



MÉTODOS LONGITUDINAIS APLICADOS AO ESTUDO DA  
MULTICAUSALIDADE ASSOCIADA A REAÇÕES ADVERSAS AO  
TRATAMENTO QUIMIOTERÁPICO PARA O CÂNCER DE MAMA

Daniela Polessa Paula

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Biomédica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Biomédica.

Orientadores: Flavio Fonseca Nobre  
Rosane Viana Jorge

Rio de Janeiro  
Dezembro de 2015

MÉTODOS LONGITUDINAIS APLICADOS AO ESTUDO DA  
MULTICAUSALIDADE ASSOCIADA A REAÇÕES ADVERSAS AO  
TRATAMENTO QUIMIOTERÁPICO PARA O CÂNCER DE MAMA

Daniela Polessa Paula

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ  
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)  
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR  
EM CIÊNCIAS EM ENGENHARIA BIOMÉDICA.

Examinada por:

---

Prof. Flavio Fonseca Nobre, Ph.D.

---

Prof. Rosane Viana Jorge, Ph.D.

---

Prof. Rosimary Terezinha de Almeida, Ph.D.

---

Prof. Gulnar Azevedo e Silva, Ph.D.

---

Prof. Newton Gonçalves de Castro, Ph.D.

RIO DE JANEIRO, RJ – BRASIL  
DEZEMBRO DE 2015

Paula, Daniela Polessa

Métodos longitudinais aplicados ao estudo da multicausalidade associada a reações adversas ao tratamento quimioterápico para o câncer de mama/Daniela Polessa Paula. – Rio de Janeiro: UFRJ/COPPE, 2015.

XIV, 118 p.: il.; 29,7cm.

Orientadores: Flavio Fonseca Nobre

Rosane Viana Jorge

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Biomédica, 2015.

Referências Bibliográficas: p. 95 – 103.

1. Modelos longitudinais.      2. Câncer de mama.
3. Reações adversas.      I. Nobre, Flavio Fonseca *et al.*  
II. Universidade Federal do Rio de Janeiro, COPPE,  
Programa de Engenharia Biomédica. III. Título.

*Aos meus pais Sueli e Devanir*

# Agradecimentos

Agradeço a Deus pelo bem da vida e pela fé para seguir a cada dia.

Aos meus pais por terem me dado apoio em todos os momentos, nas alegrias e nas aflições, e principalmente por terem me dedicado anos de suas vidas me ensinando o valor e a beleza do conhecimento e da descoberta.

Ao meu orientador Flávio, pelo apoio e pela paciência que sempre dedicou a mim em todos esses anos. Obrigada pelos conselhos nos momentos de angústia e pelas palavras certas nos momentos certos.

A minha coorientadora Rosane pelo carinho e pela paciência que teve para me ensinar um pouco desse mundo de farmacologia totalmente desconhecido para uma matemática.

Aos meus amigos do Peb, em especial do Less, pelas trocas de conhecimento, pela ajuda e pelo apoio durante esses anos. Pelos momentos de alegria nas nossas confraternizações.

A todos meus amigos e familiares que me incentivaram e por terem compreendido meus momentos de ausência.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

MÉTODOS LONGITUDINAIS APLICADOS AO ESTUDO DA  
MULTICAUSALIDADE ASSOCIADA A REAÇÕES ADVERSAS AO  
TRATAMENTO QUIMIOTERÁPICO PARA O CÂNCER DE MAMA

Daniela Polessa Paula

Dezembro/2015

Orientadores: Flavio Fonseca Nobre  
Rosane Viana Jorge

Programa: Engenharia Biomédica

Astenia, mialgia, artralgia, mucosite, dor abdominal e diarreia são algumas reações comumente relatadas por mulheres em tratamento quimioterápico para o câncer de mama. A gravidade das reações está relacionada à variabilidade interindividual no conjunto de características associadas aos pacientes. Identificar perfis de susceptibilidade para as reações adversas permite otimizar o efeito do tratamento por meio de terapêuticas individualizadas. Os objetivos deste estudo foram comparar métodos longitudinais para identificação de fatores associados à ocorrência dessas reações adversas ao longo do tratamento FAC-D (3 ciclos de ciclofosfamida, dexorrubicina e 5-fluorouracil seguidos de 3 ciclos de docetaxel), analisar impacto da relação entre perfis clínico/genético e protocolos quimioterápicos para a gravidade das reações, e o impacto da mudança de protocolo ao longo do tratamento para a significância das variáveis. As análises foram realizadas com 166 mulheres de uma coorte hospitalar que completaram os 6 ciclos do tratamento FAC-D, os níveis das reações adversas foram acompanhados ao final de cada ciclo de quimioterapia através de entrevistas.

Os modelos marginais ajustados por equações de estimativas generalizadas apresentaram melhor ajuste aos dados. As reações analisadas, tanto musculares quanto gastrointestinais, agravaram-se durante a fase Docetaxel. Os resultados indicam o protocolo como principal fator associado aos efeitos adversos ao longo de todo o tratamento e sugerem que diferentes protocolos implicam diferentes perfis de susceptibilidade. No caso do tratamento FAC-D não foram encontrados perfis comuns para as duas fases.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

LONGITUDINAL METHODS APPLIED TO THE STUDY OF MULTIPLE  
CAUSES ASSOCIATED WITH ADVERSE REACTIOS TO CHEMOTHERAPY  
FOR BREAST CANCER

Daniela Polessa Paula

December/2015

Advisors: Flavio Fonseca Nobre  
Rosane Viana Jorge

Department: Biomedical Engineering

Asthenia, myalgia, arthralgia, mucositis, abdominal pain and diarrhea are some reactions commonly reported by women undergoing chemotherapy for breast cancer. The severity of the reactions is related to interindividual variability in the set of characteristics associated with patients. To identify susceptibility profiles for adverse reactions, allows to optimize the effect of treatment by individual therapies. The objective of this study were compare longitudinal methods to identify factors associated with these adverse reactions throughout the FAC-D treatment (3 cycles of cyclophosphamide, dexorubicina and 5-fluoroural followed by 3 docetaxel cycles), analyze the impact of the relation between chemotherapy protocol and clinical / genetic profiles for severity of reactions and the impact of the protocol change for significance of the variables. Analyses were conducted with 166 women in a hospital cohort who responded to 6 cycles of FAC-D treatment, levels of adverse reactions were followed at the end of each chemotherapy cycle through interviews.

The marginal models fitted by generalized estimating equations presented better fit for the data. The analyzed reactions, both muscle as gastrointestinal, worsened during Docetaxel phase. The results indicate that the protocol is the main factor associated with adverse reactions throughout treatment and suggest that different protocols require different susceptibility profiles. In the case of FAC-D treatment common profiles were not found to both phases.

# Sumário

<b>Lista de Figuras</b>	<b>xi</b>
<b>Lista de Tabelas</b>	<b>xiii</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Objetivos</b>	<b>6</b>
<b>3 Revisão Bibliográfica</b>	<b>7</b>
3.1 Polimorfismos genéticos, alterações farmacocinéticas e reações adversas	9
3.2 Características clínicas associadas a reações . . . . .	11
3.3 Metodologias para tratamento de variáveis categóricas . . . . .	12
<b>4 Fundamentação teórica</b>	<b>14</b>
4.1 Quimioterapia antineoplásica . . . . .	14
4.1.1 Quimioterapia FAC-D . . . . .	15
4.2 Farmacogenética e quimioterapia FAC-D . . . . .	19
4.3 Variáveis categóricas . . . . .	23
4.4 Modelos lineares generalizados . . . . .	24
4.4.1 Estimação de parâmetros . . . . .	26
4.4.2 Regressão logística . . . . .	27
4.4.3 Regressão logística nominal e ordinal . . . . .	28
4.4.4 Implementação de modelos ordinais no R . . . . .	33
4.5 Modelos longitudinais . . . . .	34
4.6 Modelos marginais . . . . .	35
4.6.1 Modelos marginais para variáveis ordinais . . . . .	37
4.6.2 O método de quase-verossimilhança . . . . .	39
4.6.3 Metodologia de equações de estimação generalizadas (GEE) .	40
4.6.4 Implementação do método GEE no R . . . . .	44
4.7 Modelos lineares generalizados mistos . . . . .	46
4.7.1 Modelos mistos para respostas ordinais . . . . .	51
4.7.2 Implementação do método GLMM no R . . . . .	52



4.7.3	Comparação entre modelos marginais e modelos mistos . . . . .	52
<b>5</b>	<b>Materiais e métodos</b>	<b>54</b>
5.1	População e período do estudo . . . . .	54
5.2	Coleta de dados . . . . .	54
5.3	Análise de dados . . . . .	56
<b>6</b>	<b>Resultados</b>	<b>60</b>
6.1	Análises preliminares . . . . .	60
6.2	Resultados para os modelos de regressão logística ordinal . . . . .	63
6.2.1	Modelos ajustados para astenia . . . . .	63
6.2.2	Modelos ajustados para mialgia . . . . .	63
6.2.3	Modelos ajustados para artralgia . . . . .	66
6.2.4	Modelos ajustados para dor abdominal . . . . .	66
6.2.5	Modelos ajustados para mucosite . . . . .	66
6.2.6	Modelos ajustados para diarreia . . . . .	66
6.3	Resultados para os modelos marginais . . . . .	66
6.3.1	Modelos ajustados para astenia . . . . .	73
6.3.2	Modelos ajustados para mialgia . . . . .	73
6.3.3	Modelos ajustados para artralgia . . . . .	73
6.3.4	Modelos ajustados para dor abdominal . . . . .	75
6.3.5	Modelos ajustados para mucosite . . . . .	77
6.3.6	Modelos ajustados para diarreia . . . . .	77
6.4	Resultados para os modelos lineares generalizados mistos . . . . .	79
6.4.1	Modelos ajustados para astenia . . . . .	79
6.4.2	Modelos ajustados para mialgia . . . . .	80
6.4.3	Modelos ajustados para artralgia . . . . .	80
6.4.4	Modelos ajustados para dor abdominal . . . . .	82
6.4.5	Modelos ajustados para mucosite . . . . .	83
6.4.6	Modelos ajustados para diarreia . . . . .	84
6.5	Análise de resíduos . . . . .	85
<b>7</b>	<b>Discussão</b>	<b>86</b>
<b>8</b>	<b>Conclusões finais</b>	<b>93</b>
	<b>Referências Bibliográficas</b>	<b>95</b>
	<b>Referências Bibliográficas</b>	<b>95</b>

<b>A</b>	<b>Resíduos dos modelos marginais e mistos:</b>	<b>104</b>
A.0.1	Primeiro modelo marginal e misto para astenia: . . . . .	104
A.0.2	Primeiro modelo marginal e misto para mialgia: . . . . .	105
A.0.3	Primeiro modelo marginal e misto para artralgia: . . . . .	106
A.0.4	Primeiro modelo marginal e misto para dor abdominal: . . . . .	107
A.0.5	Primeiro modelo marginal e misto para mucosite: . . . . .	108
A.0.6	Primeiro modelo marginal e misto para diarreia: . . . . .	109
A.0.7	Demais modelos para astenia: . . . . .	110
A.0.8	Demais modelos para mialgia: . . . . .	112
A.0.9	Demais modelos para artralgia: . . . . .	113
A.0.10	Demais modelos para dor abdominal: . . . . .	115
A.0.11	Demais modelos para mucosite: . . . . .	116
A.0.12	Demais modelos para Diarreia: . . . . .	117

# Lista de Figuras

4.1	Via metabólica do docetaxel. Adaptado de KIYOTANI et al, 2008; Marsh et al., 2009. Legenda: TUBB - cadeia de beta tubulina; MAPT - gene tau proteína associada a microtúbulo; CDKN1A - gene do inibidor kinase dependente de ciclina; TP53 - gene da proteína tumoral P53. . . . .	18
4.2	Etapas de um fármaco no organismo. Adaptado de Suarez-Kurtz, 2004.	19
4.3	Respostas de indivíduos ao longo do tempo (dados hipotéticos) . . . .	47
6.1	Proporções para os níveis das reações ao longo dos ciclos . . . . .	61
6.2	Modelos ajustados para cada um dos três tipos de abordagem utilizada	62
6.3	Processo de seleção de variáveis para as três abordagens utilizadas . .	64
A.1	Resíduos do primeiro modelo marginal para astenia . . . . .	104
A.2	Resíduos do primeiro modelo misto para astenia . . . . .	105
A.3	Resíduos do primeiro modelo marginal para mialgia . . . . .	105
A.4	Resíduos do primeiro modelo misto para mialgia . . . . .	106
A.5	Resíduos do primeiro modelo marginal para artralgia . . . . .	106
A.6	Resíduos do primeiro modelo misto para artralgia . . . . .	107
A.7	Resíduos do primeiro modelo marginal para dor abdominal . . . . .	107
A.8	Resíduos do primeiro modelo misto para dor abdominal . . . . .	108
A.9	Resíduos do primeiro modelo marginal para mucosite . . . . .	108
A.10	Resíduos do primeiro modelo misto para mucosite . . . . .	109
A.11	Resíduos do primeiro modelo marginal para diarreia . . . . .	109
A.12	Resíduos do primeiro modelo misto para diarreia . . . . .	110
A.13	Resíduos do segundo modelo marginal para astenia- fase FAC . . . .	110
A.14	Resíduos do segundo modelo misto para astenia - fase FAC . . . . .	111
A.15	Resíduos do terceiro modelo marginal para astenia - fase Docetaxel .	111
A.16	Resíduos do terceiro modelo misto para astenia - fase Docetaxel . . .	112
A.17	Resíduos do terceiro modelo marginal para mialgia - fase Docetaxel .	112
A.18	Resíduos do terceiro modelo misto para mialgia - fase Docetaxel . . .	113
A.19	Resíduos do segundo modelo marginal para artralgia - fase FAC . . .	113

A.20 Resíduos do segundo modelo misto para artralgia - fase FAC . . . . .	114
A.21 Resíduos do terceiro modelo marginal para artralgia - fase Docetaxel	114
A.22 Resíduos do terceiro modelo marginal para dor abdominal - fase Do- cetaxel . . . . .	115
A.23 Resíduos do terceiro modelo misto para dor abdominal - fase Docetaxel	116
A.24 Resíduos para o terceiro modelo marginal para mucosite - fase Docetaxel	116
A.25 Resíduos para o terceiro modelo marginal para diarreia - fase Docetaxel	117
A.26 Resíduos para o terceiro modelo misto para diarreia - fase Docetaxel .	118

# Lista de Tabelas

4.1	Exemplo de estrutura de dados longitudinais . . . . .	36
5.1	Níveis de gravidade das reações . . . . .	55
5.2	Polimorfismos . . . . .	55
5.3	Variáveis clínicas e demográficas . . . . .	55
6.1	Variáveis clínicas - (amostra) . . . . .	61
6.2	Polimorfismos - (amostra) . . . . .	61
6.3	Estimativas para os parâmetros dos modelos de regressão logística ordinal para astenia . . . . .	65
6.4	Estimativas para os parâmetros dos modelos de regressão logística ordinal para mialgia . . . . .	67
6.5	Estimativas para os parâmetros dos modelos de regressão logística ordinal para artralgia . . . . .	68
6.6	Estimativas para os parâmetros dos modelos de regressão logística ordinal para dor abdominal . . . . .	69
6.7	Estimativas para os parâmetros dos modelos de regressão logística ordinal para mucosite . . . . .	70
6.8	Estimativas para os parâmetros dos modelos de regressão logística ordinal para diarreia . . . . .	71
6.9	Estimativas para os parâmetros dos modelos marginais para astenia .	74
6.10	Estimativas para os parâmetros dos modelos marginais para mialgia .	75
6.11	Estimativas para os parâmetros dos modelos marginais para artralgia	76
6.12	Estimativas para os parâmetros dos modelos marginais para dor ab- dominal . . . . .	77
6.13	Estimativas para os parâmetros dos modelos marginais para mucosite	78
6.14	Estimativas para os parâmetros dos modelos marginais para diarreia .	78
6.15	Estimativas para os parâmetros dos modelos mistos para astenia . . .	81
6.16	Estimativas para os parâmetros dos modelos mistos para mialgia . . .	82
6.17	Estimativas para os parâmetros dos modelos mistos para artralgia . .	82
6.18	Estimativas para os parâmetros dos modelos mistos para dor abdominal	83

6.19 Estimativas para os parâmetros dos modelos mistos para mucosite . .	84
6.20 Estimativas para os parâmetros dos modelos mistos para diarreia . .	84

# Capítulo 1

## Introdução

O câncer de mama é o segundo tipo de câncer mais frequente no mundo e o primeiro entre as mulheres, respondendo por 22% dos novos casos a cada ano [1]. A evolução no tratamento para o câncer de mama aliada a detecção precoce contribuiu para que as taxas de mortalidade apresentassem redução em alguns países nos últimos anos [1]. No Brasil, as taxas de mortalidade continuam altas e a doença constitui a primeira causa de morte por câncer entre as mulheres [2, 3].

As taxas de mortalidade por câncer de mama no Brasil continuam elevadas devido a diagnósticos em estágios avançados da doença [2]. Dados do Ministério da Saúde indicam que cerca de 60 por cento dos tumores de mama são diagnosticados nos estágios III ou IV [2]. O diagnóstico tardio implica a redução da sobrevida pela doença e um tratamento mais agressivo, aumentando o risco de complicações e reações adversas. As reações adversas são um importante componente que afeta a qualidade de vida dos pacientes, mas que podem por vezes ser evitadas ou atenuadas por meio de condutas terapêuticas individualizadas. Para isso, é importante o conhecimento dos fatores relacionados ao desenvolvimento das reações.

A Organização Mundial da Saúde (OMS) define reação adversa a medicamento como qualquer efeito prejudicial ou indesejável, que aparece após a administração de um medicamento em dose normalmente utilizada no homem como profilaxia, diagnóstico ou tratamento de enfermidades [4]. As reações adversas aumentam a morbidade/mortalidade e estão associadas a custos consideráveis do sistema de saúde [5]. A notificação espontânea das reações adversas tem se mostrado um método eficaz para a identificação de reações adversas [6]. No caso de tratamentos quimioterápicos, algumas reações são esperadas e podem indicar eficácia terapêutica. Em virtude disso, é possível que haja subvalorização e subnotificação. Nesses casos, quando se pretende avaliar fatores prognósticos, o registro e busca de reações adversas devem ser considerados.

As reações adversas ocorrem a partir de alterações fisiológicas que surgem da interação dos fármacos com as moléculas do organismo como receptores celulares,

enzimas e DNA. Após a administração da dose, o medicamento passa por várias etapas até a excreção. Essas etapas farmacocinéticas influenciam a concentração dos medicamentos no organismo, que está diretamente relacionada ao efeito do medicamento, podendo ser este de eficácia ou toxicidade. Para muitos medicamentos como os antineoplásicos utilizados para o tratamento do câncer, a dosagem máxima aplicada na terapêutica é estabelecida pelo limiar para o efeito tóxico [7, 8].

Os protocolos de quimioterapia baseados nos antineoplásicos, antraciclina e taxanos, como é o caso do tratamento quimioterápico FAC-D (3 ciclos de ciclofosfamida, doxorubicina e 5-fluorouracil seguidos de 3 ciclos de docetaxel), são amplamente utilizados para o tratamento de câncer de mama [9]. As antraciclina (doxorubicina e epirrubina) são antibióticos e os taxanos (docetaxel e paclitaxel) são inibidores mitóticos. Quando associados respondem por uma melhora na sobrevida livre de doença [10–12]. Porém, devido à ação inespecífica dos antineoplásicos e de seus metabólitos, eles podem agir em células de vários tecidos, como o epitélio de revestimento do aparelho gastrointestinal, tecido germinativo, tecido hematopoiético e folículo piloso, o que pode levar a reações hematológicas e não hematológicas. Dentre as reações hematológicas frequentes em tratamentos com antraciclina e taxanos podemos citar leucopenia, trombocitopenia, anemia e neutropenia. As reações não hematológicas comuns são fadiga/astenia, mialgia artralgia, febre, alopecia, hiperpigmentação, náusea, vômito, mucosite, diarreia, estomatite e anorexia [13, 14].

Embora as reações adversas a quimioterápicos citotóxicos sejam esperadas e indicadoras de sua eficácia, o principal desafio no cenário clínico é a grande variabilidade interindividual que pode alterar o risco de reações graves. A ocorrência de reações graves pode acarretar atrasos ou interrupção no tratamento decorrentes de internação ou tratamentos adicionais, e inclusive causar óbitos. As características individuais que podem afetar o risco e a gravidade das reações são:

**Idade:** As mudanças de funções metabólicas relacionadas à idade como limitações nas funções de absorção, distribuição e excreção do organismo podem aumentar a vulnerabilidade às reações adversas [15–17].

**Gênero:** Nas mulheres, as vias hormonais e a utilização de certos medicamentos podem estar associadas à maior incidência de reações [15, 17].

**Comorbidades:** Doença preexistente ou com ocorrência concomitante ao tratamento clínico para câncer, como por exemplo doenças hepáticas, renais e cardíacas podem afetar o metabolismo e resultar em acúmulo de medicamento no organismo, gerando reação.

**Interações medicamentosas:** Os fármacos utilizados concomitantemente podem alterar os efeitos uns dos outros, causando interações medicamentosas [15].

**Alergias:** Indivíduos com histórico de alergia são mais suscetíveis à ocorrência de reações [18, 19].



Fatores genéticos: Variações em genes como por exemplo, polimorfismos, podem aumentar o risco para reações [18].

Em geral, os fatores genéticos respondem por 15 a 30% das diferenças interindividuais no metabolismo dos fármacos sendo que, para certos fármacos, os fatores genéticos podem ser responsáveis por mais de 95% da variabilidade interindividual na disponibilidade do medicamento e consecutiva alteração da magnitude do seu efeito [20]. Dessa maneira, o tratamento com fármacos baseado no perfil genético do paciente pode resultar não apenas em melhor resposta ao tratamento, mas também na redução da incidência das reações adversas [5].

A notificação das reações ao longo do tratamento juntamente com o registro de características individuais possibilita o desenho de estudos com objetivo de mapear perfis de risco. A atribuição para os níveis das reações notificados deve seguir algum critério pré-estabelecido, e dependendo da escala utilizada, os níveis podem ir de assintomático a grave. Por isso, a variável que descreve a intensidade das reações geralmente é assumida ser uma variável categórica, o que deve ser levado em conta na escolha de técnicas estatísticas [21–23].

Para estudos transversais, existem diversos modelos propostos na literatura para variáveis categóricas [21, 24]. O modelo mais utilizado é o modelo de regressão logística ordinal cumulativo [25]. Existem tipos de modelos de logito cumulativo como o de chances proporcionais e o de chances proporcionais parciais. Outra maneira de relacionar as categorias, que não utiliza um logito não cumulativo, é através do modelo de categorias adjacentes [21, 24]. Esses modelos são adequados para avaliar a associação entre variáveis resposta ordinais e variáveis preditoras em estudos nos quais as observações são realizadas em um determinado instante de tempo para cada indivíduo.

Quando várias respostas de um mesmo indivíduo são obtidas ao longo de um período de tempo, como por exemplo respostas de um paciente a um tratamento, temos um estudo longitudinal. Neste caso é necessário utilizar ferramentas específicas para a modelagem, a fim de incorporar a heterogeneidade existente entre os indivíduos e a associação entre as várias respostas obtidas para um mesmo indivíduo ao longo do tempo, decorrente das medidas repetidas. Modelos lineares generalizados - MLG (McCullagh e Nelder, 1989) [25], que têm como pressuposto independência entre as observações não são apropriados para analisar dados longitudinais. Liang & Zeger (1986) [26] foram os pioneiros em propor uma maneira de tratar dados longitudinais utilizando MLG e efeitos aleatórios para tratar a correlação entre as medidas repetidas. A partir deste trabalho foram desenvolvidas metodologias para o estudo longitudinal de respostas ordinais como a metodologia de Equações de Estimativas Generalizadas (GEE), os modelos lineares generalizados mistos (GLMM) e os modelos de transição.

Esses modelos apresentam grande aplicabilidade em diversas áreas de pesquisa como agricultura, economia e biologia [27]. Na área epidemiológica por exemplo, há interesse frequente em avaliar o risco de eventos adversos. É usual classificar a resposta de interesse em duas ou mais categorias a fim de estimar a razão de chances ou o risco relativo [21, 28]. Em ensaios clínicos, respostas em escala ordinal são geralmente utilizadas para quantificar reações e sintomas de pacientes ou avaliar a eficácia de procedimentos pós-cirúrgicos [6]. A utilização de respostas ordinais tem ganhado bastante espaço em estudos longitudinais sobre qualidade de vida, indicadores de saúde e gravidade de doenças. Apesar dos modelos GEE, GLMM e modelos de transição serem bem conhecidos por estatísticos, eles ainda não são muito aplicados em pesquisas na área da saúde.

