori será obtida ao combinar a função de verossimilhança com a distribuição a priori dos parâmetros. No modelo de mistura padrão a distribuição a posteriori é definida como sendo

$$\pi(\boldsymbol{\vartheta}_q | \mathbf{D}) \propto \pi(\boldsymbol{\vartheta}_q) L(\boldsymbol{\vartheta}_q; \mathbf{D}).$$

Ao passo que no modelo defeituoso a distribuição a posteriori é dada por

$$\pi(\boldsymbol{\vartheta}_q^1 | \mathbf{D}) \propto \pi(\boldsymbol{\vartheta}_q^1) L(\boldsymbol{\vartheta}_q^1; \mathbf{D}).$$

As funções de verossimilhança  $L(\boldsymbol{\vartheta}_q; \mathbf{D})$  e  $L(\boldsymbol{\vartheta}_q^1; \mathbf{D})$  são determinadas por (18) e (19), respectivamente.

As distribuições a priori para  $\boldsymbol{\vartheta}_q$  podem ser definidas da seguinte maneira, supondo que sejam independentes a priori: para  $\alpha$  e  $\lambda$ , assumimos uma distribuição gama com média 1 e variância 100; para os outros parâmetros, assumimos uma distribuição normal com média 0 e variância 100 para cada  $\beta$  e para cada  $\gamma$ .

Por outro lado, também supondo independência a priori, as distribuições a priori para  $\boldsymbol{\vartheta}_q^1$  podem ser definidas como: assumimos uma distribuição normal com média 0 e variância 100 para cada  $\beta_{qj}$ ,  $\gamma_k$  e  $v_l$ , com  $j = 0, 1, \dots, p$ ,  $k = 0, 1, \dots, r$  e  $l = 0, 1, \dots, s$ .

Em razão da distribuição a posteriori não ter forma fechada, geramos valores da distribuição a posteriori por Monte Carlo via Cadeias de Markov (MCMC), método implementado no pacote estatístico `LaplacesDemon` (Statisticat; LLC., 2021). Dessa forma, obteremos uma amostra de tamanho  $M$  da distribuição a posteriori conjunta de  $\boldsymbol{\vartheta}_q$ , a saber,  $\boldsymbol{\vartheta}_q = (\boldsymbol{\vartheta}_q^{(1)}, \boldsymbol{\vartheta}_q^{(2)}, \dots, \boldsymbol{\vartheta}_q^{(M)})$ ,  $\boldsymbol{\vartheta}_q$  o vetor de parâmetros do modelo de mistura padrão. De forma análoga, obteremos  $\boldsymbol{\vartheta}_q^1 = (\boldsymbol{\vartheta}_q^{1(1)}, \boldsymbol{\vartheta}_q^{1(2)}, \dots, \boldsymbol{\vartheta}_q^{1(M)})$ ,  $\boldsymbol{\vartheta}_q^1$  o vetor de parâmetros do modelo defeituoso.

Geralmente, as amostras iniciais são “queimadas” porque a cadeia de Markov ainda não convergiu para uma distribuição estacionária. Essas amostras são chamadas de amostras de aquecimento ou *burn-in* e são utilizadas somente para “dar tempo” para a cadeia de Markov atingir estacionariedade. Por outro lado, como o método de MCMC é construído com base em um processo estocástico e, portanto, gera uma sequência de variáveis aleatórias dependentes no tempo, usamos amostras de salto, cujo tamanho do salto é uma constante  $J$  qualquer, para garantir que as amostras simuladas sejam

aleatórias independentes e identicamente distribuídas (PAULINO et al., 2018). As amostras *burn-in* e de salto são descartadas do processo de simulação por não serem consideradas amostras da distribuição a posteriori.

Assim, a média a posteriori de  $\boldsymbol{\vartheta}_q^1$  pode ser aproximada por

$$\widehat{\boldsymbol{\vartheta}}_q^1 = \frac{1}{M} \sum_{m=1}^M \boldsymbol{\vartheta}_q^{1(m)},$$

e a média a posteriori da fração de cura dada em (15), por

$$\widehat{p}_0 = \frac{1}{M} \sum_{m=1}^M 1 - \left( 1 - \exp \left\{ \frac{\lambda^{(m)}}{\alpha^{(m)}} \right\} \right)^{-\frac{\log(q)}{\log \left( 1 - \exp \left\{ \frac{\lambda^{(m)}}{\alpha^{(m)}} \right\} \right) - \log \left( 1 - \exp \left\{ -\frac{\lambda^{(m)}}{\alpha^{(m)}} \left( e^{\alpha^{(m)} \mu_q^{(m)}} - 1 \right) \right\} \right)}}}.$$

O método que consideramos neste trabalho para comparar os modelos concorrentes e selecionar aquele que melhor se ajustou ao conjunto de dados é baseado no Critério da Densidade Preditiva Ordenada (CPO - abreviatura em inglês para *Conditional Predictive Ordinate*) (GELFAND; SMITH; LEE, 1992).

O CPO para a  $i$ -ésima observação,  $CPO_i$ , pode ser interpretado como a probabilidade de observar  $t_i$  quando o modelo é ajustado sem essa observação. Considere  $D$  os dados com todas as observações e  $D_{-i}$  os dados sem a  $i$ -ésima observação. Então, para a  $i$ -ésima observação, o  $CPO_i$  é dado por

$$CPO_i = \int_{\Theta_q} f(t_i | \boldsymbol{\vartheta}_q^1) \pi(\boldsymbol{\vartheta}_q^1 | D_{-i}) d\boldsymbol{\vartheta}_q^1 = \left\{ \int_{\Theta_q} \frac{\pi(\boldsymbol{\vartheta}_q^1 | D)}{f(t_i | \boldsymbol{\vartheta}_q^1)} d\boldsymbol{\vartheta}_q^1 \right\}^{-1},$$

em que  $\Theta_q$  é o espaço paramétrico de  $\boldsymbol{\vartheta}_q^1$ ,  $f(t_i | \boldsymbol{\vartheta}_q^1)$  é a função densidade de probabilidade e  $\pi(\boldsymbol{\vartheta}_q^1 | D_{-i})$  é a densidade a posteriori de  $\boldsymbol{\vartheta}_q^1$  dado  $D_{-i}$ .

Segundo Gilks, Richardson e Spiegelhalter (1995), o  $CPO_i$  pode ser estimado considerando os  $M$  valores da distribuição a posteriori de  $\boldsymbol{\vartheta}_q^1$ , isto é,

$$\widehat{CPO}_i = \left\{ \frac{1}{M} \sum_{m=1}^M \frac{1}{g(t_i | \boldsymbol{\vartheta}_q^{1(m)})} \right\}^{-1},$$

sendo que  $g(t_i | \boldsymbol{\vartheta}_q^{1(m)}) = f(t_i | \boldsymbol{\vartheta}_q^{1(m)})$  se  $\delta_i = 1$  e  $S(t_i | \boldsymbol{\vartheta}_q^{1(m)})$  se  $\delta_i = 0$ ,  $i = 1, \dots, n$ .

Valores altos de  $CPO_i$  indicam que o modelo é capaz de explicar a  $i$ -ésima observação adequadamente ou, de outro modo, que o modelo se ajustou satisfatoriamente aos dados de interesse.

A medida LPML (*Log Pseudo Marginal Likelihood*) é a soma dos logaritmos dos CPOs de todas as observações, ou seja,

$$LPML = \sum_{i=1}^n \log(\widehat{CPO}_i),$$

em que quanto maior o seu valor, melhor será o ajuste do modelo. No que segue, chamaremos de CPO o LPML.

Mostramos somente a inferência para os parâmetros do modelo defeituoso, pois o processo para obter as medidas a posteriori e CPO do modelo de mistura padrão é o mesmo.

## 4 Estudo de simulação

O objetivo do nosso estudo de simulação consiste em avaliar como diferentes combinações de valores de quantis, parâmetros e tamanho de amostra afetam nos parâmetros estimados pelos

modelos (1) de mistura padrão, (2) defeituoso com covariável ligada somente à função quantílica  $\mu_q$  e (3) defeituoso com covariável associada aos parâmetros  $\mu_q$  e  $\alpha$ . A ideia é verificar se, de fato, os parâmetros dos modelos defeituosos ficam sobrecarregados ao estimar ao mesmo tempo a fração de cura e os parâmetros do tempo de vida dos suscetíveis.

O algoritmo e os procedimentos utilizados para realizar o estudo de simulação pelo método de Monte Carlo podem ser lidos na Subseção 4.1, enquanto os resultados da simulação são discutidos na Subseção 4.2.

Todo o processo de simulação (e da aplicação na Seção 5) dos modelos defeituosos e de mistura padrão aqui considerados foram ajustados por meio das rotinas do pacote R `ggcrqr`<sup>5</sup>, sob as seguintes configurações computacionais: Sistema Operacional Windows 10 Home Single Language, com Processador Intel Core i7 Octa-Core 1,8GHz e 16GB RAM. Tais condições permitiram os seguintes tempos computacionais: fixado um quantil de interesse e  $n = 100$ , foram necessários, em média, 1,87s, 2,03s e 9,79s para amostrar, respectivamente, 1.000 valores da posteriori do modelo defeituoso com covariável só em  $\mu_q^1$ , do modelo defeituoso com covariável em  $\mu_q^1$  e  $\alpha$  e do modelo de mistura padrão. Já para uma amostra 10 vezes maior (ou seja,  $n = 1000$ ), foram necessários, em média, 2,97s, 3,15s e 15,15s para amostrar 1.000 valores da posteriori do modelo defeituoso com covariável só em  $\mu_q^1$ , do modelo defeituoso com covariável em  $\mu_q^1$  e  $\alpha$  e do modelo de mistura padrão, nessa ordem.

#### 4.1 Procedimentos

Um estudo de simulação de Monte Carlo foi realizado para gerar os dados a partir do modelo com distribuição Gompertz generalizada na versão defeituosa. Avaliamos o desempenho dos modelos (1), (2) e (3) em termos do viés, da raiz do erro quadrático médio (REQM) e também por meio do CPO. Na etapa do ajuste dos modelos, foi considerada apenas uma covariável pois é o suficiente para a boa performance do algoritmo de estimação (RODRIGUES; BORGES; SANTOS, 2021). Para o cenário de simulação dado no algoritmo 1, foram os reais<sup>6</sup> valores dos parâmetros:

- para  $q = 0,2$ :  $\alpha = -\exp(-3,77)$ ;  $\lambda = \exp(-2,99)$ ;  $\beta_0 = 2,71$ ;  $\beta_1 = -0,33$ ;
- para  $q = 0,5$ :  $\alpha = -\exp(-3,77)$ ;  $\lambda = \exp(-2,99)$ ;  $\beta_0 = 3,45$ ;  $\beta_1 = -0,23$ ;
- para  $q = 0,8$ :  $\alpha = -\exp(-3,77)$ ;  $\lambda = \exp(-2,99)$ ;  $\beta_0 = 4,14$ ;  $\beta_1 = -0,16$ .

As frações de cura foram igualmente  $p_{00} = 0,26$  e  $p_{01} = 0,20$  para  $q \in \{0,2; 0,5; 0,8\}$ . Vale ressaltar ainda que os tamanhos amostrais utilizados foram  $n = 50; 100; 300; 500; 1.000; 2.000$  e que a porcentagem de censura à direita considerada foi de 27%. Para cada configuração de valores de parâmetros e tamanho de amostra,  $B = 1.000$  conjuntos de dados foram gerados. Para obter as estatísticas a posteriori, as 100.000 primeiras amostras do modelo de mistura padrão e as 200.000 primeiras amostras dos modelos defeituosos foram classificadas como amostras *burn-in* e, por isso, descartadas da análise. Saltos de tamanho 100 foram considerados nas amostras dos modelos defeituosos e de mistura a fim de que não houvesse correlação entre os valores simulados ou, ao menos, que ela fosse próxima de zero. Como resultado, obtivemos  $M = 1.000$  médias a posteriori, para as quais a média foi considerada como a estimativa pontual de cada parâmetro.

Sendo assim, os dados foram gerados pelo algoritmo 1.

<sup>5</sup> Disponível em: <<https://github.com/brsantos/ggcrqr>>.

<sup>6</sup> Os verdadeiros valores dos parâmetros utilizados no processo de simulação foram as estimativas obtidas a partir do conjunto de dados da aplicação apresentada na Seção 5 apenas com a covariável de grupo.

