

Anna Carolina Lustosa Lima

*Modelagem Conjunta de Dados Longitudinais  
e de Sobrevivência*

Belo Horizonte – MG

Outubro, 2007

Anna Carolina Lustosa Lima

*Modelagem Conjunta de Dados Longitudinais  
e de Sobrevivência*

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Mestre em Estatística.

Orientador: Prof. Enrico Antônio Colosimo

MESTRADO EM ESTATÍSTICA  
DEPARTAMENTO DE ESTATÍSTICA  
INSTITUTO DE CIÊNCIAS EXATAS  
UNIVERSIDADE FEDERAL DE MINAS GERAIS

Belo Horizonte – MG  
Outubro, 2007.

# *Sumário*

<b>1. Introdução .....</b>	<b>1</b>
<b>2. Análise de Dados Longitudinais .....</b>	<b>4</b>
2.1 Modelo Marginal .....	4
2.2 Modelo de Efeitos Aleatórios .....	5
2.3 Inferência para os Modelos Marginal e de Efeitos Aleatórios .....	5
2.4 Modelo Linear Generalizado com Efeitos Aleatórios .....	6
2.5 Inferência para o Modelo Generalizado .....	7
<b>3. Análise de Sobrevivência .....</b>	<b>8</b>
3.1 Modelo de Riscos Proporcionais de Cox .....	8
3.2 Modelo de Cox com Covariáveis Dependentes do Tempo .....	9
<b>4. Modelagem Conjunta .....</b>	<b>10</b>
4.1 O Modelo Conjunto por Wulfsohn e Tsiatis (1997) .....	10
4.2 O Modelo Conjunto por Henderson, Diggle e Dobson (2000) .....	11
4.3 O Modelo Conjunto por Diggle, Souza e Chetwynd (2007) .....	12
4.4 O Modelo em dois estágios .....	13
<b>5. Aplicação .....</b>	<b>14</b>
5.1 Descrição do experimento .....	14
5.2 Análise de dados .....	15
5.2.1 Análise de Dados Longitudinais .....	16
5.2.2 Análise de Sobrevivência .....	21
5.3 Modelo de Cox com covariável dependente do tempo.....	22
5.4 Modelo em dois estágios .....	24
<b>6. Conclusões .....</b>	<b>26</b>
<b>7. Referências Bibliográficas .....</b>	<b>27</b>
<b>ANEXOS</b>	
Anexo A .....	29
Anexo B .....	30

Papai do Céu, Obrigada!

Por tantas e tão ricas bênçãos!

Pai Generoso que junto a cada desafio me ofereceu a força, a luz e a coragem para a conquista.

Por cada uma das pessoas maravilhosas que encontrei em meu caminho e fazem parte de quem sou e de mais esta conquista. Muito sou grata pelo aperfeiçoamento de minha mente e de minha alma dos quais elas foram veículo e peças fundamentais. Peço, Meu Pai, que derrame suas bênçãos e espalhe sorrisos e sorte pela vida de cada uma delas:

Enrico, querido mestre e amigo, tantos ensinamentos, tantas parcerias, tantas conversas.... antes eu não pensava que chegaria tão longe, hoje quero saber qual o próximo passo. Seu exemplo, sua conduta, sua inteligência e capacidade muitas vezes foram meus combustíveis para essa caminhada. Meu sincero e eterno agradecimento!

Marília Sá (FIOCRUZ) e Marcel (UFJF) que aceitaram o convite de participar da banca de qualificação e demonstrando um enorme interesse pelo assunto planejaram e organizaram, juntamente com o Enrico, o WorkShop em Análise de dados Longitudinais realizado na UFJF. Uma oportunidade única que possibilitou o aprimoramento deste trabalho e meu crescimento pessoal e profissional!

Rosemeire (UFBA) que colaborou tão fortemente com este trabalho, debatendo-o no WorkShop de Juiz de Fora, sugerindo novas propostas e formas de modelagem e generosamente intermediou o meu contato com Inês Souza, uma das autoras das principais referências citadas.

Valeska (FIOCRUZ) que aceitou o convite de debater este trabalho no WorkShop e fez sugestões relevantes que contribuíram muitíssimo para a modelagem dos dados.

Dr. Roberto que enquanto diretor do Centro de Pesquisas René Rachou (CPqRR) possibilitou a conciliação deste mestrado com as atividades profissionais tão essenciais para a formação da profissional que sou hoje. Dr. Álvaro Romanha, atual diretor do CPqRR que manteve as condições cedidas pela direção anterior, incentivando o meu crescimento e possibilitando a conclusão deste trabalho.

Amigos do CPqRR, conquistas preciosas de quatro anos de trabalho e dedicação. Em especial, meus colegas de sala: Andréa, Beth Fleury e Aílson. Obrigada, meus queridos, pelo agradável, divertido e harmonioso convívio diário! Martin, Ana Karine, Ana Carolina, Célia, Anne, Cíntia, Paula, Fernando, Cristiano, Virgínia, Ivanete, Solange, obrigada pela amizade, fruto de tantas horas em comum de parceria no trabalho da pesquisa e aprendizagem da estatística! Segemar, sempre solícito, conseguiu os artigos mais impossíveis sempre que precisei, muito obrigada!

Bráulio, Pedro, Felipe, Renato, Henrique, Gustavo, João, Rodrigo, Amigos do ICEX que foram a dose diária de frescor para a mente. Um pouco de amenidade no meio da correria....

Joria, uma jóia rara, amiga enviada de São Luiz para colorir o dia e me fazer companhia neste mundo da estatística.

Luciano, Renata, Michele, Paula colegas de mestrado que fizeram muita diferença.

Elias que tanto me ajudou na programação de diversos dos modelos apresentados neste trabalho com sua incrível habilidade com o R.

“Qualquer amor já é um pouquinho de saúde, um descanso na loucura.”

Mamãe, presente todo o tempo em minha alma, você que me deu coragem a me candidatar para o mestrado, você que acreditou em mim o tempo todo e desembarçou cada nó que eu não consegui. Você, minha mãe, é meu exemplo maior de vida, tanto a admiro... Eternos são minha gratidão, amizade, respeito e Amor por você!

Tio Virgílio e Vovó Mainha, pessoas de ouro colocadas em minha vida para ensinar e iluminar! Suas preces acalmaram meu coração e me devolveram a razão nas horas de dificuldade. Sou grata por existirem. Amo vocês!

Ni, linda ermã, seus olhos lêem minha alma e me acalmam... Um dia disseram que nosso amor era grande demais, diferente de tudo, indescritível. Eles têm razão! Impossível colocar em palavras.... A sintonia, o carinho, a confiança, o respeito.... é Amor demais!!!

Vi, meu irmão. Sempre foi para mim como um pai, nas palavras, nos gestos, nos cuidados... Referência de proteção e conforto! Obrigada por esse cantinho seguro que me reservaste em seu coração! Te amo!

Lucas, minha Luz! Meu amor transborda ao escrever essas linhas...! Você que acompanhou cada passo tão de perto, que escutou cada novo parágrafo escrito, que me tirou de tantas baixas ao longo do caminho e que sorriu comigo a cada nova conquista! A você, Minha Vida, dedico minha gratidão, meu carinho, minha admiração, meu respeito, meu eterno e verdadeiro Amor!!!

Lu, chegou de mansinho, conquistou um lugar todo dela! Obrigada pelo doce “Pequena”, pelo carinho, pela preocupação, pela amizade e pelos Êhhhh!!! Você é muito especial!!!

Ale, essa piupiu é pau pra toda obra! Obrigada por abraçar essa família dessa forma!!!

Thays, amiga de verdade, pra vida toda! Obrigada pela lealdade que sempre pude contar, por me escutar e por me deixar te escutar! Te amo!

Heitor, pai zeloso, que tantas noites estudou comigo e me orientou quando meus passos ainda eram pequenos no mundo da matemática. Obrigada!

Liléia, mãe que adotei em meu coração, é um amar que não se explica! Obrigada pelo incentivo, pelas portas abertas, pela confiança e carinho.

Minha Estrela da Sorte, sou imensamente grata por tantos presentes.... por colorir minha vida deste verde dourado.... e por estar Sempre presente!!!!

Mãe Pomba, mãe maior, guia e conselheira, obrigada por me fazer a cada dia uma pessoa melhor!

Aos sete amigos, do grupo dos 7, da Tenda da Mãe Pomba, agradeço a proteção, o carinho, os conselhos, a força! Salve!

Salve a força da Cabocla Jurema! Salve a força que dá Vida! Salve, salve minha mãe!

Pelas belezas encontradas em cada sol, em cada flor, em cada sorriso....

Obrigada, Meu Pai, por tornar tão suave e prazerosa a caminhada.

Amém.

# Modelagem Conjunta de Dados Longitudinais e de Sobrevivência

## 1. Introdução

Em estudos envolvendo análise de dados longitudinais a variável resposta é medida repetidamente, ao longo de um intervalo de tempo, para cada indivíduo. Nos estudos de análise de sobrevivência a variável resposta é o tempo até a ocorrência de um evento. Muitos estudos clínicos geram conjuntamente dados longitudinais e de sobrevivência. Por exemplo, nos estudos prospectivos de análise de sobrevivência em que, grupos de pacientes são acompanhados ao longo do tempo até a ocorrência de óbito, é comum haver visitas de acompanhamento quando são medidas variáveis relacionadas à progressão da doença. Entretanto, tratar estas medidas como covariáveis dependentes do tempo no processo de sobrevivência nem sempre é satisfatório uma vez que há ocasiões cujo interesse do pesquisador é avaliar o efeito de tratamentos tanto sobre os tempos de sobrevivência quanto sobre o marcador longitudinal. Este estudo mostra também que a inclusão do marcador longitudinal como covariável no processo de sobrevivência pode ocasionar a perda de importância de covariáveis medidas na baseline e possivelmente associadas ao tratamento.

Existem modelos bem estabelecidos na literatura para analisar estes dados separadamente, mas esta forma de abordagem pode ser inadequada se a variável longitudinal for correlacionada com o tempo de sobrevivência. Por exemplo, Henderson, Diggle e Dobson (2000) mostraram que, em detrimento do modelo conjunto, utilizar os modelos separados para o processo longitudinal e de sobrevivência acarreta um vício severo na estimação de alguns componentes de variância e no coeficiente de tendência temporal. Guo e Carlin (2004) desenvolveram uma aproximação Bayesiana para o modelo conjunto e mostraram haver significativa melhora nas estimativas dos tempos de sobrevivência e de outros parâmetros.

O presente estudo pretende comparar as abordagens, separada e conjunta, aplicada ao banco de dados de um experimento biológico envolvendo camundongos, realizado com o objetivo de avaliar o efeito da co-infecção de malária e esquistossomose e da imunização induzida para malária, sobre os tempos de sobrevivência e nível de parasitemia (malária) dos camundongos. Neste experimento, 64 camundongos foram distribuídos aleatoriamente em quatro grupos experimentais. Os camundongos alocados no primeiro e segundo grupos foram ambos previamente induzidos à imunização para malária e infectados com malária e malária associada à esquistossomose, respectivamente. Os camundongos alocados nos grupos 3 e 4 não foram imunizados e foram infectados por malária e malária associada à esquistossomose, respectivamente. O pesquisador visitou os camundongos diariamente e registrou o percentual médio de parasitemia no sangue e a ocorrência ou não de morte. Esta avaliação foi realizada num período de 25 dias (do 5º ao 30º dia após a infecção). Observou-se que os camundongos dos

grupos 1 e 2 (ambos imunizados) tiveram melhor prognóstico e níveis de parasitemia baixos, enquanto que nos grupos 3 e 4 foram observados menores tempos de sobrevivência e níveis de parasitemia crescentes.