Os principais modelos utilizados para a identificação de fatores associados a reações adversas ao tratamento quimioterápico para o câncer de mama são: modelos de regressão logística [29–32], modelos de crescimento com mistura [33], modelos lineares multiníveis [34, 35] e modelos log-lineares [36]. Na literatura não existem registros de inclusão de interações entre fatores clínicos, polimorfismos e protocolos quimioterápicos como covariáveis nos modelos. A inclusão representaria um ganho na identificação de perfis relacionados ao desenvolvimento de reações, pois a natureza de múltiplas causas associadas às reações demanda um estudo integrado do paciente. Além disso, não existem registros de estudos para reações em protocolos com múltiplas fases como é o caso do protocolo FAC-D (fases FAC e Docetaxel). Como é sugerido em alguns estudos comparativos, a gravidade das reações pode ser afetada pelo tipo protocolo [15–17, 20], o que leva a hipótese de que mudanças de fase em protocolos podem alterar o nível das reações e o perfil de susceptibilidade.

Outra lacuna existente na literatura está relacionada a técnicas para avaliação da qualidade do ajuste e diagnóstico em modelos longitudinais com variáveis desfecho categóricas. Tais técnicas ainda são pouco conhecidas e utilizadas em estudos sobre reações adversas, sobretudo para quimioterapia para o câncer de mama, o que pode estar relacionado em parte, ao desenvolvimento relativamente recente de novas técnicas para avaliação de modelos para dados categóricos [21, 24]. No caso de estudos longitudinais, as ferramentas disponíveis para o tratamento de variáveis categóricas não são tão amplas como no caso de variáveis contínuas e ainda estão em desenvolvimento [22, 37].

Neste trabalho, as reações adversas estudadas foram: astenia, mialgia, artralgia, dor abdominal, mucosite e diarreia. Os níveis das reações foram considerados como variáveis categóricas de acordo com um instrumento muito utilizado na prática clínica para detecção de reações denominado CTCAE, do inglês (*Common Terminology Criteria Adverse Events*).

Um conjunto de 166 mulheres submetidas ao tratamento quimioterápico FAC-D

para o câncer de mama foram acompanhadas. Esse tratamento é composto por duas fases (FAC e Docetaxel) de três ciclos cada. Como as reações adversas relatadas pelas mulheres podem estar associadas a múltiplas causas como o próprio protocolo quimioterápico, fatores genéticos e clínicos, os relatos de diferentes níveis de toxicidade para diferentes protocolos [35, 38] levam a uma série de questionamentos sobre o impacto da mudança de quimioterápicos em fatores de risco encontrados na literatura. Mudanças de fase ao longo de um protocolo podem ser um fator de risco para as reações? Interações entre fases e variáveis clínicas ou genéticas podem ser fatores de risco para as reações? Fatores de risco para uma dada fase seriam fatores de risco em outra fase? Seria possível encontrar fatores de risco comuns em um tratamento em que há múltiplos fases?

Para responder a essas perguntas utilizamos ferramentas estatísticas para o estudo de dados longitudinais categóricos. Equações de estimativas generalizadas foram utilizadas para o ajuste de modelos marginais e modelos lineares generalizados mistos foram utilizados para introduzir o componente aleatório a nível do sujeito e verificar o quanto o componente aleatório influencia o ajuste e o conjunto das variáveis significantes.

Este trabalho está dividido em oito capítulos: No segundo capítulo, descrevemos os objetivos deste trabalho. No capítulo 3 realizamos um breve histórico sobre fatores de risco para reações adversas e sobre métodos para o tratamento de variáveis resposta categóricas. No quarto capítulo abordamos aspectos básicos sobre farmacologia e métodos e modelos comumente utilizados para o tratamento de variáveis categóricas. No capítulo 5, descrevemos os materiais e métodos utilizados. No capítulo 6 apresentamos os resultados do nosso estudo. No capítulo 7 realizamos a discussão. No capítulo 8, apresentamos as conclusões finais e propostas futuras.

# Capítulo 2

## Objetivos

O objetivo deste trabalho é comparar diferentes estratégias de modelagem longitudinal, quanto às estimativas e à qualidade do ajuste, utilizadas para identificação de características individuais (clínicas, genéticas e suas interações), relacionadas ao desenvolvimento de reações adversas ao longo da quimioterapia FAC-D para o câncer de mama.

Os objetivos secundários são avaliar de que maneira a mudança de fase altera a gravidade das reações e os perfis de susceptibilidade ao longo do tratamento, e investigar se existem características preditoras comuns às duas fases (FAC e Docetaxel).

# Capítulo 3

## Revisão Bibliográfica

O primeiro artigo sobre qualidade de vida em pacientes com câncer de mama foi publicado em 1974 [39]. Nesse artigo histórico, pacientes com câncer de mama em estágio avançado, recebendo adrenalectomia com quimioterapia, foram avaliados segundo suas respostas ao tratamento, sobrevida e qualidade de vida. Os resultados mostraram que em 64% dos pacientes, o tratamento paliativo resultou em um retorno à vida normal durante o tempo de sobrevida. Dois anos depois, em 1976, em um segundo trabalho ao abordar esse tema, Priestman e Baum [40] utilizaram a auto-avaliação dos pacientes para medir as reações ao tratamento em mulheres com câncer de mama em estágio avançado. Eles comprovaram que a técnica de auto-avaliação poderia ser utilizada para monitorar os efeitos adversos ao tratamento e comparar as toxicidades de diferentes terapêuticas. Os resultados também demonstraram que a toxicidade não estava relacionada à idade e diminuía ao longo de sucessivas administrações de drogas. No final dos anos 80, início dos anos 90, a literatura foi gradualmente complementada com trabalhos que utilizavam instrumentos para medir a qualidade de vida em pacientes com câncer de mama, o que contribuiu para difundir as técnicas de auto-avaliação como mecanismo de monitoramento de pacientes.

Montazeri [41] realizou uma revisão bibliográfica sobre a qualidade de vida em pacientes com câncer de mama. A revisão contemplou os trabalhos em língua inglesa publicados de 1974 a 2007, sendo examinados 477 trabalhos. Os resultados encontrados revelaram uma grande variedade de instrumentos para detecção de qualidade de vida nos estudos analisados. Foram encontrados impactos similares na qualidade de vida de vários procedimentos cirúrgicos apesar da mastectomia estar mais relacionada a depreciação da imagem corporal e sexualidade que o procedimento cirúrgico convencional. A maior parte dos estudos analisados indicaram que o tratamento quimioterápico está associado a mais reações adversas que afetam negativamente a qualidade de vida. Ansiedade e depressão foram reações comuns podendo ser reportadas inclusive anos após o tratamento. Fadiga, dores articulares e sintomas

de pós-menopausa também estavam entre as reações mais comuns relatadas por pacientes com câncer de mama. O reconhecimento e o monitoramento desses sintomas através de intervenções com aconselhamento, promovendo um suporte social e exercícios físicos, são fatores importantes e estavam associados a uma melhor qualidade de vida em muitos dos trabalhos analisados.

Os trabalhos sobre qualidade de vida citados acima foram fundamentais para a validação de instrumentos de mensuração que mais tarde deram origem aos critérios para a avaliação de reações adversas a tratamentos como o CTCAE (*Common Terminology Criteria Adverse Events*). Os Critérios de Terminologia Comuns para Eventos Adversos (CTCAE), anteriormente chamados de Critérios Comuns de Toxicidade (CTC ou NCI - CTC), são um conjunto de critérios para a classificação padronizada de efeitos adversos a drogas utilizadas na terapia de câncer. O sistema CTCAE é um produto do Instituto Nacional do Câncer dos EUA (NCI). Atualmente, a versão em uso é a 4.0, lançada em 2009. A maioria dos testes de drogas dos EUA e do Reino Unido codifica as suas observações com base neste sistema, que tem uma gama de níveis de 1 a 5 conforme a orientação geral: 1 - Leve, 2 - Moderado, 3 - Forte, 4 - Com risco de vida, 5 - Fatal.

Neste trabalho utilizamos o CTCAE para a avaliação das reações adversas: astenia, mialgia, artralgia, dor abdominal, mucosite e diarreia, ao longo do tratamento FAC-D. Os tratamentos quimioterápicos de uma forma geral causam toxicidades esperadas pois atingem as células do organismo como um todo, além das células cancerosas. Tecidos de rápida regeneração como por exemplo células da medula óssea, gônadas e folículos pilosos são particularmente afetados pela quimioterapia e desencadeiam reações esperadas como alopecia, variações em células sexuais e algumas reações sanguíneas. Algumas das reações mais comuns são neutropenia, leucopenia, anemia, náusea, vômito, mucosite, fadiga/astenia, mialgia e artralgia [13, 14].

A variabilidade das reações em tratamentos quimioterápicos pode estar associada a diversos fatores relacionados a variações interindividuais. Por exemplo, variações fisiológicas como idade, IMC; variações patológicas, como presença de comorbidades; variações de hábitos de vida como tabagismo e consumo de álcool e variações genéticas como presença de polimorfismos [15–18]. Além disso, é possível que ocorram alterações na gravidade das reações ao longo do tratamento decorrentes de variações nas dosagens e tipos de quimioterápicos utilizados.

Como o objetivo deste trabalho é identificar possíveis perfis associados a reações adversas, assumindo como perfil o conjunto e interações entre fatores clínicos, genéticos e protocolos quimioterápicos, a revisão realizada neste capítulo pautou-se em artigos que relacionavam o efeito das possíveis combinações entre essas variáveis para as reações adversas. O que foi encontrado na literatura e será exposto a seguir,

foram trabalhos que tratavam separadamente de polimorfismos, variáveis clínicas e protocolos quimioterápicos. Em particular, foram encontrados estudos que relatam que polimorfismos podem ter efeitos diferenciados na metabolização de diferentes tipos de quimioterápicos. Esse resultado é importante para o nosso estudo, em que há mudança de protocolo, devido ao fato de interações entre polimorfismos e protocolos poderem ter desdobramentos em metabolização e, conseqüentemente, em reações adversas. Além disso, como um dos objetivos específicos deste estudo é comparar metodologias para variáveis categóricas, foi incluído nesta revisão um breve histórico sobre metodologias utilizadas e avanços para o tratamento estatístico deste tipo de variável.

### **3.1 Polimorfismos genéticos, alterações farmacocinéticas e reações adversas**

Com relação aos fatores genéticos, variações genotípicas individuais podem causar impacto na sensibilidade aos antineoplásicos e na capacidade de distribuição, metabolização e excreção, afetando a eficácia e a segurança dos tratamentos quimioterápicos [42].

Alguns trabalhos relacionam a presença de polimorfismos a uma dificuldade maior de metabolismo do quimioterápico Docetaxel quando comparado a outra fase [? ]. Portanto, a interação entre esses dois fatores pode implicar a ocorrência de reações adversas a níveis graves em tratamentos em que há mudança de protocolo. Essa é uma importante interação a ser testada nos modelos, que não observamos nos estudos descritos na literatura e foi incluída em nosso trabalho. Alguns estudos citados abaixo relacionam o metabolismo de quimioterápicos à presença dos polimorfismos:

Marre et al, em 1996, [43] demonstraram que em culturas de hepatócitos e microsomas hepáticos humanos, inibidores típicos da subfamília CYP3A, como eritromicina, cetoconazol, nifedipina, midazolam e troleandomicina, inibiram o metabolismo do docetaxel. Na presença de indutores clássicos desta subfamília, como a rifampicina e a dexametasona, o metabolismo do docetaxel foi induzido, sugerindo que a subfamília CYP3A é a responsável pelo metabolismo do docetaxel. Além disso, os autores encontraram uma correlação estatisticamente significativa (0,7698;  $p < 0,0001$ ) entre a taxa de metabolismo do docetaxel e a atividade eritromicina N-demetilase, um marcador da atividade CYP3A3/4. Este estudo in vitro mostrou que a variabilidade interindividual da oxidação do docetaxel está relacionada com a variabilidade na expressão da subfamília CYP3A.

Shou et al [44], em 1998 identificaram que o docetaxel sofre metabolismo tanto pela isoforma CYP3A4 quanto pela CYP3A5, sendo que esta última apresenta menor afinidade por este agente antineoplásico. Miyoshi et al, em 2005, [45], analisaram a expressão de CYP3A4, por imunohistoquímica em células de biópsia de câncer de mama localmente avançado e recorrente, e avaliaram a relação com a resposta ao tratamento com docetaxel. Pacientes com tumores negativos para CYP3A4 apresentaram taxa de resposta maior do que os pacientes com expressão de CYP3A4 (n = 16) (67% e 19%, respectivamente p<0,01). A contribuição da atividade das enzimas CYP3A4 e CYP3A5 para o metabolismo do docetaxel foi investigada in vivo, com avaliação de parâmetros farmacocinéticos, como depuração, e de efeitos tóxicos. A presença dos alelos *CYP3A4\*1B* e *CYP3A5\*1A* estava associada com aumento de 64% na depuração do docetaxel (n=92, p=0,00015) independente de sexo e da atividade da CYP3A [46]. Goh et al [47], em 2002, examinaram a correlação entre o alelo CYP3A5\*3 (inativo) e a depuração do docetaxel em pacientes asiáticos diagnosticados com câncer, mas não encontraram associação significativa.

Tsai et al. [48] analisaram em 2009, os efeitos colaterais após o tratamento quimioterápico com docetaxel em pacientes com câncer de mama e suas associações aos polimorfismos CYP3A4, CYP3A5 e ABCB1. Pacientes com genótipo *CYP3A5\*1/s\*3* e *ABCB1 2677G/G* demonstraram maiores chances para níveis mais graves de febre e fadiga.

O transportador transmembranar ABCB1, também conhecido como PgP, foi primeiramente descrito em células neoplásicas, sendo identificado como a proteína responsável pela resistência ao tratamento com quimioterápico. Tal fenótipo é decorrente do efluxo de diversos fármacos mediados pela PgP. Esta proteína é codificada pelo gene *ABCB1*, que sofre ação de vários inibidores e indutores. O gene *ABCB1* apresenta mais de 100 polimorfismos, dos quais três (*C1236T*; *G2677A/T* e *C3435T*) têm sido alvo da maioria dos estudos por afetarem a atividade enzimática ou a especificidade por substratos [49].

Com relação à ciclofosfamida, o metabolismo parece ser dependente de *GSTP1* e de *CYP2B6*. Este gene, *CYP2B6*, codifica um membro da superfamília de enzimas do citocromo P450. As proteínas do citocromo P450 monooxigenases são as que catalisam diversas reações envolvidas no metabolismo do fármaco e da síntese de colesterol, esteróides e outros lípidos. A enzima é conhecida por metabolizar alguns xenobióticos, tais como a ciclofosfamida, drogas anti-cancerosas e ifosfamida.

Zhang et al [50], em 2011, analisaram o efeito de polimorfismos em *GSTP1* no metabolismo de ciclofosfamida para câncer de mama. Pacientes com genótipo *GSTP1 (105)Ile/Val* ou *(105)Ile/Ile* estavam associados a melhores respostas ao tratamento quimioterápico que as mulheres com genótipo *(105)Val/Val*. O metabolismo não estava relacionado com receptor de estrogênio, progesterona, idade, estágio



do tumor, status de menopausa e dosagem utilizada no tratamento.

Com relação às reações gastrointestinais, em um estudo de 2009, Rocha et al, [51] identificaram que a substituição de lisina por arginina na posição 262 na proteína CYP2B6 estava associada com aumento da mucosite oral ( $p= 0,00067$ ).

## 3.2 Características clínicas associadas a reações

Dentre as reações adversas estudadas neste trabalho, a fadiga/astenia é relatada como uma das mais incidentes em tratamentos quimioterápicos [29, 34, 35]. Em artigo de revisão da literatura de 2002, Jong et al. [34] selecionaram artigos publicados de 1989 a 2001 relacionados à fadiga em pacientes com câncer de mama submetidos à quimioterapia. Em muitos estudos avaliados pelos autores, fadiga permaneceu estável ao longo do tratamento quimioterápico, apesar do senso comum de que fadiga aumenta ao longo do tratamento. Os primeiros dois dias após a quimioterapia foram identificados como o pior período. A influência de fatores como dor, qualidade do sono e depressão foram identificados em vários estudos, apesar de não estar claro quando esses sintomas causam fadiga ou vice-versa. Fatores como mudança de peso, sintomas de menopausa, apoio social e familiar e mudanças bioquímicas foram mencionadas na literatura como variáveis associadas à fadiga. Foram relatados resultados conflituosos e a necessidade de mais estudos sobre o tema. Os modelos utilizados em grande parte dos trabalhos selecionados por eles foram modelos lineares multiníveis.

Em seu artigo de 2007, Donovan et al. [33] utilizaram um modelo de crescimento com mistura para identificar preditores para fadiga em um grupo de 261 mulheres. Ao utilizar o modelo de crescimento com mistura, eles consideraram a fadiga como variável contínua e extraíram dois grupos, um com baixos níveis fadiga e outro com altos níveis fadiga. Mulheres que não eram casadas estavam associadas a um baixo incômodo, alto índice de massa corporal e maior fadiga. Pouca atividade física também estava relacionada a maior fadiga.

Em 2012, Minton et al. [29] analisaram a ocorrência de fadiga em mulheres que completaram o tratamento quimioterápico para o câncer de mama. Foi utilizado o modelo de regressão logística para identificar os preditores. Eles encontraram como preditores para fadiga a ansiedade, a depressão, a dor e a insônia. Não foram encontradas associações entre variáveis demográficas e clínicas e fadiga.

Quanto à mialgia e artralgia, algumas referências apontam fatores de risco para ocorrência dessas reações [36, 52, 53]. Em 1998, Kunitoh et al. [52] abordaram a toxicidades neuromusculares, entre elas, mialgia e artralgia, em pacientes em tratamento quimioterápico. Nesse estudo, eles utilizaram o coeficiente de correlação de Spearman para determinar a correlação entre duas variáveis. Mialgia e artralgia

apresentaram associação com a dose cumulativa do quimioterápico paclitaxel, ministrado durante os ciclos. Idade, sexo, peso, tratamento quimioterápico anterior, função renal ou hepática e sítios de metástase não apresentaram correlação com toxicidade. Em 2003, Garrison et al [53] identificaram também a associação entre a dosagem e níveis mais graves de mialgia e artralgia mas, esse estudo foi focado em medicamentos para atenuar esses sintomas como os corticosteróides. Em 2013, Castel et al [36] realizam um estudo sobre artralgia em pacientes um ano após o tratamento quimioterápico. Através de modelos de regressão Poisson foram identificados fatores de risco para incidência de artralgia em mulheres em pós-menopausa. A gravidade dos sintomas de menopausa e a existência de comorbidades estavam associadas a níveis maiores de artralgia.

Para as reações gastrointestinais existem evidências de associações entre variáveis demográficas e a incidência dessas reações em alguns trabalhos que citaremos a seguir.

Stein et al. [30], em 1995, através de modelos de regressão logística, identificaram fatores associados à incidência de níveis graves de diarreia em quimioterapia com fluorouracil. Idade avançada e sexo foram identificados como preditores. Em 1998, McCarthy et al.[31] identificaram, através de modelos de regressão logística, que fumo, consumo de álcool, diabetes, higiene oral e herpes foram variáveis associadas à maior incidência de mucosite oral em pacientes em tratamento quimioterápico. Em 2007, Maroun et al. [32] realizaram um estudo em que identificaram fatores de risco associados à incidência de diarreia induzida pela quimioterapia. Segundo os resultados, os níveis de diarreia foram mais graves nos primeiros ciclos de quimioterapia. A duração do ciclo maior que três semanas, a coexistência de neutropenia e de outros sintomas como estomatite, vômito, anorexia e anemia foram preditores para o modelo de regressão logística. A incidência de grau 1 ou 2 de diarreia no primeiro ciclo foi um fator protetor.

### **3.3 Metodologias para tratamento de variáveis categóricas**

Os trabalhos citados mostram alguns casos de utilização de ferramentas estatísticas sofisticadas, desenvolvidas para análise de dados categóricos longitudinais [33, 34], mas a grande parte dos estudos utiliza ferramentas clássicas de uso consolidado na área de epidemiologia como regressão logística e testes não paramétricos [29–32, 54, 55].

A utilização de métodos estatísticos clássicos na maioria dos trabalhos que tratam de perfis de risco para reações adversas pode estar relacionada ao desenvolvimento

relativamente recente de técnicas para o tratamento de variáveis categóricas. Em particular, a análise de dados categóricos longitudinais teve início na década de 70 com os trabalhos de Harville (1977) e Laird & Ware (1982).

Laird & Ware utilizaram a classe mais geral de modelos lineares mistos, inicialmente descrita por Harville, para descrever uma classe de modelos lineares mistos para respostas contínuas. Eles desenvolveram modelos de efeitos aleatórios em que as observações repetidas ao longo do tempo para um indivíduo são medidas correlacionadas. As variações entre as medidas de um indivíduo são descritas por termos não observáveis que variam aleatoriamente entre os indivíduos, podendo ser acrescentado no modelo como um componente aleatório. Essa classe descrita por Laird & Ware incluía como casos particulares a ANOVA univariada para medidas repetidas e os modelos para curvas de crescimento em dados longitudinais (Fitzmaurice et al., 2009).

Foi em meados da década de 80, paralelamente ao desenvolvimento dos modelos lineares mistos para variáveis resposta contínuas, que um grande avanço surgiu na metodologia para analisar dados longitudinais ordinais. Após a publicação do trabalho de Liang & Zeger (1986), teve início uma série de publicações sobre modelos para estudos longitudinais, em particular estudos envolvendo modelos de quase-verossimilhança. No estudo de Liang & Zeger, eles propuseram uma forma relativamente simples de tratar dados longitudinais utilizando modelos lineares generalizados (MLG) ajustados através de Equações de Estimação Generalizadas (EEG). Essas equações de estimação são uma extensão natural do método de quase-verossimilhança proposto por Wedderburn (1974) para modelar respostas multivariadas em modelos lineares generalizados. A ausência de distribuições multivariadas não-gaussiana com a mesma estrutura de correlação da distribuição normal multivariada transformou o trabalho de Liang & Zeger em um dos mais citados da Estatística desde 1986. Nos últimos anos vários modelos que estendem os modelos lineares generalizados, inclusive incluindo efeitos aleatórios, foram propostos para modelar respostas longitudinais ordinais.

A escolha do método apropriado para o tratamento dos dados categóricos é de extrema importância para o desdobramento correto do estudo [24]. Em particular na área de epidemiologia, em que interpretações e resultados de modelos podem levar a melhorias em tratamentos e qualidade de vida dos pacientes.

# Capítulo 4

## Fundamentação teórica

### 4.1 Quimioterapia antineoplásica

A terapia quimioterápica antineoplásica para o câncer surgiu durante a segunda Guerra Mundial, a partir de uma tentativa de utilizar as propriedades mielossupressoras de um veneno usado na guerra (mostarda nitrogenada), para o tratamento do linfoma não-Hodgkin. Quase sete décadas após o fim da guerra, a quimioterapia antineoplásica evoluiu significativamente, possuindo nos dias atuais um grande variedade de agentes químicos e biológicos [56].

Em 1968 e 1975 ocorreram os primeiros registros dos benefícios da quimioterapia adjuvante sistêmica (administração de antineoplásicos após o procedimento cirúrgico) para o câncer de mama. Os trabalhos de Fisher [57, 58], NSABP B-01, que utilizou tiotepa no pós-operatório, e NSABP B-05, que utilizou antineoplásico melfalano durante 5 dias consecutivos, a cada 6 semanas, por 2 anos, relataram aumento na sobrevida dos pacientes. Com o uso do tiotepa, a taxa de sobrevida após 5 anos atingiu 57%, contra 24% no grupo que recebeu placebo. No estudo NSABP B-05, o tratamento falhou em 9,7% dos pacientes que receberam melfalano e no grupo que não recebeu o antineoplásico, a recorrência foi de 22%.

Bonadonna e colaboradores [59], em 1976, realizaram um estudo para avaliar o efeito da administração de uma combinação de ciclofosfamida, metotrexato e 5-fluoracil (CMF) mensalmente, por um ano, comparativamente a um grupo sem tratamento adjuvante submetido apenas à mastectomia radical. Dois anos e 4 meses após, verificou-se uma taxa de 24% de recorrência da doença em pacientes controle, enquanto que no grupo que recebeu a quimioterapia, a taxa de recorrência foi de 5,7% ( $p < 10^{-6}$ ). A partir deste estudo, o tratamento adjuvante com CMF foi consolidado para pacientes com câncer de mama inicial [60].

A substituição do metotrexato por antraciclina ocorreu após estudos na década de 1980 que demonstraram um ganho na sobrevida de 4% por 10 anos com regimes

com antraciclina quando comparados a regimes sem antraciclina [60].