---

**Algoritmo 1** Algoritmo de geração de dados pelo modelo defeituoso

---

- 1: Determine valores desejáveis para  $\gamma_0$ ,  $v_0$ ,  $q$  e  $\beta_q = (\beta_{q0}, \beta_{q1})^\top$ ;
- 2: Defina a proporção de dados censurados, dado por  $pc_0$  e  $pc_1$ , para  $x = 0$  e  $x = 1$ , respectivamente;
- 3: Para a  $i$ -ésima observação, gere  $x_i \sim \text{Bernoulli}(0.5)$ , e calcule  $p_{0x_i}$ , em que  $p_{00}$  e  $p_{01}$  são dados por (15) quando  $x_i = 0$  e  $x_i = 1$ , respectivamente;
- 4: Gere  $u_i \sim \text{Uniforme}(0, 1)$ . Se  $u_i < p_{0x_i}$ , faça  $w_i = \infty$ ; caso contrário, gere  $u_{1i} \sim U(0, 1 - p_{0x_i})$  e calcule

$$w_i = \frac{1}{\alpha} \log \left[ 1 - \frac{\alpha}{\lambda} \log \left( 1 - u_{1i}^{(1/\theta_{x_i})} \right) \right], \text{ com } x_i = 1 \text{ ou } x_i = 0,$$

e  $\theta_{x_i}$  é dado na Equação (12) com  $\mu_q^1 = \exp(\beta_0)$ , se  $x_i = 0$  e  $\mu_q^1 = \exp(\beta_0 + \beta_1)$ , se  $x_i = 1$ .

- 5: Gere  $c_i \sim U(0, \tau_i)$ , em que  $\tau_i$  é definido para ter aproximadamente  $p_{cx_i}$  proporção de dados censurados.
  - 6: Determine  $t_i = \min(w_i, c_i)$ . Se  $t_i = w_i$ , faça  $\delta_i = 1$ , caso contrário  $\delta_i = 0$ ;
  - 7: Repita os passos 3 a 6 para todo  $i = 1, \dots, n$ . Os dados para o  $i$ -ésimo indivíduo é  $\{t_i, x_i, \delta_i\}$ ,  $i = 1, \dots, n$ .
- 

## 4.2 Resultados e discussões

Como mencionado na Subseção 3.4, o Critério da Densidade Preditiva Ordenada é um método útil para diagnóstico e comparação de modelos. Por isso, resumimos na Figura 1 os 1.000 valores de CPO dos modelos defeituosos e de mistura em forma de boxplots para toda configuração de  $q \in \{0, 2; 0, 5; 0, 8\}$  e tamanhos amostrais considerados na simulação. De forma geral, os modelos defeituosos obtiveram os maiores valores de CPO independente de  $q$  e do tamanho da amostra, com vantagem tênue do modelo defeituoso com covariável relacionada somente à função quantílica  $\mu_q$ . Por outro lado, é curioso observar que, embora a variabilidade dos dados tenha se reduzido à medida que o tamanho da amostra aumentou, a presença de muitos valores atípicos nos boxplots do modelo de mistura padrão quando  $n = 1.000$  e  $n = 2.000$  expôs um possível problema de adequação desse modelo aos dados.

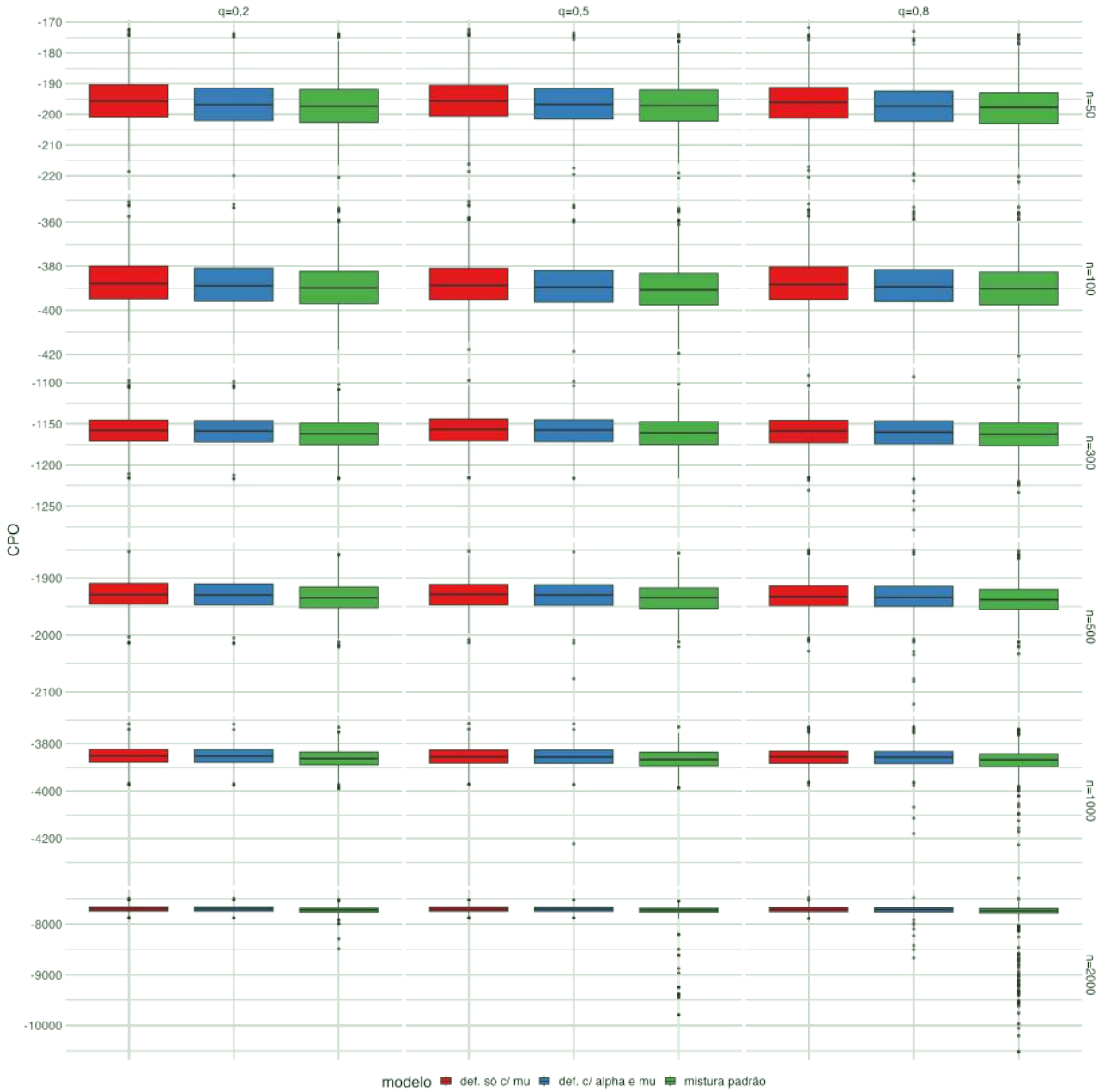


Figura 1 – Boxplots representativos do Critério da Densidade Preditiva Ordenada (CPO) dos modelos defeituoso com covariável só em  $\mu_q$  e defeituoso com covariável em  $\mu_q$  e  $\alpha$  e de mistura padrão para  $q \in \{0,2; 0,5; 0,8\}$ .

Na Figura 2, temos as estimativas das médias a posteriori (e seus respectivos valores de REQm) apenas dos parâmetros comparáveis entre os três modelos em discussão - para ver as médias estimadas e REQm de todos os parâmetros de cada um dos modelos individualmente, consulte o Apêndice 6. Para todos os  $q$  considerados, ficou evidente que as estimativas pontuais obtidas para os modelos defeituosos com covariável só em  $\mu_q$  e com covariável em  $\mu_q$  e  $\alpha$  foram substancialmente próximas dos seus valores verdadeiros, e se aproximaram ainda mais com o aumento do tamanho amostral, o que caracterizou a quase nulidade de viés nesse contexto. Ao observarmos a raiz do erro quadrático médio dos modelos defeituosos, a medida se mostrou relativamente maior para os tamanhos amostrais 50 e 100, mas diminuiu rapidamente conforme os tamanhos das amostras aumentaram. Nessa mesma figura, vale ainda destacar o comportamento desfavorável das estimativas pontuais do modelo de mistura padrão: em geral, viés e REQm altos especialmente nos maiores tamanhos de amostra.

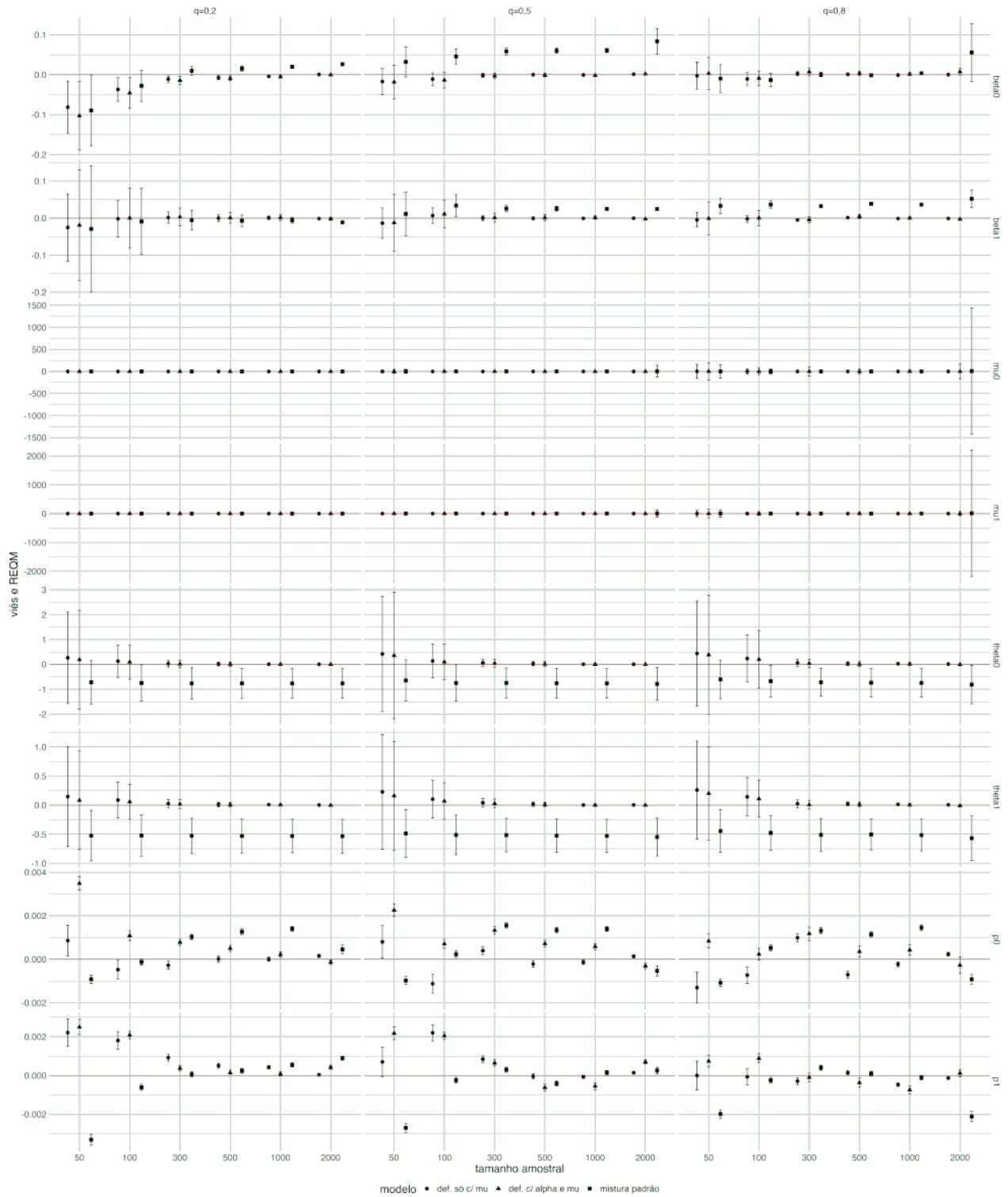


Figura 2 – Médias a posteriori, representadas pelos pontos (modelo defeituoso com covariável só em  $\mu_q$ ), triângulos (modelo defeituoso com covariável em  $\mu_q$  e  $\alpha$ ) e quadrados (modelo de mistura padrão), para as diferentes combinações de  $q \in \{0,2; 0,5; 0,8\}$ , valores de parâmetros e tamanhos de amostra  $n = 50; 100; 300; 500; 1.000; 2.000$ . A barra ao redor do ponto, triângulo e quadrado retrata o REQM, enquanto a linha vermelha centrada em zero representa ausência de viés.