Em específico pretende-se comparar em termos metodológicos e de interpretação os seguintes modelos:

**M1:** Modelos para dados longitudinais;

**M2:** Modelo de Cox para a resposta de sobrevivência;

**M3:** Modelo de Cox com covariável dependente do tempo (variável longitudinal);

**M4:** Modelo Conjunto para dados longitudinais e de sobrevivência.

A análise conjunta de dados longitudinais de sobrevivência tem sido explorada recentemente por alguns autores. Wulfshon e Tsiatis (1997) desenvolveram uma aproximação para o modelo conjunto que utiliza o modelo de Cox para o processo de sobrevivência e a resposta longitudinal é tratada como covariável contínua cujo valor real é dado por um modelo linear de efeitos aleatórios. A estimação dos parâmetros do modelo proposto é realizada por um algoritmo EM que também foi utilizado por Henderson, Diggle e Dobson (2000). Estes autores formularam um modelo generalizado para a análise conjunta em que um sub-modelo linear de efeitos aleatórios é assumido para os dados longitudinais, um sub-modelo semi-paramétrico de riscos proporcionais para os dados de sobrevivência, e assume que estes dois sub-modelos podem ser ligados por um processo latente normal bivariado. Uma característica interessante do modelo proposto por estes autores é que na ausência de associação entre o marcador longitudinal e os tempos de sobrevivência, os resultados da análise conjunta deverão ser os mesmos que obteríamos a partir da análise separada de cada um dos componentes através de métodos tradicionais.

Outros autores também utilizaram esta forma generalizada para o modelo conjunto. Por exemplo, Wang e Taylor (2001) desenvolveram uma aproximação que considera o processo latente de ligação como um processo estocástico integrado Ornstein-Uhlenbeck. Esta escolha foi motivada pela necessidade de obter um modelo mais flexível para o processo longitudinal. Brown, Ibrahim e DeGruttola (2005) também sugeriram uma alteração no modelo para a resposta longitudinal, eles sugerem o uso do modelo B-Splines Cúbico em lugar de um modelo linear de efeitos aleatórios.

Song, Davidian e Tsiatis (2002) avaliaram o modelo conjunto assumindo apenas que os coeficientes aleatórios pertencem a uma família de densidades suavizadas e não estritamente a distribuição normal. Guo e Carlin (2004) desenvolveram uma aproximação Bayesiana para o modelo conjunto proposto por Henderson, Diggle e Dobson (2000).

Tseng, Hsieh e Wang (2005) abordaram o caso em que a suposição de riscos proporcionais no modelo de Cox não é verificada e considera o modelo de tempo de tempos de falha acelerado como alternativa para a modelagem dos dados de eventos.



Diggle, Souza e Chetwynd (2007) desenvolveram uma aproximação completamente paramétrica que considera uma transformação logarítmica para o tempo de sobrevivência, e que a distribuição conjunta destes tempos de sobrevivência transformados e as medidas longitudinais é normal multivariada.

Ratcliffe et. al. (2004) propuseram um modelo conjunto para dados em conglomerados em que os dois sub-modelos são ligados através das fragilidades no nível do grupo.

Zheng e Heagerty (2005) e Xu e Zeger (2001) desenvolveram aproximações de modelagem conjunta mas, com o foco na estimação dos parâmetros do processo de sobrevivência.

Este trabalho foi estruturado da seguinte forma. A Seção 2 aborda a metodologia de análise de dados longitudinais e apresenta os modelos marginal, de efeitos aleatórios e o modelo linear generalizado com efeitos aleatórios. A Seção 3 apresenta de forma resumida os conceitos de análise de sobrevivência e o modelo de Cox (1972) com e sem covariáveis dependentes do tempo. Alguns modelos conjuntos e algumas das suas possíveis formas de estimação são apresentados na Seção 4. A descrição detalhada do experimento, a modelagem dos dados utilizando os métodos tradicionais de análise separada, e a análise através do modelo em dois estágios estão apresentadas na Seção 5. A Seção 6 discorre sobre as principais conclusões deste trabalho.

## 2. Análise de Dados Longitudinais

Esta seção introduz os conceitos associados à análise de dados longitudinais, a notação utilizada na literatura e apresenta os modelos marginal, de efeitos aleatórios e o modelo linear generalizado com efeitos aleatórios. Também conhecido como **medidas repetidas**, este tipo de estudo envolve bancos de dados em que a variável resposta,  $Y$ , é medida repetidamente em cada indivíduo da amostra, ao longo do tempo. Suponha um conjunto de  $m$  indivíduos avaliados  $n$  vezes ao longo do tempo, por simplicidade de notação assumiremos que o número de observações da medida longitudinal é constante para todos os indivíduos, embora os métodos sejam igualmente aplicáveis a dados desbalanceados que ocorrem com frequência na prática.  $Y_i^T = (Y_{i1}, \dots, Y_{in})$  é o vetor de respostas do  $i$ -ésimo indivíduo, então  $Y_{ij}$  representa a resposta do  $i$ -ésimo indivíduo no  $j$ -ésimo tempo.

Em análise de dados longitudinais assumimos que, entre indivíduos, as respostas são independentes, mas para o mesmo indivíduo elas são correlacionadas.

Diggle, Liang e Zeger (1996) mostraram que, ao modelar dados longitudinais pelo modelo linear normal as estimativas dos coeficientes de regressão,  $\hat{\beta}'s$ , permanecem consistentes porém, estas são ineficientes uma vez que a variância dos coeficientes de regressão estimados,  $Var(\hat{\beta})$ , fica viciada.

Uma forma de corrigir este problema é através da estimação dos componentes de variância ou seja, modelar a estrutura de covariância dos dados. O modelo marginal e o modelo de efeitos aleatórios apresentam duas formas de estimação destes componentes e serão apresentados a seguir.

### 2.1 Modelo Marginal

No modelo marginal a média e a variância de  $Y$  são modeladas separadamente. Assumimos que  $X_i$  são as covariáveis incluídas na análise e  $\beta$ 's seus respectivos coeficientes de regressão, a resposta,  $Y_i$ , é resultado de um processo Gaussiano  $n$ -variado com média  $(X_i^T \beta)$  e matriz de covariância  $V_i$ . A variância de  $V_i$  incorpora a estrutura de correlação das medidas do mesmo indivíduo. Portanto, é necessário atribuir forma específica à  $V_i$ .

As formas mais comuns de  $V_i$  são: **Correlação Uniforme** neste caso, assumimos que a correlação entre as medidas do mesmo indivíduo é constante; **Correlação Serial**, esta forma assume que a correlação diminui quando as medidas se afastam no tempo; e, **Não-estruturada** neste caso, é especificada uma covariância diferente para cada par de observações do indivíduo. A matriz de correlação dos resíduos padronizados, obtidos pelo ajuste do modelo linear normal para a resposta

longitudinal versus o tempo, desconsiderando as demais variáveis, é útil para determinar a estrutura de correlação das medidas do mesmo indivíduo (Diggle, Liang e Zeger, 1996).

O modelo marginal pode ser escrito como:

$$Y_i = X_i^T \beta + \varepsilon_i, \quad i = 1, 2, \dots, m$$

em que,  $Y_i \sim N_n(X_i^T \beta, V_i)$ , e  $\varepsilon_i \sim N(0, V_i)$ .

## 2.2 Modelo de Efeitos Aleatórios

Neste modelo a resposta é assumida ser uma função linear das variáveis explicativas com coeficientes de regressão que variam entre indivíduos, isto é, este modelo permite que os coeficientes sejam fixos ou aleatórios. Os coeficientes aleatórios explicam a dependência entre as respostas do mesmo indivíduo, esta variabilidade reflete a heterogeneidade devido a fatores não mensurados ou não mensuráveis.

A premissa básica do modelo de efeitos aleatórios é que a heterogeneidade natural entre indivíduos é modelada em um subconjunto de coeficientes de regressão. O modelo de efeitos aleatórios assume a seguinte forma:

$$Y_i = X_i^T \beta + Z_i b_i + \varepsilon_i, \quad i = 1, 2, \dots, m \quad (1)$$

em que,  $X_i^T \beta$ : representa o **componente fixo**;  $X_i$  são as covariáveis incluídas na análise e  $\beta$ 's seus respectivos coeficientes de regressão,  $Z_i b_i$ : representa o **componente aleatório**;  $b_i$ 's são os coeficientes aleatórios que variam para cada indivíduo; e,  $Z_i$  é, em geral, um subconjunto de  $X_i$  de covariáveis dependentes do tempo.  $\varepsilon_i \sim N(0, \sigma^2 I_n)$ .

No modelo de efeitos aleatórios assumimos que as respostas do  $i$ -ésimo indivíduo têm distribuição Normal  $n$ -variada com média  $(X_i^T \beta)$  e matriz de covariância  $\Sigma = (Z_i V_i Z_i^T + \sigma^2 I)$  em que  $V_i$  é a matriz de covariâncias das medidas do mesmo indivíduo. Os coeficientes aleatórios,  $b_i$ , são independentes e têm distribuição Normal com média zero e matriz de covariância  $V_i$ .

## 2.3 Inferência para os Modelos Marginal e de Efeitos Aleatórios

Os coeficientes de regressão do modelo marginal e do modelo de efeitos aleatórios podem ser estimados através do método da máxima verossimilhança. Entretanto, este método produz um estimador viciado para os componentes de variância (Diggle, Liang e Zeger, 1996).

O método da máxima verossimilhança restrita corrige este problema e é dada por:

$$L_R(\theta) = |X^T H^{-1} X|^{-1/2} \cdot L_v(\theta)$$

em que,  $\theta$  representa o conjunto de todos os parâmetros a serem estimados no modelo,  $X$  é a matriz das covariáveis incluídas na análise,  $H = \sigma^2 I$  e  $L_v(\theta)$  é a função de verossimilhança usual:

$$L_v(\theta) = \prod_{i=1}^m (2\pi)^{-n/2} |V_i|^{-1/2} (\sigma^2)^{-n/2} \cdot \exp\left\{-\frac{1}{2\sigma^2} (Y_i - X_i \beta)^T V_i^{-1} (Y_i - X_i \beta)\right\}$$

O estimador de máxima verossimilhança restrita possui as mesmas propriedades que o estimador de máxima verossimilhança. Maiores detalhes sobre a este método podem ser encontrados em Diggle, Liang e Zeger, 1996.

## 2.4 Modelo Linear Generalizado com Efeitos Aleatórios

Um modelo linear generalizado com efeitos aleatórios é definido a seguir. Seja  $Y_i^T = [Y_{i1}, \dots, Y_{im}]$  o vetor de  $n$  observações da variável resposta que pode ser escrito como:

$$Y_i = \mu_i + \varepsilon_i, \quad i = 1, 2, \dots, m \quad (2)$$

Seja  $g(\cdot)$  uma função monótona, um link, tal que  $g(\mu)$  possa ser escrita como um modelo linear:

$$g(\mu_i) = \eta_i = X_i \beta + Z_i b_i \quad (3)$$

Se, condicional a  $\mu_i$ , os componentes de  $Y_i$  são independentemente distribuídos, e se sua distribuição pertence à família exponencial, então (2) e (3) definem um modelo linear generalizado com efeitos aleatórios.

Assumimos que os coeficientes aleatórios  $b_i$  são não correlacionados, com esperança zero e  $\text{cov}(b_i) = \sigma_i^2 I_{q_i}$ , em que  $q_i$  é o número de coeficientes aleatórios do  $i$ -ésimo indivíduo. Estes coeficientes são independentes do vetor de erros,  $\varepsilon_i$ .