A quimioterapia sistêmica ainda desempenha um importante papel no tratamento para o câncer de mama, já que as células tumorais da mama são muito sensíveis aos agentes disponíveis na quimioterapia. Uma meta-análise realizada por Laporte e colaboradores [61], em 2009, incluiu mais de 20.000 pacientes com acompanhamento de 5 anos, num total de 12 estudos. Este estudo demonstrou um claro benefício em termos de sobrevida livre de doença (Hazard ratio – HR = 0,81; IC95% = 0,74 – 0,89) e da sobrevida global (HR = 0,81; IC95% = 0,72 – 0,91), quando o docetaxel foi adicionado ao tratamento. Esses estudos foram importantes para consolidar os antineoplásicos associados a melhores resultados e validar o uso mais frequente de ciclofosfamida, 5-fluoruracil, doxorrubicina e docetaxel em tratamentos para o câncer de mama.

#### 4.1.1 Quimioterapia FAC-D

Neste estudo, avaliamos um conjunto de 166 mulheres submetidas ao tratamento FAC-D para o câncer de mama. O tratamento FAC-D é composto por 3 ciclos de 5-Fluorouracil 500 mg/m<sup>2</sup>, Doxorrubicina 50 mg/m<sup>2</sup> e Ciclofosfamida 500 mg/m<sup>2</sup> IV a cada 21 dias, seguidos de 3 ciclos de Docetaxel 100 mg/m<sup>2</sup> IV a cada 21 dias.

##### Descrição dos quimioterápicos

**5-fluorouracil:** O 5-fluorouracil (5-FU) é um análogo da pirimidina que age como antimetabólito da uracila, um componente do ácido ribonucleico. A 5-FU é convertida intracelularmente em seu metabólito ativo, desoxinucleotídeo, que interfere com a síntese de DNA (e em menor proporção na síntese de RNA) bloqueando a metilação do ácido desoxiuridílico para ácido timidílico, pela inibição da enzima celular timidilato sintetase. O 5-fluorouracil (5-FU) por inibir a conversão de deoxiuridina monofosfato (dUMP) para deotimidina monofosfato (dTMP) e formar um complexo ternário com a timidilato sintase (TS) inibe a síntese de DNA, como mostrado na figura.2 (AMSTUTZ et al., 2011). Além disso, a 5-FU pode também inibir a incorporação da timina no DNA. Uma vez que a timina é essencial para formação de DNA e RNA, a deficiência de timina criada pela 5-FU leva à morte celular. Os efeitos da 5-FU são, portanto, mais marcantes nos tecidos com maior taxa de divisão celular. As reações adversas mais incidentes são náusea (10 – 30%), vômito, estomatite, mucosite, leucopenia e neutropenia, que são mais comuns no período entre 9 e 14 dias após a administração [62].

**Doxorrubicina:** A doxorrubicina foi isolada de *Streptomyces peucetius*, uma espécie de actinobactéria, em 1960. Este fármaco é um membro da classe das antraciclina, usado no tratamento de tumores sólidos e hematológicos há mais de 40 anos. Seu pico de ação ocorre de 10 a 15 dias após administração intravenosa, con-

forme sugerido pelo Nadir (menor valor de contagem) de células do sangue, com recuperação da atividade hematopoiética após 21 dias.

Entre as reações adversas, destaca-se, pela gravidade, a cardiotoxicidade, que pode ocorrer tardiamente, após meses ou até anos, em casos de doses acumuladas acima de 450 a 550 mg/m<sup>2</sup>. Outras reações adversas são náuseas e vômitos graves e mucosite podendo progredir para úlceras [62].

**Ciclofosfamida:** A Ciclofosfamida é um dos agentes alquilantes mais usados para o tratamento de leucemia mielóide aguda e crônica, leucemia linfóide crônica, câncer de mama, linfoma de Burkitt, câncer cervical, doença de Hodgkin, estádios III e IV e lúpus.

O fármaco sofre ativação metabólica (hidroxilação) pelo sistema de oxidase de função mista do citocromo P450 hepático e sua ação depende de transportadores que facilitam a entrada dos metabólitos nas células-alvo. A ciclofosfamida é um inibidor não específico do ciclo celular, podendo atuar em qualquer fase do processo de divisão celular.

O metabólito sem efeito antineoplásico da ciclofosfamida, a 2-decloroetilciclofosfamida, assim como o composto menos tóxico cloroacetaldeído, são formados por uma reação de N-decloroetilação, pela ação enzimática das CYP3A4/3A5 nos hepatócitos. Entretanto, aproximadamente 70% da dose é oxidada na posição C4 pelas isoformas CYP2A6, 2B6, 2C8, 2C9, 2C19, 3A4 e 3A5, gerando o metabólito ativo 4-hidroxiciclofosfamida (4-OH-CPA) [63]. Outra via de detoxificação inclui a conjugação da ciclofosfamida com glutathione por várias glutathione S-transferases (GSTs; GSTA1, GSTM1, GSTP1 e GSTT1).

As reações adversas mais comuns relacionadas à ciclofosfamida são náusea, vômito, alopecia, imunossupressão e disfunção menstrual. Outras reações bem conhecidas em doses convencionais são cistite hemorrágica, pneumonia intersticial (inflamação dos septos alveolares) e retenção de líquidos, além de ser considerada teratogênica. O período do Nadir ocorre em torno do 9<sup>o</sup> ao 12<sup>o</sup> dias, com início da recuperação no 15<sup>o</sup> dia.

#### **Docetaxel:**

O docetaxel tem como mecanismo de ação impedir a despolimerização dos fusos mitóticos na divisão celular levando a célula a entrar em apoptose. Este fármaco foi aprovado em 1996 para o tratamento do câncer de mama metastático refratário ao tratamento com antraciclinas e mais tarde foi aprovado para tratamento de câncer de pulmão de células não pequenas em estágios III ou IV, refratário aos regimes com platinas [? ]. Este agente antineoplásico também apresenta atividade contra outros tipos de tumores como ovariano, tumores de cabeça e pescoço, melanoma maligno e leucemias [46]. A dose recomendada para pacientes em tratamento de câncer está na faixa de 60 a 100mg/m<sup>2</sup> em infusão intravenosa por 1 hora uma vez a

cada 21 dias [64]. O docetaxel se liga a proteína plasmática (>98%), principalmente glicoproteína ácida alfa-1 e albumina.

Uma importante limitação do uso do docetaxel é a ampla e não preditiva variabilidade interindividual da toxicidade (toxicidade hematológica grave) e da eficácia. A grande variabilidade interindividual na farmacocinética do docetaxel, nos parâmetros de depuração e, conseqüentemente, a exposição do fármaco, calculada através da área sob a curva (ASC) pode explicar as diferentes graduações das RAM [46, 64]. O metabolismo do docetaxel envolve uma oxidação inicial da cadeia lateral *tert*-butil pelas enzimas CYP3A4/5, levando à inativação do antineoplásico e redução do efeito terapêutico. O tempo de meia vida de eliminação é 11 horas [62]. Tran e colaboradores [65], em 2006, analisaram a farmacocinética do docetaxel em 58 pacientes com câncer de mama e mostraram associação entre os parâmetros de depuração e ASC aos genótipos de CYP3A4, CYP3A5, ABCB1 e GSTP1, com conseqüente alteração nos perfis de toxicidade. Quatro pacientes CYP3A4\*1B (\*1A/\*1B) que também eram CYP3A5\*1/\*3 apresentaram maior depuração e menor área sob a curva com dose normalizada do docetaxel do que pacientes com genótipo selvagem. Neste estudo de Tran e colaboradores [65], a neutropenia febril ocorreu com maior frequência em pacientes com genótipo GSTP1\*A/\*B do que em pacientes \*A/\*A ( $p=0,037$ ), enquanto a neutropenia de grau 3 ocorreu com mais frequência em pacientes que apresentaram o genótipo ABCB1 3435TT do que nos outros genótipos ( $p=0,046$ ).

A figura 4.1.1 mostra as enzimas metabólicas e transportadores envolvidos na farmacocinética do docetaxel.

O docetaxel é um substrato para bomba de efluxo Glicoproteína-P (Pg-P), também chamada de ABCB1 ou de MDR1, transportador responsável pelo fenótipo de resistência a múltiplos fármacos [65]. O docetaxel também é substrato do transportador ABCC2 ou MRP2, conhecido como Proteína 2, expressa na membrana canalicular dos hepatócitos e na membrana apical das células endoteliais do túbulo proximal renal e também associada à resistência a múltiplos fármacos [66]. O transporte de docetaxel para o interior dos hepatócitos é realizado pelo transportador SLCO1B3 que se localiza na membrana basolateral e que está associado com aumento de reações adversas do docetaxel [67]. As reações adversas mais incidentes são astenia grave (61 a 66%), alterações neurosensoriais graves (parestesia, disestesia e dor), anormalidades nas unhas (11,4 a 18,5%), alopecia (56,3 a 97,8%), prurido, rash cutâneo, retenção de fluido (20,1 a 35,1%), diarreia (32,8 – 42,6%), náusea (38,8 – 80,5%), vômito (22,3 – 44,5%), estomatite (41 a 51%), febre, anemia (89,1 – 93,6%), anemia grave (< 8g/dL de hemoglobina) (4,3 – 9,2%), leucopenia (<1000 leucócitos/mm<sup>3</sup>) (31,6 – 43,7%), neutropenia (< 500 neutrófilos/mm<sup>3</sup>) (75,4 – 85,6%), neutropenia febril (5,2 – 24,7%), edema (47 a 64,1%) e efusão pleural [62].

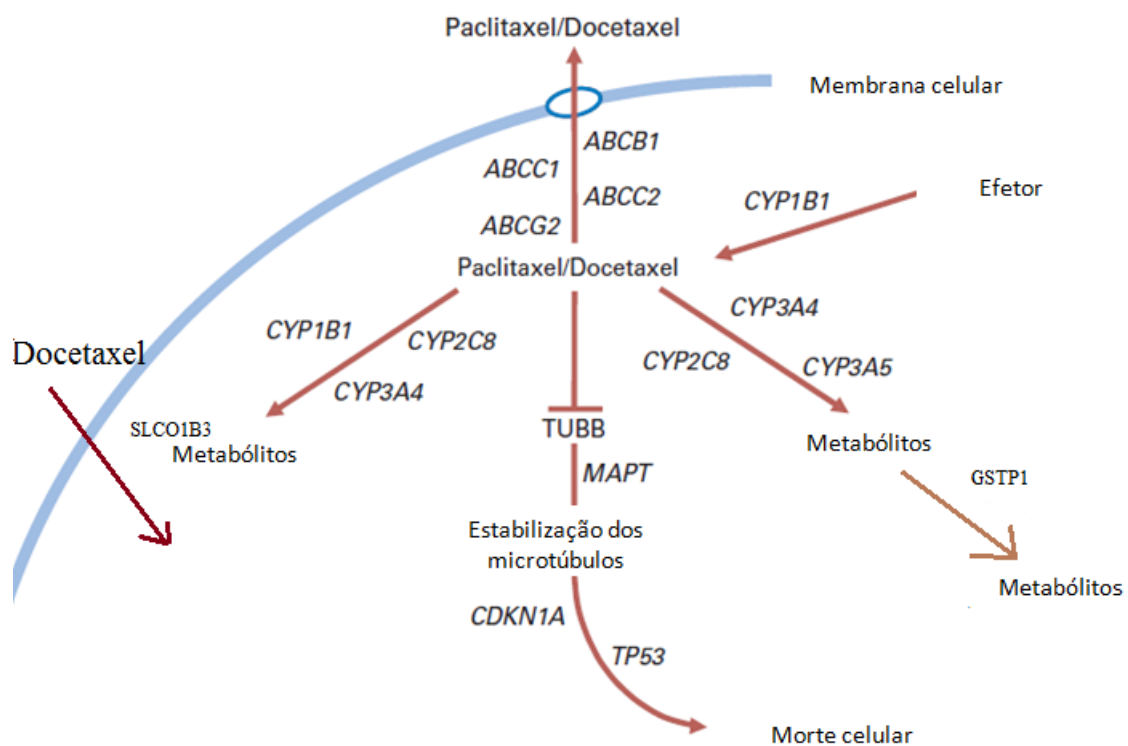


Figura 4.1: Via metabólica do docetaxel. Adaptado de KIYOTANI et al, 2008; Marsh et al., 2009. Legenda: TUBB - cadeia de beta tubulina; MAPT - gene tau proteína associada a microtúbulo; CDKN1A - gene do inibidor kinase dependente de ciclina; TP53 - gene da proteína tumoral P53.



## 4.2 Farmacogenética e quimioterapia FAC-D

A farmacogenética estuda as influências genéticas sobre a farmacocinética e farmacodinâmica com intuito de individualizar a terapia por meio da identificação de alterações genéticas relacionadas a eficácia e a segurança dos medicamentos. Informações genéticas são utilizadas na identificação do risco de doença, na escolha de agentes de tratamento e na definição da posologia [42].

A farmacocinética trata dos processos de absorção (acesso do medicamento à circulação sanguínea sistêmica); distribuição (transferência do medicamento do sangue para os tecidos); e eliminação (exclusão do medicamento do organismo), seja por excreção (na urina, nas fezes, no leite materno etc.) ou por biotransformação (alterações químicas da molécula do medicamento por ação de enzimas orgânicas). A farmacodinâmica trata das interações do medicamento com seus alvos, denominados receptores [68]. A figura 4.2 mostra as etapas de um fármaco no organismo.

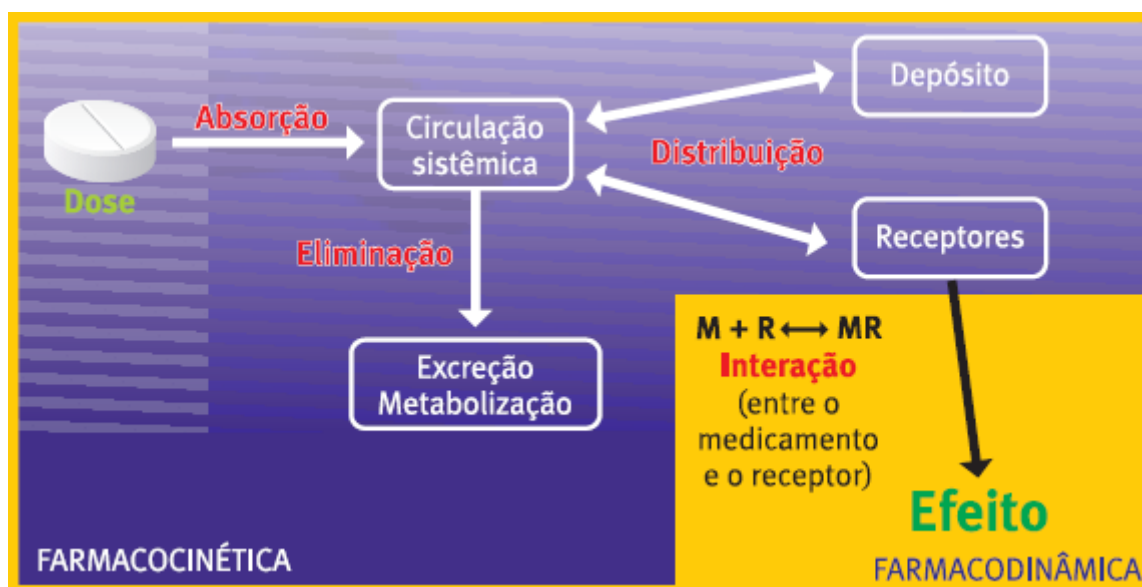


Figura 4.2: Etapas de um fármaco no organismo. Adaptado de Suarez-Kurtz, 2004.

A farmacogenética tem um papel importante na redução do “fator surpresa” associados com a ocorrência das reações adversas sendo usada na prática clínica para auxiliar os profissionais de saúde na terapia personalizada [69]. Em geral, os fatores genéticos respondem por 15 a 30% das diferenças interindividuais no metabolismo dos fármacos e sua resposta, sendo que, para certos fármacos, os fatores genéticos podem ser responsáveis por mais de 95% da variabilidade interindividual na disponibilidade do medicamento e consecutiva alteração da magnitude do seu efeito. Assim, o tratamento com fármacos baseado no perfil genético do paciente pode resultar não apenas em melhor resposta ao tratamento, mas também na redução da incidência das reações adversas [5].

Diversos estudos são realizados na tentativa de identificar fatores de risco associados à ocorrência de complicações durante o tratamento oncológico. Variações individuais na sensibilidade aos fármacos antineoplásicos ou na capacidade de metabolização dos mesmos podem afetar a eficácia e a segurança de tratamentos quimioterápicos, já que esses fármacos possuem janela terapêutica estreita, com potencial risco de toxicidade [42]. No âmbito da farmacogenética, informações genotípicas podem abranger os SNP, haplótipos, sequências microssatélites ou sequências de repetição (loci polimórficos presentes em DNA que consiste em unidades repetitivas de 1 a 6 pares de bases), inserção e/ou deleção e números de cópias variantes [42]. Apresentaremos agora uma breve descrição sobre as informações genéticas utilizadas neste trabalho.

### **CYP3A5:**

O gene *CYP3A5* faz parte da subfamília *CYP3A*, juntamente com *CYP3A4*, *CYP3A7* e *CYP3A43* e 3 pseudogenes *CYP3AP1*, *CYP3AP2* e *CYP3AP3*, todos localizados no cromossomo 7 posição q21-g22.1. Nesta subfamília, as enzimas com relevância clínica são CYP3A4, CYP3A5 e CYP3A7 devido a uma expressão hepática predominante e semelhantes especificidades aos substratos, o que dificulta a avaliação de suas contribuições individuais. A enzima CYP3A7 é expressa, predominantemente, na fase fetal, tendo sua expressão diminuída após o nascimento, enquanto que com a CYP3A4 ocorre o contrário. A expressão hepática da enzima CYP3A4 varia em até 40 vezes, sendo que essa variância pode ser atribuída a fatores ambientais (fumo, dieta e medicação), hormonais ou fatores genéticos [70, 71].

Marre e colaboradores [43] (1996) mostraram que culturas de hepatócitos e microsossomos hepáticos humanos, na presença de inibidores típicos da subfamília CYP3A, como eritromicina, cetoconazol, nifedipina, midazolam e troleandomicina, inibiram o metabolismo do docetaxel. Na presença de indutores clássicos desta subfamília, como a rifampicina e a dexametasona, o metabolismo do docetaxel foi induzido, sugerindo que a subfamília CYP3A é a responsável pelo metabolismo do docetaxel. Além disso, os autores encontraram uma correlação estatisticamente significativa (0,7698;  $p < 0,0001$ ) entre a taxa de metabolismo do docetaxel e a atividade eritromicina N-demetilase, um marcador da atividade CYP3A3/4. Este estudo in vitro mostrou que a variabilidade interindividual da oxidação do docetaxel está relacionada com a variabilidade na expressão da subfamília CYP3A. Em 1998, Shou e colaboradores [44], identificaram que o docetaxel sofre metabolismo tanto pela isoforma CYP3A4 quanto pela CYP3A5, sendo que esta última apresenta menor afinidade por este agente antineoplásico. Miyoshi et al [45], em 2005, analisaram a expressão de CYP3A4, por imunohistoquímica em células de biópsia de câncer de mama localmente avançado ( $n = 21$ ) e recorrente ( $n = 10$ ), e avaliaram a relação com a resposta ao tratamento com docetaxel. Pacientes com tumores negativos para

CYP3A4 (n = 15) apresentaram taxa de resposta maior do que os pacientes com expressão de CYP3A4 (n = 16) (67% e 19%, respectivamente  $p < 0,01$ ). A contribuição da atividade das enzimas CYP3A4 e CYP3A5 para o metabolismo do docetaxel está sendo investigada in vivo, com avaliação de parâmetros farmacocinéticos, como depuração, e de efeitos tóxicos. A presença dos alelos *CYP3A4\*1B* e *CYP3A5\*1A* foi associada com aumento de 64% na depuração do docetaxel (n=92, p=0,00015) independente de sexo e da atividade da CYP3A [46]. Goh e colaboradores [47], em 2002, examinaram a correlação entre o alelo CYP3A5\*3 (inativo) e a depuração do docetaxel em pacientes asiáticos diagnosticados com câncer, mas não encontraram associação significativa.

### **CYP2B6:**

Lang e colaboradores [72], em 2001, foram os primeiros a descrever nove polimorfismos no gene CYP2B6, dos quais cinco resultam em trocas de aminoácidos: *CYP2B6\*2* (R22C), *CYP2B6\*3* (S259R), *CYP2B6\*4* (K262R), *CYP2B6\*5* (R487C), *CYP2B6\*6* (Q172H e K262R), e *CYP2B6\*7* (Q172H, K262R e R487C). As isoformas *CYP2B6\*4*, *CYP2B6\*5* e *CYP2B6\*6* resultam em alteração da atividade enzimática. Em 2009, Rocha e colaboradores [51], identificaram que a substituição de lisina por arginina na posição 262 na proteína CYP2B6 estava associada com aumento da mucosite oral (p= 0,00067).

### **CBR3:**

A enzima limitante da metabolização da doxorrubicina é a CBR, que pode apresentar as isoformas 1 e 3. As duas isoformas são estudadas, porém somente o SNP G11A do gene *CBR3* está associado com RAM. Em 2008, Fan e colaboradores [73] estudaram a farmacocinética e a farmacodinâmica de 101 pacientes asiáticas com câncer de mama em uso de doxorrubicina. O variante CBR3 11G<sub>1</sub>A foi associado com menor razão ASCdoxorrubicinol/ASCdoxorrubicina (p = 0,009, GG vs AA) e menor expressão de CBR3 nos tecidos tumorais (p = 0,001, GG vs AA), maior redução do tamanho do tumor (p = 0,0015, GG vs AA) e maior percentual de redução de leucócito e plaquetas no período de Nadir (teste de tendência  $p \leq 0,03$ ).

### **GSTP1:**

Uma das enzimas mais importantes da fase II do metabolismo é a Glutathione S-transferase (GST). Há quatro subfamílias principais de genes:  $\alpha$ , localizada no cromossomo 6;  $\mu$ , localizada no cromossomo 1;  $\theta$ , localizada no cromossomo 22 e  $\phi$ , localizada no cromossomo 11 [74]. Cada subfamília é composta de vários genes, alguns com polimorfismos genéticos. A enzima GSTP1 parece atuar como fator preditor para a ocorrência de RAM ao tratamento com ciclofosfamida e docetaxel. A subfamília GSTP da classe  $\phi$  tem um gene polimórfico denominado *GSTP1*, presente nas células humanas [75]. A enzima GSTP1 é a GST mais amplamente distribuída [75], sendo particularmente abundante no pulmão, no esôfago e na placenta

[76]. Um dos polimorfismos do gene *GSTP1*, no códon 105, leva à substituição do aminoácido isoleucina (Ile) pela valina (Val) no sítio ativo da enzima (GSTP1I105V) [77]. O alelo Val é associado à menor atividade quando comparado ao alelo Ile [78]. Assim, o genótipo GSTP1 Ile/Ile (AA) apresenta maior atividade enzimática quando comparado com os genótipos GSTP1 Ile/Val (AG) ou Val/Val (GG). Zhong e colaboradores [79], em 2006, estudaram 102 pacientes chineses de etnia Han, com lúpus eritematoso sistêmico, que foram tratados com 500 a 750mg/m<sup>2</sup> de ciclofosfamida IV, em bolus em associação com MESNA. Antes e após a aplicação de ciclofosfamida, foi realizado o monitoramento das RAM hematológicas, do trato gastro intestinal (TGI) e a incidência de infecção. A leucopenia foi avaliada entre os dias 7 e 10 após a administração. Após análise por regressão múltipla, os autores identificaram que a presença de *GSTP1 Ile105Val* ou *GSTP1 Val105Val* é fator de risco para mielosupressão e RAM do TGI (OR=5,7; IC95% 2,2 – 14,7 e OR= 3,8; IC95%= 1,4 – 10,0, respectivamente) em comparação ao genótipo selvagem.

#### **ABCB1:**

O transportador transmembranar ABCB1, também conhecido como PgP, foi primeiramente descrito em células neoplásicas, sendo identificado como a proteína responsável pela resistência ao tratamento com quimioterápico. Tal fenótipo é decorrente do efluxo de diversos fármacos mediados pela PgP. Esta proteína é codificada pelo gene *ABCB1*, que sofre ação de vários inibidores e indutores como destacado na tabela 6. O gene *ABCB1* apresenta mais de 100 polimorfismos, dos quais três (*C1236T*; *G2677A/T* e *C3435T*) têm sido alvo da maioria dos estudos por afetarem a atividade enzimática ou a especificidade por substratos [49].