Ainda que o modelo de mistura padrão se comporte dessa maneira, é esperado que seus resultados estejam aquém dos modelos defeituosos, posto que os dados foram simulados a partir de um modelo defeituoso. Mesmo assim, o desempenho melhor dos modelos defeituosos significa que não ter

incluído um parâmetro extra para a proporção de curados não foi capaz de sobrecarregar os demais parâmetros da distribuição Gompertz generalizada. Outro detalhe importante desse estudo de simulação é que, entre os modelos com abordagem defeituosa, ambos apresentaram desempenhos semelhantes tanto nas medidas de CPO e REQM quanto no que se refere a viés.

## 5 Aplicação

Com este estudo, esperamos investigar o impacto de algumas variáveis de caracterização; tais quais faixa etária ( $<30$  e  $\geq 30$ ), grupo (gestante ou puérpera) e vacina contra a COVID-19; de sintoma, saturação e perda de paladar; e de comorbidade, obesidade, no tempo de internação até o óbito de gestantes e puérperas devido a complicações da doença de COVID-19.

Consideramos a base de dados do Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe), sistema oficial para o registro dos casos e óbitos por Síndrome Respiratória Aguda Grave (SRAG) no Brasil disponibilizado pelo Ministério da Saúde, no portal openDataSUS<sup>7</sup>. Foram 9.936 registros de gestantes e puérperas de 10 a 55 anos hospitalizadas com SRAG por COVID-19, confirmada por teste de RT-PCR, datados desde o início da pandemia até março de 2022. Foram considerados como morte por causa específica o óbito causado estritamente por complicações da doença de COVID-19 (1.198 - 12% da amostra total) e como censura à direita todos os outros casos (8.738 - 88% da amostra total).

O tempo, em dias, de hospitalização até o óbito pela COVID-19 foi definido como o evento de interesse e, como dito anteriormente, algumas variáveis foram observadas para analisar seu efeito sobre o tempo de sobrevivência de grávidas e puérperas internadas em decorrência da COVID-19 grave.

Na Subseção 5.1, a análise descritiva foi utilizada para caracterizar as variáveis em frequências absolutas e relativas e estimar as funções de sobrevivência por meio do estimador não-paramétrico Kaplan-Meier (COLOSIMO; GIOLO, 2006). Na Subseção 5.2 são apresentados os resultados dos ajustes aos dados de gestantes e puérperas hospitalizadas com SRAG por COVID-19 dos modelos de mistura padrão, defeituoso com covariável associada apenas ao parâmetro  $\mu_q$  e defeituoso com covariável relacionada aos parâmetros  $\mu_q$  e  $\alpha$ . Na Subseção 5.3, selecionamos o melhor modelo segundo a medida de CPO.

O *software* livre R (TEAM, 2022) também foi utilizado para a análise estatística aqui descrita.

### 5.1 Análise descritiva

Ao avaliar as variáveis de caracterização, um total de 8.084 (81,36%) eram gestantes e 1.852 (18,64%) eram puérperas. Dessas mulheres, 4.519 (45,48%) tinham menos do que 30 anos e 5.417 (54,52%) tinham 30 ou mais anos. Por fim, 9.225 (92,84%) gestantes e puérperas deste estudo não tomaram a vacina em combate à COVID-19, enquanto apenas 711 (7,16%) receberam o imunizante. Ao avaliar as variáveis de sintomas, 4.094 (41,20%) apresentaram dessaturação e 1.215 (12,23%), perda de paladar. Ao avaliar a variável de comorbidade, 716 (7,21%) foram classificadas como obesas.

As figuras 3 e 4 mostram as estimativas de Kaplan-Meier (K-M) das variáveis de caracterização, sintoma e comorbidade aqui supracitadas. À primeira vista, podemos perceber nas figuras 3 e 4 que uma fração da população de mulheres sob investigação foi curada - portanto, utilizar uma abordagem que considera a cura é de fato apropriada para a problemática em questão. Apesar disso, notamos nas figuras 3(a) e 3(b) que a pandemia de COVID-19 parece ter sido mais letal em mulheres no ciclo puerperal e em gestantes e puérperas cuja idade é de 30 anos ou mais. Por outro lado, é de se constatar na Figura 3(c) que aquelas que foram vacinadas contra o vírus receberam o melhor prognóstico.

Além disso, a presença de alguns sintomas pode servir como um indicativo para o agravamento da doença, uma vez que podem diminuir a probabilidade de sobrevivência da paciente internada. São os casos dos sintomas relacionados à insuficiência respiratória, como saturação, que pode ser verificada na Figura 4(a). Na contramão, existem os sintomas que estão associados aos casos leves de COVID-19,

<sup>7</sup> Site: <<https://opendatasus.saude.gov.br/group/dados-sobre-srag>>.



como a perda de paladar (LECHIEN et al., 2021). Na Figura 4(b), observamos que a manifestar durante a infecção pelo vírus diminui o risco da morte materna. Muito diferente da situação em que gestantes e puérperas são obesas, como pode ser visto na Figura 4(c).

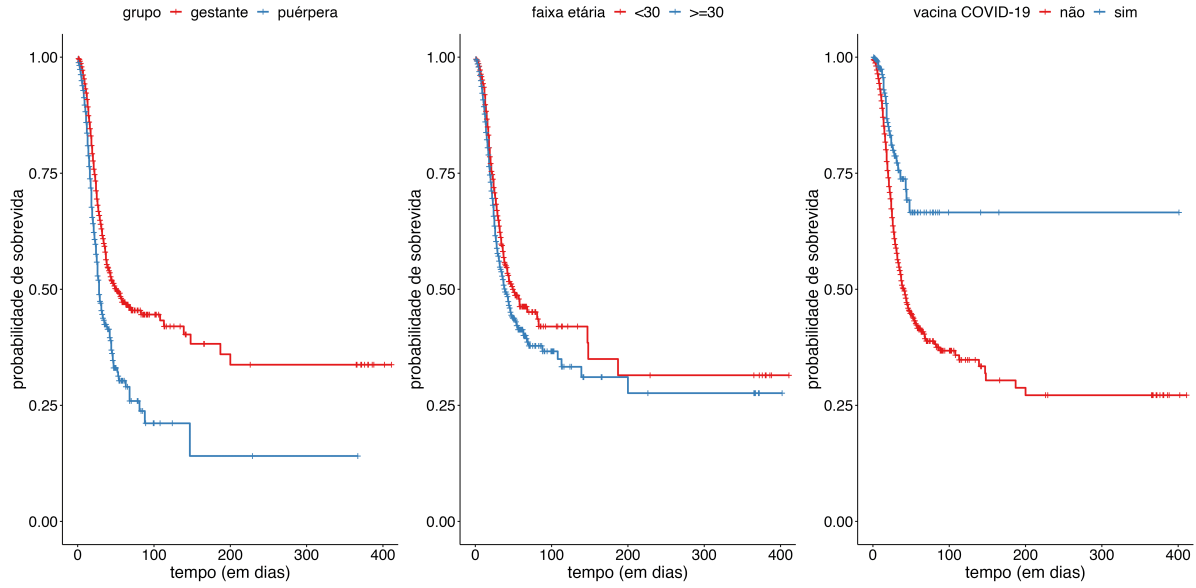


Figura 3 – Curva de sobrevida estimada via K-M para gestantes e puérperas internadas com SRAG por COVID-19, por (a) grupo, (b) faixa etária e (c) vacina contra a COVID-19.

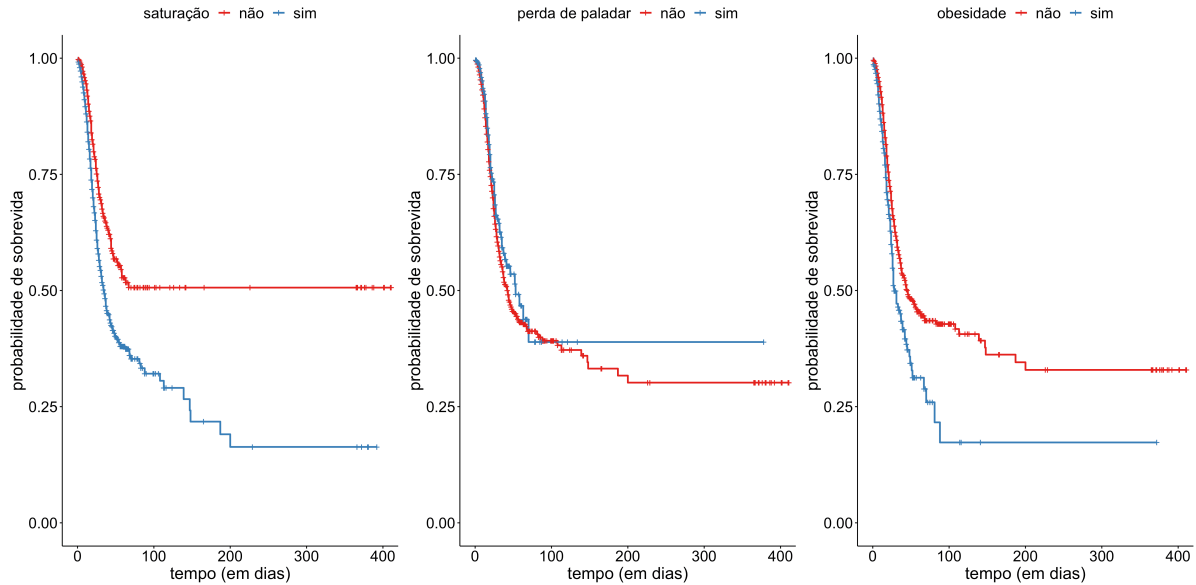


Figura 4 – Curva de sobrevida estimada via K-M para gestantes e puérperas internadas com SRAG por COVID-19, por (a) saturação, (b) perda de paladar e (c) obesidade.

## 5.2 Modelagem

A seguir, são apresentados os resultados acerca dos ajustes dos modelos de mistura padrão (Subsubseção 5.2.1), defeituoso com covariável associada somente ao parâmetro  $\mu_q$  (Subsubseção 5.2.2) e defeituoso que considera covariável nos parâmetros  $\mu_q$  e  $\alpha$  (Subsubseção 5.2.3).

### 5.2.1 Modelo de mistura padrão

Ao considerar o modelo de mistura padrão, todas as covariáveis (faixa etária, grupo, vacina de COVID-19, saturação, perda de paladar e obesidade) foram incluídas na função quantílica  $\mu_q$  definida em (4), para os indivíduos suscetíveis, e na função da fração de cura definida em (5), para os indivíduos imunes. No entanto, como pode ser visto na Figura 5, as covariáveis faixa etária, obesidade, perda de paladar e vacina não foram significantes do ponto de vista estatístico, uma vez que o zero está contido nos intervalos com 95% de credibilidade HPD correspondentes. Dessas, devido à sua importância clínica, apenas vacina se manteve no modelo de mistura padrão considerado como final, ou seja, o modelo foi ajustado sem as covariáveis faixa etária, obesidade e perda de paladar. Cabe mencionar que a covariável perda de paladar<sup>8</sup> também não foi importante estatisticamente para a fração de cura, e por isso não foi incluída na fração de cura do ajuste final.