No caso particular quando  $Y_i$  segue uma distribuição normal e  $g(\cdot)$  é uma função link identidade, o modelo linear generalizado se reduz ao modelo linear de efeitos aleatórios (1) apresentado na Seção 2.2.

A análise de um modelo linear generalizado com efeitos aleatórios pode ser aplicada em modelos logísticos para respostas longitudinais dicotômicas. Neste caso, assumimos que os efeitos aleatórios seguem uma distribuição normal com média zero e a distribuição condicional de  $Y$  é binomial. Schall (1991) apresenta este modelo aplicado a um banco de dados com resposta binária e apresenta detalhadamente o algoritmo de estimação.

## 2.5 Inferência para o Modelo Generalizado

A estimação dos parâmetros,  $\beta$ , e dos parâmetros especificando a distribuição dos efeitos aleatórios,  $b$ , podem ser realizadas através do método da máxima verossimilhança restrita (Schall, 1991) baseada na distribuição marginal das observações,  $Y$ . Schall (1991), descreve dois algoritmos para a estimação destes modelos. O primeiro é aplicado ao caso particular em que a resposta  $Y$  tem distribuição Normal e o segundo para o caso generalizado.

### 3. Análise de Sobrevivência

Nesta seção introduziremos os conceitos básicos de análise de sobrevivência e a notação utilizada na literatura. Os modelos de Cox, com e sem covariável dependente do tempo, e suas formas de estimação são também apresentados nesta seção.

Suponha um conjunto de  $m$  indivíduos distribuídos aleatoriamente em grupos de tratamento,  $X_i$  é a variável indicadora de grupo. Seja  $T_i$  o tempo máximo de acompanhamento do  $i$ -ésimo indivíduo. Este tempo pode ser classificado de duas formas:  $T_i$  é um **tempo de falha** se indica o tempo em que o evento de interesse foi observado; ou,  $T_i$  é um **tempo de censura** se ao final do acompanhamento o evento não ocorreu.

De forma geral, o vetor  $(T_i, \delta_i, X_i)$  representa os dados de sobrevivência do  $i$ -ésimo indivíduo em que  $T_i$  é o tempo máximo de acompanhamento,  $\delta_i$  uma função indicadora de falha e  $X_i$  representa as covariáveis associadas a este indivíduo.

A taxa de falha no intervalo  $[t, t+\Delta t)$  é definida como a probabilidade de que a falha ocorra neste intervalo, dado que não ocorreu antes de  $t$ , dividida pelo comprimento do intervalo. Desta forma, se  $\Delta t \rightarrow 0$ ,  $\lambda(t)$  representa a taxa de falha instantânea no tempo  $t$ , condicional à sobrevivência até o tempo  $t$ , e é expressa por:

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

Modelos tanto paramétricos quanto semi-paramétricos estão disponíveis para modelar dados de sobrevivência (Colosimo e Giolo, 2006). Os modelos paramétricos Exponencial e Weibull são comumente usados por sua simplicidade e pela fácil interpretação de seus parâmetros. Na prática, entretanto, os modelos semi-paramétricos de riscos proporcionais (Cox, 1972) são amplamente usados pois, eles não impõe uma forma particular para a curva de sobrevivência.

#### 3.1 Modelo de Riscos Proporcionais de Cox

No modelo semi-paramétrico de riscos proporcionais de Cox a função de risco assume a forma:

$$\lambda_i(t) = \lambda_0(t) \cdot \exp\{X_i^T \beta\} \quad (4)$$

em que,  $\lambda_0(t)$ , a função de taxa de falha básica e o vetor  $X_i = (X_{1i}, X_{2i}, \dots, X_{pi})^T$  e  $\beta$  representam as  $p$  covariáveis medidas no  $i$ -ésimo indivíduo e seus correspondentes coeficientes de regressão.

O modelo de Cox, sem covariável dependente do tempo (4), assume que a razão das taxas de falha de dois indivíduos distintos na amostra é constante para todo o tempo de acompanhamento.

A estimação dos parâmetros deste modelo é baseada no **método da máxima verossimilhança parcial** (Collet, 2003; Cox, 1975):

$$L(\beta) = \prod_{i=1}^m \left( \frac{\exp\{X_i^T \beta\}}{\sum_{j \in R(t_i)} \exp\{X_j^T \beta\}} \right)^{\delta_i}, \quad (5)$$

em que,  $R(t_i)$  é o conjunto dos índices das observações sob risco no tempo  $t_i$ .

A função de verossimilhança parcial (5) assume que os tempos de sobrevivência são contínuos e, conseqüentemente, não pressupõe a possibilidade de empates nos tempos observados. Na prática, empates podem ocorrer nos tempos de falha ou de censura devido à escala de medida. Breslow (1972) e Peto (1972) propuseram uma modificação na função de verossimilhança parcial para acomodar as observações empatadas quando estas ocorrem. Considere  $s_i$  o vetor formado pela soma das covariáveis para os indivíduos que falharam no mesmo tempo  $t_i$  ( $i = 1, \dots, k$ ) e  $d_i$  o número de falhas neste tempo. A função de verossimilhança parcial proposta por estes autores tem a seguinte forma:

$$L(\beta) = \prod_{i=1}^k \left( \frac{\exp\{s_i^T \beta\}}{\sum_{j \in R(t_i)} \exp\{X_j^T \beta\}} \right)^{d_i}. \quad (6)$$

Esta aproximação é adequada quando o número de empates em qualquer tempo não é grande. Quando não há empates a expressão (6) se reduz à (5).

### 3.2 Modelo de Cox com Covariáveis Dependentes do Tempo

Considere a possibilidade que entre as  $p$  covariáveis medidas em cada indivíduo,  $X_1, \dots, X_p$  existam algumas que foram medidas  $n$  vezes, ao longo do tempo. Assim temos  $(T_i, \delta_i, X_i(t_j))$  o vetor que representa os dados de sobrevivência do  $i$ -ésimo indivíduo no  $j$ -ésimo tempo;  $X_i(t_j)$  o vetor de covariáveis do  $i$ -ésimo indivíduo avaliado no tempo  $j$  ( $j = 1, \dots, n_j$ ).

Neste modelo a função de risco assume a forma:

$$\lambda_i(t_j) = \lambda_0(t_j) \cdot \exp\{X_i^T(t_j) \cdot \beta\}. \quad (7)$$

Então, o modelo de Cox com variável dependente do tempo não supõe riscos proporcionais pois a razão das funções de risco de dois indivíduos passa a depender do tempo.

Assim, a função de verossimilhança parcial para o modelo de Cox com covariável dependente do tempo pode ser escrita como:

$$L(\beta) = \prod_{i=1}^m \left( \frac{\exp\{X_i^T(t_i) \cdot \beta\}}{\sum_{j \in R(t_i)} \exp\{X_j^T(t_i) \cdot \beta\}} \right)^{\delta_i}. \quad (8)$$

Andersen e Gill (1982) mostraram que, sob certas condições de regularidade, os estimadores produzidos pela função de verossimilhança parcial (6 e 8) são consistentes e assintoticamente normais.

## 4. Modelagem Conjunta

Nas Seções 2 e 3 apresentamos formas de modelagem separada de dados longitudinais e de sobrevivência, respectivamente. Nesta seção iremos apresentar, de forma sucinta, algumas das propostas de modelagem conjunta encontradas na literatura.

### 4.1 O Modelo Conjunto por Wulfsohn e Tsiatis (1997)

A modelagem conjunta de dados longitudinais e de sobrevivência tem sido explorada recentemente por muitos autores. O modelo conjunto proposto por Wulfsohn e Tsiatis (1997) tem sido referência para muitos deles. Estes autores utilizaram um modelo de Cox para o processo de sobrevivência e a resposta longitudinal é tratada como covariável contínua cujo valor real é dado por um modelo linear de efeitos aleatórios.

Considere  $m$  indivíduos, seja  $Y_i$  uma variável dependente do tempo, e cada indivíduo tem  $n$  medidas da covariável longitudinal nos tempos  $(t_{ij} : t_{ij} \leq T_i; i = 1, \dots, m; j = 1, \dots, n)$ .  $T_i$  é o tempo máximo de acompanhamento do paciente e  $\delta_i$  uma função indicadora de falha.

Assumimos que  $Y_i$  pode ser ajustada por um modelo linear de efeitos aleatórios, dado pela equação (1):

$$Y_i = X_i^T \beta + Z_i b_i + \varepsilon_i, \quad i = 1, 2, \dots, m$$

em que,  $Y \sim (X_i^T \beta; \Sigma)$ ,  $\Sigma = (Z_i V_i Z_i^T + \sigma^2 I)$ ,  $V_i$  é a matriz de covariâncias das medidas do mesmo indivíduo.  $\varepsilon_i \sim N(0, \sigma^2 I_n)$ , o vetor de erros independentes. Os coeficientes aleatórios,  $b_i$ , são independentes e têm distribuição Normal com média zero e matriz de covariância  $V_i$ .  $\varepsilon_i \perp b_i$ .

O risco de falha é então ajustado pelo modelo semi-paramétrico de Cox e o valor ajustado pelo modelo linear de efeitos aleatórios utilizado como valor real do marcador longitudinal, em lugar da medida observada  $Y_{ij}$ .

$$\begin{aligned} \lambda_i(t | \beta, b_i, Y_i, X_i, Z_i) &= \\ &= \lambda_0(t) \cdot \exp\{\gamma(X_i^T \beta + Z_i b_i)\} \end{aligned}$$

em que,  $\lambda_0(t)$  é a função de risco baseline (não-paramétrica) que é avaliada a cada tempo de falha,  $\gamma$  é o parâmetro que mede a associação entre o tempo de sobrevivência e a variável longitudinal.

Neste modelo, assumimos que o tempo de censura é não informativo, sendo realizado independentemente do tempo de sobrevivência e da medida do marcador longitudinal.



Estes autores propuseram um algoritmo EM para estimar os parâmetros de interesse maximizando a função de verossimilhança dos dados observados. Isto é feito através da iteração entre o passo E, em que se calcula a esperança da log-verossimilhança dos dados completos, condicional aos dados observados e as estimativas atuais dos parâmetros; e o passo M, em que novas estimativas para os parâmetros são calculadas maximizando sua log-verossimilhança esperada.

## 4.2 O Modelo Conjunto por Henderson, Diggle e Dobson (2000)

Henderson, Diggle e Dobson (2000) propuseram um modelo generalizado para a análise conjunta que pode ser dividido em dois sub-modelos e um processo de ligação latente. Para as medidas longitudinais é assumido um sub-modelo linear com efeitos aleatórios, e para o processo de sobrevivência um sub-modelo semi-paramétrico de riscos proporcionais. O modelo conjunto assume que dado o processo latente normal bivariado,  $W(t) = \{W_1(t), W_2(t)\}$ , e as covariáveis avaliadas, os processos longitudinal e de sobrevivência são condicionalmente independentes.

Os sub-modelos podem ser escritos na seguinte forma:

- (1) Sub-modelo para as medidas longitudinais:

$$Y_{ij} = \mu_i(t_{ij}) + W_{1i}(t_{ij}) + \varepsilon_{ij}$$

em que,  $\mu_i(t_{ij})$  é a resposta média e,  $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$  é uma seqüência de erros mutuamente independentes. Assumimos que a resposta média pode ser descrita por um modelo linear:

$$\mu_i(t) = X_{1i}(t)' \beta_1$$

em que, os vetores  $X_{1i}(t)$  e  $\beta_1$  representam variáveis explicativas, que podem ser dependentes do tempo, e seus correspondentes coeficientes de regressão.