Apesar de dois polimorfismos no gene *ABCB1*- *C1236T* e *C3435T*- serem silenciosos (ou seja, a substituição das bases nucleotídicas não promove mudanças na sequência dos aminoácidos durante a formação da proteína ABCB1), observam-se mudanças na estabilidade dos RNAm, na quantidade de proteína traduzida e na função do transportador [80]. Sugere-se que como os códons gerados utilizam RNAt que são menos abundantes, os resultados são proteínas com conformações diferentes podendo alterar a afinidade pelos seus substratos [81]. Segundo Lal e colaboradores [82], a presença dos SNP *ABCB1 C1236T*, *G2677T/A* e *C3435T* diminui a depuração da doxorubicina. Quanto aos trabalhos envolvendo pacientes em uso do docetaxel, Bosch e colaboradores [?] observaram associação entre a presença do variante *1236T* e a diminuição da depuração, Tsai e colaboradores [48] observaram que o homocigoto selvagem *2677GG* está associado a maior incidência de febre e neutropenia febril, e Tran e colaboradores verificaram aumento do risco de neutropenia grau 3 para o homocigoto variante *3435TT* [65].

#### **SLCO1B3:**

O transportador SLCO1B3 pertence à família de transportadores orgânicos *orga-*

*nic anion transporter 1* (OAT1) e é codificada pelo gene *SLCO1B3*. Esse transportador é exclusivamente expresso na região basal do hepatócito, onde é responsável pelo influxo de várias drogas aniônicas, incluindo xenobióticos e compostos endógenos, tais como ácidos biliares e bilirrubina [83].

Apresentaremos agora os aspectos teóricos das metodologias estatísticas que utilizamos para a realização das análises.

### 4.3 Variáveis categóricas

O desenvolvimento de novos métodos para análise de dados categóricos tem sido influenciado pela necessidade crescente de avaliação de variáveis categóricas em várias áreas, como ciências sociais, biomédicas, saúde pública, ecologia, educação e controle de qualidade industrial. O surgimento de técnicas para o tratamento desse tipo de variável teve início em 1960 e continua em evolução. Apenas recentemente essas técnicas têm alcançado níveis de sofisticação semelhantes a metodologias para o tratamento de variáveis contínuas [24, 37].

Uma variável categórica tem como escala de medida um conjunto de categorias. Escalas categóricas estão frequentemente presentes em saúde pública. Por exemplo, o diagnóstico para câncer de mama baseado em mamografia utilizando categorias normal, benigno, suspeito e maligno. Resposta ao tratamento utilizando categorias satisfatória e insatisfatória. Avaliação da qualidade de vida por meio dos níveis ruim, satisfatório, bom e ótimo.

Variáveis categóricas podem ser classificadas como nominais, quando as categorias não possuem nenhum tipo de ordenação natural como estado civil (solteiro, casado ou viúvo), presença de comorbidades (sim ou não), vínculo empregatício (empregado ou desempregado) ou ordinais, quando as categorias possuem algum tipo de ordenação como nível de escolaridade (fundamental, médio e superior), classe social (baixa, média e alta), qualidade de vida (ruim, satisfatório, bom e ótimo). A maneira como a variável é classificada determina os métodos estatísticos apropriados para as análises.

De acordo com Agresti, [24] existe uma hierarquia para variáveis categóricas. Variáveis ordinais são de nível intermediário e variáveis nominais são de níveis inferiores e possuem menos ferramentas para análises e modelagem. Métodos estatísticos adequados para variáveis de um nível podem ser utilizados para níveis superiores mas, de maneira geral, é aconselhável aplicar métodos apropriados para cada tipo de variável.

Variáveis categóricas podem também ser classificadas como dependentes (variável resposta) ou independentes (variável explicativa). A relação entre variáveis dependentes e independentes pode ser descrita por meio de modelos probabilísticos. A

família de modelos lineares generalizados traz importantes ferramentas para modelagem de variáveis categóricas.

## 4.4 Modelos lineares generalizados

Modelos lineares generalizados (MLG) são uma extensão do modelo linear clássico que contemplam variáveis resposta com distribuições não normais. Nelder e Wedderburn em 1972 [84], introduziram os MLG e mostraram que uma série de modelos comumente estudados separadamente podem ser reunidos sob a família MLG. Alguns exemplos de MLG são: modelo clássico de regressão com erro normal, modelo clássico de análise de variância e covariância com erro normal, modelo logístico para proporções, modelo de regressão de Poisson, modelo de Box & Cox e alguns modelos de séries temporais.

Um modelo linear clássico é da forma:

$$y_i = \mathbf{x}_i' \beta + \epsilon_i \quad (4.1)$$

onde  $y_i$  é o valor observado da variável resposta  $Y_i$  para a  $i$ -ésima unidade de investigação, ( $i=1,2,\dots,n$ ),  $\mathbf{x}_i$  é um vetor coluna de covariáveis ou variáveis explicativas para a unidade de investigação  $i$ , e  $\beta$  é o vetor de parâmetros desconhecidos a ser estimado. Os  $\epsilon_i$  são assumidos variáveis independentes, aleatórias, com distribuição normal com média zero e variância constante ( $E(\epsilon_i) = 0$ ,  $Var(\epsilon_i) = \sigma^2$ ). O valor esperado de  $Y_i$ , denotado por  $\mu_i$  é dado por:

$$E(Y_i) = \mu_i = \mathbf{x}_i' \beta \quad (4.2)$$

No entanto, existem alguns problemas para os quais o modelo linear tradicional não é apropriado. Por exemplo, na modelagem de contagens e proporções, a suposição de normalidade pode não ser adequada, assim como a homocedasticidade.

Nesses casos, o modelo linear generalizado pode ser utilizado para modelar a relação entre as variáveis. O modelo linear generalizado descreve a dependência entre a resposta média  $\mu_i = E(Y_i)$  e as variáveis explicativas através de três componentes. Um componente aleatório, um componente sistemático e uma função de ligação.

a) O componente aleatório de um MLG consiste de uma variável resposta  $Y$  com observações independentes ( $y_1, y_2, \dots, y_N$ ) de uma mesma distribuição na família exponencial. Essa família tem função de densidade de probabilidade da forma:

$$f(y_i, \theta_i, \phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi)\} \quad (4.3)$$

Muitas distribuições pertencem à família exponencial como as distribuições Poisson e Binomial. O parâmetro  $\theta_i$  é conhecido como parâmetro natural ou canônico. A função  $a(\phi)$  é tal que  $a(\phi) = \phi/w_i$  onde  $w_i$  é um peso conhecido,  $\phi$  é o parâmetro de dispersão e  $b(\cdot)$  e  $c(\cdot)$  são funções conhecidas.

b) O componente sistemático de um MLG especifica as variáveis explicativas utilizadas como função do preditor linear. O componente sistemático portanto, relaciona um vetor  $(\eta_1, \dots, \eta_N)$  às variáveis explicativas através de uma combinação linear:

$$\eta_i = \sum_j x_{ij}\beta_j, \quad (4.4)$$

onde  $x_{ij}$  é o valor do preditor  $j$ ,  $j = (1, \dots, p)$  para a unidade  $i$  e  $\beta_j$  é o  $j$ -ésimo componente do vetor de parâmetros  $\beta = (\beta_1, \dots, \beta_p)'$ . Essa combinação linear de variáveis explicativas é denominada preditor linear.

c) A função de ligação é uma função que relaciona  $E(Y)$  ao componente sistemático. Seja  $\mu_i = E(Y_i)$ ,  $i = 1, \dots, N$ , a função de ligação  $g$  é uma função monótona e diferenciável tal que  $\eta_i = g(\mu_i)$  ou ainda:

$$g(\mu_i) = \sum_j x_{ij}\beta_j, \quad (4.5)$$

Quando o parâmetro natural e o preditor linear coincidem, a função de ligação é chamada função de ligação canônica. Neste caso especial tem-se que

$$\theta_i = \eta_i = \sum_{j=1}^p x_{ij}\beta_j, \quad (4.6)$$

Os estimadores para os coeficientes de um modelo linear generalizado são encontrados via método da máxima verossimilhança, o que traz um ganho com relação ao processo de inferência uma vez que permite a construção de testes de hipóteses para os parâmetros e para a qualidade do ajuste do modelo.

#### 4.4.1 Estimação de parâmetros

A família exponencial tem a propriedade que seus elementos satisfazem condições de regularidade suficientes para assegurar que o máximo global do logaritmo da função de verossimilhança possa ser encontrado através de um sistema de equações. Dessa maneira podemos encontrar os estimadores de máxima verossimilhança do modelo.

Para um modelo linear generalizado com respostas independentes, o logaritmo da função de verossimilhança é dado por:

$$L(\theta; y) = \sum_{i=1}^N y_i - b(\theta_i)/a(\phi) + \sum_{i=1}^n c(y_i, \phi) \quad (4.7)$$

Os estimadores de máxima verossimilhança para um modelo linear generalizado são dados pela solução do sistema de equações:

$$U(\beta) = \frac{\partial L}{\partial \beta} = 0 \quad (4.8)$$

Temos então, que a função escore é dada por:

$$U_j(\beta_j) = \sum_{i=1}^N \frac{\partial L}{\partial \beta_j} = \sum_{i=1}^N \frac{\partial l(\theta_i, y_i)}{\partial \beta_j}, \quad j = 1, \dots, p \quad (4.9)$$

mas,  $L = \log(f(\theta_1, \dots, \theta_N))$  onde  $\theta_i = \int V_i^{-1} d\mu_i = h(\mu_i)$ ,  $V_i$  é a função de variância,  $h(\mu_i)$  é uma função conhecida de  $\mu_i$  e,  $\mu_i = g^{-1}(\eta_i)$ . Portanto  $\eta_i = \sum_{j=1}^p x_{ij}\beta_j$ .

Pela regra da cadeia temos:

$$U_j(\beta_j) = \sum_{i=1}^N \frac{\partial l}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}, \quad j = 1, \dots, p \quad (4.10)$$

onde  $\partial l / \partial \theta_i = [y_i - b'(\theta_i)] / a_i(\phi) = [y_i - \mu_i] / a_i(\phi)$  e  $a_i(\phi) = \phi / w_i = \partial \mu_i / \partial \theta_i = \partial^2 b(\theta_i) / \partial \theta_i^2 = V_i$  com  $\partial \mu_i / \partial \eta_i$  dependendo da função de ligação  $g$  e  $\partial \eta_i / \partial \beta_j = x_{ij}$ .

Substituindo as derivadas em 4.12 temos que os estimadores de máxima verossimilhança para um modelo linear generalizado podem ser encontrados através da equação escore:

$$U_j(\beta_j) = \sum_{i=1}^N \frac{\partial \mu_i}{\partial \eta_i} \frac{[y_i - \mu_i] x_{ij}}{a_i(\phi) V_i} = 0, \quad j = 1, \dots, p \quad (4.11)$$



ou na forma matricial:

$$U(\beta) = \frac{\partial L(\beta, y)}{\partial \beta} = \phi^{-1} X' W^{1/2} V^{-1/2} (y - \mu) \quad (4.12)$$

onde  $X$  é uma matriz  $N \times p$ ,  $W = \text{diag}(W_1, W_2, \dots, W_N)$  é a matriz de pesos com componentes  $W_i = (\partial \mu_i / \partial \eta_i)^2 w_i / V_i$ ,  $V = \text{diag}(V_1, \dots, V_N)$ ,  $y = (y_1, \dots, y_N)'$ , e  $\mu = (\mu_1, \dots, \mu_N)'$ .

As equações de score em geral são não lineares, e nesse caso são resolvidas por processos iterativos como os de Newton-Raphson.

Dentre os modelos lineares generalizados um dos mais utilizados em estudos epidemiológicos é o modelo de regressão logística. Para variáveis resposta com mais de duas categorias, existem extensões do modelo de regressão logística para as variáveis nominais e ordinais. Apresentaremos a seguir o modelo de regressão logística e algumas de suas extensões que utilizaremos neste trabalho.

#### 4.4.2 Regressão logística

A regressão logística é um dos modelos lineares generalizados mais difundido na literatura para a modelagem da relação entre respostas binárias e uma ou mais variáveis explicativas que podem ser qualitativas ou quantitativas.

Consideremos uma amostra de  $n$  indivíduos, para cada um dos quais existe uma resposta associada  $y_i$  dada por:

$$y_i = \begin{cases} 1 & \text{se a resposta para o } i\text{-ésimo indivíduo e sucesso.} \\ 0 & \text{se a resposta para o } i\text{-ésimo indivíduo e fracasso.} \end{cases}$$

Supondo que  $y_i$  segue distribuição Bernoulli com parâmetro  $\pi_i$  e para cada um dos  $n$  indivíduos haja observações sobre  $p$  variáveis explicativas,  $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})$ . O modelo de regressão logística é dado por

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \mathbf{x}_i' \beta \quad (4.13)$$

onde  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$  é o vetor de parâmetros do modelo.

Uma medida muito utilizada em várias áreas é a razão de chances (odds ratio (OR)). Para introduzir o conceito de razão de chances, suponhamos o caso mais simples em que no modelo 4.13 tenhamos apenas uma covariável, e que essa possa assumir os valores 0 e 1, então a razão de chances é definida por:

$$OR = \frac{\pi_i(1)/(1 - \pi_i(1))}{\pi_i(0)/(1 - \pi_i(0))} \quad (4.14)$$

onde  $\pi_i(x)$  é  $P(y_i = 1)$  para cada valor possível de  $x$ .

Por exemplo, se  $y_i$  representa a presença  $y_i = 1$  ou ausência  $y_i = 0$  de câncer de mama e  $x_i = 1(0)$  representa se a mulher está (não) em menopausa, um valor para  $OR=3$  pode ser interpretado da seguinte maneira: “a chance de uma mulher em menopausa ter câncer de mama é três vezes a chance da mulher que não está em menopausa”.

De uma maneira geral, a interpretação da razão de chances é realizada da mesma maneira quando temos mais que uma covariável no modelo. No caso de covariáveis contínuas a  $OR$  é definida como a razão entre as chances obtidas quando a covariável é acrescida de uma unidade.

Os modelos de regressão logística são muito utilizados para modelar relação existente entre as covariáveis e a variável resposta e também para classificar indivíduos dentro de categorias (presente/ausente) da variável resposta, segundo a probabilidade de pertencer a cada uma delas.

### 4.4.3 Regressão logística nominal e ordinal

Se a variável resposta é categórica com mais de duas categorias existem duas opções para modelos lineares generalizados. A primeira abordagem é uma generalização do modelo de regressão logística para respostas nominais ou ordinais com mais de duas categorias. A segunda abordagem é baseada na modelagem de frequências ou contagens para covariáveis ou variáveis resposta com distribuição de Poisson. Esses modelos também são chamados modelos log-lineares. Nesta seção, apresentaremos a primeira abordagem que será utilizada neste trabalho.

Para a regressão logística ordinal ou nominal uma variável é tomada como variável resposta e todas as outras são variáveis explicativas. Nos modelos log-lineares todas as variáveis são tratadas da mesma forma. A escolha de qual abordagem utilizar em uma particular situação depende se uma variável é claramente uma resposta (por exemplo, o resultado de um estudo prospectivo) ou se as variáveis tem o mesmo status (por exemplo em um estudo de corte transversal). Além disso, a escolha pode depender de como os resultados sejam apresentados e interpretados. Regressão logística nominal e ordinal fornecem estimativas de razão de chances que são relativamente fáceis de interpretar. Modelos log-lineares são aconselháveis para testar hipóteses sobre interações complexas mas as estimativas dos parâmetros são mais difíceis de serem interpretadas.

## Regressão logística nominal

Modelos de regressão logística nominal são utilizados quando não existe uma ordem natural entre as categorias de resposta. Esse modelo é uma extensão do modelo logístico para respostas dicotômicas, e consiste de uma combinação de modelos logitos estimados simultaneamente. Para  $y$ , uma variável categórica com  $J$  categorias,  $j=1,2,\dots,J$ , tal que para cada indivíduo  $i$  definimos:

$$y_{ij} = \begin{cases} 1 - \text{se o indivíduo } i \text{ esta na categoria } j. \\ 0 - \text{caso contrario.} \end{cases}$$

Denotando  $\pi_j$  a probabilidade do indivíduo estar na categoria  $j$ , escolhemos uma categoria como referência. Suponha que essa categoria seja a primeira. Os logitos para as outras categorias são então definidos por:

$$\text{logit}(\pi_1) = \log\left(\frac{\pi_j}{\pi_1}\right) = \mathbf{x}^T \beta_j, \quad \text{para } j = 2, \dots, J \quad (4.15)$$

onde  $\mathbf{x} = (1, x_1, \dots, x_p)$  e  $\beta_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{pj})^T$ .

As  $(J-1)$  expressões de logito em 4.15 são simultaneamente usadas para estimar os parâmetros  $\beta_j$ . Uma vez que as estimativas  $b_j$  tenham sido obtidas, os preditores lineares  $x^T b_j$  podem ser calculados, e as probabilidades estimadas para cada categoria.. De 4.15 temos:

$$\hat{\pi}_j = \hat{\pi}_1 \exp(x^T b_j), \quad j = 2, \dots, J \quad (4.16)$$

mas  $\hat{\pi}_1 + \hat{\pi}_2 + \dots + \hat{\pi}_J = 1$  então

$$\hat{\pi}_1 = \frac{1}{1 + \sum_{j=2}^J \exp(x^T b_j)} \quad (4.17)$$

e

$$\hat{\pi}_j = \frac{\exp(x^T b_j)}{1 + \sum_{j=2}^J \exp(x^T b_j)}, \quad j = 2, \dots, J \quad (4.18)$$

É mais fácil e usual interpretar os efeitos de fatores explicativos em termos de razão de chances que através dos parâmetros  $\beta$ . Por simplicidade, consideremos uma variável resposta com  $J$  categorias e uma variável explicativa binária  $x$  que denota quando a exposição à um fator é presente ( $x=1$ ) ou ausente ( $x=0$ ). As razões de

chances para exposição para a resposta  $j$  ( $j=2,\dots,J$ ) relativa à categoria de referência  $j=1$  é:

$$OR_j = \frac{\pi_{jp}/\pi_{1p}}{\pi_{ja}/\pi_{1a}} \quad (4.19)$$

onde  $\pi_{jp}$  e  $\pi_{ja}$  denotam as probabilidades da categoria de resposta  $j$  ( $j=1,\dots,J$ ) de acordo com ao fator de exposição (presente ou ausente, respectivamente). Para o modelo

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \beta_{0j} + \beta_{1j}x, \quad j = 2, \dots, J \quad (4.20)$$

então

$$\log\left(\frac{\pi_{ja}}{\pi_{1a}}\right) = \beta_{0j}, \quad x = 0 \quad (4.21)$$

$$\log\left(\frac{\pi_{jp}}{\pi_{1p}}\right) = \beta_{0j} + \beta_{1j}, \quad x = 1 \quad (4.22)$$

O logaritmo da razão de chances pode ser dado por:

$$\log(OR_j) = \log\left(\frac{\pi_{jp}}{\pi_{1p}}\right) - \log\left(\frac{\pi_{ja}}{\pi_{1a}}\right) = \beta_{1j} \quad (4.23)$$

Portanto  $OR_j = \exp(\beta_{1j})$  que é estimada por  $OR_j = \exp(b_{1j})$ . Se  $\beta_{1j} = 0$  então  $OR_j = 1$  que corresponde ao não efeito do fator de exposição. Os intervalos de confiança também podem ser construídos. Por exemplo o intervalo de confiança 95% para  $OR_j$  são dados por:  $\exp[b_{1j} \pm 1.96 \times sd(b_{1j})]$  onde  $sd(b_{1j})$  denota o desvio padrão de  $b_{1j}$ . Intervalos de confiança que não incluem a unidade correspondem à valores de  $\beta$  diferente de zero.

Assim como na regressão logística para variáveis binárias, as variáveis explicativas nos modelos de regressão logística nominal podem ser categóricas ou contínuas. A escolha da categoria de referência para a variável resposta afetará a estimativa do parâmetro  $b$  mas não as probabilidades estimadas  $\hat{\pi}$  ou os valores ajustados.

### Regressão logística ordinal

Se existe uma ordem natural entre as categorias de resposta então isso pode ser

levado em conta na especificação do modelo. Respostas ordinais são comuns em pesquisa de mercado, pesquisas de opinião e área de saúde. Em algumas situações é conveniente considerar variáveis resposta contínuas em classes através de pontos de corte. As classes então formadas podem ser consideradas níveis de uma variável ordinal. Por exemplo, para severidade de doenças podemos criar categorias como leve, moderado e grave. Para variáveis resposta com categorias ordinais podem ser utilizados diferentes modelos que citaremos em seguida.

### Modelo de logito cumulativo

Para  $y$ , uma variável categórica ordinal com  $J$  categorias,  $j=1,2,\dots,J$ , tal que para cada indivíduo  $i$  definimos:

$$y_{ij} = \begin{cases} 1 - \text{se a resposta para o } i\text{-ésimo indivíduo e } \leq j \\ 0 - \text{caso contrário} \end{cases}$$

isto é

$$y_{ij} = \begin{cases} 1 - \text{se } y_{ij} \leq j \\ 0 - \text{se } y_{ij} > j \end{cases}$$

Nesse modelo a forma de representar as probabilidades acumuladas é dada pela expressão:

$$P(y_{ij} \leq j) = \pi_1 + \dots + \pi_j, \quad j = 1, 2, \dots, J$$

As chances cumulativas para a  $j$ -ésima categoria são dadas por:

$$\frac{P(y_{ij} \leq j)}{P(y_{ij} > j)} = \frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J} \quad (4.24)$$

O modelo de logito cumulativo é dado por:

$$\log\left(\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J}\right) = x^T \beta_j \quad (4.25)$$

onde  $\mathbf{x} = (1, x_1, \dots, x_p)$  e  $\beta_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{pj})^T$ .

### Modelo de chances proporcionais

Se o preditor linear  $x^T \beta_j$  tem o intercepto  $\beta_{0j}$  dependente da categoria  $j$  mas, as

outras variáveis explicativas não dependem de  $j$  o modelo é dado por:

$$\log\left(\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J}\right) = \beta_{0j} + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (4.26)$$

Esse modelo é chamado modelo de chances proporcionais. É baseado na hipótese de que os efeitos das covariáveis  $x_1, \dots, x_p$  são os mesmos para todas as categorias da escala logarítmica.

### Modelo logito de categorias adjacentes

Uma alternativa para modelos de chances cumulativas é considerar as razões de probabilidades para categorias sucessivas por exemplo:

$$\frac{\pi_1}{\pi_2}, \frac{\pi_2}{\pi_3}, \dots, \frac{\pi_{J-1}}{\pi_J}$$

O modelo de logito para categoria adjacente é:

$$\log\left(\frac{\pi_j}{\pi_{j+1}}\right) = \beta_{0j} + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (4.27)$$

quando o efeito de cada variável explicativa é assumido ser o mesmo para todas as categorias adjacentes. Os parâmetros  $\beta_k$  são usualmente interpretados como razão de chances usando  $OR = \exp(\beta_k)$ .

### Modelo logito de razão de continuidade

Outra alternativa é modelar a razão de probabilidades no logito de razão de continuidade:

$$\frac{\pi_1}{\pi_2}, \frac{\pi_1 + \pi_2}{\pi_3}, \dots, \frac{\pi_1 + \dots + \pi_{J-1}}{\pi_J}$$

ou:

$$\frac{\pi_1}{\pi_1 + \dots + \pi_J}, \frac{\pi_2}{\pi_3 + \dots + \pi_J}, \dots, \frac{\pi_{J-1}}{\pi_J}$$

A equação:

$$\log\left(\frac{\pi_j}{\pi_{j+1} + \dots + \pi_J}\right) = x^T \beta_j \quad (4.28)$$

modela as chances da resposta pertencer a uma categoria  $j$  condicionada a chance de pertencer a categoria maior que  $j$ . Esse modelo pode ser mais fácil de ser interpretado que o modelo de chances proporcionais se as probabilidades de categorias individuais  $\pi_j$  são de interesse.

#### 4.4.4 Implementação de modelos ordinais no R

Para a implementação do modelo ordinal no R, utilizamos o modelo de chances proporcionais “polr” do pacote “MASS”, que produz ajuste de modelos de logito cumulativo para dados ordinais. Esse modelo foi implementado pela primeira vez por Brian Ripley et al em 1998. As estimativas oferecidas pelo R são baseadas no enfoque dado por Agresti [24].

O modelo de chances proporcionais polr considera uma variável resposta ordinal  $Y$  com  $j=1, \dots, J$  categorias e um conjunto de  $x_1, \dots, x_p$  de covariáveis explicativas tais que o modelo de logito cumulativo é dado por:

$$\text{logit}[P(Y \leq j)] = \beta_{0j} - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_p x_p \quad (4.29)$$

A razão de chances associada ao evento  $Y > j$  para uma dada covariável  $x_k$  é dada pela exponencial do coeficiente  $\beta_k$ .