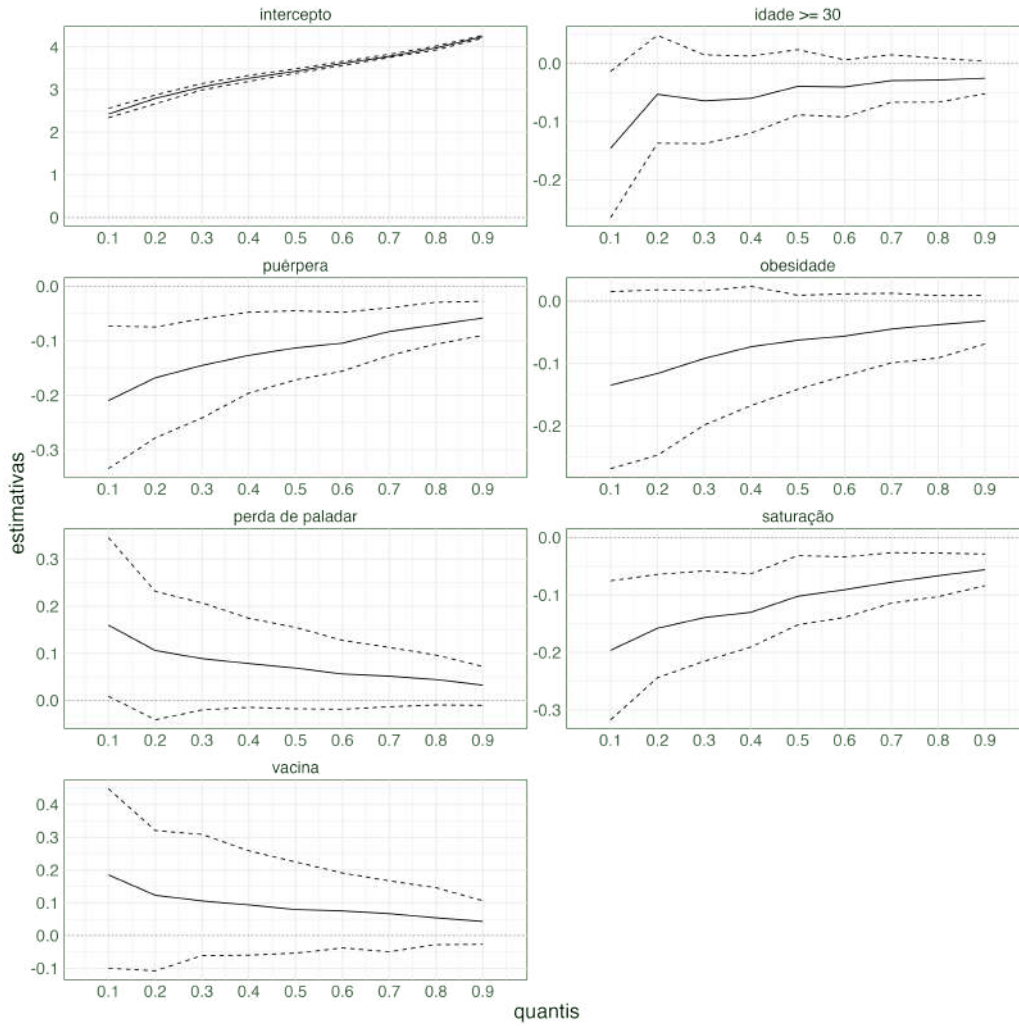


Figura 5 – Estimativas a posteriori dos parâmetros de regressão  $\beta$  para os diferentes quantis  $q$  do modelo de mistura padrão com todas as covariáveis incluídas nas funções quantílica  $\mu_q$  e da fração de cura. As linhas sólidas representam a média posteriori e as linhas tracejadas, o intervalo de com 95% credibilidade HPD.

Logo, o modelo de mistura padrão foi reajustado e a função quantílica  $\mu_q$  foi dada por

$$\mu_q = \exp \left\{ \beta_{q0} + \beta_{q1} \times \text{grupo}_{\text{puérpera}} + \beta_{q2} \times \text{saturação}_{\text{sim}} + \beta_{q3} \times \text{vacina}_{\text{sim}} \right\},$$

em que  $q = \{0, 1; 0, 2; 0, 3; \dots; 0, 9\}$  e os coeficientes  $\beta_{qi}$ ,  $i = 0, 1, 2, 3, 4$ , variando em  $q$ .

<sup>8</sup> Média a posteriori do parâmetro  $\gamma$  associado à perda de paladar: 0,083; IC HPD de 95%: (-0,529; 0,695).

A função da fração de cura, por sua vez, é dada por

$$p_0 = a/1 + a,$$

em que

$$a = \exp \left\{ \gamma_0 + \gamma_1 \times \text{idade}_{>=30} + \gamma_2 \times \text{grupo}_{\text{puérpera}} + \gamma_3 \times \text{obesidade}_{\text{sim}} + \gamma_4 \times \text{saturação}_{\text{sim}} + \gamma_5 \times \text{vacina}_{\text{sim}} \right\}.$$

As médias a posteriori e os intervalos de credibilidade HPD de 95% (IC HPD 95%) estimados para os parâmetros  $\alpha$  e  $\lambda$  e para os coeficientes relacionados à  $p_0$  (fração de cura) são apresentados na Tabela 1. A partir desses resultados, podemos obter as estimativas da fração de cura para cada uma das covariáveis avaliadas neste estudo.

parâmetro	média a posteriori	IC HPD de 95%
$\alpha$	1,16e-16	(1,14e-16; 1,17e-16)
$\lambda$	0,047	(0,047; 0,047)
$\gamma_0$	0,252	(-0,045; 0,577)
$\gamma_1$	-0,575	(-0,836; -0,279)
$\gamma_2$	-1,041	(-1,575; -0,552)
$\gamma_3$	-1,233	(-1,913; -0,582)
$\gamma_4$	-0,853	(-1,180; -0,551)
$\gamma_5$	1,463	(0,717; 2,050)

Tabela 1 – Estimativas a posteriori para os parâmetros  $\alpha$  e  $\lambda$  e para os coeficientes de  $p_0$  do modelo de mistura padrão final.

Os coeficientes regressores da função quantílica são representados graficamente. Na Figura 6, é possível observar os valores dos coeficientes  $\beta$  do modelo final sob os diferentes quantis  $q$  tal que  $q \in \{0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9\}$ . É interessante perceber que os coeficientes variam ao longo dos quantis, sugerindo que ser gestante ou puérpera, apresentar problemas de saturação ou tomar a vacina contra a doença de COVID-19 implica em efeitos distintos nos tempos de sobrevivência das pacientes suscetíveis ao óbito. Entretanto, todos os coeficientes pareceram se aproximar de zero nos maiores quantis (a saber, 0,7, 0,8 e 0,9), indicando uma menor diferença entre os tempos de falha nesses quantis. Também é importante notar que, com a exceção da vacina, todos os intervalos de credibilidade HPD, dessa vez, não incluíram o valor zero para quaisquer valores de  $q$ .

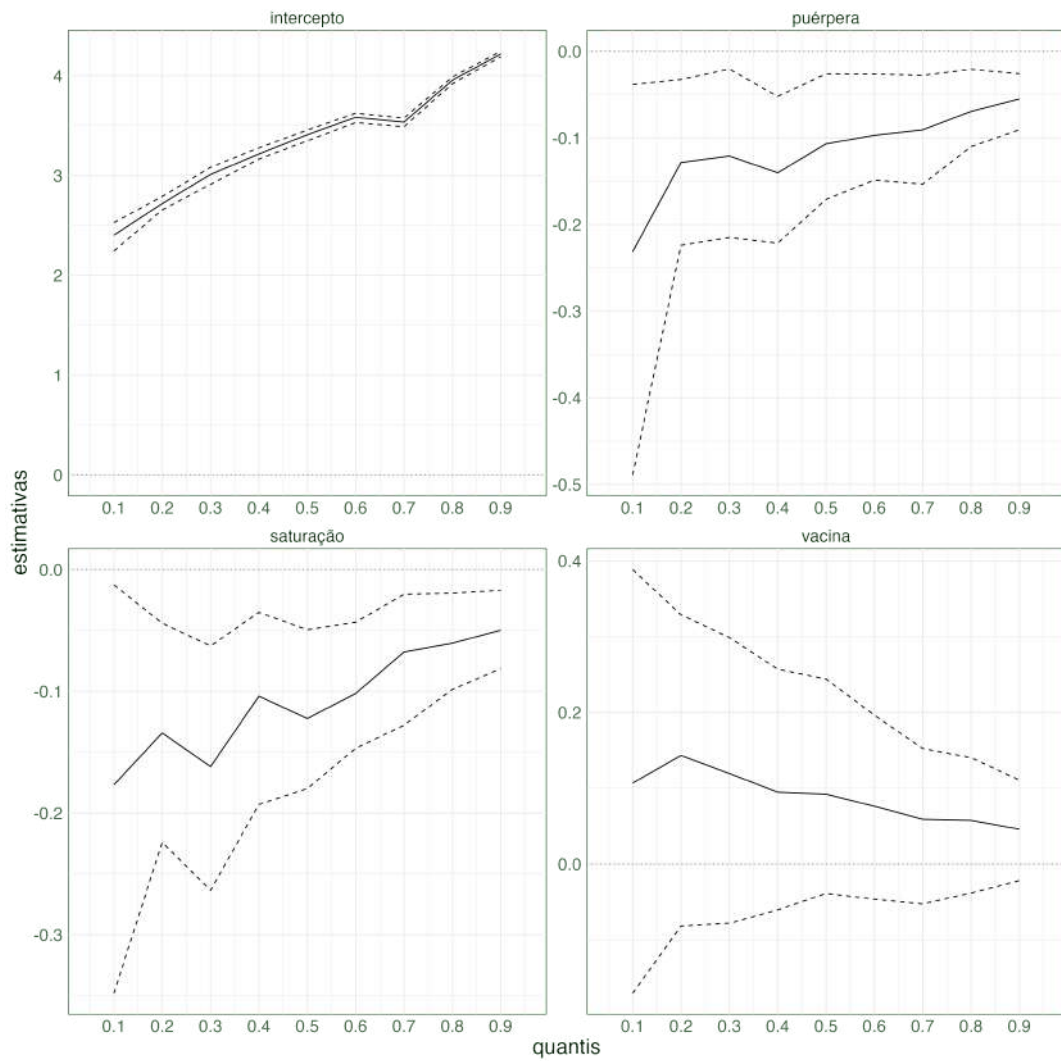


Figura 6 – Estimativas a posteriori dos parâmetros de regressão  $\beta$  para os diferentes quantis  $q$  do modelo de mistura padrão final. As linhas sólidas representam a média a posteriori e as linhas tracejadas, o intervalo com 95% de credibilidade HPD.

Com o suporte das estimativas a posteriori da fração de cura, construímos gráficos da densidade a posteriori de cada combinação de grupo (gestante ou puérpera) com as demais covariáveis consideradas no modelo de mistura padrão final. Podemos verificar na Figura 7 que gestantes e puérperas que foram imunizadas com a vacina de COVID-19 receberam um melhor prognóstico para a doença, assim como as que não apresentaram saturação abaixo de 95% (valor considerado como estável) e nem foram classificadas como obesas. Por outro lado, a idade não pareceu interferir na cura.

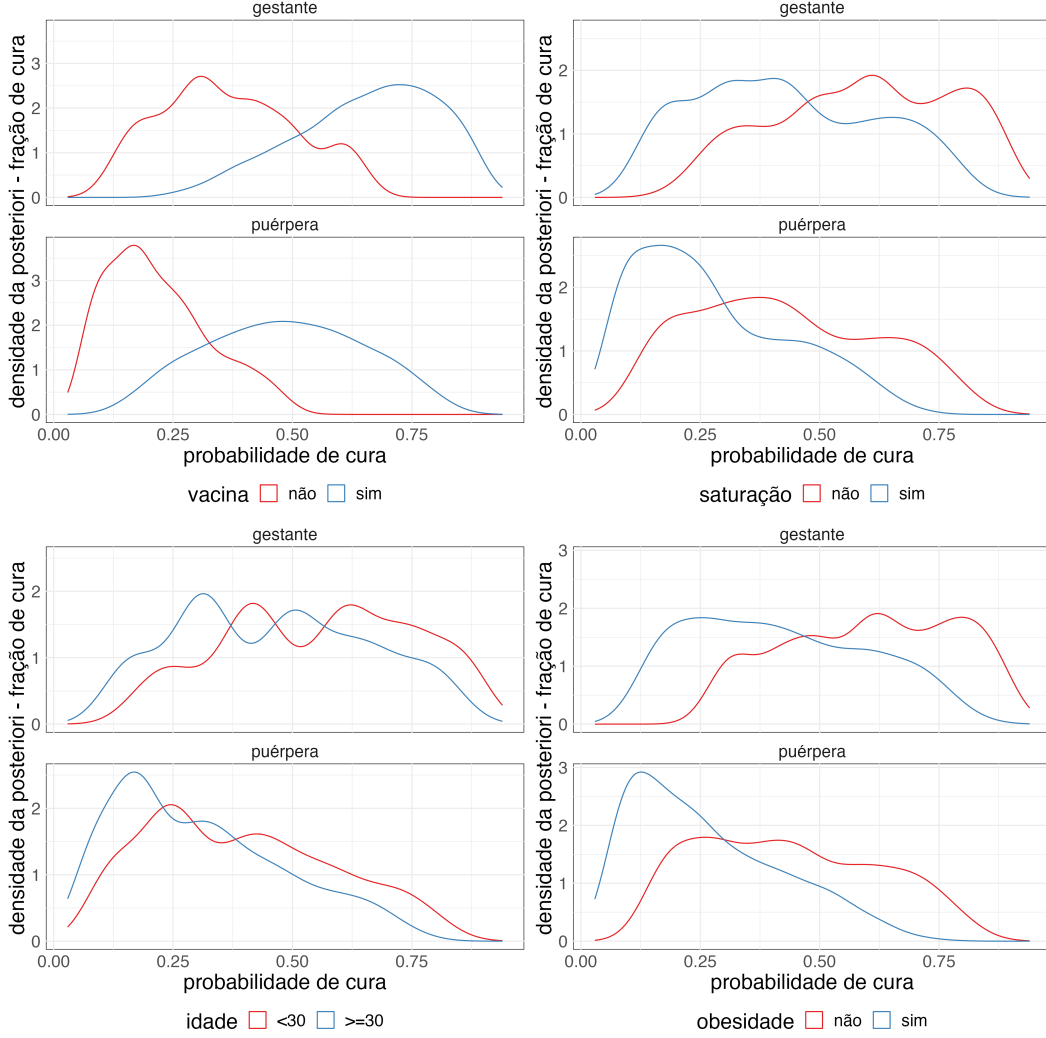


Figura 7 – Densidades a posteriori da fração de cura para cada combinação de grupo (gestante ou puérpera) e outras covariáveis consideradas no modelo de mistura padrão final.