- (2) Sub-modelo para o processo de sobrevivência no tempo  $t$  é dado pelo modelo semi-paramétrico multiplicativo:

$$\lambda_i(t) = \lambda_0(t) \exp\{X_{2i}(t)' \beta_2 + W_{2i}(t)\}$$

em que, o componente não-paramétrico  $\lambda_0(t)$  têm forma não especificada. Os vetores  $X_{2i}(t)$  e  $\beta_2$  podem, ou não, ter elementos em comum com  $X_{1i}(t)$  e  $\beta_1$ .

Para o processo de ligação latente assumimos que,

$$W_1(t) = U_1 + U_2 t$$

em que,  $(U_1, U_2)$  têm distribuição normal bivariada com médias zero, com respectivas variâncias  $\sigma_1^2$  e  $\sigma_2^2$  e coeficiente de correlação  $\rho$ . E,  $W_2(t)$  tem a forma:

$$W_2(t) = \gamma_1 U_1 + \gamma_2 U_2 + \gamma_3 (U_1 + U_2 t) + U_3$$

em que,  $U_3 \sim N(0, \sigma_3^2)$  é independente de  $(U_1, U_2)$ . Neste modelo, os parâmetros  $\gamma_1$ ,  $\gamma_2$  e  $\gamma_3$  medem a associação induzida pelo intercepto, inclinação e valor de  $W_1(t)$ , respectivamente.  $U_3$  modela a fragilidade ortogonal do processo de medidas.

Estes autores descrevem uma extensão do algoritmo EM proposto por Wulfsohn e Tsiatis (1997) para a estimação dos parâmetros do modelo através do método da máxima verossimilhança.

### 4.3 O Modelo Conjunto por Diggle, Souza e Chetwynd (2007)

Assim como nos modelos descritos acima, considere  $m$  indivíduos independentes, seja  $Y_i$  uma covariável dependente do tempo, e cada indivíduo tem  $n_i$  medidas da covariável longitudinal nos tempos  $(t_{ij}; t_{ij} \leq T_i)$ .  $T_i$  é o tempo máximo de acompanhamento do paciente e  $\delta_i$  uma função indicadora de falha. Estes autores desenvolveram uma versão completamente paramétrica para o modelo conjunto. O modelo proposto considera uma transformação logarítmica para o tempo de sobrevivência,  $T$ , e então considera a distribuição conjunta de  $S = \log(T)$  e as medidas longitudinais,  $Y$ , como:

$$[Y, S] \sim NMV(\mu, \Sigma),$$

$$\text{em que, } \mu = (\mu_y, \mu_s), \text{ e } \Sigma = \begin{bmatrix} V(\theta) & g(\phi) \\ g'(\phi) & \eta^2 \end{bmatrix}.$$

Este modelo considera que os dados de indivíduos diferentes são independentes. O tempo de evento,  $T_i$ , pode ser observado de forma exata, censurado à directa, ou com censura intervalar. Além disso,  $T_i$  pode, ou não, determinar o fim da sequência de medidas longitudinais.

A inferência neste modelo é baseada no método da máxima verossimilhança, calculando a contribuição de cada indivíduo para a verossimilhança através da fatoração da distribuição conjunta em  $[Y, S] = [Y] * [S | Y]$ .

#### 4.4 O modelo em dois estágios

As aproximações apresentadas acima são bastante interessantes em diversos aspectos, no entanto, a estimação dos parâmetros destes modelos é computacionalmente complexa e envolve estimação a partir de algoritmos EM. Infelizmente, ainda não há programas disponíveis para o ajuste destes modelos. Consideramos o modelo em dois estágios como uma aproximação para o modelo conjunto. O primeiro estágio deste modelo consiste no ajuste do modelo de efeitos aleatórios (1) apresentado na Seção 2.2 para os dados longitudinais e, no segundo estágio utilizamos os valores ajustados por este modelo como covariável dependente do tempo no modelo de Cox (7). A idéia se assemelha àquela proposta por Wulfsohn e Tsiatis (1997), no entanto a estimação proposta por estes autores baseia-se na função de verossimilhança da distribuição conjunta de  $Y$  e  $T$  através da fatoração  $L_{\theta}(Y, T) = L_{\theta}(Y) \cdot L_{\theta}(T | Y)$ .

## 5. Aplicação

### 5.1 Descrição do experimento

Modelos experimentais, utilizando o *S. mansoni* (esquistossomose) em associação com espécies de *Plasmodium* (malária) de roedores, Lewin et. al. (1981) mostraram que as parasitoses mistas levaram a um aumento de parasitemia devida ao protozoário.

Com o objetivo de aprofundar os conhecimentos sobre o efeito da associação da infecção do protozoário da malária *P.berguei* com o helminto causador da esquistossomose *S.mansoni*, sobre a parasitemia por *P.berguei* e sobre a sobrevivência dos camundongos, foi conduzido, no Instituto de Ciências Biológicas (ICB-UFMG) em parceria com o Centro de Pesquisas René Rachou (CPqRR-FIOCRUZ), um experimento em que 80 camundongos foram alocados aleatoriamente em 4 grupos de tratamentos.

Estes camundongos foram acompanhados durante 25 dias. O tempo zero indica o tempo em que os camundongos foram infectados com *P. berguei*. A partir do 5º dia após a infecção o percentual de parasitas no sangue do camundongo torna-se grande o suficiente para ser detectado pelo teste de esfregaço sanguíneo (teste para determinar a parasitemia), este dia então foi usado como padrão para o início da contagem que foi realizada diariamente até o 30º dia, quando todos os camundongos que haviam sobrevivido foram sacrificados. Valores altos de parasitemia indicam alto grau de infecção do camundongo, enquanto que valores observados iguais a zero indicam a ocorrência de “cura” do camundongo. Os camundongos que morreram ao longo do período de acompanhamento (5º ao 30º dia após a infecção), tiveram o dia do óbito registrado.

Os 80 camundongos foram distribuídos aleatoriamente entre 4 grupos de 20 camundongos cada, no entanto, ocorreram mortes de camundongos antes do início da medida da parasitemia devido a fatores distintos do objetivo do estudo sendo, portanto, estes eliminados do banco de dados. O banco final dispõe de 64 camundongos assim distribuídos:

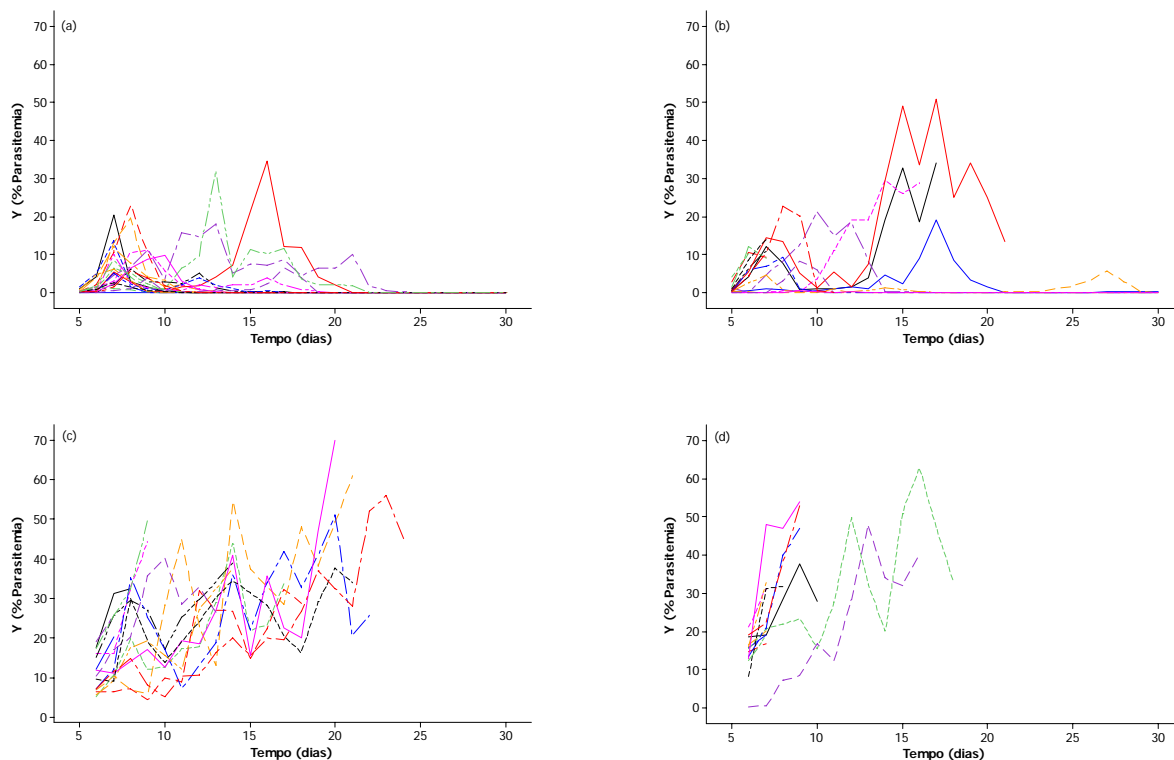
- Grupo 1:** 20 camundongos infectados por *P. berguei* que foram previamente imunizados à malária por infecções sucessivas seguidas de tratamento.
- Grupo 2:** 16 camundongos infectados cronicamente por *S. mansoni* e posteriormente infectados por *P. berguei* que foram previamente imunizados à malária por infecções sucessivas seguidas de tratamento.
- Grupo 3:** 15 camundongos infectados por *P. berguei*.
- Grupo 4:** 13 camundongos infectados cronicamente por *S. mansoni* e posteriormente infectados por *P. berguei*.

## Indução à imunidade à *P. berguei*:

Os camundongos dos grupos 1 e 2 foram imunizados contra a malária por *P. berguei* através de sucessivas infecções experimentais seguidas de cura utilizando tratamento com cloroquina (100mg/kg, via oral). O processo de indução de imunidade à malária iniciou-se 18 semanas após a infecção por *S. mansoni*. Para comprovar o sucesso da imunização, um grupo de 10 camundongos foi posteriormente infectado com  $1 \times 10^6$  hemácias parasitadas por *P. berguei* e acompanhado para a constatação da ausência de parasitemia.

### 5.2 Análise de dados

Com base no banco de dados deste experimento definimos  $O_{ij} = (X1_i, X2_i, Y_{ij}, T_i, \delta_i)$  os valores observados das medidas do *i*-ésimo camundongo no *j*-ésimo tempo.  $X1_i$  e  $X2_i$ , representam as medidas das covariáveis indicadoras: coinfeção e imunização, respectivamente;  $Y_{ij}$ , representa o percentual médio de parasitemia do *i*-ésimo camundongo no *j*-ésimo dia após a infecção;  $T_i$ , representa o tempo máximo de acompanhamento do *i*-ésimo camundongo,  $T_i \leq 30$ ; e,  $\delta_i$  é o indicador de falha.



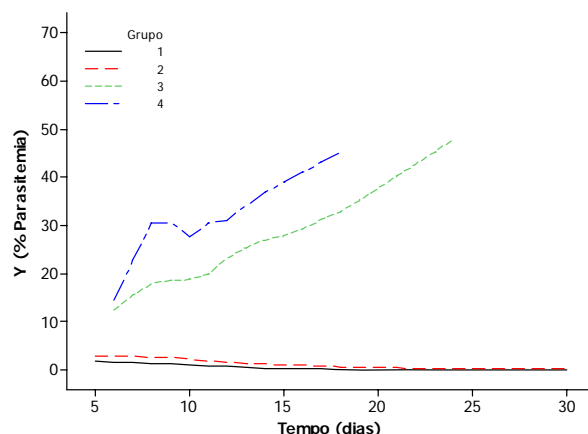
**Figura 1** - Gráfico de Perfis do % médio de parasitemia para os camundongos por grupo experimental (a) grupo 1, (b) grupo 2, (c) grupo 3 e (d) grupo 4.