É possível extrair de um modelo polr no R os valores ajustados, o grau de liberdade dos resíduos e a deviance, que auxiliam em testes de hipóteses para o ajuste e para comparação de modelos. A inferência para os coeficientes utiliza métodos padrão como testes de Wald e da razão de verossimilhanças sob a hipótese  $H_0 : \beta_k = 0$ . A anova pode ser implementada para esse objeto, o que permite a realização do teste da razão de verossimilhança para a comparação de modelos mais simples com outros mais complexos.

A desvantagem do modelo polr no R é não permitir testes para a proporcionalidade das chances. Para isso é necessário utilizar outros modelos, como por exemplo vglm, com menos recursos relacionados a testes para coeficientes e para o ajuste.

## 4.5 Modelos longitudinais

Estudos com respostas repetidas tem aplicações em várias áreas. Na área de saúde em particular, respostas categóricas repetidas podem ocorrer por exemplo em estudos longitudinais para a avaliação do efeito de uma droga, ou do estado de saúde ao longo do tratamento [24, 37]. É importante salientar que o problema considerado sobre a denominação de análise de dados longitudinais difere daquele comumente tratado na literatura estatística sobre o nome de análise de séries temporais pelo fato de que nesta análise em geral, dispõe-se de muitas observações realizadas em uma única unidade amostral, enquanto que naquela são analisadas várias unidades amostrais observadas em diferentes instantes de tempo [85].

Em estudos longitudinais, medidas realizadas em uma mesma pessoa em diferentes instantes de tempo podem ser mais similares que respostas de pessoas diferentes, porque respostas podem ser afetadas por características individuais. A suposição de independência entre as respostas nesse caso não é razoável. Por isso, é necessário estabelecer uma estrutura de correlação para descrever a associação entre as medidas observadas para a variável resposta [24, 37].

Outra situação em que a suposição de independência não é razoável, ocorre quando as medidas são realizadas em indivíduos relacionados. Por exemplo, medidas de peso em mulheres com mesma idade, selecionadas em regiões específicas de um país. Nesse caso, as regiões podem ser consideradas unidades primárias de amostragem ou clusters, e as mulheres sub-amostras dentro de cada unidade primária de amostragem. Mulheres de uma mesma região podem apresentar maior similaridade por compartilharem condições ambientais e socioeconômicas semelhantes [24, 37].

Para modelar dados longitudinais e dados em cluster é necessário incluir uma estrutura de correlação. Existem duas maneiras comuns de modelagem para esses tipos de dados [37]. Uma abordagem envolve a modelagem da estrutura de correlação de maneira explícita. Esse método inclui muitos modelos como análise de variância para medidas repetidas e equações de estimação generalizadas (GEE). Os procedimentos para estimação e inferência para esses modelos são, em princípio, análogos aos modelos lineares generalizados para resultados independentes [37]. Liang e Zeger [26], em 1986, desenvolveram um método baseado em uma extensão dos modelos lineares generalizados para observações longitudinais. Eles introduziram uma classe de equações de estimação com as quais podemos obter estimativas consistentes dos parâmetros e suas variâncias, sob condições regulares. Esses métodos são semi-paramétricos, e as equações de estimação são construídas sem a especificação completa da distribuição conjunta das observações dos sujeitos. Em vez disso, é especificada a verossimilhança para as distribuições marginais (univariadas) e uma matriz de covariância de trabalho para os vetores das observações.



A abordagem alternativa para modelagem é baseada em uma estrutura hierárquica presente no desenho do estudo. Esta abordagem é denominada modelagem multinível. Por exemplo, suponhamos que existam medidas repetidas para diferentes indivíduos alocados em grupos experimentais. As medidas repetidas de cada indivíduo podem ser caracterizadas como nível 1. Os diferentes indivíduos, nível 2. Os grupos experimentais, nível 3. Para modelos multiníveis, os efeitos dos níveis podem ser descritos por parâmetros fixos ou por variáveis aleatórias. Se o preditor linear do modelo tem ambos efeitos fixos e aleatórios, denominamos modelo misto. A estrutura de correlação entre as observações é atribuída aos efeitos aleatórios. Tal fato pode tornar a correlação mais fácil de ser interpretada em modelos multiníveis que em modelos com estrutura de correlação explícita. Eles são mais apropriados quando os parâmetros de correlação são de interesse. Os modelos de efeitos aleatórios foram primeiramente desenvolvidos por Harville [86], em 1977 e Laird [87], em 1982. Apesar de existir uma literatura extensa sobre modelos multiníveis para dados normais, ainda são poucas as técnicas e softwares para implementação de modelos multiníveis para variáveis categóricas [24, 37].

A estrutura de um estudo longitudinal é composta por um conjunto de indivíduos observados em diferentes ocasiões. Para um estudo longitudinal com  $n$  indivíduos, onde  $i=1, \dots, n$  denota o  $i$ -ésimo indivíduo, suponha que cada um deles é observado em  $T_i$  instantes de tempo quanto a uma variável resposta  $y$  e a um vetor  $\mathbf{x}$ , de dimensão  $p$ , de variáveis explicativas. Para  $Y_{it}$  a observação da variável resposta  $y$  para o  $i$ -ésimo indivíduo no tempo  $t$  onde  $t = 1, \dots, T_i$ , temos em notação matricial:

$\mathbf{Y}_i = (y_{i1}, \dots, y_{iT_i})^T$ , o vetor de observações para o  $i$ -ésimo indivíduo.

$\mathbf{Y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_{iT_i}^T)^T$ , o vetor ( $n_T \times 1$ ) de todas as observações para todos os indivíduos, onde  $n_T = \sum_{i=1}^n T_i$ .

$\mathbf{X}_i$ , a matriz de observações ( $T_i \times p$ ) de  $p$  covariáveis do  $i$ -ésimo indivíduo.

$\mathbf{X}$ , a matriz de observações ( $n_T \times p$ ) que contém as informações das covariáveis de todos os indivíduos.

A tabela 4.5 mostra um exemplo de uma estrutura de dados longitudinais.

Abordaremos agora os aspectos teóricos de modelos para dados longitudinais categóricos com enfoque em modelos marginais e modelos lineares generalizados mistos, que serão utilizados neste trabalho.

## 4.6 Modelos marginais

Um modelo marginal é um modelo que descreve a esperança marginal de uma variável resposta contínua ou discreta como uma função de variáveis preditoras [26]. Esse modelo é adequado quando o interesse principal é realizar inferências sobre a média populacional. Pela esperança marginal podemos obter a resposta média sobre

Tabela 4.1: Exemplo de estrutura de dados longitudinais

Indivíduo	Tempo	Resposta	Variáveis explicativas
1	1	$y_{11}$	$x_{111} \dots x_{11p}$
1	2	$y_{12}$	$x_{121} \dots x_{12p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$
1	$T_1$	$y_{T1}$	$x_{1T_11} \dots x_{1T_1p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$
i	1	$y_{i1}$	$x_{i11} \dots x_{i1p}$
i	2	$y_{i2}$	$x_{i21} \dots x_{i2p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$
i	$T_i$	$y_{iT_i}$	$x_{iT_i1} \dots x_{iT_ip}$
$\vdots$	$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$
n	1	$y_{n1}$	$x_{n11} \dots x_{n1p}$
n	2	$y_{n2}$	$x_{n21} \dots x_{n2p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \vdots$
n	$T_n$	$y_{nT_n}$	$x_{nT_n1} \dots x_{nT_np}$

a sub população que compartilha um valor comum para uma variável independente do modelo [27].

O modelo marginal, também conhecido como modelo “média-populacional”, é obtido por meio da introdução da parametrização da covariância a nível individual (ou correlação). Essa correlação é estimada pela média sobre a informação de todos os indivíduos. Na modelagem dos efeitos das covariáveis nas esperanças marginais é possível especificar um modelo de associação entre as observações de cada sujeito.

Atualmente modelos marginais têm recebido muita atenção, e existem implementações de modelos marginais disponíveis em diversos softwares estatísticos como o R e o SPSS [88].

As dificuldades que podem ser identificadas na análise de um modelo marginal estão relacionadas à dependência das observações ao longo do tempo, à grande dispersão das distribuições marginais que pode surgir da variação entre os indivíduos, e a não determinação de frequências esperadas a partir das hipóteses marginais [89].

O modelo marginal tem as seguintes hipóteses:

**i-** A esperança marginal da variável resposta  $Y_{it}$ , para um sujeito  $i$ , no tempo  $t$ ,  $E(Y_{it}) = \mu_{it}$  depende das variáveis explicativas  $x_{it}$  através de alguma função de ligação  $g$  conhecida,  $g(\mu_{it}) = x'_{it}\beta$ .

**ii-** A variância marginal está associada à média marginal através de uma função conhecida  $H$  e um parâmetro de escala  $\phi$ ,  $Var(Y_{it}) = H(\mu_{it})\phi$ .

**iii-** A correlação entre respostas em tempos  $j$  e  $t$  é uma função das médias marginais e, em alguns casos é necessário estimar um parâmetro adicional  $\alpha$ ,  $Cor(Y_{ij}, Y_{it}) = \rho(\mu_{ij}, \mu_{it}; \alpha)$ .

Os coeficientes de regressão do modelo marginal  $\beta$  têm interpretações semelhantes aos parâmetros de um modelo de regressão logística ordinal. Modelos marginais para dados correlacionados são análogos aos modelos lineares generalizados para respostas categóricas quanto a interpretação de coeficientes e odds ratio.

Para exemplificar a interpretação dos parâmetros em um modelo marginal, consideremos o problema de avaliar ao longo do tempo a dependência de um diagnóstico positivo para câncer de mama em função da detecção de microcalcificações em mamografias. Seja  $x_{it}$  a variável que indica o aparecimento de microcalcificações na  $t$ -ésima mamografia,  $x_{it} = 1$  quando a microcalcificação é identificada e,  $x_{it} = 0$ , quando não identificada. Seja  $Y_{it}$  a variável que indica o diagnóstico para o câncer de mama no tempo  $t$ ,  $Y_{it} = 1$  quando o diagnóstico é positivo para o câncer de mama e,  $Y_{it} = 0$  quando o diagnóstico é negativo. Assumindo  $E(Y_{it}) = \mu_{it}$ , um modelo marginal para  $Y_{it}$  é obtido como:

- $\text{logit}(\mu_{it}) = \log(\mu_{it}/(1 - \mu_{it})) = \log(P(Y_{it} = 1)/P(Y_{it} = 0)) = \beta_0 + \beta_1 x_{it}$
- $\text{Var}(Y_{it}) = \mu_{it}(1 - \mu_{it})$
- $\text{Cor}(Y_{it}, Y_{ij}) = \alpha$

O coeficiente de regressão transformado  $\exp(\beta_0)$  é a razão entre a probabilidade de uma mulher ter diagnóstico positivo para o câncer de mama e a probabilidade de uma mulher ter diagnóstico negativo na sub-população de mulheres em que não foram detectadas microcalcificações. O parâmetro  $\exp(\beta_1)$  é a razão de chances de diagnóstico positivo entre as mulheres com microcalcificações detectadas e as sem microcalcificações.

Os modelos para dados binários ou qualquer outro tipo de dado categorizado podem diferir entre si em termos de efeito de média, podendo ser esta populacional ou grupo/sujeito-específico. Os modelos marginais se referem às distribuições marginais das médias sobre grupos ou clusters da população.

Na seção anterior foram discutidos aspectos principais da teoria de modelos lineares generalizados para o tratamento de variáveis respostas categóricas binárias e ordinais. A seguir será realizada uma breve revisão sobre a teoria de modelos marginais para dados longitudinais ordinais.

#### 4.6.1 Modelos marginais para variáveis ordinais

Suponha que  $n$  indivíduos sejam selecionados aleatoriamente de uma população e que para cada indivíduo  $i$  são registradas  $T_i$  observações sobre uma variável resposta

ordinal  $Y$  com  $K$  categorias,  $k=1, \dots, K$ . Para cada observação  $t$ ,  $t=1, \dots, T_i$  são registrados para o sujeito  $i$  além da resposta  $Y_{it}$ , os valores associados para o vetor  $\mathbf{X}_{it}$  (px1) de variáveis explicativas. O modelo marginal para a resposta ordinal  $Y_{it}$  referente a variável resposta para a pessoa  $i$  no tempo  $j$  pode ser definido de uma forma geral por meio de uma extensão do modelo de logito cumulativo apresentado no início do capítulo para dados transversais:

$$\log\left[\frac{P(Y_{it} \leq k)}{1 - P(Y_{it} \leq k)}\right] = \beta_{0k} + \mathbf{X}_{it}^T \beta_{\mathbf{k}} \quad k = 1, \dots, K - 1 \quad (4.30)$$

onde  $\beta_{\mathbf{k}} = (\beta_{1k}, \beta_{2k}, \dots, \beta_{pk})^T$  é o vetor de parâmetros do modelo.

O modelo 5.6 é uma extensão do modelo de chances proporcionais onde os coeficientes  $\beta_k$  variam de acordo com os  $k$  pontos de corte de resposta. Se os coeficientes  $\beta_k$  não variam com os pontos de corte, o modelo 5.6 se reduz ao modelo de chances proporcionais:

$$\log\left[\frac{P(Y_{it} \leq k)}{1 - P(Y_{it} \leq k)}\right] = \beta_{0k} + X_{it}^T \beta \quad k = 1, \dots, K - 1 \quad (4.31)$$

onde  $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ .

No modelo definido em 4.31, as mudanças nos  $K-1$  logitos cumulativos ao longo do tempo estão relacionadas com as variáveis preditoras. Apesar do modelo incluir  $K-1$  interceptos  $\beta_{0k}$ , ele assume que o efeito das variáveis preditoras é o mesmo para os  $K-1$  logitos, o que é equivalente a assumir que o efeito das variáveis preditoras sobre as chances cumulativas são proporcionais.

Se para o indivíduo  $i$  temos  $T_i$  observações, seja  $Y_{k_1, \dots, k_{T_i}}(i) = 1$ , se as respostas para a pessoa  $i$  assumem as categorias  $k_1, \dots, k_{T_i}$  e  $Y_{k_1, \dots, k_{T_i}}(i) = 0$ , caso contrário. Marginalmente, assumimos que esse indicadores são ensaios de uma distribuição multinomial com parâmetros  $\pi_{k_1, \dots, k_{T_i}}(x_i) = P(Y_{k_1, \dots, k_{T_i}}(i) = 1)$ . Então, o logaritmo da verossimilhança é proporcional à:

$$\sum_{i=1}^n \sum_{k_1=1}^c \sum_{k_2=1}^c \dots \sum_{k_{T_i}=1}^c y_{k_1, \dots, k_{T_i}}(i) \log[\pi_{k_1, \dots, k_{T_i}}(x_i)] \quad (4.32)$$

onde  $y_{k_1, \dots, k_{T_i}}(i)$  denota a realização de  $Y_{k_1, \dots, k_{T_i}}(i)$ . É difícil maximizar este logaritmo de verossimilhança porque o modelo 4.32 se refere às probabilidades marginais e  $\pi_k(x_{it}) = P(Y_{it} = k)$  e  $\pi_{k_1, \dots, k_{T_i}}(x_i)$  são probabilidades multinomiais no logaritmo da verossimilhança. Então não é possível substituir diretamente na função de verossimilhança e maximizar usando métodos padrões. Dada a dificuldade em especificar

completamente a distribuição conjunta, pode-se aplicar uma outra metodologia baseada em uma generalização multivariada da quase-verossimilhança e considerar as equações de estimação generalizadas e suas estratégias de estimação que são computacionalmente mais simples para o processo de estimação em modelos marginais com variáveis desfecho categóricas [24].

A seguir abordaremos os conceitos do método de quase-verossimilhança e a generalização multivariada deste método, que dá origem às equações de estimação generalizadas.

## 4.6.2 O método de quase-verossimilhança

O método de quase-verossimilhança foi proposto por Wedderburn em 1974 [90], e posteriormente trabalhado extensivamente por McCullagh & Nelder em 1989 [25]. Esse é um método usado em regressão que requer poucas informações sobre a distribuição da variável dependente ou variável resposta. Na análise de verossimilhança devemos especificar a forma da distribuição. Na quase-verossimilhança, especificamos somente a relação entre a média da resposta das covariáveis e a covariância ou, equivalentemente, especifica-se o modelo para a média e uma função que expresse a dependência entre variância e a média.

Esse método será apresentado agora de maneira breve, inicialmente com modelos para uma simples resposta e depois com o método para respostas multivariadas.

Para um sujeito  $i$ ,  $i=1, \dots, n$ , seja  $y_i$  o valor observado para  $Y_i$ , com  $E(Y_i) = \mu_i$  e função de variância  $v(\mu_i)$ , e seja  $x_{ij}$  o valor da variável explicativa  $j$ . Para a função de ligação  $g$ , o preditor linear é  $\eta_i = g(\mu_i) = \sum_j \beta_j x_{ij} = x_i' \beta$ . Os estimadores de quase-verossimilhança para  $\beta$  são soluções das equações quase-escores:

$$u(\beta) = \sum_i \left( \frac{\partial \mu_i}{\partial \beta} \right)' v(\mu_i)^{-1} (y_i - \mu_i) = 0 \quad (4.33)$$

onde  $\mu_i = g^{-1}(x_i' \beta)$ . Essas equações, também denominadas equações de estimação, são as mesmas equações de verossimilhança usadas para estimação de parâmetros em modelos lineares generalizados, mas no caso de modelos lineares generalizados, a variável resposta pertence à família exponencial e, nós substituímos  $\partial \mu_i / \partial \beta_j$  por:

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij} \quad (4.34)$$

Dessa maneira, as equações de verossimilhança para estimação em modelos lineares generalizados dependem da média e variância de  $y_i$  e, da função de ligação

$g$  que determina  $\partial\mu_i/\partial\beta_j$ . Wedderburn então sugeriu, em 1974, a utilização das equações de estimação para qualquer função de ligação e de variância, mesmo se eles não correspondem a um membro particular da família exponencial.

### 4.6.3 Metodologia de equações de estimação generalizadas (GEE)

No caso univariado, o método da quase-verossimilhança especifica um modelo para  $\mu = E(Y)$  e uma função de variância  $v(\mu)$  descrevendo como  $Var(Y)$  depende de  $\mu$ . Medidas repetidas geram resposta multivariada  $(Y_1, \dots, Y_T)$  onde T pode variar de sujeito para sujeito. A metodologia de equações de estimação generalizadas (GEE) consiste na aplicação do método de quase-verossimilhança para a distribuição marginal de cada  $Y_t$ . A metodologia também requer uma estrutura de correlação entre  $Y_t$ 's. Os estimadores serão soluções dessas equações. Liang e Zeger, [26] em 1986, foram os pioneiros ao propor esse método para modelos marginais.

A abordagem GEE utiliza uma estrutura conhecida de covariância para  $(Y_1, \dots, Y_T)$ , especificando uma função de variância e um padrão de correlação para cada duas variáveis, sem assumir uma distribuição multivariada particular. Os estimadores do modelo GEE são válidos mesmo sem uma especificação adequada da estrutura de correlação. Além disso são consistentes (isto é convergem em probabilidade para os parâmetros) pois a consistência depende do primeiro momento mas não do segundo [24]. Especificamente consideramos o modelo válido no sentido que a escolha da função de ligação e o preditor linear descreve verdadeiramente como  $E(Y_t)$ ,  $t=1, \dots, T$ , depende dos preditores.

Na prática a escolha de um modelo nunca é exatamente correto. Esse resultado entretanto é útil por sugerir que a estrutura de correlação não necessariamente afetará a qualidade dos estimadores utilizados no modelo, mesmo quando nenhuma informação sobre a estrutura de correlação é conhecida. Uma simples implementação do método GEE trata  $Y_t$  como variáveis duas a duas independentes. Uma alternativa é utilizar a estrutura de correlação permutável, essa estrutura trata  $cor(Y_t, Y_s)$  como idênticas para todo s e t. Essa hipótese é mais flexível e realista que a hipótese de independência. Uma outra estrutura de correlação mais realista é a correlação não estruturada que permite a especificação de correlações diferentes para cada par. Quando T é grande, entretanto, essa abordagem sofre perda de eficiência devido à quantidade de parâmetros adicionais a serem estimados.

Em teoria, a escolha de uma estrutura de correlação pode beneficiar ou melhorar a eficiência de um estimador. Entretanto, Liang e Zeger, [26] notaram que estimadores baseados em estrutura de correlação independente pode surpreendentemente ter uma boa eficiência quando a correlação verdadeira é fraca ou moderada. É possível

checar a sensibilidade do modelo comparando resultados para diferentes estruturas de correlação. Segundo Agresti, [24], quando as correlações são moderadas todas as estruturas de correlação fornecem estimativas e erros padrões similares. A menos que exista uma diferença considerável entre as correlações, é recomendável o uso da estrutura de correlação permutável pois ela reconhece a dependência através de um único parâmetro extra.

A abordagem GEE é mais atrativa para dados categóricos por conta da simplicidade computacional quando comparada à abordagem de máxima verossimilhança. As vantagens incluem a não necessidade de uma distribuição multivariada, e a consistência dos estimadores mesmo sem o conhecimento de uma estrutura de correlação adequada. Entretanto, o método GEE tem limitações. Devido ao fato de não especificar precisamente a distribuição conjunta, não temos uma função de verossimilhança. Por isso, os métodos baseados em verossimilhança não são aplicáveis para testar o ajuste, comparar modelos e realizar inferência sobre os parâmetros. Utilizamos para inferência a estatística de Wald, construída através da distribuição assintótica normal dos estimadores juntamente com sua matriz de covariância estimada. Entretanto, a menos que o tamanho amostral seja grande os erros padronizados tendem a subestimar e ter maior variabilidade que os verdadeiros erros [91]. Outras abordagens de testes têm sido publicados na literatura a fim de contornar esse problema através de testes de score análogos para avaliar o efeito dos preditores [92].

Na seção 4.6.2 consideramos a metodologia de quase-verossimilhança para respostas univariadas e descrevemos as equações de estimação. Agora descreveremos as equações de estimação generalizadas (GEE) com detalhes.

Para um sujeito  $i$ , seja  $y_i = (y_{i1}, \dots, y_{iT_i})'$  e  $\mu_i = (\mu_{i1}, \dots, \mu_{iT_i})'$  onde  $\mu_{it} = E(Y_{it})$ . O número  $T_i$  de respostas pode variar por cluster (sujeito). Seja  $x_{it}$  o vetor ( $p \times 1$ ) de variáveis explicativas para  $y_{it}$  e o preditor linear  $\eta_{it} = g(\mu_{it}) = x'_{it}\beta$ , para uma função de ligação  $g$ . A notação permite casos em que variáveis explicativas possam variar para medidas repetidas. O modelo refere à distribuição marginal em cada tempo  $t$  no lugar da distribuição conjunta. Seja  $X_i$  a matriz  $T_i \times p$  de valores preditores para o cluster (sujeito)  $i$ .

Assumimos que  $y_{it}$  tem função de probabilidade de massa da seguinte forma:

$$f(y_{it}; \theta_{it}, \phi) = \exp[y_{it}\theta_{it} - b(\theta_{it})]/\phi + c(y_{it}, \phi) \quad (4.35)$$

onde  $\phi$  é conhecido, esse é um membro da família exponencial com parâmetro natural  $\theta_{it}$ . Para a família exponencial, como vimos no capítulo 3:

$$\mu_{it} = E(Y_{it}) = b'(\theta_{it}), \quad v(\mu_{it}) = \text{var}(Y_{it}) = b''(\theta_{it})\phi \quad (4.36)$$

O método GEE também assume uma matriz de correlação de trabalho  $R(\alpha)$  para  $Y_i$  dependendo dos parâmetros  $\alpha$ . Seja  $b_i(\theta) = (b(\theta_{i1}), \dots, b(\theta_{iT_i}))$  e seja  $B_i$  a matriz diagonal com elementos da diagonal iguais à  $b_i''(\theta)$ . Então a matriz de covariância de trabalho para  $Y_i$  é dada por:

$$V_i = B_i^{1/2} R(\alpha) B_i^{1/2} \phi \quad (4.37)$$

Quando  $R(\alpha)$  é a verdadeira matriz de correlação para os  $Y_i$ 's,  $V_i$ ,  $V_i = \text{cov}(Y_i)$  é igual a matriz de covariância dos  $Y_i$ 's.

A escolha da matriz de correlação de trabalho,  $R(\alpha)$  leva a distintas análises, e a escolha de  $R(\alpha)$  próxima da verdadeira correlação aumenta a eficiência dos estimadores. As principais matrizes de estruturas de correlação de trabalho são:

1- Independência: Considera respostas independentes para um mesmo cluster (sujeito):

$$R(\alpha) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad (4.38)$$

nesse caso o modelo é conhecido como modelo de regressão linear múltipla.

2- Auto-regressiva (AR1): Se a distribuição da variável resposta é normal, essa estrutura de correlação corresponde ao modelo auto-regressivo de primeira ordem.