### 5.2.2 Modelo defeituoso com covariável só em $\mu_q^1$

O modelo defeituoso a ser descrito nesta subsubseção considerou covariáveis ligadas apenas à função quantílica  $\mu_q^1$  enunciada em (16). Assim,

$$\mu_q^1 = \exp \left\{ \beta_{q0} + \beta_{q1} \times \text{idade}_{\geq 30} + \beta_{q2} \times \text{grupo}_{\text{puérpera}} + \beta_{q3} \times \text{obesidade}_{\text{sim}} + \beta_{q4} \times \text{perda de paladar}_{\text{sim}} + \beta_{q5} \times \text{saturação}_{\text{sim}} + \beta_{q6} \times \text{vacina}_{\text{sim}} \right\}, \quad (20)$$

em que  $q = \{0, 1; 0, 2; 0, 3; \dots; 0, 9\}$  e os coeficientes  $\beta_{qi}$ ,  $i = 0, 1, 2, 3, 4, 5, 6$ , variam em  $q$ .

Nesse cenário, todos os intervalos de credibilidade HPD não abrangeram o valor zero para nenhum valor de  $q$  (Figura 8), conferindo significância para todas as covariáveis consideradas no ajuste do modelo especificado na Equação (20). Na Figura 8, também é possível observar as estimativas dos efeitos das covariáveis nos diferentes quantis, como também o comportamento heterogêneo dos efeitos ao longo desses quantis. Pacientes internadas com SRAG por COVID-19 que atenderam à faixa etária maior ou igual a 30 anos ou que eram puérperas ou obesas ou que apresentaram dessaturação tiveram tempo de sobrevivência menor em relação àquelas com características contrárias. Em contrapartida, para a covariável perda de paladar e vacina de COVID-19, o efeito foi positivo sobre o tempo de sobrevivência dessas mulheres.

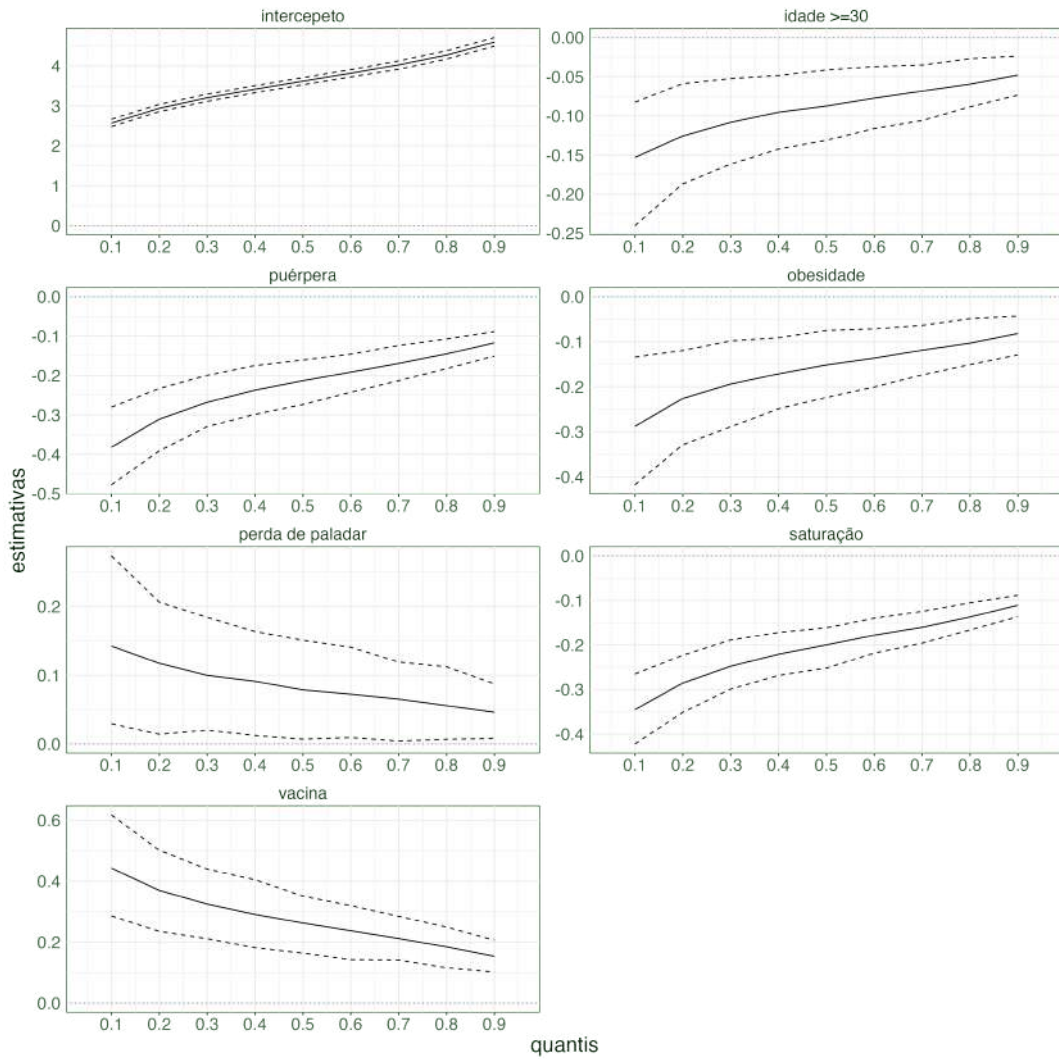


Figura 8 – Estimativas a posteriori dos parâmetros de regressão  $\beta$  para os diferentes quantis  $q$  do modelo defeituoso com covariável só em  $\mu_q^1$ . As linhas sólidas representam a média a posteriori e as linhas tracejadas, o intervalo com 95% de credibilidade HPD.

Na Tabela 2 estão as estimativas das médias a posteriori e os intervalos de credibilidade HPD de 95% dos coeficientes da regressão associados à  $\alpha$  e  $\lambda$ .

parâmetro	média a posteriori	IC HPD de 95%
$\gamma_0$	-3,779	(-3,937; -3,604)
$v_0$	-3,015	(-3.134; -2.888)

Tabela 2 – Estimativas a posteriori para os coeficientes  $\gamma_0$  e  $v_0$  do modelo defeituoso com covariável só em  $\mu_q^1$ .

Ao avaliar os efeitos das covariáveis sobre a taxa de cura considerando o grupo a que a paciente pertence (Figura 9), as vacinas ofereceram maior probabilidade de cura tanto para as mulheres gestantes quanto para as mulheres em período puerperal. O mesmo aconteceu se a paciente não sofreu com saturação baixa, pois sob essa condição sua probabilidade de cura foi relativamente maior. Contudo, ter idade menor, maior ou igual a 30 anos, ser ou não obesa e não apresentar o sintoma da perda de paladar pareceram não fazer diferença na fração de cura, visto as curvas quase sobrepostas nos gráficos das densidades dessas covariáveis.

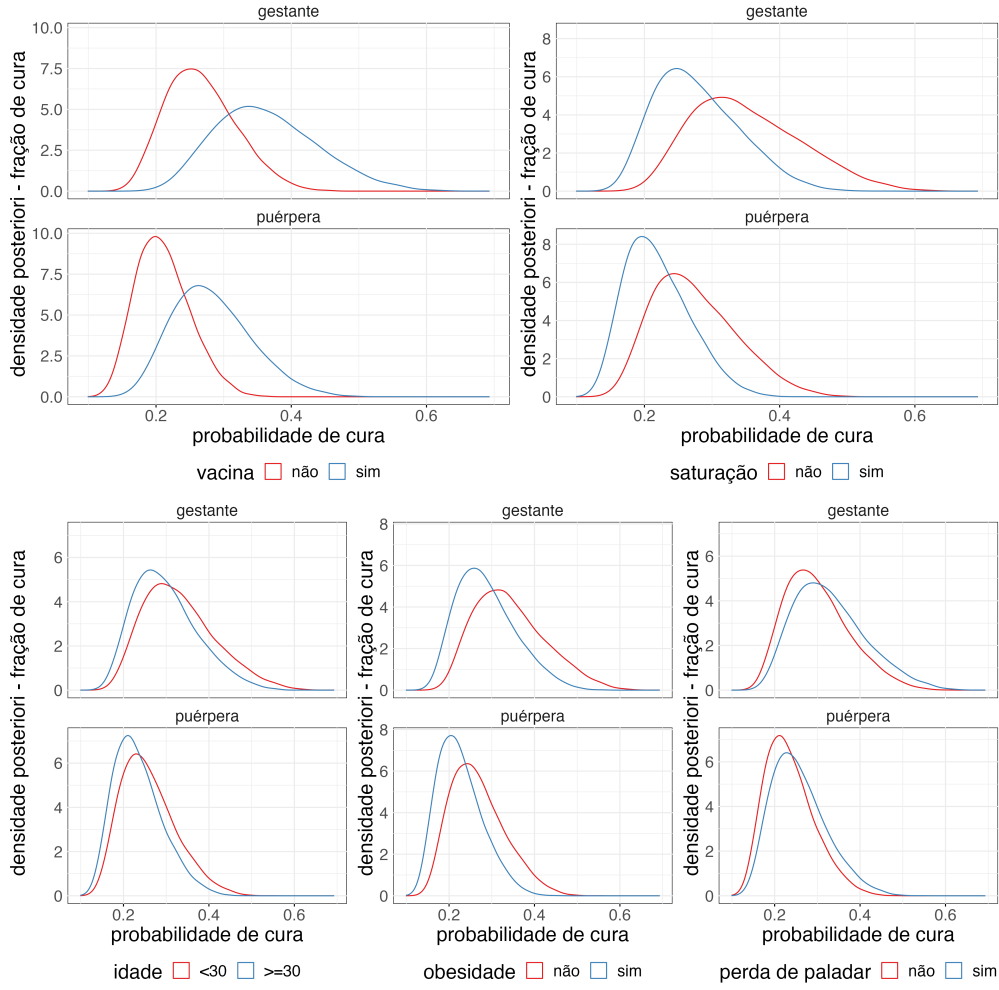


Figura 9 – Densidades a posteriori da fração de cura para cada combinação de grupo (gestante ou puérpera) e outras covariáveis consideradas no modelo defeituoso com covariável só em  $\mu_q^1$ .