A Figura 1, contém os gráficos de perfis dos dados associados a cada camundongo por grupo experimental. O grupo 1 (Figura 1 (a)) é composto por camundongos que foram imunizados e infectados apenas por malária, correspondendo ao grupo de melhor prognóstico uma vez que apresenta os mais baixos valores de parasitemia e, a partir do 25º dia após a infecção até o tempo final do estudo todos os camundongos apresentaram parasitemia zero (indicativo de cura) sendo portanto, todos censurados no 30º dia após a infecção. No grupo 2 (Figura 1 (b)) estão alocados os camundongos que foram imunizados e coinfectados. Neste grupo temos a presença de 6 camundongos que mantiveram o percentual de parasitemia baixo até o final do experimento e foram censurados no 30º dia após a infecção, 8 que morreram com níveis de parasitemia diversos e 2 casos que tiveram óbito no 12º dia mas que apresentavam indicador de cura no dia anterior à morte. Os camundongos alocados nos grupos 3 e 4 (Figura 1 (c) e (d)) não foram imunizados e todos eles morreram antes do fim do estudo, sendo que podemos observar valores altos de parasitemia precedendo a morte. Observa-se que no grupo 4, em que os camundongos foram coinfectados, o tempo de sobrevivência de apenas 2 animais foi superior a 11 dias, enquanto que no grupo 3 (animais com infecção simples) o tempo de sobrevivência parece ter sido maior. Nos grupos experimentais 3 e 4 as medidas de parasitemia do 5º dia após a infecção foram perdidas.

### 5.2.1 Análise de Dados Longitudinais

Primeiramente, iremos focar a análise deste banco de dados sob o aspecto da estrutura de medidas repetidas. Então, a variável resposta a ser considerada,  $Y_{ij}$ , é o percentual médio de parasitemia do  $i$ -ésimo camundongo no  $j$ -ésimo dia após a infecção, com  $i= 1, \dots, 64$  e  $j= 5, \dots, n_i$ ;  $5 \leq n_i \leq 30$ .

A Figura 2 apresenta a curva suavizada (método Lowess – Cleveland, 1985) do percentual de parasitemia médio para cada um dos grupos experimentais. Observa-se que o nível de parasitemia dos camundongos que foram imunizados (grupos 1 e 2) é aparentemente inferior ao nível dos camundongos não imunizados (grupos 3 e 4). Além disso, observa-se que os camundongos dos grupos 3 e 4 apresentam valores de parasitemia crescentes ao longo do acompanhamento, enquanto que os camundongos alocados nos grupos 1 e 2, apresentaram valores mais baixos de parasitemia em todo o período de acompanhamento. Os gráficos de perfis do percentual médio de parasitemia para cada um dos grupos experimentais foram traçados e apresentados na Figura 1.



**Figura 2** – Gráfico da curva de alisamento (Lowess) do % médio de parasitemia para os grupos experimentais.

Um modelo de regressão linear simples foi ajustado para a variável resposta: percentual de parasitemia (Y) versus o tempo após a infecção, desconsiderando-se o grupo experimental e o camundongo. Os resíduos padronizados foram calculados com o objetivo de avaliar a variabilidade da resposta no tempo (Diggle, Liang e Zeger, 1996). A partir destes resíduos, foi traçado um gráfico de perfis para cada grupo experimental. Estes gráficos estão apresentados na Figura 3, Anexo A e pode-se observar que, exceto no grupo 3 em que, a amplitude de variação dos resíduos parece constante, a variância dos resíduos parece aumentar com o tempo.

A matriz de correlação dos resíduos padronizados entre os tempos distintos estão apresentados no Anexo A, estas medidas foram calculadas no intuito de avaliar a estrutura de correlação das medidas no tempo, ou seja, se a correlação entre os tempos decresce quando as medidas se afastam sugerindo uma correlação serial ou se possuem outra forma como a correlação uniforme ou não-estruturada. Com base nos resultados encontrados observamos que a correlação tende a decrescer quando as medidas se afastam no tempo, logo a forma de correlação serial parece ser adequada para ajuste dos dados.

O modelo marginal e o modelo de efeitos aleatórios são as principais formas de modelagem de dados longitudinais e foram apresentadas na Seção 2. Uma série de modelos, utilizando estas técnicas, foram ajustados ao banco de dados e seguem descritos a seguir:

**ML1** – Modelo Marginal com Correlação Uniforme: este modelo considera que a correlação entre as medidas do mesmo indivíduo é constante no tempo.

**ML2** – Modelo Marginal com Correlação Serial: este modelo considera que a correlação entre as medidas do mesmo indivíduo decresce quando estas se afastam no tempo.

**ML3** – Modelo de Efeitos Aleatórios (efeito aleatório no intercepto): este modelo considera um intercepto distinto para cada camundongo,  $\beta_{0i}$ .

**ML4** – Modelo de Efeitos Aleatórios (efeito aleatório no intercepto e no Tempo): este modelo permite que tanto o intercepto, quanto a inclinação com o tempo variem entre camundongos.

O Modelo Marginal com Correlação Não-estruturada, ao ser ajustado não convergiu por ser tratar de um modelo complexo devido ao seu grande número de parâmetros a ser estimado.

A Tabela 1 apresenta um resumo destes ajustes em que são apresentados os coeficientes de regressão estimados ( $\hat{\beta}$ 's) e os valores p associados ao teste de Wald. Nesta tabela são apresentadas duas medidas de qualidade do ajuste. São elas:

$$\text{AIC: Akaike Information Criterion} \quad AIC = -2[l(\hat{\theta}) - \dim(\theta)]$$

$$\text{BIC: Bayesian Information Criterion} \quad BIC = -2\left[l(\hat{\theta}) - \frac{N \log(\dim(\theta))}{2}\right]$$

em que,  $\theta$  representa o conjunto de parâmetros estimados em cada modelo.

Estas medidas são usadas para comparar modelos a partir da informação do quanto a função de verossimilhança foi capaz de absorver dos dados, penalizada pelo grau de complexidade do modelo. Valores menores destas medidas indicam melhores ajustes.

A partir dos resultados apresentados na Tabela 1, podemos confirmar que o modelo marginal com correlação serial (ML2) é o que melhor se ajustou aos dados uma vez que, apresentou os menores valores de AIC e BIC. Neste modelo todas as covariáveis foram significativas. A coinfeção tem efeito de aumentar o grau de parasitemia em cerca de 3%, enquanto que a imunização reduz em média 22% o nível de parasitemia do camundongo. O efeito do tempo é de crescimento na parasitemia (em média 0,18% ao dia).

**Tabela 1** – Ajuste de modelos para Dados Longitudinais

Modelo	Covariáveis no modelo	Estimativas	se( $\beta$ )	p-valor	AIC	BIC
ML1	Intercepto	$\beta_0 = 23,16$	1,08	0,0000	6592,02	6621,14
	X1: coinfeção	$\beta_1 = 3,47$	1,12	0,0021		
	X2: imunização	$\beta_2 = -22,39$	1,15	0,0000		
	T: tempo	$\beta_3 = 0,056$	0,04	0,1181		
ML2	Intercepto	$\beta_0 = 21,69$	1,50	0,0000	6120,13	6149,24
	X1: coinfeção	$\beta_1 = 2,99$	1,29	0,0209		
	X2: imunização	$\beta_2 = -23,45$	1,42	0,0000		
	T: tempo	$\beta_3 = 0,183$	0,07	0,0143		
ML3	Intercepto	$\beta_0 = 23,16$	1,08	0,0000	6592,02	6621,14
	X1: coinfeção	$\beta_1 = 3,47$	1,12	0,0030		
	X2: imunização	$\beta_2 = -22,39$	1,15	0,0000		
	T: tempo	$\beta_3 = 0,056$	0,04	0,1180		
ML4	Intercepto	$\beta_0 = -0,72$	2,51	0,7733	6222,88	6261,69
	X1: coinfeção	$\beta_1 = 0,71$	1,20	0,5573		
	X2: imunização	$\beta_2 = -7,91$	1,26	0,0000		
	T: tempo	$\beta_3 = 2,18$	0,40	0,0000		



Os perfis médios ajustados a partir destes modelos estão apresentados na Figura 4, Anexo B. A exploração destes perfis revelou, entretanto, que a aproximação é ainda muito grosseira quando comparada aos perfis médios observados (Figura 2).

Modelos que considerem perfis individuais (efeitos aleatórios) e a interação do efeito das covariáveis com o tempo são especialmente interessantes de se avaliar neste tipo de situação clínica. Os modelos ML5, ML6 e ML7 são modelos lineares com efeitos aleatórios no intercepto e no tempo, e foram ajustados com o objetivo de avaliar a presença deste possível efeito de interação da coinfeção com o tempo, da imunização com o tempo, e de ambos efeitos de interação, respectivamente. O resumo dos resultados destes ajustes estão apresentados na Tabela 2, observamos que ambos os efeitos de interação são significativos. O modelo ML7 mostrou-se então o mais eficiente segundo os critérios de AIC e BIC, apresentando os menores valores para estas medidas. Além disso, ao observar os perfis ajustados para estes modelos apresentados na Figura 5, Anexo B, verificamos que este modelo é aquele que mais se aproxima da média observada (Figura 2). O ajuste do modelo com interação entre as covariáveis: coinfeção e imunização, foi considerado mas, seus resultados foram omitidos por não ter apresentado efeito significativo.

**Tabela 2** – Ajuste de Modelos de Efeitos Aleatórios com Interações

Modelo	Covariáveis no modelo	Estimativas	se( $\beta$ )	p-valor	AIC	BIC
ML5	Intercepto	$\beta_0 = 3,52$	2,81	0,2112	6215,72	6259,38
	X1: coinfeção	$\beta_1 = -11,41$	4,40	0,0119		
	X2: imunização	$\beta_2 = -7,42$	1,24	0,0000		
	T: tempo	$\beta_3 = 1,34$	0,48	0,0056		
	X1*T: interação do efeito de coinfeção com o tempo	$\beta_4 = 2,22$	0,79	0,0048		
ML6	Intercepto	$\beta_0 = -10,44$	2,32	0,0000	6174,71	6218,37
	X1: coinfeção	$\beta_1 = 0,38$	1,21	0,7560		
	X2: imunização	$\beta_2 = 12,42$	2,66	0,0000		
	T: tempo	$\beta_3 = 4,15$	0,37	0,0000		
	X2*T: interação do efeito de imunização com o tempo	$\beta_4 = -3,92$	0,45	0,0000		
ML7	Intercepto	$\beta_0 = -6,66$	2,05	0,0012	6158,72	6207,22
	X1: coinfeção	$\beta_1 = -7,23$	2,12	0,0012		
	X2: imunização	$\beta_2 = 11,35$	2,29	0,0000		
	T: tempo	$\beta_3 = 3,39$	0,32	0,0000		
	X1*T: interação do efeito de coinfeção com o tempo	$\beta_4 = 1,58$	0,36	0,0000		
	X2*T: interação do efeito de imunização com o tempo	$\beta_5 = -3,76$	0,37	0,0000		

Observamos, porém, que o ajuste de um modelo linear talvez não seja a melhor alternativa para o ajuste destas medidas longitudinais, uma vez que aqueles camundongos, que sobreviveram os trinta dias de acompanhamento do experimento, tiveram o nível de parasitemia altamente oscilatório no

início do experimento e reduz para zero até o final do experimento, indicando cura da malária. Isto ocorreu com todos os camundongos alocados no grupo de tratamento 1 (Figura 1a) e em 6 dos 16 alocados no grupo 2, ambos os grupos imunizados contra a malária (Figura 1b).