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & 0 & \dots & 0 \\ \alpha & 1 & \alpha & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \alpha & 1 \end{bmatrix} \quad (4.39)$$

3- Não estruturada: Todos os valores de correlação são diferentes. Para  $n_i$  observações realizadas para o cluster (sujeito)  $i$ , temos:



$$R(\alpha) = \begin{bmatrix} 1 & \alpha_{12} & 0 & \dots & \alpha_{1n_i} \\ \alpha_{12} & 1 & \dots & \dots & \alpha_{2n_i} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_{1n_i} & \alpha_{2n_i} & \dots & \dots & 1 \end{bmatrix} \quad (4.40)$$

4- Permutável: A correlação entre quaisquer duas variáveis é a mesma.

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha & \dots & \alpha \\ \alpha & 1 & \alpha & \dots & \alpha \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha & \alpha & \dots & \alpha & 1 \end{bmatrix} \quad (4.41)$$

5-M-dependente: A correlação  $cor(Y_{it}, Y_{it+l}) = \alpha_l$ , para  $l=1, \dots, M$ , depende das  $M$  observações anteriores. No caso de  $n_i$  observações para o cluster (sujeito)  $i$ , se  $M = n_i - 1$  temos:

$$R(\alpha) = \begin{bmatrix} 1 & \alpha_1 & \dots & \dots & \alpha_{n_i-1} \\ \alpha_1 & 1 & \dots & \dots & \alpha_{n_i-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \alpha_{n_i-1} & \alpha_{n_i-2} & \dots & \dots & 1 \end{bmatrix} \quad (4.42)$$

6- Fixa: É especificada pelo pesquisador. Para  $n_i$  observações realizadas para o cluster (sujeito)  $i$ , temos:

$$R(\alpha) = \begin{bmatrix} 1 & r_{12} & \dots & \dots & r_{1n_i} \\ r_{12} & 1 & \dots & \dots & r_{2n_i} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{1n_i} & r_{2n_i} & \dots & \dots & 1 \end{bmatrix} \quad (4.43)$$

onde  $r_{st}$  é especificado pelo pesquisador.

Para a escolha de uma estrutura de matriz de correlação de trabalho é aconselhável a realização de análise exploratória. No caso de dados binários as medidas de correlação são baseadas nas odds ratio [93].

Uma vez escolhida a matriz de correlação de trabalho, o próximo passo é a construção das equações estimadas generalizadas. Abordaremos agora essa construção.

Seja  $\Delta_i$  a matriz diagonal com os elementos  $\partial\theta_{it}/\partial\eta_{it}$  na diagonal principal para

$t = 1, \dots, T_i$ . Seja  $D_i = \partial\mu_i/\partial\beta = B_i\Delta_iX_i$  a matriz  $T_i \times p$  com os elementos expressando  $\partial\mu_{it}/\partial\beta_j$  na forma de  $(\partial\mu_{it}/\partial\theta_{it})(\partial\theta_{it}/\partial\eta_{it})(\partial\eta_{it}/\partial\beta_j)$

De 4.33 para modelos lineares generalizados univariados temos que as equações de estimação de quase-verossimilhança têm a forma:

$$u(\beta) = \sum_i \left(\frac{\partial\mu_i}{\partial\beta}\right)' v(\mu_i)^{-1}(y_i - \mu_i(\beta)) = 0 \quad (4.44)$$

onde  $\mu_i = \mu_i(\beta) = g^{-1}(x_i'\beta)$ . As equações análogas no caso multivariado são dadas pelo conjunto de equações estimadas generalizadas:

$$\sum_{i=1} D_i' V_i^{-1}(y_i - \mu_i(\beta)) = 0 \quad (4.45)$$

Os estimadores  $\hat{\beta}$  das equações estimadas generalizadas são soluções dessas equações.

#### 4.6.4 Implementação do método GEE no R

Para a implementação do modelo marginal no R, utilizamos a função “ordgee” do pacote “geese”, o qual produz um ajuste para as equações estimadas generalizadas para dados agrupados ordinais. Esse objeto foi implementado por Yan em 2003. As estimativas oferecidas pelo R são baseadas no enfoque dado por Heagerty e Zeger em 1996 [92].

Esses autores consideram um vetor de medidas ordinais  $O_i$  para o  $i$ -ésimo cluster (sujeito) e  $O_{it}$  a  $t$ -ésima observação no tempo do  $i$ -ésimo sujeito e  $\mathbf{X}_{it}$  o vetor de covariáveis associadas a  $O_{it}$ . Podemos representar a medida ordinal  $O_{it} = k$ , onde  $k \in [0, \dots, c]$  e  $c$  é o número de categorias da variável resposta, através das variáveis indicadoras cumulativas:

$$Y_{itk} = I(O_{it} > k) \quad (4.46)$$

onde  $k \in [0, 1, 2, \dots, c - 1]$ . E portanto, para o modelo de odds proporcional para as médias marginais é assumido que:

$$\text{logit}[E(Y_{itk})] = \beta_{0k} + \mathbf{X}_{it}^T \beta \quad (4.47)$$

com  $t = (1, \dots, t_i)$  e  $k = 1, \dots, c - 1$  e  $t = 1, \dots, t_i$ .

Na utilização do pacote `geese` é possível considerar dois tipos de modelos que são modelos com intercepto constante e modelos com interceptos não constantes em relação ao tempo. Além disso é possível escolher mudanças nas estruturas das matrizes de correlação de trabalho. Para o intercepto constante:

$$\text{logit}[E(Y_{itk})] = \beta_{0k} + \mathbf{X}_{it}^T \beta \quad (4.48)$$

e, para o intercepto não constante:

$$\text{logit}[E(Y_{itk})] = \beta_{0kt} + \mathbf{X}_{it}^T \beta \quad (4.49)$$

com  $i = (1, \dots, n)$ ,  $k = 1, \dots, c - 1$  e  $t = 1, \dots, T_i$ .

Utilizar o modelo com intercepto não constante pode ser particularmente útil em situações em que o objetivo é obter estimativas sensíveis a mudanças no tempo. Em muitas situações, cada cluster (sujeito) é considerado como um conjunto de medidas repetidas. Em outras situações, cada cluster é o conjunto de unidades amostrais que são consideradas similares de acordo com determinadas características.

No R, não há muitos testes de hipóteses disponíveis para a função `ordgee`, em virtude do método para a estimação ser baseado em equações de quase verossimilhança. O teste de Wald pode ser aplicado para testar a significância dos coeficientes, mas anova e testes de razão de verossimilhança não podem ser aplicados. De uma maneira geral, na literatura há alguns estudos sobre técnicas de diagnóstico e comparações de modelos que são possíveis reproduzir no R.

Em um estudo de 2001, Pan[94] propôs uma medida equivalente à medida AIC para as equações de estimação generalizadas. Esta medida foi denominada QIC (Quasi-likelihood under the independence model criterion) e pode ser utilizada para comparar modelos de maneira semelhante à medida AIC, menores valores de QIC representam melhores modelos. Em 2009, Hin e Wang [95] propuseram um critério que apresentou uma melhoria com relação ao QIC o qual eles denominaram CIC (correlation information criterion). A interpretação do CIC é análoga ao QIC e ao AIC. Essas medidas podem ser utilizadas também para a escolha da matriz de correlação mais adequada aos dados. O cálculo do QIC e CIC para modelos ajustados pela função `ordgee` no R pode ser feito por meio de implementação de uma rotina simples e foi realizado neste trabalho.

Em um estudo de 2008, Sohee et al [96] propuseram gráficos de resíduos para comparar ajustes de modelos GEE baseado nos resíduos de Pearson. Em 2010, Lin [97] propôs um teste para a bondade do ajuste para modelos GEE que representou uma generalização do teste construído por Pan em 2002 [98] para bondade do ajuste

em modelos longitudinais para dados binários.

## 4.7 Modelos lineares generalizados mistos

Outra classe de modelos bastante utilizada em estudos longitudinais são os modelos mistos que incluem no modelo de efeitos fixos, efeitos aleatórios a nível do indivíduo. Para um estudo longitudinal, não é razoável supor que medidas feitas para um mesmo indivíduo ao longo do tempo sejam não correlacionadas da mesma maneira que não é razoável supor homogeneidade para as respostas dos indivíduos. Por meio dos modelos mistos é possível modelar a heterogeneidade entre indivíduos e incluir a covariância entre as respostas repetidas.

Para exemplificar, suponha que as respostas para 8 indivíduos em 7 diferentes instantes de tempo sejam apresentadas como na figura 4.7. Neste exemplo, podemos visualizar que um modelo linear que permita um intercepto e um coeficiente angular aleatórios possibilitaria descrever a resposta para cada indivíduo separadamente, o que pode representar um ganho para o ajuste se pensarmos que modelos lineares para indivíduos diferentes tendem a ser representados por retas com diferentes interceptos e inclinações.

Os modelos lineares generalizados mistos, (GLMM em inglês), são uma combinação da classe de modelos lineares mistos (MLM), (Laird e Ware, 1982) e dos modelos lineares generalizados, (Nelder e Wessernburn, 1972). Os modelos lineares mistos são uma família de modelos que inclui como casos particulares o modelo linear clássico, a ANOVA univariada para medidas repetidas e os modelos para curvas de crescimento [24]. A metodologia descrita por Laird e Ware foi baseada no trabalho de Harville (1977) e consiste em um modelo de dois níveis para medidas repetidas. No primeiro nível do modelo são introduzidos parâmetros populacionais, efeitos individuais e a variação intra-indivíduo. A variação entre indivíduos é introduzida no segundo nível do modelo.

Um modelo que contenha uma média geral considerada uma constante  $\mu$ , e um termo referente ao erro, assumido como aleatório, pode ser considerado um modelo misto. Mas, a denominação de modelo linear misto é geralmente reservada a modelos lineares que contenham efeitos fixos além de  $\mu$  e qualquer termo aleatório além do erro. Um modelo linear misto para respostas contínuas é descrito na literatura como:

$$Y_i = X_i\beta + Z_ib_i + \epsilon_i \quad (4.50)$$

em que  $Y_i$  é um vetor  $n_i$  dimensional de respostas,  $\beta$  é um vetor  $px1$  de parâmetros desconhecidos.  $X_i$  é a matriz de dimensão  $n_i \times p$  de variáveis explicativas associadas

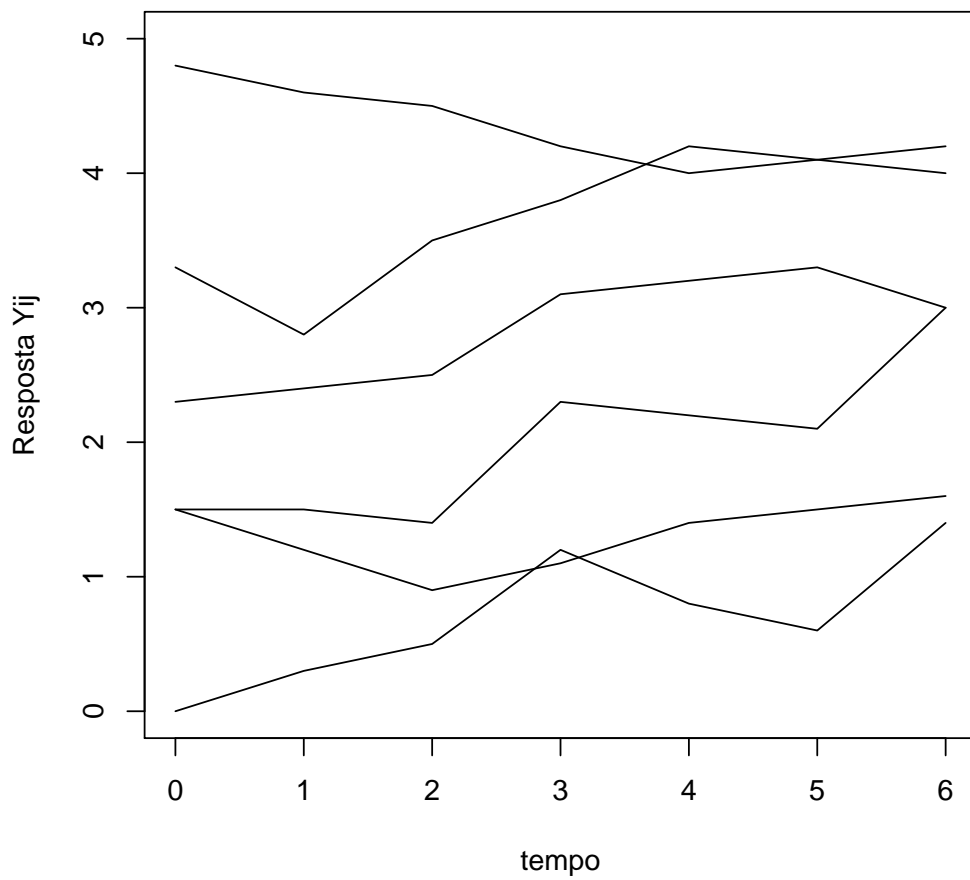


Figura 4.3: Respostas de indivíduos ao longo do tempo (dados hipotéticos)

aos efeitos fixos  $\beta$ ,  $Z_i$  é a matriz de dimensão,  $n_i \times q$  de variáveis explicativas associadas aos efeitos aleatórios  $b_i$ ,  $q \times 1$  e  $\epsilon_i$ ,  $n_i \times 1$  é o vetor de erros aleatórios. Geralmente  $Z_i$  é uma submatriz de  $X_i$ . Em geral supõe-se que tanto os erros aleatórios como os efeitos aleatórios são normalmente distribuídos, ou seja,

$$b_i \sim N(0, D) \quad (4.51)$$

e

$$\epsilon_i \sim N(0, \sum_i) \quad (4.52)$$

onde  $\sum_i = \sigma^2 I$ .

Os efeitos aleatórios no intercepto representam a heterogeneidade natural entre os indivíduos decorrente de fatores não medidos.

A esperança e variância para  $Y_i$  são dadas por  $E(Y_i) = X_i^T \beta$  e  $V(Y_i) = Z_i D Z_i^T + \sum_i$ , onde  $D$  e  $\sum_i$  são desconhecidos e podem ser substituídos pelas matrizes estimadas  $G$ , referente aos efeitos aleatórios e  $R_i$ , correspondente a correlação serial. Essas matrizes podem ser obtidas no processo de ajuste do modelo.

No ajuste do modelo misto é preciso avaliar a estrutura de covariância que melhor se adapta aos dados, nesta estrutura são incorporados os efeitos fixos e os efeitos aleatórios associados aos indivíduos. Os tipos de estrutura mais comumente utilizados são:

Uniforme: Apresenta homogeneidade tanto nas variâncias quanto nas covariâncias.

$$V(y_i) = V(y_i) = Z_i D Z_i^T + \sum_i = \begin{bmatrix} \sigma^2 + \tau & \tau & \dots & \tau \\ \tau & \sigma^2 + \tau & & \tau \\ \vdots & & \ddots & \vdots \\ \tau & \tau & \dots & \sigma^2 + \tau \end{bmatrix} \quad (4.53)$$

nesse caso temos que  $D = \tau$  com  $\tau > 0$ , e  $Z_i = 1_{n_i}$  é um vetor ( $n_i \times n_i$ ) com todos os elementos iguais a 1 e  $\sum_i = \sigma^2 I_{n_i}$ .

Auto-regressiva (AR-1):

$$V(y_i) = \begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^n \\ \phi & 1 & \phi^2 & \dots & \phi^n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi^n & \phi^{n-1} & \dots & \phi & 1 \end{bmatrix} \quad (4.54)$$

Nesta estrutura tem-se  $D = 0$  e  $\sum_i$  é uma matriz gerada por um modelo em que os erros aleatórios das medidas realizadas no  $i$ -ésimo indivíduo têm a estrutura  $e_{ij} = \phi e_{i,j-1} + \delta_{ij}$ , em que  $\delta_{ij} \sim N(0, \tau^2)$  são não correlacionados com  $e_{il}$   $l=1,2,\dots,j-1$ . Esta estrutura é função de  $\theta = (\phi, \sigma^2)$ , com  $\sigma^2 = \tau^2/(1 - \phi^2)$  e  $|\phi| < 1$  para garantir a estacionariedade.

Marginalmente, os  $Y_i$  são independentes e seguem uma distribuição normal com média  $X_i^T \beta$  e matriz de covariâncias, dada por  $Z_i D Z_i^T + \sum_i$ . Assim, a função de verossimilhança para o modelo descrito em 4.50 é dada por:

$$L(\beta, \sigma^2, D) = \prod_{i=1}^n | \sum_i |^{-1/2} \exp \frac{1}{2} (Y_i - X_i \beta)^T \sum_i (Y_i - X_i \beta)$$

Laird e Ware sugerem em seu trabalho a obtenção das estimativas dos parâmetros do modelo linear misto através de máxima verossimilhança. A estimação por máxima verossimilhança entretanto, pode fornecer estimativas viesadas para os componentes da variância. Isso se deve ao fato de que a função de verossimilhança completa envolve o vetor  $\beta$  de parâmetros fixos, que precisa ser estimado juntamente com os componentes de variância, e a perda de graus de liberdade na estimação dos efeitos fixos não é levada em consideração na estimação por máxima verossimilhança dos componentes de variância. Uma solução é utilizar o método de máxima verossimilhança restrita (Patterson e Thompson, 1971), que pode diminuir o viés dos estimadores de máxima verossimilhança para os componentes de variância. Para esse método a seguinte função de verossimilhança é considerada:

$$L(\beta, \sigma^2, D) = \prod_{i=1}^n | \sum_i |^{-1/2} | X_i^T \sum_i X_i |^{-1/2} \exp \frac{1}{2} (Y_i - X_i \beta)^T \sum_i (Y_i - X_i \beta)$$

Existem muitas situações em que o vetor de respostas não pode ser modelado a partir de uma distribuição normal como é hipótese dos modelos lineares mistos. Nesses casos a distribuição do vetor de respostas podem ser estendidas de forma a pertencer a família exponencial. Dessa maneira podemos estender os modelos lineares mistos para uma classe mais ampla de modelos conhecida como modelos

lineares mistos generalizados (GLMM). A hipótese básica ao GLMM para dados longitudinais é a suposição de heterogeneidade entre os indivíduos na população estudada, em um subconjunto dos coeficientes de regressão a partir de um modelo linear generalizado (Fitzmaurice et al., 2011).

Assim, para  $Y_{it}$  a resposta para o  $i$ -ésimo indivíduo, na  $t$ -ésima ocasião, assumimos que a distribuição condicional da resposta  $Y_{it}$  dado o vetor  $q \times 1$  de efeitos aleatórios  $b_i$  é um membro da família exponencial, e que  $b_i$  e  $Y_{it}$  são condicionalmente independentes. Os modelos lineares generalizados mistos têm as seguintes hipóteses:

**i-**  $E(Y_{it}|b_i) = \mu_{it}$  depende dos efeitos fixos e aleatórios através de uma função de ligação  $g(\mu_{it}) = \eta_{it} = x_{it}^T \beta + z_{it}^T b_i$ .

**ii-** A variância condicional está associada à média condicional através de uma função conhecida  $v$  e um parâmetro de escala  $\phi$ ,  $Var(Y_{it}|b_i) = v(\mu_{it})\phi$ .

**iii-** Os efeitos  $b_i$  seguem uma distribuição normal multivariada com média 0 e matriz de variância e covariância  $D$ .

A introdução dos efeitos aleatórios tem implicações na interpretação dos coeficientes de regressão no modelo misto que a torna diferente dos modelos marginais. A interpretação dos  $\beta$ 's nessa classe de modelo é a nível do indivíduo, o coeficiente mede a mudança na resposta do indivíduo  $i$ , a cada unidade que aumentamos ou diminuimos de  $x_{it}$ . A maneira como interpretam-se os coeficientes é o que difere os modelos mistos dos modelos marginais cujo objetivo é realizar inferência sobre a média da população (Fitzmaurice et al, 2011).

O teste da razão de verossimilhança também pode ser utilizado para verificar se  $b_i$  é estatisticamente significativo. O teste pode ser construído para testar sob  $H_0$ ,  $\sigma_b^2 = 0$ , onde  $\sigma_b$  é a variância de  $b_i$ . Se o teste apontar para um efeito estatisticamente significativo, isto implica que a inclusão de  $b_i$  no modelo é importante para explicar a variabilidade entre indivíduos decorrente de medidas repetidas.

Além de avaliar se o efeito de  $b_i$  é estatisticamente significativo é de frequente interesse também expressá-lo, nos modelos em que apenas o intercepto é aleatório, em termos de uma correlação intra-classe (Snijders e Bosker, 1999). O coeficiente de correlação intra-classe indica a proporção não explicada da variância a nível do indivíduo. Ele indica a proporção não explicada da variância a nível do indivíduo. O coeficiente de correlação intra-classe (ICC) é expresso por:

$$ICC = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_\epsilon^2} \quad (4.55)$$

onde  $\hat{\sigma}_b^2$  é a variância estimada para o efeito aleatório e,  $\hat{\sigma}_\epsilon^2$ , a variância estimada dos erros do modelo.



### 4.7.1 Modelos mistos para respostas ordinais

Suponha que  $n$  indivíduos sejam selecionados aleatoriamente de uma população e que para cada indivíduo  $i$  são registradas  $T_i$  observações sobre uma variável resposta ordinal  $Y$  com  $K$  categorias,  $k=1, \dots, K$ . Seja  $Y_{it}$  a resposta da variável  $Y$  para o sujeito  $i$  no tempo  $t$ . O modelo de efeitos mistos para as probabilidades acumuladas pode ser descrito por meio do seguinte logito (Hederker e Mermelstein, 2000):

$$\log\left[\frac{P(Y_{it} \leq k | b_i)}{1 - P(Y_{it} \leq k | b_i)}\right] = \beta_{0k} + X_i^T \beta_k + Z_i^T b_i \quad k = 1, 2, \dots, K - 1 \quad (4.56)$$

em que  $X_i$  representa a matriz de variáveis preditoras para os efeitos fixos,  $Z_i$  uma matriz para os efeitos aleatórios, sendo um subconjunto de  $X_i$ .

Se o vetor de parâmetros  $\beta_k$  não depender de  $k$ , o modelo pode ser descrito a partir do modelo de odds proporcionais, ou seja:

$$\log\left[\frac{P(Y_{it} \leq k | b_i)}{1 - P(Y_{it} \leq k | b_i)}\right] = \beta_{0k} + X_i^T \beta + Z_i^T b_i \quad k = 1, 2, \dots, K - 1 \quad (4.57)$$

A partir da expressão 4.57, uma especificação geral para o modelo de efeitos mistos para respostas ordinais pode ser dada por:

- 1-  $\eta_{it} = \gamma_k + X_i^T \beta + Z_i^T b_i$
- 2- Condicional ao vetor de efeitos aleatórios  $b_i$ , os  $Y_{it}$  são independentes e têm distribuição multinomial;
- 3- Os efeitos  $b_i$  são assumidos seguir distribuição normal bivariada, com média 0 e matriz 2x2 de covariância  $D$  (para um modelo com intercepto e coeficiente angular aleatórios).

Em nosso caso, os logitos acumulados para a variável  $Y_{it}$  possuem três categorias e podem ser expressos por:

$$\log\left[\frac{P(Y_{it} \leq 1 | b_i)}{1 - P(Y_{it} \leq 1 | b_i)}\right] = \beta_{01} + X_i^T \beta + Z_i^T b_i \quad (4.58)$$

$$\log\left[\frac{P(Y_{it} \leq 2 | b_i)}{1 - P(Y_{it} \leq 2 | b_i)}\right] = \beta_{02} + X_i^T \beta + Z_i^T b_i \quad (4.59)$$

Como os coeficientes da regressão não dependem do índice  $k$ , o efeito de cada covariável é o mesmo entre os  $k - 1$  logitos. Desta maneira, a chance de uma resposta em uma categoria maior que  $k$ , para algum  $k$  fixo, é dada por  $\exp(\beta)$

para cada unidade de mudança da covariável  $x_i$ . Assim, para uma resposta ordinal contendo 3 categorias, o modelo descreve simultaneamente o efeito de  $x_i$  sob todas as comparações entre as probabilidades (Hedeker e Mermelstein, 2000) .

Os parâmetros do modelo misto podem ser estimados por meio de métodos de aproximação numérica como o método de Laplace, Quadratura gaussiana ou método de quase-verossimilhança penalizada.

### 4.7.2 Implementação do método GLMM no R

Utilizamos a função `clmm` do pacote `ordinal`, implementado por Rune Haubo Bojesen Christensen em 2011, para o ajuste dos modelos mistos no R. Rune considera um modelo de logito cumulativo misto que assume como resposta uma variável ordinal  $Y_{it}$  com  $k = 1, \dots, K$  categorias. O modelo é definido por:

$$\log\left[\frac{P(Y_{it} \leq k)}{1 - P(Y_{it} \leq k)}\right] = \beta_{0k} - X_i^T \beta - Z_i^T b_i \quad k = 1, 2, \dots, K - 1 \quad (4.60)$$

onde  $X_i$  representa a matriz de variáveis preditoras para os efeitos fixos,  $Z_i$  é um subconjunto de  $X_i$  e representa uma matriz para os efeitos aleatórios.