### 5.2.3 Modelo defeituoso com covariável em $\mu_q^1$ e $\alpha$

Uma segundo modelo defeituoso foi ajustado aos dados de gestantes e puérperas internadas com SRAG por COVID-19. Dessa vez, além de atribuir covariáveis na função quantílica  $\mu_q^1$ , também ligamos as covariáveis ao parâmetro  $\alpha$  (vide Equação (17)) com a tentativa de melhorar a explicação dos dados. Essa discussão pôde ser vista na Subseção 3.3. A função quantílica para essa versão foi a mesma definida em (20), ao passo que  $\alpha$  foi dado por

$$\alpha = -\exp \left\{ \gamma_0 + \gamma_1 \times \text{grupo}_{\text{puérpera}} \right\}.$$

As estimativas das médias a posteriori e os intervalos de credibilidade HPD de 95% obtidos para os coeficientes relacionados à  $\alpha$  e  $\lambda$  são mostrados na Tabela 3.

parâmetro	média a posteriori	IC HPD de 95%
$\gamma_0$	-3,718	(-3,903; -3,554)
$\gamma_1$	-0,409	(-0,691; -0,097)
$v_0$	-3,026	(-3,145; -2,908)

Tabela 3 – Estimativas a posteriori para os coeficientes  $\gamma_0$ ,  $\gamma_1$  e  $v_0$  do modelo defeituoso com covariável em  $\mu_q^1$  e  $\alpha$ .



Sob a perspectiva dos intervalos de credibilidade mostrados na Figura 10, foi notável a significância das estimativas das covariáveis faixa etária, obesidade e saturação, cujos efeitos foram negativos, e perda de paladar e vacina de COVID-19, que impactaram favoravelmente para o risco de óbito. Porém, ao atentar-se à covariável grupo, não houve evidências de que quantis acima de 0,6 dos tempos de sobrevivência das pacientes puérperas fossem diferentes das pacientes gestantes, a categoria de referência. Isso significa que não houve evidências de diferença entre esses grupos para os maiores tempos de sobrevivência.

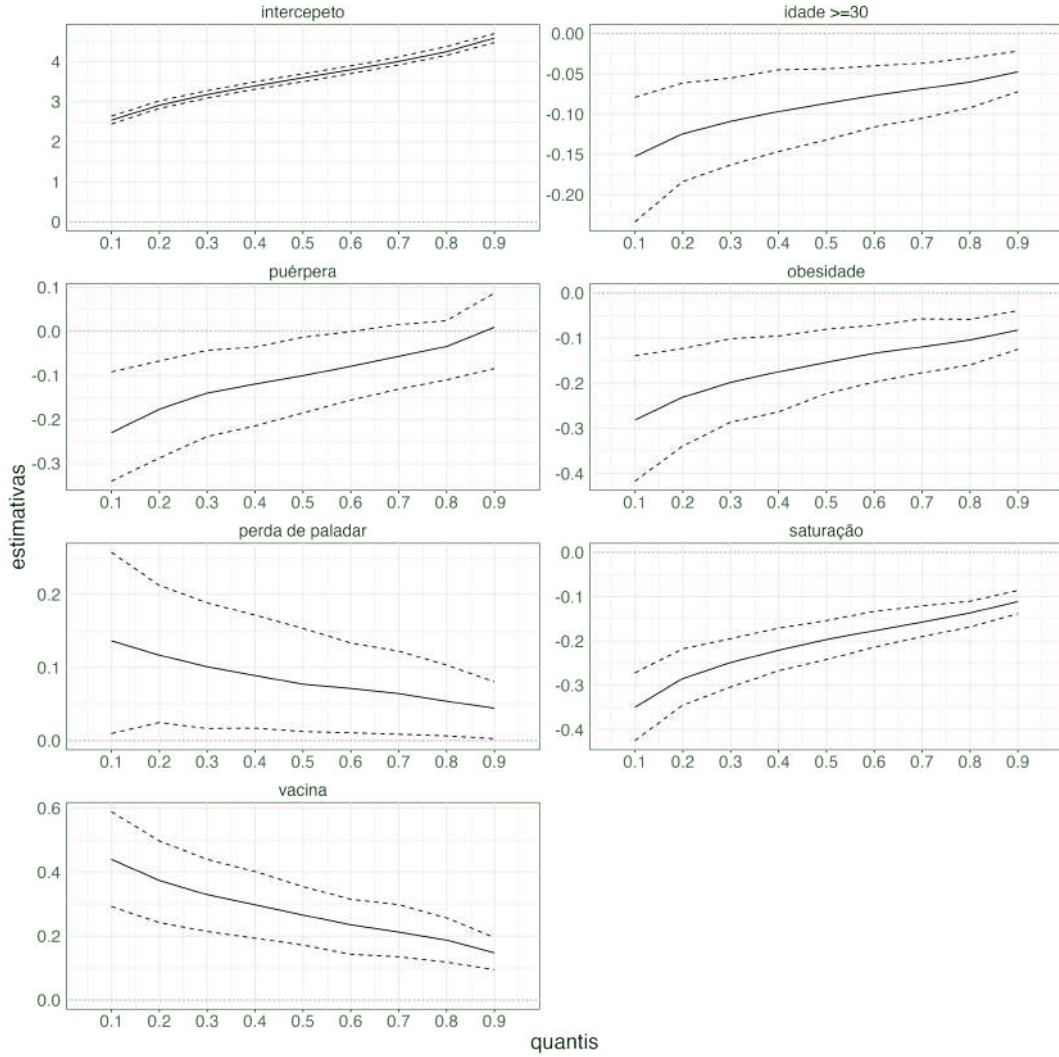


Figura 10 – Estimativas a posteriori dos parâmetros de regressão  $\beta$  para os diferentes quantis  $q$  do modelo defeituoso com covariável em  $\mu_q^1$  e  $\alpha$ . As linhas sólidas representam a média a posteriori e as linhas tracejadas, o intervalo com 95% de credibilidade HPD.

Ainda assim, nos gráficos das densidades a posteriori da fração de curados apresentados na Figura 11, mulheres grávidas tiveram maior probabilidade de cura em comparação às mulheres no período pós-parto em qualquer uma das covariáveis analisadas - resultado que está em consonância com o gráfico de Kaplan-Meier exposto na Figura 3(a). A Figura 11 nos permite ainda deduzir que, tal como no modelo defeituoso com covariável atribuída apenas em  $\mu_q$ , grávidas ou puérperas terem tomado a vacina ou não terem tido saturação anormal de oxigênio tiveram maior probabilidade de cura. Da mesma forma que faixa etária, obesidade e perda de paladar, que também não sinalizaram qualquer interferência na fração de cura, não importando se a paciente era gestante ou se estava no período de puerpério.

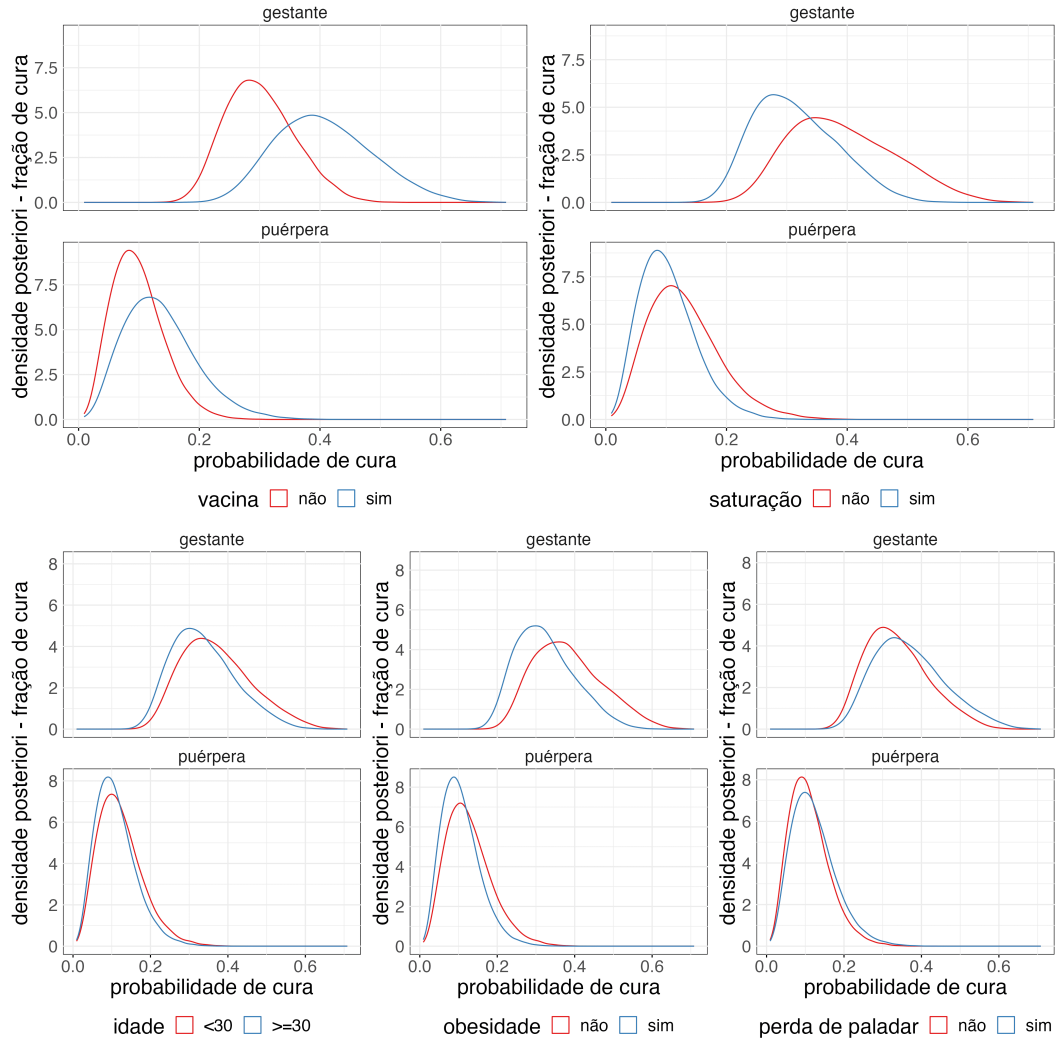


Figura 11 – Densidades a posteriori da fração de cura para cada combinação de grupo (gestante ou puérpera) e outras covariáveis consideradas no modelo defeituoso com covariável em  $\mu_q^1$  e  $\alpha$ .

### 5.3 Seleção do modelo

Entre os modelos ajustados aos dados de COVID-19 grave na população materna, usamos a metodologia do CPO para escolher aquele cujo ajuste foi o melhor. Como podemos ver na Tabela 4, o modelo defeituoso em que covariáveis foram atribuídas à função quantílica  $\mu_q^1$  e ao parâmetro  $\alpha$  obteve o maior valor de CPO para todos os valores de quantis observados. Sendo assim, esse foi o modelo considerado como o mais adequado para a nossa análise.

Além disso, é importante dizer que a larga vantagem para o modelo de mistura padrão evidencia o benefício de não precisar incluir um parâmetro para a fração de curados e, ainda sim, não comprometer os outros parâmetros envolvidos na modelagem.

$q$	modelo		
	mistura padrão	def. c/ covariável em $\mu_q$	def. c/ covariável em $\mu_q$ e $\alpha$
0, 1	-6382,499	-6243,424	-6239,460
0, 2	-6353,860	-6242,880	-6239,267
0, 3	-6358,875	-6242,797	-6239,252
0, 4	-6356,500	-6243,028	-6239,453
0, 5	-6347,160	-6243,215	-6239,283
0, 6	-6346,502	-6243,105	-6239,959
0, 7	-6344,344	-6243,300	-6240,039
0, 8	-6345,358	-6243,147	-6239,671
0, 9	-6349,388	-6243,238	-6240,159

Tabela 4 – Critério da Densidade Preditiva Ordenada (CPO) dos modelos de mistura padrão, defeituoso com covariável só em  $\mu_q^1$  e defeituoso com covariável em  $\mu_q^1$  e  $\alpha$  para diferentes valores de quantis.