Propomos, como alternativa a estes modelos apresentados nas Tabelas 1 e 2, o ajuste de um modelo linear generalizado com efeitos aleatórios com a função de ligação binomial, para a resposta transformada  $Y'$ , em que  $Y'_{ij} = 1(Y_{ij} = 0)$ , ou seja,  $Y'_{ij} = 1$  indica que o percentual de parasitemia do  $i$ -ésimo camundongo ( $i = 1, \dots, 64$ ) no  $j$ -ésimo dia de avaliação ( $j = 5, \dots, n_i; 5 \leq n_i \leq 30$ ) é igual a zero. Assim,  $Y'$  pode ser interpretado como um indicador de cura do camundongo. Os resultados do ajuste destes modelos estão apresentados na Tabela 3.

**Tabela 3** – Ajuste de Modelos Lineares Generalizados com Efeitos Aleatórios para  $Y'$

Modelo	Covariáveis no modelo	Estimativas	se( $\beta$ )	p-valor	AIC	BIC
ML8	Intercepto	$\beta_0 = 33,47$	52345,69	0,9995	*	*
	X1: coinfeção	$\beta_1 = 0,97$	0,74	0,1907		
	X2: imunização	$\beta_2 = -28,86$	52345,69	0,9996		
	T: tempo	$\beta_3 = -0,34$	0,02	0,0000		
ML9	Intercepto	$\beta_0 = 30,87$	3049583,8	1,0000	*	*
	X1: coinfeção	$\beta_1 = 0,60$	43,1	0,9890		
	X2: imunização	$\beta_2 = -26,11$	3049583,8	1,0000		
	T: tempo	$\beta_3 = -0,34$	3,0	0,9117		
ML10	Intercepto	$\beta_0 = 28,01$	52317,33	0,9996	*	*
	X1: coinfeção	$\beta_1 = -3,92$	0,79	0,0000		
	X2: imunização	$\beta_2 = -21,12$	52317,33	0,9997		
	T: tempo	$\beta_3 = -0,05$	4556,75	1,0000		
	X1*T: interação do efeito de coinfeção com o tempo	$\beta_4 = 0,36$	0,03	0,0000		
ML11	Intercepto	$\beta_0 = 36,59$	33470,38	0,9991	*	*
	X1: coinfeção	$\beta_1 = -17,70$	3,89	0,0000		
	X2: imunização	$\beta_2 = -13,49$	33470,38	0,9997		
	T: tempo	$\beta_3 = -0,50$	2655,31	0,9998		
	X1*T: interação do efeito de coinfeção com o tempo	$\beta_4 = 1,52$	0,36	0,0000		
ML12	Intercepto	$\beta_0 = 7,82$	<b>1,05</b>	<b>0,0000</b>	7186,0	7215,2
	X1: coinfeção	$\beta_1 = -4,21$	<b>1,33</b>	<b>0,0024</b>		
	T: tempo	$\beta_3 = -0,48$	<b>0,06</b>	<b>0,0000</b>		
	X1*T: interação do efeito de coinfeção com o tempo	$\beta_4 = 0,33$	<b>0,07</b>	<b>0,0000</b>		
ML13	Intercepto	$\beta_0 = 20,47$	2,77	<b>0,0000</b>	11574,4	11613,3
	X1: coinfeção	$\beta_1 = -13,75$	3,89	<b>0,0008</b>		
	T: tempo	$\beta_3 = -1,42$	0,24	<b>0,0000</b>		
	X1*T: interação do efeito de coinfeção com o tempo	$\beta_4 = 1,11$	0,38	<b>0,0031</b>		

O ajuste dos modelos ML8, ML10 e ML12 inclui um efeito aleatório no intercepto e os modelos ML9, ML11 e ML13 incluem efeito aleatório no intercepto e no Tempo.

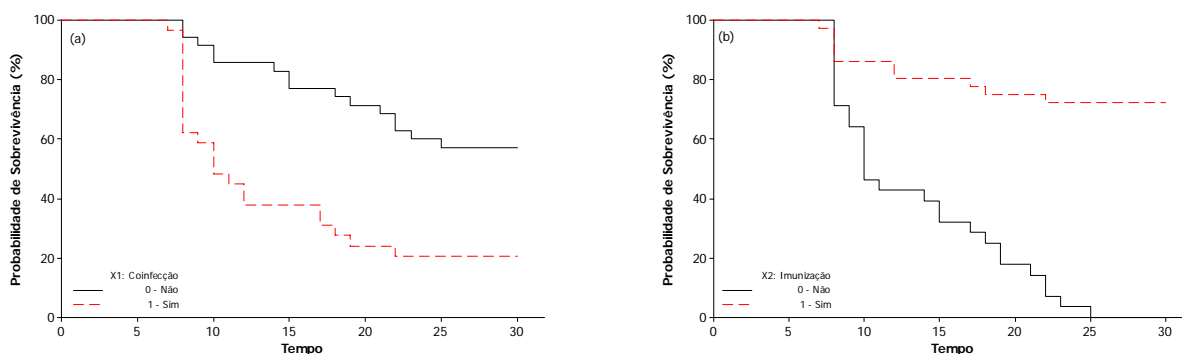
Os modelos ML12 e ML13 destacados apresentam como significativos os efeitos da Coinfecção, do Tempo e da interação: Coinfecção com o Tempo. A interpretação dos coeficientes destes efeitos é bastante intuitiva. Podemos observar que a Coinfecção diminui significativamente a chance de cura do camundongo e quanto maior o Tempo no estudo maior a chance de cura. Estes resultados são concordantes com o esperado, uma vez que permanecem no estudo por mais tempo justamente os camundongos que obtiveram cura. Estes modelos apontam os mesmos resultados como significativos, mas apresentam grande discrepância nos valores dos coeficientes estimados. Observando que os valores de AIC de BIC do modelo ML12 foram consideravelmente menores que o modelo ML13, considerados o ML12 como um melhor ajuste.

### 5.2.2 Análise de Sobrevivência

Agora, iremos focar a estrutura deste banco de dados sob o aspecto de análise de sobrevivência. Então, a variável resposta,  $T$ , é o tempo transcorrido entre a infecção e a morte camundongo, sendo que houve censura no tempo  $T = 30$  dias, quando todos os camundongos que sobreviveram foram sacrificados.

Com o objetivo de se obter uma idéia do efeito das covariáveis sobre o comportamento dos dados foram calculadas, inicialmente, as estimativas para a função de sobrevivência pelo método de Kaplan-Meier para as covariáveis  $X_1$  e  $X_2$ . A Figura 6a apresenta as curvas de sobrevivência estimadas para a covariável  $X_1$  (coinfecção). Comparando estas curvas pelo teste de Wilcoxon, observa-se que os camundongos infectados apenas com malária (infecção simples) possuem maior probabilidade de sobrevivência em relação aos que foram coinfectados ( $p < 0,0001$ ).

As curvas de sobrevivência estimadas pelo método de Kaplan-Meier para a covariável  $X_2$  (imunização) estão apresentadas na Figura 6b. Observa-se que os camundongos que foram imunizados possuem maior probabilidade de sobrevivência em relação aos que não foram imunizados. O teste de Wilcoxon confirmou que há diferença significativa ( $p < 0,0001$ ) entre as curvas de sobrevivência dos camundongos segundo a imunização.



**Figura 6** – Curvas de sobrevivência estimadas pelo método de Kaplan-Meier para a covariável X1: Co-Infecção (a) e X2: Imunização (b).

O modelo de riscos proporcionais de Cox (1972), apresentado na Seção 3, foi ajustado aos dados. A Tabela 4 apresenta um resumo dos resultados obtidos para o ajuste de modelos de Cox para as covariáveis X1 e X2.

**Tabela 4** – Ajuste do Modelo de Cox

Modelo	Covariáveis no modelo	Estimativas ( $\hat{\beta}$ )	se( $\beta$ )	p-valor	exp $\{\beta\}$	IC 95% (exp $\{\beta\}$ )	
						LI	LS
MS1	X1: coinfeccção	1,14	0,338	< 0,001	3,13	1,61	6,07
MS2	X2: imunização	-1,99	0,391	< 0,001	0,137	0,0637	0,295
MS3	X1: coinfeccção	1,39	0,355	< 0,001	3,997	1,9915	8,022
	X2: imunização	-2,20	0,410	< 0,001	0,111	0,0498	0,249

A partir dos resultados apresentados na Tabela 4, observa-se que ambas as covariáveis Coinfeccção (X1) e Imunização (X2) são significativas para explicar os tempos de sobrevivência dos camundongos. Os camundongos imunizados têm menor chance de morte e a presença de coinfeccção aumenta a chance de mortalidade em cerca de 4 vezes. A interação entre as covariáveis X1 e X2 foi também considerada, mas com a presença desta interação o modelo não convergiu.

### 5.3 Modelo de Cox com covariável dependente do tempo

O modelo de Cox com covariável dependente do tempo é um método bastante conhecido e utilizado na literatura. Portanto, consideramos este ajuste como a primeira aproximação para a modelagem conjunta de dados longitudinais e de sobrevivência. A Tabela 5 apresenta um resumo dos resultados obtidos através do ajuste de deste modelo para as covariáveis X1, X2 e parasitemia, Y, que é dependente do tempo.

**Tabela 5** – Ajuste do Modelo de Cox com covariável dependente do tempo

Modelo	Covariáveis no modelo	Estimativas ( $\hat{\beta}$ )	se( $\beta$ )	p-valor	exp{ $\beta$ }	IC 95% (exp{ $\beta$ })	
						LI	LS
MS4	Y: parasitemia	0,072	0,009	<0,00001	1,07	1,06	1,09
MS5	X1: coinfeção	0,977	0,373	0,009	2,66	1,28	5,52
	Y: parasitemia	0,069	0,009	< 0,0001	1,07	1,05	1,09
MS6	X2: imunização	-0,146	0,515	0,780	0,86	0,32	2,37
	Y: parasitemia	0,070	0,013	<0,0001	1,07	1,05	1,10
MS7	X1: coinfeção	1,052	0,383	0,006	2,86	1,35	6,07
	X2: imunização	-0,481	0,523	0,380	0,62	0,21	1,79
	Y: parasitemia	0,060	0,013	< 0,0001	1,06	1,04	1,09
MS8	X1: coinfeção	2,723	0,796	0,001	15,22	3,20	72,42
	X2: imunização	-1,237	0,742	0,095	0,29	0,07	1,24
	Y: parasitemia	0,085	0,020	< 0,0001	1,09	1,05	1,13
	X1*Y	-0,062	0,024	0,009	0,94	0,90	0,99
	X2*Y	0,042	0,030	0,160	1,04	0,98	1,11
MS9	X1: coinfeção	2,579	0,758	0,001	13,19	2,99	58,21
	X2: imunização	-0,543	0,553	0,330	0,58	0,20	1,72
	Y: parasitemia	0,092	0,019	< 0,0001	1,10	1,06	1,14
	X1*Y	-0,056	0,022	0,011	0,95	0,91	0,99
<b>MS10</b>	<b>X1: coinfeção</b>	<b>2,417</b>	<b>0,717</b>	<b>0,001</b>	<b>11,21</b>	<b>2,75</b>	<b>45,72</b>
	<b>Y: parasitemia</b>	<b>0,101</b>	<b>0,017</b>	<b>&lt; 0,0001</b>	<b>1,11</b>	<b>1,07</b>	<b>1,14</b>
	<b>X1*Y</b>	<b>-0,054</b>	<b>0,021</b>	<b>0,011</b>	<b>0,95</b>	<b>0,91</b>	<b>0,99</b>

Observa-se que a presença da covariável longitudinal no modelo acarreta a perda de significância da imunização (X2). Ou seja, ao incluir a parasitemia (Y) como covariável no modelo, esta absorve a informação do efeito de X2, tornando esta última não significativa. Como o objetivo principal do experimento é avaliar o efeito de tratamento sob a sobrevivência e sobre a parasitemia do camundongo, o modelo de Cox com covariável dependente do tempo torna-se aparentemente ineficaz. O modelo MS10 destacado aponta a coinfeção (X1), a parasitemia (Y) e a interação entre estas duas como significativas para estimar o tempo de sobrevivência dos camundongos.