A interpretação das razões de chances para os modelos mistos ajustados pela função `clmm` no R é análoga a interpretação das razões de chances nos modelos marginais ajustados via `ordgee` no R, no sentido de que as razões de chances associadas ao evento  $Y > k$  são exponenciais dos coeficientes associados aos efeitos fixos.

A inferência para os coeficientes de um modelo `clmm` no R utiliza métodos padrão como testes de Wald e da razão de verossimilhanças sob a hipótese  $H_0 : \beta_k = 0$  . A anova pode ser aplicada a essa função, o que permite a realização do teste da razão de verossimilhança para a comparação de modelos mais simples com outros mais complexos. Apesar dos resíduos não poderem ser extraídos diretamente de um modelo `clmm`, existe um crescente número de trabalhos com sugestões para a construção de resíduos e análise de diagnóstico em modelos mistos como por exemplo [99, 100] que podem ser implementados em rotinas em R.

### 4.7.3 Comparação entre modelos marginais e modelos mistos

As duas metodologias (modelos marginais e modelos mistos) diferem quanto a maneira como tratam a dependência entre as observações ao longo do tempo.

No modelo marginal, é usual estimar o vetor de parâmetros utilizando Equações Estimadas Generalizadas (GEE) propostas por Liang e Zeger (1986), ou seja, usando uma matriz de correlação de trabalho e, assumindo que esta correlação de trabalho

seja a mesma para todos os indivíduos. O procedimento de estimação GEE para modelos marginais é mais utilizado devido ao fato de estar disponível nos principais pacotes de análise estatística.

Quanto ao modelo misto com variável resposta ordinal, existem poucos pacotes, e com algumas limitações, porém é crescente o desenvolvimento de implementações nos principais softwares. No modelo misto, a dependência entre observações no mesmo indivíduo é levada em conta na verossimilhança marginal pela junção da função de densidade condicional ao efeito individual, o que não ocorre no modelo marginal (Molenberghs e Verbeke (2000)).

# Capítulo 5

## Materiais e métodos

### 5.1 População e período do estudo

As participantes deste estudo fizeram parte do projeto “Polimorfismos genéticos e evolução clínica, resposta terapêutica e reações adversas em pacientes submetidas ao tratamento do câncer de mama”, aprovado pelo Comitê de Ética em Pesquisa do INCA em fevereiro de 2009, sob registro nº 129/08. Este projeto teve como um dos objetivos documentar a ocorrência das principais complicações associadas ao tratamento antineoplásico em pacientes com câncer de mama, bem como variáveis terapêuticas e clínicas relacionadas ao tumor e ao indivíduo.

O estudo longitudinal foi realizado em uma coorte hospitalar de pacientes com câncer de mama, diagnosticadas e tratadas no Hospital do Câncer III/INCA, no período de fevereiro de 2009 a setembro de 2011. A coorte incluiu pacientes com indicação de cirurgia curativa, com posterior tratamento quimioterápico adjuvante.

Foram elegíveis mulheres com câncer de mama unilateral, não metastático, com indicação de tratamento cirúrgico curativo e submetidas posteriormente à quimioterapia adjuvante. Foram incluídas as pacientes que aceitaram participar da pesquisa e assinaram o Termo de Consentimento Livre e Esclarecido. Foram excluídas da pesquisa mulheres com câncer de mama contralateral prévio, câncer de mama bilateral sincrônico, que realizaram qualquer tratamento oncológico prévio, e aquelas que não apresentaram condições de responder ao questionário de estudo.

### 5.2 Coleta de dados

As mulheres incluídas no estudo foram acompanhadas ao final de cada ciclo de quimioterapia para a avaliação de reações adversas. Os dados foram obtidos por meio de entrevistas semi-estruturadas e foram disponibilizados para a realização deste trabalho. Para classificação da gravidade da reação foram utilizados os critérios

do Nacional Cancer Institute (EUA) - *Comum Terminology Criteria Adverse Events (CTCAE)*, no qual as reações são graduadas em quatro níveis associados a reações leves (1), moderadas (2), incapacitantes (3) e graves (4) [101, 102]. A descrição dos graus é dada na tabela 5.1.

As reações que apresentaram incidência de graus 3 e 4 superior a 10% (astenia, mialgia, artralgia, diarreia, dor abdominal e mucosite) foram consideradas clinicamente relevantes e incluídas neste estudo. Os níveis das reações foram posteriormente agrupados em 3 classes devido a baixa frequência observada em alguns níveis e a necessidade para o ajuste dos modelos. As classes foram definidas como leve (0 e 1), moderada (2) e grave (3 e 4).

As variáveis clínicas e polimorfismos incluídos neste estudo estão descritos nas tabelas 5.2 e 5.3, respectivamente.

Tabela 5.1: Níveis de gravidade das reações

Grau	Classificação
1	Leve, assintomático. A intervenção não é indicada.
2	Moderada. Quando exige intervenção mínima, local e não invasiva.
3	Grave ou clinicamente significativa mas, não imediatamente com risco de vida.
4	Grave, com consequências fatais, urgência de intervenção.
5	Fatal, quando relacionadas à morte do paciente.

Tabela 5.2: Polimorfismos

Gene	rs	Classificação
<b>CYP2B6</b>	3745274	Selvagem (AA)/homo ou heterovariante(GG ou GA)
<b>CYP2B6</b>	3211371	Selvagem (GG)/homo ou heterovariante(GT ou TT)
<b>CYP3A5</b>	776746	Selvagem (AA)/homo ou heterovariante(AG ou GG)
<b>ABCB1</b>	1128503	Selvagem (AA)/homo ou heterovariante(AG ou GG)
<b>ABCB1</b>	1045642	Selvagem (GG)/homo ou heterovariante(AG ou AA)
<b>GSTP1</b>	1695	Selvagem (GG)/homo ou heterovariante(AA ou AG)
<b>CBR3</b>	8133052	Selvagem (CC)/homo ou heterovariante(CT ou TT)
<b>SLCO1B3</b>	11045585	Selvagem (CC)/homo ou heterovariante(CT ou TT)

Tabela 5.3: Variáveis clínicas e demográficas

Variável	Definição	Classificação
Idade	Idade ao diagnóstico de câncer de mama	<55 ou ≥55
IMC	$\text{peso}/\text{altura}^{\{2\}}$	<30 ou ≥30
Menopausa	Status de menopausa	não ou sim
Dislipidemia	Dislipidemia pré-existente ao tratamento	sim ou não
Diabetes	Diabetes pré-existente ao tratamento	sim ou não
Hipertensão	Hipertensão pré-existente ao tratamento	sim ou não

### 5.3 Análise de dados

O tratamento quimioterápico FAC-D é composto por duas fases, cada uma com 3 ciclos (FAC- 3 ciclos de ciclofosomida, dexorrubicina e 5-fluoroural, e Docetaxel- 3 ciclos de Docetaxel). Para a realização das análises, selecionamos as pacientes que concluíram todos os ciclos de quimioterapia, totalizando 166 mulheres.

Foi realizado um estudo descritivo para as variáveis clínicas e polimorfismos incluídos. Testes Qui-quadrado ao nível de significância 0.05 foram realizados para identificar a associação entre as variáveis. Em seguida realizamos o teste de McNemar, a nível de significância 0.05, a fim de testar a hipótese de alteração na proporção dos níveis das reações entre os ciclos das fases FAC e Docetaxel, e também entre os ciclos dentro de cada fase.

Aplicando teste McNemar, verificamos que a alteração na proporção dos níveis das reações dentro da fase FAC (ciclos 1, 2 e 3) e dentro da fase Docetaxel (ciclos 4,5 e 6) foi não significativa (p-valores >0.05). Como a alteração da proporção entre os ciclos de cada fase foi não significativa, uma análise preliminar foi realizada por meio de modelos de regressão logística ordinal. O intuito foi avaliar a significância das variáveis clínicas e polimorfismos para as reações em cada fase do tratamento. Os modelos de regressão logística ordinal para a fase FAC e para a fase Docetaxel foram construídos utilizando as frequências observadas em todos os ciclos de cada fase.

Para cada reação e fase de tratamento foram ajustados modelos de regressão logística ordinal utilizando como covariáveis o conjunto de polimorfismos e variáveis clínicas. Como critério de seleção, foram ajustados modelos univariados para cada uma das covariáveis em virtude dos métodos de seleção como stepwise não convergirem para a maioria dos modelos ajustados. As covariáveis com p-valor <0.1 foram adicionadas ao modelo completo. O modelo reduzido foi ajustado considerando apenas as variáveis significativas (p-valor <0.05) do modelo completo. O impacto de interações entre covariáveis foi testado nos modelos longitudinais. Foram utilizados modelos de logito cumulativo de chances proporcionais. A qualidade do ajuste foi testada utilizando o teste da razão de verossimilhanças.

O modelo de regressão logística ordinal de chances proporcionais, ajustado pela função polr do pacote mass do R, foi definido como:

$$\log\left(\frac{\pi_1 + \dots + \pi_j}{\pi_{j+1} + \dots + \pi_J}\right) = \beta_{0j} - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \dots - \beta_n x_n \quad j = 1, 2, \dots, J - 1 \quad (5.1)$$

onde  $x_1, x_2, \dots, x_n$  são as variáveis selecionadas dos modelos univariados.

Como a variável resposta tem três níveis, nível 1 se a reação é leve, 2 se moderada e 3, se grave, os modelos ajustados foram:

$$\log\left(\frac{\pi_1}{\pi_2 + \pi_3}\right) = \beta_{01} - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \dots - \beta_n x_n \quad (5.2)$$

e

$$\log\left(\frac{\pi_1 + \pi_2}{\pi_3}\right) = \beta_{02} - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \dots - \beta_n x_n \quad (5.3)$$

Além dos modelos ordinais de chances proporcionais, modelos marginais e modelos lineares generalizados mistos foram utilizados para avaliar a associação entre as variáveis explicativas (variáveis clínicas, polimorfismos e protocolos quimioterápicos) e os níveis das reações relatados pelas mulheres.

Os modelos marginais foram ajustados utilizando o pacote geepack do software R. Esse pacote permite o ajuste utilizando estruturas de correlação de independência ou permutável e intercepto constante ou não constante ao longo do tempo.

O modelo marginal para dados longitudinais utilizado na função ordgee do pacote geepack, como descrito em 4.47 é dado por:

$$\text{logit}[E(Y_{itk})] = \beta_{0k} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \dots + \beta_n x_n \quad k = 1, \dots, K - 1 \quad (5.4)$$

onde  $Y_{itk} = I(O_{it} > k)$ ,  $O_{it}$  é a  $t$ -ésima resposta no tempo do  $i$ -ésimo sujeito e  $x_1, x_2, \dots, x_n$  são as variáveis e interações selecionadas dos modelos univariados.

Os índices  $t$  iguais a  $t_1, t_2, t_3, t_4, t_5$  e  $t_6$  referem-se aos tempos 1,2,3,4,5 e 6, correspondentes aos ciclos de quimioterapia. O índice  $k$  é correspondente ao nível  $k=1$  se a reação é leve,  $k=2$  se moderada e  $k=3$ , se grave. Os modelos marginais ajustados foram então:

$$\log\left[\frac{P(Y_{it} > 1)}{1 - P(Y_{it} > 1)}\right] = \beta_{01} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \dots + \beta_n x_n \quad (5.5)$$

e

$$\log\left[\frac{P(Y_{it} > 2)}{1 - P(Y_{it} > 2)}\right] = \beta_{02} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \dots + \beta_n x_n \quad (5.6)$$

O primeiro modelo marginal foi ajustado para todo o tratamento. Foram utilizadas as respostas dos 6 ciclos avaliados. Definiu-se uma variável indicativa do medicamento utilizado em cada ciclo, assumindo o valor 2 para os ciclos da fase FAC e 1 para os ciclos da fase Docetaxel. Foram ajustados modelos longitudinais univariados utilizando cada uma das variáveis clínicas, polimorfismos e a variável indicativa do medicamento como variáveis independentes. Também foram testadas as interações entre medicamento e polimorfismos, medicamento e variáveis clínicas e entre variáveis clínicas e polimorfismos. As variáveis ou interações com p-valor < 0.1 pelo teste de Wald entraram no modelo completo. As variáveis e interações do modelo completo que apresentaram p-valor < 0.05 pelo teste de Wald, entraram no modelo reduzido.

O segundo modelo marginal ajustado foi o modelo marginal para a fase FAC, e o terceiro, o modelo marginal para a fase Docetaxel. O intuito foi avaliar o impacto da mudança de protocolo sob a significância de variáveis clínicas, polimorfismos e interações. Para isso, foram utilizados apenas os 3 ciclos de cada fase, em cada modelo. As variáveis e interações incorporadas nos modelos completos foram aquelas que nos modelos univariados apresentaram p-valor < 0.1 pelo teste de Wald. As variáveis e interações do modelo completo que apresentaram p-valor < 0.05 pelo teste de Wald, entraram no modelo reduzido., analogamente ao primeiro modelo marginal.

Os modelos mistos foram ajustados utilizando o pacote ordinal do software R. Utilizamos a função `clmm` para o ajuste dos modelos. O procedimento de seleção de variáveis e interações que entraram no modelo final foi análogo ao utilizado nos modelos marginais.

Analogamente aos modelos marginais, nos modelos com efeitos aleatórios também ajustamos três modelos.

O primeiro modelo misto considerou as respostas das reações ao longo de todo o tratamento (6 ciclos). O segundo modelo misto utilizou as respostas da fase FAC e o terceiro utilizou as respostas da fase Docetaxel.

O modelo misto para dados longitudinais ajustado utilizando a função `clmm` do pacote ordinal do software R é dado por:

$$\log\left[\frac{P(Y_{it} \leq k | b_i)}{1 - P(Y_{it} \leq k | b_i)}\right] = \beta_{0k} - X_i^T \beta - Z_i^T b_i \quad k = 1, 2, \dots, K - 1 \quad (5.7)$$

em que  $X_i$  representa a matriz de variáveis preditoras para os efeitos fixos,  $Z_i$  uma matriz para os efeitos aleatórios, sendo um subconjunto de  $X_i$ .

Em nosso caso, a variável  $Y_{it}$  assume três categorias e consideramos apenas o caso do intercepto aleatório devido a quantidade de parâmetros a serem estimados. Os modelos ajustados foram:



$$\log\left[\frac{P(Y_{it} \leq 1 | b_i)}{1 - P(Y_{it} \leq 1 | b_i)}\right] = \beta_{01} - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \dots - \beta_n x_n - b_i \quad (5.8)$$

$$\log\left[\frac{P(Y_{it} \leq 2 | b_i)}{1 - P(Y_{it} \leq 2 | b_i)}\right] = \beta_{02} - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \beta_4 x_4 - \dots - \beta_n x_n - b_i \quad (5.9)$$

Um estudo dos resíduos dos modelos longitudinais foi realizado com o intuito de comparar o comportamento dos resíduos dos modelos marginais e dos modelos mistos. O software R 2.3 foi utilizado para todas as análises.

# Capítulo 6

## Resultados

Apresentaremos neste capítulo os resultados das análises realizadas neste estudo. Iniciaremos com a descrição da coorte por meio das análises exploratórias para as variáveis incluídas. Em seguida apresentaremos os resultados dos modelos ajustados para cada reação adversa.

### 6.1 Análises preliminares

As características clínicas e polimorfismos das mulheres incluídas no estudo estão descritas nas tabelas 6.1 e 6.1.

A figura 6.1 apresenta a variação para as proporções dos níveis leve, moderado e grave das reações ao longo dos ciclos de quimioterapia.

As proporções de mulheres que relatam níveis leves (0-1) de astenia, mialgia e artralgia são menores nos últimos três ciclos (fase Docetaxel) quando comparados aos três ciclos iniciais (fase FAC) de quimioterapia. O contrário ocorre para os níveis moderados (2) e graves (3-4) das reações. Nesses casos observamos um aumento das proporções durante a fase Docetaxel. O teste McNemar confirmou a mudança na proporção dos níveis dessas reações entre qualquer ciclo da fase FAC e qualquer outro da fase Docetaxel (p-valor<0.05).

Dentre as reações gastrointestinais, podemos observar no ciclo 4, um aumento na proporção de níveis graves e moderados e uma diminuição nas proporções de nível leve, quando comparamos aos ciclos da fase FAC. O teste McNemar confirmou a mudança na proporção dos níveis dessas reações entre os ciclos da fase FAC e o ciclo 4 (p-valor<0.05).

A associação entre as reações foi analisada por meio de testes Qui-quadrado. Durante a fase FAC, mialgia e artralgia estão associadas (p-valor<0.05). Na fase Docetaxel, a associação é significativa (p-valor <0.05) para os pares de reações: mialgia e artralgia, astenia e dor abdominal e mucosite e dor abdominal.

Tabela 6.1: Variáveis clínicas - (amostra)

Variáveis clínicas	n	%	Variáveis clínicas	n	%
<b>Idade (anos)</b>			<b>Menopausa</b>		
< 55	87	52.4	Não	107	64.4
>= 55	76	45.8	Sim	51	30.7
Não respondeu	3	1.8	Não respondeu	8	4.9
<b>IMC</b>			<b>Diabetes</b>		
< 30	68	41	Sim	23	13.8
>= 30	26	15.7	Não	137	82.5
Não respondeu	72	43.3	Não respondeu	6	3.6
<b>Dislipidemia</b>			<b>Hipertensão</b>		
Sim	87	52.5	Sim	91	54.8
Não	74	44.5	Não	67	41.6
Não respondeu	5	3	Não respondeu	6	3.6

Tabela 6.2: Polimorfismos - (amostra)

Genes	rs	Número de alelos variantes					
		0(1)		1 or 2(2)		Não avaliado	
		n	%	n	%	n	%
<b>CYP2B6</b>	3745274	66	39.7	73	43.9	27	16.3
<b>CYP2B6</b>	3211371	54	32.5	95	57.2	17	10.2
<b>CYP3A5</b>	776746	9	5.4	147	88.5	10	6
<b>ABCB1</b>	1128503	59	35.5	91	54.8	16	9.6
<b>ABCB1</b>	1045642	70	42.2	77	46.3	19	11.4
<b>GSTP1</b>	1695	65	39.1	93	56	8	4.8
<b>CBR3</b>	8133052	54	32.5	92	41.6	20	12
<b>SLCO1B3</b>	11045585	117	70.5	37	21.1	12	7.2

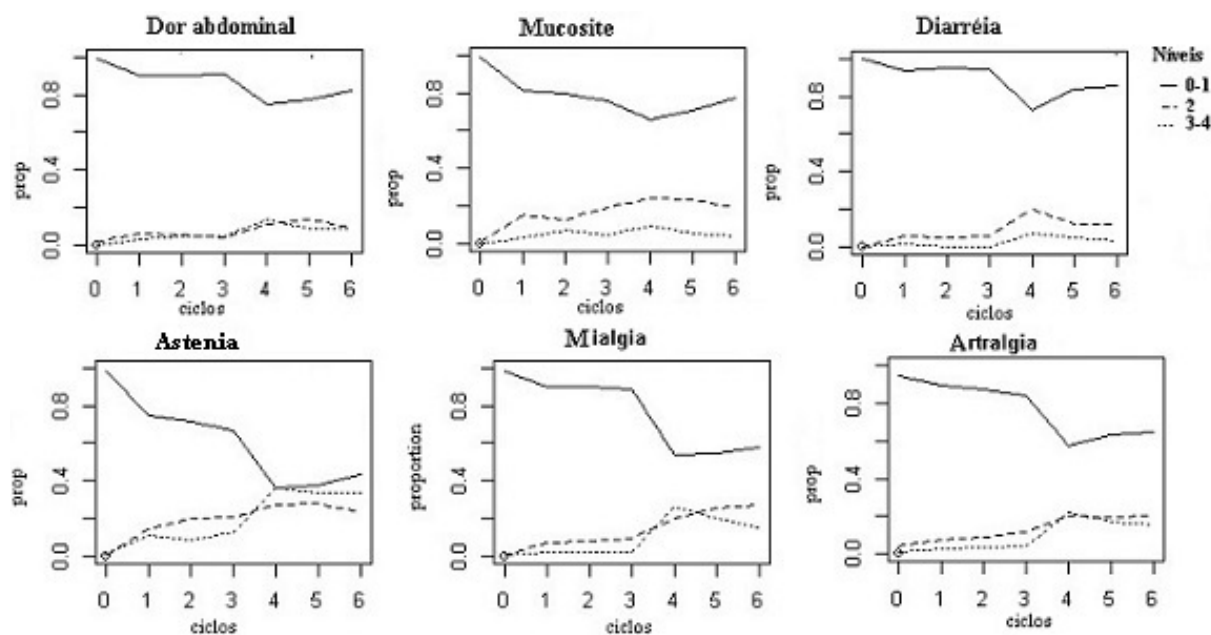


Figura 6.1: Proporções para os níveis das reações ao longo dos ciclos

Apesar do teste McNemar identificar mudança entre ciclos da fase FAC e da fase Docetaxel, dentro de uma mesma fase o teste não conseguiu identificar alteração nas proporções das reações (p-valor  $>0.05$ ). Baseado no resultado do teste, foi realizado inicialmente um estudo de cada fase por meio de modelos de regressão logística ordinal, para essa abordagem consideramos todas as respostas para as reações em cada fase e ajustamos um modelo para a fase FAC e outro para a fase Docetaxel.

Em seguida ajustamos modelos marginais por meio de equações de estimação generalizadas. Esses modelos fornecem a relação entre uma resposta média para as reações e as covariáveis incluídas. Eles diferem dos modelos de regressão logística ordinal por serem modelos para o tratamento de dados longitudinais, e por permitirem a inclusão de uma estrutura de correlação entre as respostas ao longo do tempo, que no caso dos modelos logístico ordinais são tratadas como independentes. Para essa abordagem foram ajustados três modelos marginais para cada reação. O primeiro, para todo o tratamento, utilizou todas as respostas dos seis ciclos. O segundo, para a fase FAC, utilizou as respostas dos três primeiros ciclos. O terceiro, para a fase Docetaxel, utilizou as respostas dos três últimos ciclos.

O último tipo de modelo utilizado foi o modelo linear generalizado misto, que é um modelo utilizado para dados longitudinais que difere dos modelos marginais por permitir a inclusão de efeitos aleatórios a nível de cada sujeito. Isso possibilita que diferenças individuais possam ser incorporadas na modelagem, o que pode representar um ganho na qualidade do ajuste. De maneira análoga aos modelos marginais, para essa abordagem ajustamos três modelos mistos para cada reação. O primeiro pra todo o tratamento, o segundo e o terceiro para as fases FAC e Docetaxel, respectivamente.

O diagrama 6.1 mostra, para cada um dos três tipos de abordagem utilizada neste trabalho, a quantidade de modelos ajustados.

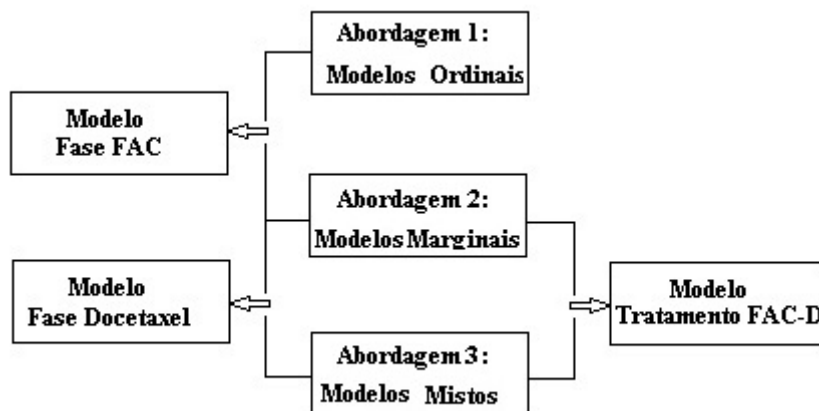


Figura 6.2: Modelos ajustados para cada um dos três tipos de abordagem utilizada

A seleção de variáveis foi realizada da mesma maneira para as três abordagens.

Modelos univariados foram ajustados para cada reação. As covariáveis que apresentaram  $p\text{-valor} < 0.1$  foram incluídas no modelo completo. Apenas as covariáveis que apresentaram  $p\text{-valor} \leq 0.05$  entraram no modelo reduzido. A diferença entre as abordagens concerne nas covariáveis utilizadas em cada passo da modelagem. Para os modelos de regressão logística, foram utilizadas apenas os polimorfismos e variáveis clínicas como covariáveis. Para os modelos marginais e mistos, foram incluídas as interações entre essas variáveis nos modelos para a fase FAC e Docetaxel. Para os modelos ajustados para todo o tratamento foi introduzida a variável protocolo e as interações relacionadas a ela.