## 6 Considerações finais e propostas futuras

Desenvolvemos neste trabalho o modelo de regressão quantílica de mistura padrão com distribuição Gompertz generalizada e ainda comparamo-lo com o modelo de regressão quantílica defeituoso proposto por [Rodrigues, Borges e Santos \(2021\)](#). O interesse maior foi avaliar se os parâmetros do modelo defeituoso ficavam sobrecarregados ao acomodarem informações tanto dos indivíduos suscetíveis quanto dos indivíduos imunes. Um estudo de simulação de Monte Carlo com os dados gerados pelo modelo de [Rodrigues, Borges e Santos \(2021\)](#) mostrou que a não adição de um parâmetro a mais para a fração de cura nos modelos defeituosos possibilitou melhores estimativas em comparação ao modelo de mistura. A aplicação dos modelos de mistura padrão e defeituosos (com covariável apenas na função quantílica e com covariável tanto na função quantílica quanto em um parâmetro de escala) foi realizada em um conjunto de dados de gestantes e puérperas de 10 a 55 anos internadas devido a complicações da doença de COVID-19. Ao considerar um modelo de regressão quantílica paramétrico, não só foi possível avaliar o impacto de variáveis de caracterização, sintoma e comorbidade como fatores de prognóstico na sobrevida de mulheres no ciclo gravídico-puerperal e na fração de curados, como também fomos capazes de estudar a influência dessas variáveis nos diferentes parâmetros quantílicos. Dessa maneira, mostramos, por exemplo, como modelos de regressão quantílica que abordam a fração de cura podem se tornar ferramentas úteis para médicos ginecologistas e obstetras reconhecerem e previnirem os reveses causados pelo novo coronavírus em prejuízo para a população materna.

Como trabalho futuro, vale ainda dizer que um segundo estudo de simulação será realizado, porém considerando a geração dos dados a partir do modelo de mistura padrão (ver algoritmo de estimação no Apêndice 6). Nessa situação, repetiremos todos os passos simulatórios desenvolvidos para o modelo defeituoso e avaliaremos, mais uma vez, os três modelos aqui discutidos.

O conjunto de dados e todo código computacional utilizados neste estudo estão disponíveis neste link: [https://github.com/ornscar/mono\\_sobrev\\_reg\\_quantilica.git](https://github.com/ornscar/mono_sobrev_reg_quantilica.git).

## Agradecimentos

À professora, orientadora e amiga, não necessariamente nessa ordem, Agatha Sacramento Rodrigues, a melhor pior pessoa do mundo que já conheci. Obrigada por me mostrar os caminhos mais bonitos da Estatística e por ter sido alento até aqui.

## Referências

- BALKA, J.; DESMOND, A. F.; MCNICHOLAS, P. D. Review and implementation of cure models based on first hitting times for wiener processes. *Lifetime data analysis*, Springer, v. 15, n. 2, p. 147–176, 2009. Citado na página 3.
- 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 na página 3.
- BITTENCOURT, S. D. d. A. et al. Adequacy of public maternal care services in brazil. *Reproductive health*, BioMed Central, v. 13, n. 3, p. 257–265, 2016. Citado na página 2.
- 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 na página 3.
- BORGES, P. et al. A cure rate survival model under a hybrid latent activation scheme. *Statistical methods in medical research*, SAGE Publications Sage UK: London, England, v. 25, n. 2, p. 838–856, 2016. Citado na página 3.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. [S.l.]: Editora Blucher, 2006. Citado na página 15.
- 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 na página 3.
- DINIZ, D.; BRITO, L.; RONDON, G. Maternal mortality and the lack of women-centered care in brazil during covid-19: Preliminary findings of a qualitative study. *The Lancet Regional Health–Americas*, Elsevier, v. 10, 2022. Citado na página 2.
- DONG, E.; DU, H.; GARDNER, L. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, Elsevier, v. 20, n. 5, p. 533–534, 2020. Citado na página 2.
- EL-GOHARY, A.; ALSHAMRANI, A.; AL-OTAIBI, A. N. The generalized gompertz distribution. *Applied mathematical modelling*, Elsevier, v. 37, n. 1-2, p. 13–24, 2013. Citado na página 5.
- GELFAND, A. E.; SMITH, A. F.; LEE, T.-M. Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association*, Taylor & Francis, v. 87, n. 418, p. 523–532, 1992. Citado na página 10.
- GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. *Markov chain Monte Carlo in practice*. [S.l.]: CRC press, 1995. Citado na página 10.
- GUPTA, C. et al. Semiparametric bayesian estimation of quantile function for breast cancer survival data with cured fraction. *Biometrical Journal*, Wiley Online Library, v. 58, n. 5, p. 1164–1177, 2016. Citado 2 vezes nas páginas 3 e 4.
- HONORE, B.; KHAN, S.; POWELL, J. L. Quantile regression under random censoring. *Journal of econometrics*, Elsevier, v. 109, n. 1, p. 67–105, 2002. Citado na página 4.
- HUGHES, L. Coronavirus (covid-19) and pregnancy: what maternal-fetal medicine subspecialists need to know 4.11. 20 (this is an update from draft posted on 3-17-2020). *Washington, DC: The Society for Maternal-Fetal Medicine*, 2020. Citado na página 2.
- LECHIEN, J. R. et al. Prevalence and 6-month recovery of olfactory dysfunction: a multicentre study of 1363 covid-19 patients. *Journal of Internal Medicine*, Wiley Online Library, v. 290, n. 2, p. 451–461, 2021. Citado na página 16.

- MAZA-ARNEDO, F. et al. Maternal mortality linked to covid-19 in latin america: Results from a multi-country collaborative database of 447 deaths. *The Lancet Regional Health-Americas*, Elsevier, v. 12, p. 100269, 2022. Citado na página 2.
- NARISSETTY, N.; KOENKER, R. Censored quantile regression survival models with a cure proportion. *Journal of Econometrics*, Elsevier, v. 226, n. 1, p. 192–203, 2022. Citado na página 4.
- NOUFAILY, A.; JONES, M. Parametric quantile regression based on the generalized gamma distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Wiley Online Library, v. 62, n. 5, p. 723–740, 2013. Citado na página 4.
- OPAS. *Folha informativa sobre COVID-19*. 2022. Disponível em: <<https://www.paho.org/pt/covid19>>. Acesso em: 27/07/2022. Citado na página 2.
- PAULINO, C. D. et al. *Estatística Bayesiana*. 2a. ed. Lisboa: Fundação Calouste Gulbenkian, 2018. Citado na página 10.
- PENG, L. Quantile regression for survival data. *Annual review of statistics and its application*, NIH Public Access, v. 8, n. 1, p. 413, 2021. Citado na página 3.
- PENG, L.; HUANG, Y. Survival analysis with quantile regression models. *Journal of the American Statistical Association*, Taylor & Francis, v. 103, n. 482, p. 637–649, 2008. Citado na página 4.
- PORTNOY, S. Censored regression quantiles. *Journal of the American Statistical Association*, Taylor & Francis, v. 98, n. 464, p. 1001–1012, 2003. Citado na página 4.
- PRENTICE, R. L. Linear rank tests with right censored data. *Biometrika*, Oxford University Press, v. 65, n. 1, p. 167–179, 1978. Citado na página 3.
- RODRIGUES, A.; BORGES, P.; SANTOS, B. A defective cure rate quantile regression model for male breast cancer data. *arXiv preprint arXiv:2105.03699*, 2021. Citado 4 vezes nas páginas 4, 6, 11 e 25.
- SANTOS, B. R. d. *Modelos de regressão quantílica*. Tese (Doutorado) — Universidade de São Paulo, 2012. Citado na página 4.
- SCUDILIO, J. et al. Defective models induced by gamma frailty term for survival data with cured fraction. *Journal of Applied Statistics*, Taylor & Francis, v. 46, n. 3, p. 484–507, 2019. Citado 2 vezes nas páginas 3 e 7.
- SILVA, R. d. A. *Modelos de fração de cura com fatores latentes competitivos e fragilidade*. Tese (Doutorado) — Universidade de São Paulo, 2011. Citado na página 3.
- Statisticat; LLC. *LaplacesDemon: Complete Environment for Bayesian Inference*. [S.l.], 2021. R package version 16.1.6. Disponível em: <<https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software>>. Citado na página 9.
- TEAM, R. C. R: A language and environment for statistical computing. Vienna, Austria, Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>. Citado na página 15.
- 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 na página 3.
- UNFPA Brasil. *A razão da mortalidade materna no Brasil aumentou 94% durante a pandemia. Fundo de População da ONU alerta para grave retrocesso*. 2022. Disponível em: <<https://brazil.unfpa.org/pt-br/news/razao-da-mortalidade-materna-no-brasil-aumentou-94-durante-pandemia-fundo-de-populacao-da-onu>>. Acesso em: 04/12/2022. Citado na página 2.
- VELTEN, F. B. Modelo de regressão quantílica weibull exponenciada com fração de cura. 2021. Citado na página 4.

- WHO, G. Statement on the second meeting of the international health regulations (2005) emergency committee regarding the outbreak of novel coronavirus (2019-ncov). *World Health Organization*, 2020. Citado na página [2](#).
- WHO, W. H. O. Rolling updates on coronavirus disease (covid-19). WHO Geneve, Switzerland, 2020. Citado na página [2](#).
- WU, Y.; YIN, G. Cure rate quantile regression accommodating both finite and infinite survival times. *Canadian Journal of Statistics*, Wiley Online Library, v. 45, n. 1, p. 29–43, 2017. Citado na página [4](#).
- XUE, X.; XIE, X.; STRICKLER, H. D. A censored quantile regression approach for the analysis of time to event data. *Statistical methods in medical research*, SAGE Publications Sage UK: London, England, v. 27, n. 3, p. 955–965, 2018. Citado na página [4](#).
- YIN, G.; ZENG, D.; LI, H. Power-transformed linear quantile regression with censored data. *Journal of the American Statistical Association*, Taylor & Francis, v. 103, n. 483, p. 1214–1224, 2008. Citado na página [4](#).
- YING, Z.; JUNG, S.-H.; WEI, L.-J. Survival analysis with median regression models. *Journal of the American Statistical Association*, Taylor & Francis, v. 90, n. 429, p. 178–184, 1995. Citado na página [4](#).



## Apêndice A

Gráficos suplementares da Seção 4.

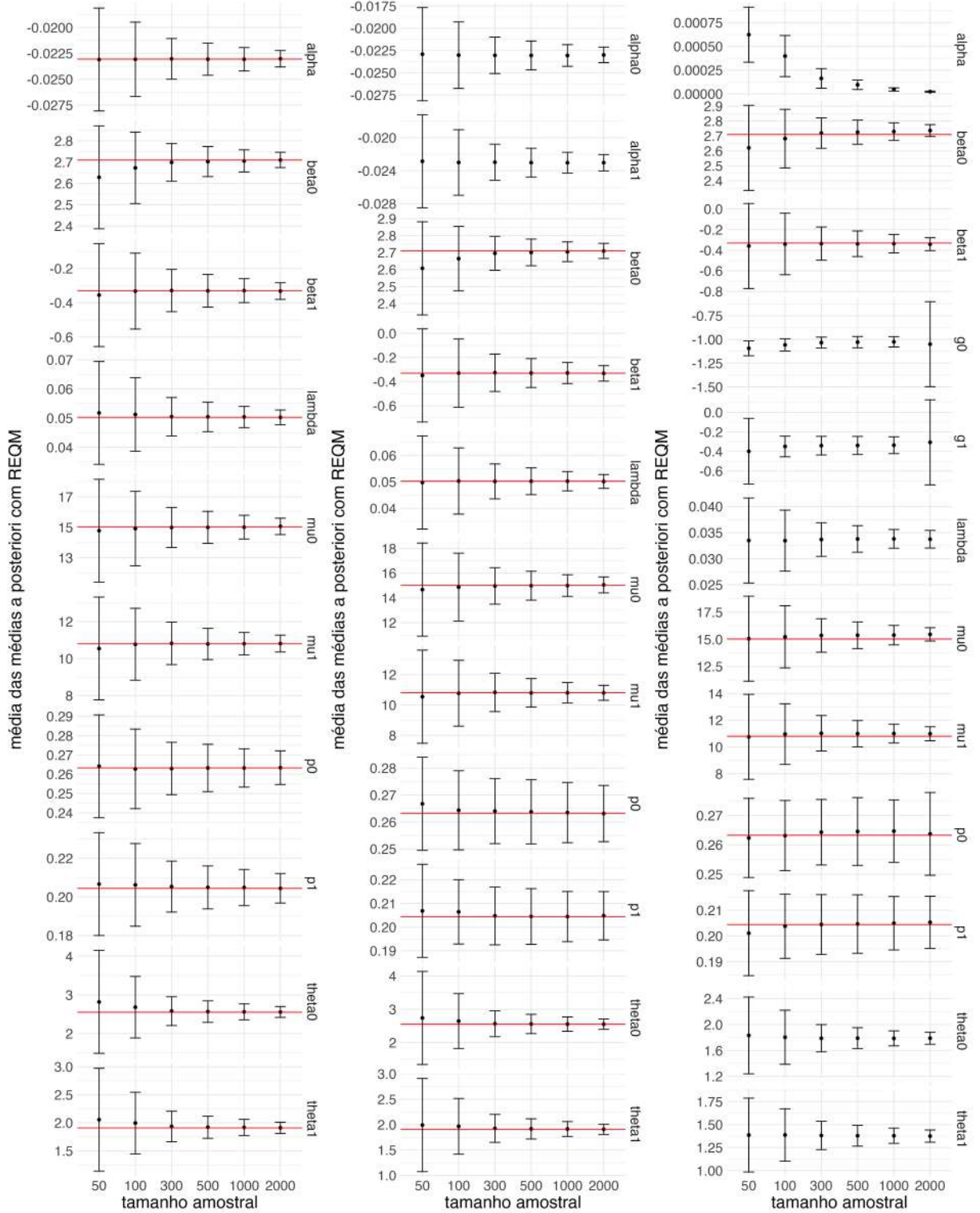


Figura 12 – Médias a posteriori, representadas pelos pontos, do modelo (a) defeituoso só com covariável em  $\mu_q^1$ , (b) defeituoso com covariável em  $\mu_q^1$  e  $\alpha$  e (c) de mistura padrão para as diferentes combinações de valores de parâmetros e tamanhos de amostra  $n = 50; 100; 300; 500; 1.000; 2.000$ , quando  $q = 0, 2$ . A barra ao redor do ponto retrata o REQM, enquanto a linha vermelha representa o valor verdadeiro do parâmetro.