O ajuste do modelo de Cox com a variável longitudinal dicotomizada, Y<sup>2</sup>, como covariável dependente do tempo também foi considerado, mas essa medida não se mostrou significativa, portanto os resultados deste ajuste foram omitidos.

## 5.4 Modelo em dois estágios

Uma proposta simples para a análise conjunta seria realizar a análise em dois estágios, em que o primeiro estágio consiste no ajuste de um modelo adequado para a resposta longitudinal e o segundo estágio ajusta um modelo de sobrevivência considerando como covariável dependente do tempo os valores ajustados no estágio anterior.

Consideramos o modelo linear, com efeitos aleatórios no intercepto e no tempo, com a presença das covariáveis de tratamento X1 e X2, do tempo e das interações X1\*Tempo e X2\*Tempo (ML7) como o melhor ajuste para as medidas de parasitemia. Os valores ajustados por este modelo foram então, considerados como uma covariável dependente do tempo no modelo de Cox. O resumo do ajuste deste modelo está apresentado na Tabela 6.

**Tabela 6 – Ajuste do Modelo em dois estágios**

Modelo	Covariáveis no modelo	Estimativas ( $\hat{\beta}$ )	se( $\beta$ )	p-valor	exp $\{\beta\}$	IC 95% (exp $\{\beta\}$ )	
						LI	LS
MC1	Ajuste ML7	0,101	0,014	< 0,0001	1,11	1,08	1,14
MC2	X1: coinfeção Ajuste ML7	0,773	0,373	0,038	2,17	1,04	4,50
		0,095	0,013	< 0,0001	1,10	1,07	1,13
MC3	X2: imunização Ajuste ML7	1,571	0,652	0,016	4,81	1,34	17,27
		0,142	0,022	< 0,0001	1,15	1,10	1,21
MC4	X1: coinfeção	0,485	0,410	0,240	1,62	0,73	3,63
	X2: imunização	1,192	0,713	0,094	3,30	0,82	13,32
	Ajuste ML7	0,128	0,025	< 0,0001	1,14	1,08	1,19
MC5	X1: coinfeção	2,939	1,091	0,007	18,89	2,23	160,22
	X2: imunização	0,041	0,990	0,970	1,04	0,15	7,25
	Y: parasitemia	0,177	0,035	< 0,0001	1,19	1,11	1,28
	X1* Ajuste ML7	-0,097	0,034	0,005	0,91	0,85	0,97
	X2* Ajuste ML7	0,077	0,040	0,056	1,08	1,00	1,17
MC6	X1: coinfeção	2,557	0,942	0,007	12,90	2,04	81,67
	X2: imunização	1,247	0,749	0,096	3,48	0,80	15,10
	Y: parasitemia	0,182	0,034	< 0,0001	1,20	1,12	1,28
	X1* Ajuste ML7	-0,079	0,029	0,007	0,92	0,87	0,98
MC7	X1: coinfeção	3,027	0,969	0,002	20,64	3,09	137,85
	Y: parasitemia	0,151	0,028	< 0,0001	1,16	1,10	1,23
	X1* Ajuste ML7	-0,084	0,031	0,007	0,92	0,86	0,98

Como resultado destes ajustes observamos que o efeito das medidas longitudinais ajustadas pelo modelo ML7 é sempre altamente significativo para estimar o tempo de sobrevivência dos camundongos, quanto maior o nível de parasitemia do camundongo maior o risco de mortalidade. Assim como no modelo MS10 a interação entre a coinfeção (X1) e a medida longitudinal (aqui, seus valores preditos pelo modelo ML7) é significativa para estimar o tempo de sobrevivência dos camundongos.

Os efeitos das covariáveis de tratamento não se mostraram significativos quando ambas são consideradas no modelo. No entanto, quando consideradas separadamente ambas as covariáveis de tratamento são significativas. A coinfeção aumenta a chance de óbito do camundongo como esperado, entretanto o efeito da imunização também foi de aumento da chance de óbito, o inverso do esperado.

Desta forma, o modelo em dois estágios parece ineficaz para oferecer melhorias para a modelagem em relação à análise separada.

## 6. Conclusões

Embora esta área de pesquisa na estatística ainda seja recente, algumas opções para a modelagem conjunta são apresentadas na literatura disponível como apresentado na Seção 1 e alguns exemplos detalhados na Seção 4. Em geral, os modelos propostos envolvem a fatoração da distribuição conjunta  $F(Y,T)$  em  $F(Y)*F(T|Y)$ . A maximização das funções de máxima verossimilhança envolve algoritmos EM, e computacionalmente a estimação dos parâmetros dos modelos conjuntos é bastante complexa. Infelizmente, os modelos conjuntos já implementados por estes autores ainda não foram tornados públicos e não encontramos nenhuma versão de modelagem disponível.

A modelagem separada utilizando o modelo linear com efeitos aleatórios para a resposta longitudinal e o modelo de Cox para o tempo de sobrevivência apontou ambas as covariáveis, coinfeção e imunização, como significativas para prever tanto o tempo de sobrevivência dos camundongos, quanto o grau de parasitemia destes camundongos.

Na análise conjunta, a partir do modelo de Cox com variável dependente do tempo ou do modelo em dois estágios (Seções 5.3 e 5.4), observamos que a presença da medida longitudinal, ou seu valor predito, como covariável dependente do tempo incorpora parte da informação das covariáveis de tratamento e eventualmente algumas delas ficam não significativas.

Embora intuitivamente o modelo em dois estágios pareça uma aproximação melhor para o ajuste do modelo conjunto que o modelo de Cox com a variável longitudinal observada, os erros associados às estimativas dos coeficientes no modelo em dois estágios (Tabela 6) são, em geral, maiores que os observados no modelo de Cox com a variável longitudinal observada (Tabela 5). Desta forma, consideramos o modelo em dois estágios como uma aproximação ineficaz para o modelo conjunto.

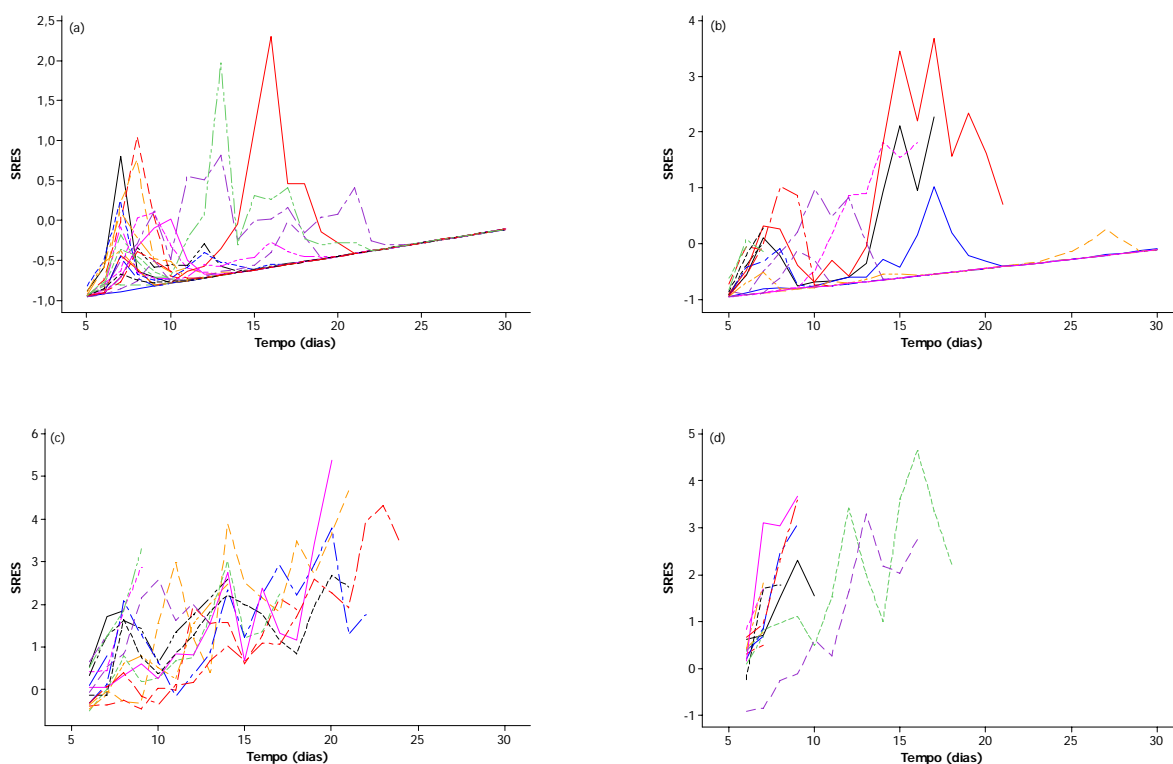


## 7. Referências Bibliográficas

- Andersen, P.K., Gill, R. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *Annals of Statistics*, **10**, 1100-1200.
- Breslow, N. (1972). Contribuição à Discussão do artigo de D.R. Cox. *Journal of the Royal Statistical Society B*, **34**, 216-217.
- Brown, E.R., Ibrahim, J.G., DeGruttola, V. (2005). A Flexible B-Spline Model for Multiple Longitudinal Biomarkers and Survival. *Biometrics*, **61**, 64-73.
- Collet, D. (2003). *Modelling Survival Data in Medical Research*, 2ed., Chapman and Hall, London.
- Colosimo, E.A., Giolo, S.R. (2006). *Análise de Sobrevivência Aplicada*. Edgard Blücher, São Paulo.
- Cox, D.R. (1972). Regression Models and Life Tables (with discussion). *Journal Royal Statistical Society, B*, **34**, 187-220.
- Cox, D.R. (1975). Partial Likelihood. *Biometrika*, **62**, 269-276.
- Diggle, P.J., Liang, K.Y., Zeger, S.L. (1996). *Analysis of Longitudinal Data*. Oxford University Press, New York.
- Diggle, P.J., Souza, I., Chetwynd, A.G. (2007) Joint Modelling of Repeated Measurements and time-to-event Outcomes: the Fourth Armitage Lecture. *Statistics in Medicine*, **1**, 1-19.
- Guo, X., Carlin, P. (2004). Separate e Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages. *The American Statistician*, **58**, 16-24.
- Henderson, R., Diggle, P., Dobson, A. (2000) Joint Modeling of Longitudinal Measurements and Event Time Data. *Biostatistics*, **1**, 465-480.
- Lewin, M., Last, C., Targett, G., Doenhoff, M.J. (1981). Factors affecting the acquisition of resistance against *Schistosoma mansoni* in the mouse. *Annals of Tropicals Medicine and Parasitology*, **75**, 79-86.
- Peto, R. (1972). Contribuição à discussão do artigo de D.R. Cox. *Journal of the Royal Statistical Society B*, **34**, 205-207.
- Ratcliffe, S.J., Guo, W., Have, T.R.T. (2004). Joint Modeling of Longitudinal and Survival data via a Common Frailty. *Biometrics*, **60**, 892-899.
- Schall, R. (1991). Estimation in Generalized Linear Models with Random Effects. *Biometrika*, **78**, 719-727.

- Song, X., Davidian, M., Tsiatis, A.A. (2002). A Semiparametric Likelihood Approach to Joint Modeling of Longitudinal and Time-to-Event Data. *Biometrics*, **58**, 742-753.
- Tseng, Y.K., Hsieh, F., Wang, J.L. (2005). Joint Modeling of Accelerate Failure Time and Longitudinal Data. *Biometrika*, **92**, 587-603.
- Xu, J., Zeger, S.L. (2001). Joint Analysis of Longitudinal Data comprising Repeated measures and Time to Events. *Royal Statistical Society*, **50**, 375-387.
- Wang, Y., Taylor, J.M.G. (2001). Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome. *The American Statistical association*, **96**, 895-905.
- Wulfsohn, M.S., Tsiatis, A.A. (1997). A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics*, **53**, 330-339.
- Zheng, Y., Heagerty, P.J. (2005). Partly Conditional Survival Models for Longitudinal Data. *Biometrics*, **61**, 379-391.

## ANEXO A



**Figura 3** - Gráfico de Perfis dos resíduos padronizados para os camundongos por grupo experimental (a) grupo 1, (b) grupo 2, (c) grupo 3 e (d) grupo 4.

### Matriz de correlação dos resíduos padronizados por tempo

	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29		
6	0,747																										
7	0,457	0,802																									
8	0,042	0,816	0,808																								
9	0,128	0,851	0,757	0,912																							
10	0,357	0,616	0,419	0,562	0,798																						
11	0,362	0,648	0,434	0,446	0,616	0,842																					
12	0,307	0,678	0,417	0,53	0,724	0,775	0,824																				
13	0,1	0,569	0,344	0,536	0,644	0,635	0,682	0,871																			
14	0,054	0,62	0,379	0,562	0,6	0,751	0,781	0,714	0,764																		
15	0,09	0,589	0,421	0,453	0,466	0,531	0,661	0,681	0,664	0,801																	
16	-0,006	0,65	0,384	0,485	0,581	0,592	0,682	0,787	0,761	0,809	0,922																
17	0,126	0,675	0,505	0,552	0,541	0,501	0,574	0,642	0,681	0,819	0,913	0,886															
18	0,095	0,704	0,444	0,499	0,542	0,705	0,782	0,716	0,665	0,922	0,852	0,868	0,892														
19	0,191	0,804	0,434	0,604	0,654	0,62	0,682	0,696	0,69	0,95	0,81	0,882	0,86	0,91													
20	0,202	0,826	0,396	0,596	0,694	0,676	0,718	0,743	0,711	0,951	0,718	0,846	0,768	0,859	0,971												
21	0,218	0,682	0,31	0,441	0,466	0,74	0,9	0,778	0,592	0,938	0,757	0,724	0,647	0,901	0,863	0,907											
22	0,105	0,612	0,288	0,5	0,451	0,328	0,428	0,439	0,466	0,782	0,612	0,565	0,646	0,845	0,917	0,841	0,95										
23	0,07	0,525	0,22	0,321	0,232	0,133	0,371	0,325	0,358	0,886	0,496	0,431	0,562	0,862	0,978	0,979	0,942	1									
24	-0,078	0,523	0,216	0,318	0,226	0,131	0,365	0,32	0,354	0,884	0,493	0,429	0,56	0,861	0,977	0,978	0,94	0,999	1								
25	-0,091	-0,155	-0,222	-0,182	-0,157	-0,097	-0,053	-0,098	-0,077	-0,103	-0,076	-0,074	-0,094	-0,087	-0,08	-0,07	-0,037	0,023	0,339	0,994							
26	-0,091	-0,155	-0,222	-0,182	-0,157	-0,097	-0,053	-0,098	-0,077	-0,103	-0,076	-0,074	-0,094	-0,087	-0,08	-0,07	-0,037	0,023	0,339	0,994	1						
27	-0,094	-0,16	-0,229	-0,187	-0,163	-0,1	-0,053	-0,099	-0,079	-0,09	-0,075	-0,068	-0,069	-0,069	-0,068	-0,064	-0,038	0,021	0,337	0,993	0,999	0,999					
28	-0,094	-0,159	-0,229	-0,187	-0,163	-0,1	-0,053	-0,099	-0,079	-0,091	-0,075	-0,068	-0,07	-0,07	-0,069	-0,064	-0,038	0,021	0,337	0,993	0,999	0,999	1				
29	-0,124	-0,181	-0,271	-0,225	-0,21	-0,116	-0,053	-0,075	-0,077	0,283	-0,014	0,129	0,579	0,421	0,269	0,125	-0,059	-0,028	0,106	0,411	0,418	0,418	0,45	0,449			
30	-0,091	-0,121	-0,187	-0,157	-0,152	-0,079	*	-0,034	-0,046	0,363	0,023	0,179	0,683	0,506	0,336	0,173	-0,047	-0,043	-0,053	-0,044	-0,04	-0,04	-0,004	-0,006	0,891		

## ANEXO B

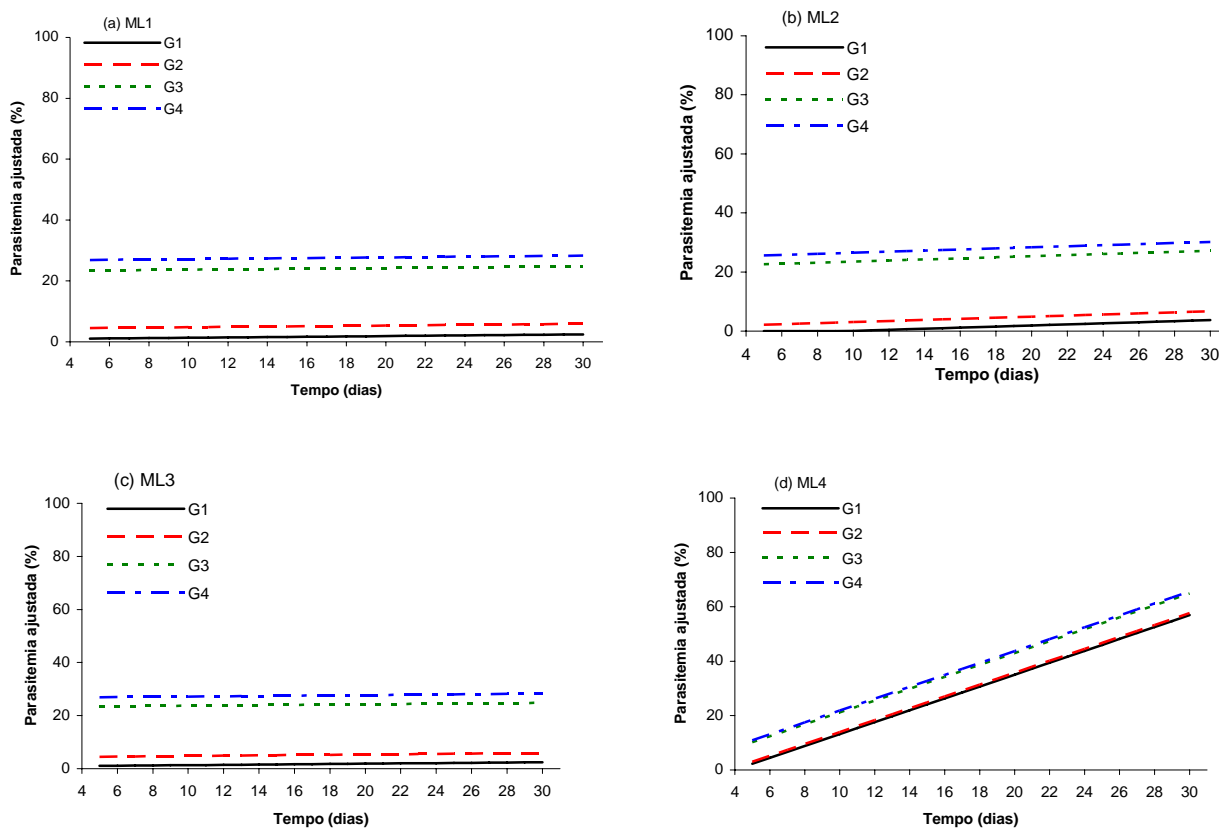


Figura 4 - Gráfico de Perfis médios ajustados pelos modelos longitudinais: (a) ML1, (b) ML2, (c) ML3 e (d) ML4.

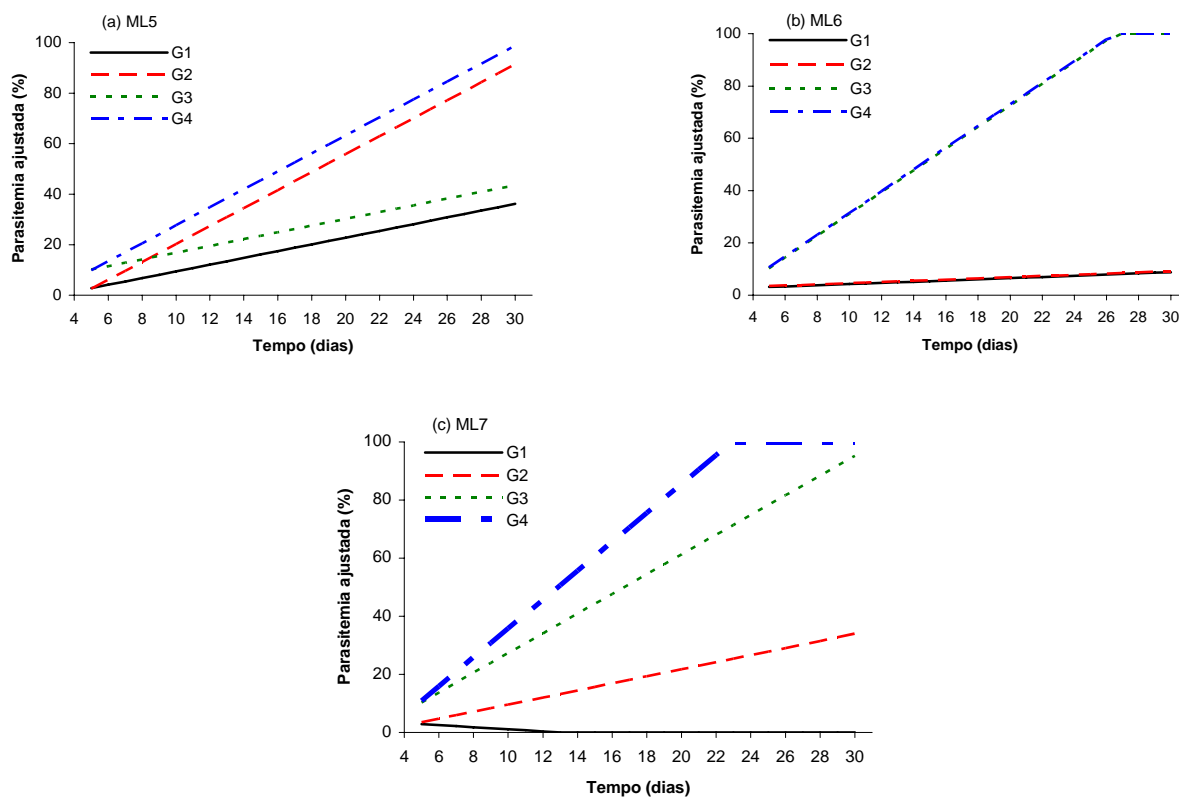


Figura 5 - Gráfico de Perfis médios ajustados pelos modelos longitudinais: (a) ML5, (b) ML6 e (c) ML7.