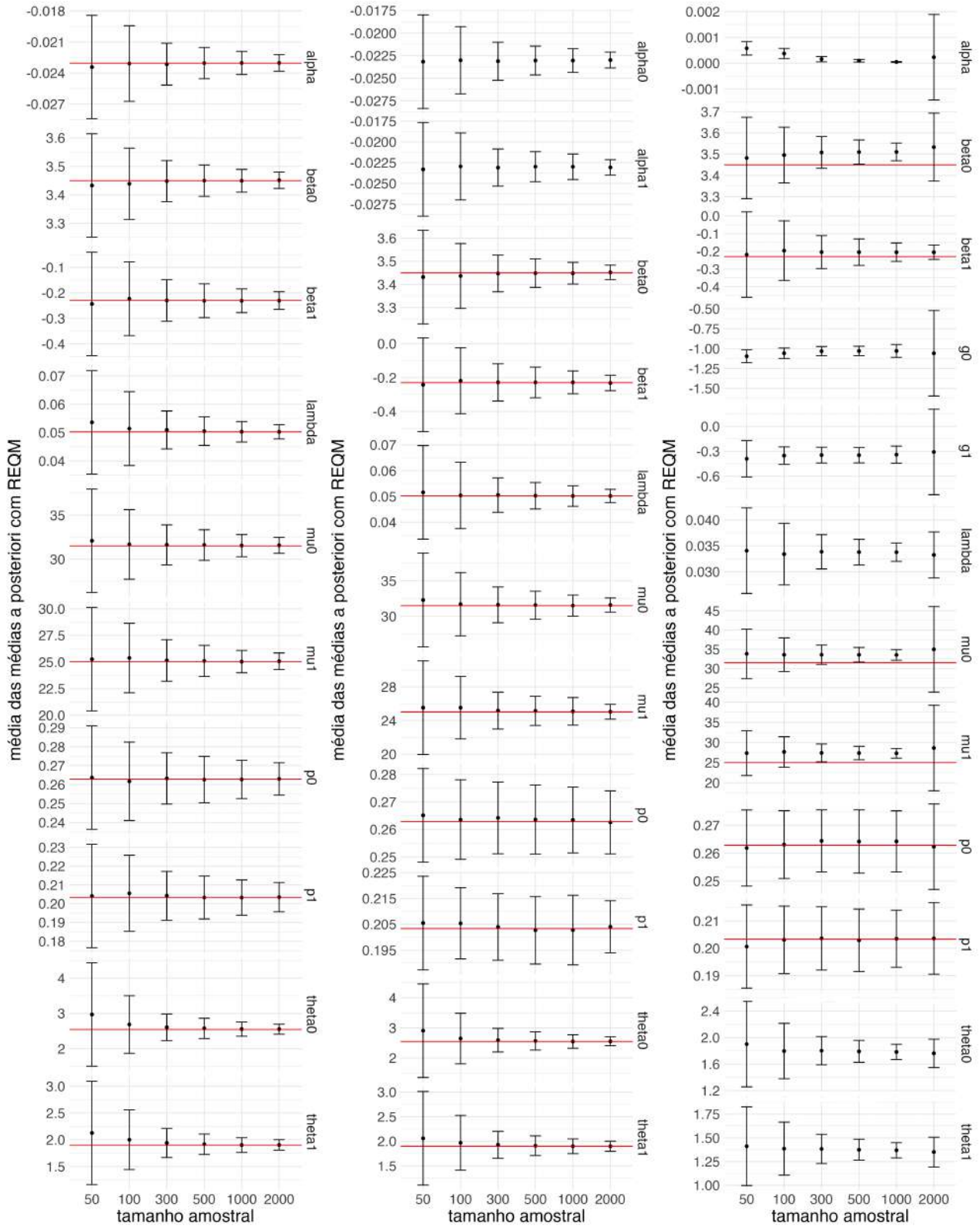


Figura 13 – Médias a posteriori, representadas pelos pontos, do modelo (a) defeituoso só com covariável em  $\mu_q^1$ , (b) defeituoso com covariável em  $\mu_q^1$  e  $\alpha$  e (c) de mistura padrão para as diferentes combinações de valores de parâmetros e tamanhos de amostra  $n = 50; 100; 300; 500; 1.000; 2.000$ , quando  $q = 0,5$ . A barra ao redor do ponto retrata o REQM, enquanto a linha vermelha representa o valor verdadeiro do parâmetro.

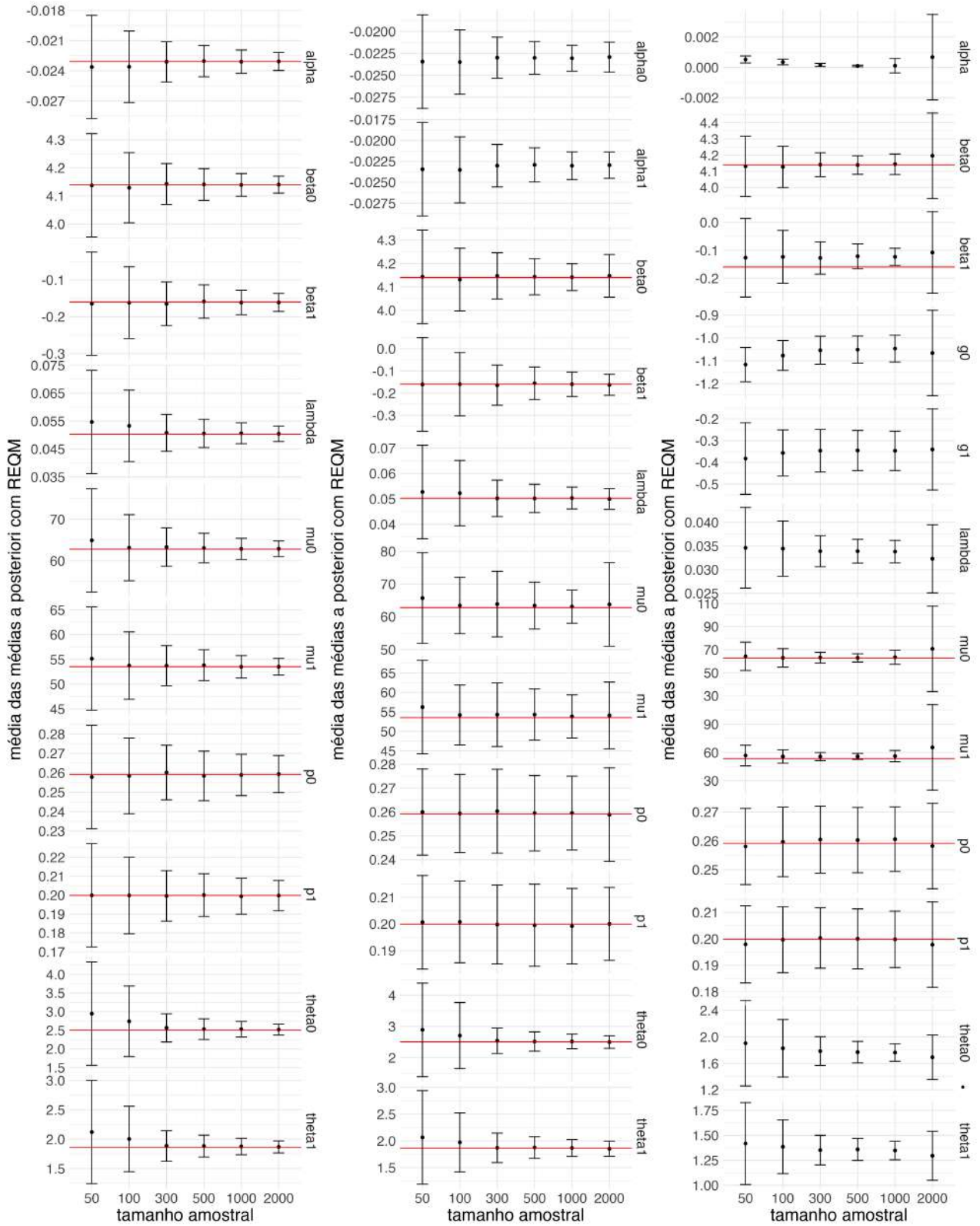


Figura 14 – Médias a posteriori, representadas pelos pontos, do modelo (a) defeituoso só com covariável em  $\mu_q^1$ , (b) defeituoso com covariável em  $\mu_q^1$  e  $\alpha$  e (c) de mistura padrão para as diferentes combinações de valores de parâmetros e tamanhos de amostra  $n = 50; 100; 300; 500; 1.000; 2.000$ , quando  $q = 0,8$ . A barra ao redor do ponto retrata o REQM, enquanto a linha vermelha representa o valor verdadeiro do parâmetro.

## Apêndice B

Algoritmo para gerar dados a partir do modelo de mistura padrão.

---

**Algoritmo 2** Algoritmo de geração de dados pelo modelo mistura padrão

---

- 1: Determine valores desejáveis para  $\alpha$ ,  $\lambda$ ,  $q$ ,  $\beta = (\beta_0, \beta_1)^\top$  e  $\gamma = (\gamma_0, \gamma_1)^\top$ ;
- 2: Defina a proporção de dados censurados, dado por  $pc_0$  e  $pc_1$ , para  $x = 0$  e  $x = 1$ , respectivamente;
- 3: Para a  $i$ -ésima observação, gere  $x_i \sim \text{Bernoulli}(0.5)$ , e calcule  $p_{0x_i}$ , em que  $p_{00}$  e  $p_{01}$  são dados por (5) quando  $x_i = 0$  e  $x_i = 1$ , respectivamente;
- 4: Gere  $u_i \sim \text{Uniforme}(0, 1)$ . Se  $u_i < p_{0x_i}$ , faça  $w_i = \infty$ ; caso contrário, gere  $u_{1i} \sim U(0, 1 - p_{0x_i})$  e calcule

$$w_i = \frac{1}{\alpha} \log \left[ 1 - \frac{\alpha}{\lambda} \log \left( 1 - u_{1i}^{(1/\theta_{x_i})} \right) \right], \text{ com } x_i = 1 \text{ ou } x_i = 0,$$

e  $\theta_{x_i}$  é dado na Equação (3.2) com  $\mu_q = \exp(\beta_0)$ , se  $x_i = 0$  e  $\mu_q = \exp(\beta_0 + \beta_1)$ , se  $x_i = 1$ .

- 5: Gere  $c_i \sim U(0, \tau_i)$ , em que  $\tau_i$  é definido para ter aproximadamente  $p_{cx_i}$  proporção de dados censurados.
  - 6: Determine  $t_i = \min(w_i, c_i)$ . Se  $t_i = w_i$ , faça  $\delta_i = 1$ , caso contrário  $\delta_i = 0$ ;
  - 7: Repita os passos 3 a 6 para todo  $i = 1, \dots, n$ . Os dados para o  $i$ -ésimo indivíduo é  $\{t_i, x_i, \delta_i\}$ ,  $i = 1, \dots, n$ .
-