O diagrama 6.1 representa o processo de seleção de variáveis para as três abordagens.

## 6.2 Resultados para os modelos de regressão logística ordinal

A seguir são apresentados os resultados dos modelos de regressão logística ordinal ajustados para cada reação. Os dois modelos ordinais ajustados para cada reação podem ser representados por

$$\log\left[\frac{P(Y \leq k)}{1 - P(Y \leq k)}\right] = \beta_{0k} - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \dots - \beta_n x_n \quad k = 1, 2 \quad (6.1)$$

onde  $k$  assume os valores 1 e 2, relativos aos níveis das reações e as covariáveis  $x_n$  são selecionadas dos modelos univariados.

### 6.2.1 Modelos ajustados para astenia

A tabela 6.3 apresenta as estimativas para os coeficientes das variáveis incluídas nos modelos ajustados para astenia.

O teste da razão de verossimilhanças para comparação entre o modelo completo e o modelo reduzido para as duas fases foi não significativo para a hipótese de diferença entre os modelos,  $p\text{-valores}$  de 0.14 e 0.2 para as fases FAC e Docetaxel respectivamente, indicando que não diferença entre os modelos.

### 6.2.2 Modelos ajustados para mialgia

Os resultados para os modelos ajustados para mialgia estão apresentados na tabela 6.4.

Os testes de razão de verossimilhanças para os modelos completo e reduzido

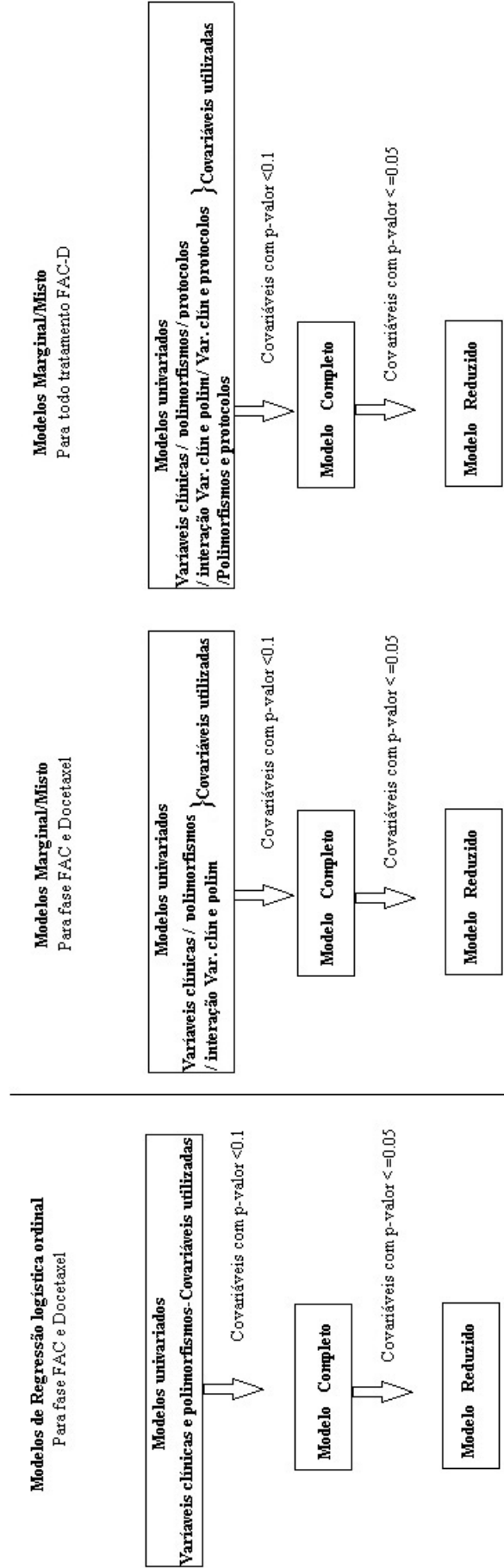


Figura 6.3: Processo de seleção de variáveis para as três abordagens utilizadas

Tabela 6.3: Estimativas para os parâmetros dos modelos de regressão logística ordinal para astenia

	Fase FAC			Fase Docetaxel		
	modelo completo	modelo reduzido	OR	modelo completo	modelo reduzido	OR
CYP2B6 (rs3745274) (0)	Referente					
CYP2B6 (rs3745274) (1 ou 2)	0.03	0.04	1.94			
CYP2B6 (rs3211371) (0)	Referente					
CYP2B6 (rs3211371) (1 ou 2)	0.005	0.006	0.36	0.02	0.02	0.69
ABCB1 (rs1128503) (0)	Referente					
ABCB1 (rs1128503) (1 ou 2)	0.01	0.02	2.29			
ABCB1 (rs1045642) (0)	Referente					
ABCB1 (rs1045642) (1 ou 2)	0.002	0.005	0.34			
GSTP1(rs1695)(0)	Referente					
GSTP1(rs1695)(1 ou 2)	0.0004	0.0003	3.54	0.05	0.13	1.25
SLCO1B3(rs11045585) (0)	Referente					
SLCO1B3(rs11045585) (1 ou 2)	0.01	0.009	0.37			
IMC(<30)	Referente					
IMC (>=30)	0.001	0.001	2.54			
Menopausa (Não)	Referente					
Menopausa (Sim)	0.01	0.002	0.41			
Hipertensão (não)	Referente					
Hipertensão (sim)	0.07					
Dislipidemia (não)	Referente					
Dislipidemia (sim)	0.4					

foram não significativos (p-valores 0.18 e 0.2 para as fases FAC e Docetaxel respectivamente), indicando que não diferença entre os modelos.

### 6.2.3 Modelos ajustados para artralgia

A tabela 6.5 apresenta os resultados dos modelos ajustados para artralgia.

Os testes da razão de verossimilhança para a comparação dos modelos completo e reduzido nas fases FAC e Docetaxel resultaram não significativos (p-valores 0.18 e 0.27), , indicando que não diferença entre os modelos.

### 6.2.4 Modelos ajustados para dor abdominal

A tabela 6.6 apresenta os resultados dos modelos ajustados para dor abdominal.

Os testes de razão de verossimilhança para a comparação dos modelos completo e reduzido nas fases FAC e Docetaxel resultaram não significativos (p-valores 0.33 e 0.21), indicando que não diferença entre os modelos.

### 6.2.5 Modelos ajustados para mucosite

A tabela 6.7 apresenta os resultados dos modelos ajustados para mucosite.

Os testes de razão de verossimilhança para a comparação dos modelos completo e reduzido nas fases FAC e Docetaxel resultaram não significativos (p-valores 0.25 e 0.67), indicando que não diferença entre os modelos.

### 6.2.6 Modelos ajustados para diarreia

A tabela 6.8 apresenta os resultados dos modelos ajustados para diarreia.

Os testes de razão de verossimilhança para a comparação dos modelos completo e reduzido nas fases FAC e Docetaxel resultaram não significativos (p-valores 0.22 e 0.31), indicando que não diferença entre os modelos.

## 6.3 Resultados para os modelos marginais

Os três modelos marginais ajustados para cada reação podem ser representados por:

$$\log\left[\frac{P(Y_{it} > k)}{1 - P(Y_{it} > k)}\right] = \beta_{0k} - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \dots - \beta_n x_n \quad k = 1, 2 \quad (6.2)$$

onde k assume os valores 2 e 3, relativos aos níveis das reações e as covariáveis  $x_n$  são selecionadas dos modelos univariados. O primeiro modelo foi ajustado para todo o tratamento, o segundo e terceiro modelo foram ajustados para a fase FAC



Tabela 6.4: Estimativas para os parâmetros dos modelos de regressão logística ordinal para mialgia

	Fase FAC		Fase Docetaxel	
	modelo completo	modelo reduzido	modelo completo	modelo reduzido
	p-valor	OR	p-valor	OR
CYP2B6 (rs3745274) (0)			Referente	
CYP2B6 (rs3745274) (1 ou 2)			0.02	0.03
CYP2B6 (rs3211371) (0)			Referente	
CYP2B6 (rs3211371) (1 ou 2)			0.02	0.01
ABCB1 (rs1128503) (0)	Referente			
ABCB1 (rs1128503) (1 ou 2)	0.007	0.01	3.43	0.002
ABCB1 (rs1045642) (0)	Referente			
ABCB1 (rs1045642) (1 ou 2)	0.01	0.01	0.32	
CYP3A5(rs1695)(0)			Referente	
CYP3A5(rs1695)(1 ou 2)			0.18	
IMC(<30)	Referente			
IMC (>=30)	0.09			
Idade			0.001	0.0004
Hipertensão (não)	Referente			
Hipertensão (sim)	0.0003	0.0004	0.19	
Diabetes (não)			Referente	
Diabetes (sim)			0.01	0.008
				2.39

Tabela 6.5: Estimativas para os parâmetros dos modelos de regressão logística ordinal para artralgia

	Fase FAC		Fase Docetaxel	
	modelo completo	modelo reduzido	modelo completo	modelo reduzido
	p-valor	p-valor	p-valor	OR
CYP2B6 (rs3745274) (0)	Referente			
CYP2B6 (rs3745274) (1 ou 2)	0.15	0.08	0.28	1.15
CYP2B6 (rs3211371) (0)	Referente			
CYP2B6 (rs3211371) (1 ou 2)	0.18	0.13		
ABCB1 (rs1128503) (0)	Referente			
ABCB1 (rs1128503) (1 ou 2)	0.12			
SLCO1B3 (rs1045642) (0)	Referente			
SLCO1B3 (rs1045642) (1 ou 2)	0.01	0.15		
CYP3A5(rs11045585)(0)	Referente			
CYP3A5(rs11045585)(1 ou 2)	0.02	0.09	0.42	
IMC(<30)	Referente			
IMC (>=30)	0.002	0.005	2.51	
Idade	0.16		0.009	0.02
Hipertensão (não)	Referente			
Hipertensão (sim)	0.009	0.002	0.31	0.03
Diabetes (não)	Referente			
Diabetes (sim)	0.03	0.02	0.13	0.01
Menopausa (não)			Referente	
Menopausa (sim)			0.03	0.07
				1.7

Tabela 6.6: Estimativas para os parâmetros dos modelos de regressão logística ordinal para dor abdominal

	Fase FAC				Fase Docetaxel			
	modelo completo	p-valor	modelo reduzido	OR	modelo completo	p-valor	modelo reduzido	OR
CYP2B6 (rs3745274) (0)	Referente							
CYP2B6 (rs3745274) (1 ou 2)	0.17							
CYP2B6 (rs3211371) (0)			Referente					
CYP2B6 (rs3211371) (1 ou 2)			0.02	0.04	0.04	0.62		
ABCB1 (rs1128503) (0)	Referente							
ABCB1 (rs1128503) (1 ou 2)	0.05		0.16					
CBR3 (rs1045642) (0)	Referente							
CBR3 (rs1045642) (1 ou 2)	0.04	0.05	2.46					
CYP3A5(rs11045585)(0)	Referente							
CYP3A5(rs11045585)(1 ou 2)	0.16							
Dislipidemia (não)			Referente					
Dislipidemia (sim)			0.01	0.14	0.14	0.52		
Hipertensão (não)	Referente							
Hipertensão (sim)	0.02	0.02	0.39	0.16				
Diabetes (não)			Referente					
Diabetes (sim)			0.008	0.004	0.004	2.8		
Menopausa (não)	Referente							
Menopausa (sim)	0.01	0.01	4.16	0.24				

Tabela 6.7: Estimativas para os parâmetros dos modelos de regressão logística ordinal para mucosite

	Fase FAC				Fase Docetaxel			
	modelo completo	modelo reduzido	modelo completo	modelo reduzido	modelo completo	modelo reduzido	modelo completo	modelo reduzido
	p-valor	p-valor	OR	OR	p-valor	p-valor	OR	OR
ABC1 (rs1128503) (0)	Referente							
ABC1 (rs1128503) (1 ou 2)	0.08	0.03	2.02					
SLCO1B3 (rs1045642) (0)				Referente				
SLCO1B3(rs1045642) (1 ou 2)				0.07	0.05	1.54		
Idade				0.18				
Dislipidemia (não)	Referente							
Dislipidemia (sim)	0.08	0.04	0.16	0.45				
Hipertensão (não)	Referente							
Hipertensão (sim)	0.18							
Diabetes (não)	Referente							
Diabetes (sim)	0.002	0.006	3.23	0.01	0.01	2.05		
Menopausa (não)	Referente							
Menopausa (sim)	0.12							

Tabela 6.8: Estimativas para os parâmetros dos modelos de regressão logística ordinal para diarreia

	Fase FAC		Fase Docetaxel	
	modelo completo	modelo reduzido	modelo completo	modelo reduzido
	p-valor	p-valor	p-valor	OR
CYP2B6 (rs3745274) (0)	Referente			
CYP2B6 (rs3745274) (1 ou 2)	0.05	0.11	2.27	
CYP2B6 (rs3211371) (0)	Referente			
CYP2B6 (rs3211371) (1 ou 2)	0.12		0.0004	0.35
CYP3A5(rs776746)(0)	Referente			
CYP3A5(rs776746)(1 ou 2)	0.004	0.003	0.02	
SLCO1B3(rs11045585)(0)			Referente	
SLCO1B3(rs11045585)(1 ou 2)			0.03	0.05
CBR3(rs8133052)(0)			Referente	
CBR3(rs8133052)(1 ou 2)			0.15	
Idade	0.003	0.02	0.88	0.01
IMC(<30)	Referente		0.003	0.96
IMC (>=30)	0.004	0.005	6.49	0.01
Hipertensão (não)	Referente		0.01	2.05
Hipertensão (sim)	0.02	0.02	0.08	

e Docetaxel, respectivamente. O propósito de ajustar modelos para as duas fases foi verificar o efeito de mudança de protocolo para a significância das variáveis. Em termos de modelagem, as variáveis dependentes (reações), são as mesmas para os três modelos. As variáveis independentes para o segundo e terceiro modelo foram as variáveis e interações entre polimorfismos e variáveis clínicas. Para o primeiro modelo, além dessas variáveis, foi incluída uma variável indicativa do protocolo quimioterápico e as interações relacionadas a ela. Somente o nível da reação e protocolo foram consideradas variáveis tempo dependente. Modelos univariados foram ajustados inicialmente e, variáveis e interações que apresentaram  $p\text{-valor} < 0.1$  foram incluídas no modelo completo. Um novo modelo reduzido foi ajustado nos casos em que havia covariáveis não significativas nos modelos completos  $p\text{-valor} > 0.05$ . A comparação dos modelos foi realizada utilizando o Correlation Information Criterion (CIC) proposto em [95]. A bondade do ajuste foi verificada por meio de uma estatística de teste proposta em [97].

Os modelos foram ajustados com intercepto constante e estrutura de correlação permutável. A escolha para a estrutura de correlação foi baseada na análise descritiva. Neste caso, a correlação é a mesma entre quaisquer dois ciclos e apenas um parâmetro de correlação é estimado. As estimativas para os parâmetros de correlação dos primeiros modelos foram 1.27, 1.67, 1.51, 2.76, 2.01, 1.06 com  $p\text{-valores}$  associados de 0.02, 0.001, 0.003,  $<0.05$ ,  $<0.05$ , 0.002 para os modelos com variáveis dependentes astenia, mialgia, artralgia, dor abdominal, diarreia e mucosite, respectivamente.

Os parâmetros de correlação estimados para os modelos da fase FAC foram: 3.03, 1.54, 2.31, 4.17, 1.62, 0.68, com valores  $p$  associados 0.03, 0.0001, 0.02, 0.001, 0.01, 0.28 para as reações astenia, mialgia, artralgia, dor abdominal, mucosite e diarreia, respectivamente. Para a diarreia, a estimativa do parâmetro de correlação não foi significativa, indicando uma estrutura de correlação inadequada.

A estimativa para os parâmetros de correlação para os modelos da fase Docetaxel foram: 2.55, 2.86, 2.56, 2.36, 3.14, -3.26 com  $p\text{-valores}$  associado  $<0.05$ ,  $<0.05$ ,  $<0.05$ ,  $<0.05$ , 0.0001, 0.96 para as reações astenia, mialgia, artralgia, dor abdominal, diarreia e mucosite, respectivamente. Nesta fase, o parâmetro de correlação não resultou também significativo para o modelo para a diarreia. Não foi possível ajustar modelos para a diarreia com outras estruturas de correlação devido à baixa frequência de níveis moderados e graves observadas tanto na fase FAC e Docetaxel.

### 6.3.1 Modelos ajustados para astenia

A tabela 6.9 apresenta os resultados dos modelos finais para astenia. Os modelos univariados ajustados utilizando todos os ciclos, apresentaram como significativas as variáveis, protocolo, IMC e polimorfismos em CYP2B6 (rs3745274). Dessas variáveis, apenas a variável protocolo foi significativa no modelo final (p-valor  $\leq 0.05$ ), as variáveis IMC e CP2B61 apresentaram p-valores 0.4 e 0.9 respectivamente. Um novo modelo foi ajustado, utilizando apenas a variável protocolo. Comparativamente, o modelo utilizando apenas a variável protocolo apresentou melhor ajuste com relação ao anterior CIC=1.98 versus CIC=6.20 e foi considerado como modelo final. O teste para a bondade do ajuste foi não significativo p-valor=0.15, o que sugere um bom ajuste aos dados.

Para os modelos univariados ajustados para os ciclos da fase FAC, polimorfismos no gene GSTP1(rs1695) foi a única variável significativa. O modelo final apresentou CIC=1.83 e o teste para a bondade do ajuste resultou não significativo, com p-valor=0.18.

Na fase Docetaxel, apenas a interação entre IMC e polimorfismos no gene CBR3 foi significativa nos modelos univariados, o modelo final apresentou CIC=6.16 e o teste para a bondade do ajuste não foi significativo, p-valor=0.1.

### 6.3.2 Modelos ajustados para mialgia

A tabela 6.10 apresenta os resultados para os modelos para mialgia. Os modelos univariados com todos os ciclos apresentaram como significativas as variáveis protocolo e interação entre menopausa e polimorfismos no gene, ambas permaneceram significativas no modelo final. O valor da medida CIC foi 6.12 e o teste para a bondade do ajuste foi significativo p-valor=0.01.

Para a fase FAC, nenhuma variável foi significativa para os modelos univariados. Na fase Docetaxel, polimorfismos no gene ABCB1 foi a única variável significativa, o modelo final apresentou CIC=1.88 e, o teste para a bondade do ajuste resultou não significativo (p-valor=0.12).

### 6.3.3 Modelos ajustados para artralgia

Os resultados para os modelos para artralgia encontram-se na tabela 6.11. Os modelos univariados utilizando todos os ciclos apresentaram como variáveis significativas protocolo e interação entre IMC e polimorfismos no gene ABCB1 (rs1045642) que permaneceram significativas no modelo final. O valor da medida CIC foi 2.51 e o teste para a bondade do ajuste resultou não significativo p-valor=0.08.

Para a fase FAC, as variáveis significativas para os modelos univariados foram

Tabela 6.9: Estimativas para os parâmetros dos modelos marginais para astenia

	Modelo 1*		Modelo 2 **		Modelo 3***	
	OR	p-valor	OR	p-valor	OR	p-valor
Protocolo FAC	Referente					
Protocol Docetaxel	3.7	(<0.05)				
GSTP1(rs1695) (0)	Referente					
GSTP1(rs1695) (1 or 2)	4.69	0.03				
IMC(<30)	Referente					
IMC (>=30)	8.1	0.01				
CBR3(rs8133052) (0) ou BMI (<30)	Referente					
CBR3(rs8133052) (1 or 2) e BMI (>=30)	0.25	0.03				

\*modelo incluindo todos os ciclos

\*\* modelo incluindo os ciclos da fase FAC

\*\*\* modelo incluindo os ciclos da fase Docetaxel



Tabela 6.10: Estimativas para os parâmetros dos modelos marginais para mialgia

	Modelo 1*		Model 3 ***	
	OR	p-valor	OR	p-valor
Protocolo FAC	Referente			
Protocolo Docetaxel	7.17	<0.05		
ABCB1 (rs1128503) (0) ou menopausa (não)	Referente			
ABCB1 (rs1128503) (1 or 2) e menopausa (sim)	0.77	0.01		
ABCB1 (rs1128503) (0)			Referente	
ABCB1 (rs1128503) (1 or 2)			0.35	0.02

\*modelo incluindo todos os ciclos

\*\*\* modelo incluindo os ciclos da fase Docetaxel

IMC, hipertensão e polimorfismos no gene ABCB1(rs1045642). Dessas variáveis apenas IMC e hipertensão foram significativas no modelo final que apresentou CIC=2.91. O modelo ajustado apenas com IMC e hipertensão como covariáveis apresentou CIC=9.6 , e por esse motivo não foi selecionado como modelo final. O teste para a bondade do ajuste resultou não significativo p-valor=0.11.

Os modelos univariados para a fase Docetaxel apresentaram como variáveis significativas as interações entre idade e polimorfismos em SLCO1B3, dislipidemia e polimorfismos em CYP2B6 e polimorfismos em ABCB1. No modelo final apenas a interação entre idade e SLCO1B3 foi significativa, o valor para CIC foi 11.63. Um novo modelo foi ajustado apenas com essa interação, o valor para CIC foi 4.47 por isso apresentou melhor ajuste que o anterior e foi considerado como modelo final. Apesar disso, o teste para a bondade do ajuste foi significativo (p-valor=0.04).

### 6.3.4 Modelos ajustados para dor abdominal

Os resultados para os modelos para dor abdominal estão apresentados na tabela 6.12. Os modelos univariados para todo o tratamento apresentaram como significativas as variáveis protocolo e interação entre IMC e polimorfismos no gene CBR3, dessas variáveis apenas protocolo foi significativa no modelo final que apresentou valor para CIC de 13.7. Um novo modelo foi ajustado considerando apenas protocolo, o valor para CIC foi de 2.06 e, por isso ele foi considerado como modelo final. O teste para bondade do ajuste resultou não significativo, p-valor=0.18.

Os modelos univariados para a fase FAC não resultaram em nenhuma variável significativa. Para a fase Docetaxel, os modelos univariados apresentaram como significativa as interações entre IMC e polimorfismos no gene CBR3 e hipertensão e polimorfismos no gene CYP2B61, dessas interações apenas a interação entre hipertensão e polimorfismos em CYP2B61 foi significativa para o modelo final, que apresentou

Tabela 6.11: Estimativas para os parâmetros dos modelos marginais para artralgia

	Model 1*		Model 2 **		Model 3***	
	OR	p-valor	OR	p-valor	OR	p-valor
Protocolo FAC	Referente					
Protocolo Docetaxel	9.9	<0.05				
ABCB1 (rs1045642) (0 ou hipertensão (não)	Referente					
ABCB1 (rs1045642) (1 or 2) e hipertensão (sim)	0.03	0.01				
IMC (<30)	Referente					
IMC (>=30)	13.2	0.01				
Hipertensão (não)	Referente					
Hipertensão (sim)	0.1	0.02				
SLCO1B3(rs11045585) (0) ou Idade (<55)	Referente					
SLCO1B3(rs11045585) (1 or 2) e Idade (>=55)	0.77	0.01				

\*modelo incluindo todos os ciclos

\*\* modelo incluindo os ciclos da fase FAC

\*\*\* modelo incluindo os ciclos da fase Docetaxel

Tabela 6.12: Estimativas para os parâmetros dos modelos marginais para dor abdominal

	Modelo 1*		Modelo 3 ***	
	OR	p-valor	OR	p-valor
Protocolo FAC	Referente			
Protocolo Docetaxel	2.66	0.04		
CBR3(rs8133052) (0)	Referente			
CBR3(rs8133052) (1 ou 2)			6.95	0.01
Hipertensão (não)	Referente			
Hipertensão (sim)			0.12	0.008
CYP2B6(rs3745274) (0) ou Hipertensão (não)	Referente			
CYP2B6(rs3745274) (1 ou 2) e Hipertensão (sim)			10.5	0.04

\*modelo incluindo todos os ciclos

\*\*\* modelo incluindo os ciclos da fase Docetaxel

CIC=10.87. Um novo modelo foi ajustado apenas com essa interação e, o valor de CIC observado foi 7.05 e por isso ele foi considerado como modelo final. O teste para a bondade do ajuste resultou não significativo (p-valor=0.06).

### 6.3.5 Modelos ajustados para mucosite

Os resultados para os modelos ajustados para mucosite estão apresentados

na tabela 6.13. Os modelos univariados ajustados para todos os ciclos apresentaram como significativas protocolo e interação entre idade e polimorfismos em CBR3, dessas variáveis, apenas protocolo permaneceu significativa no modelo final, que teve o valor para CIC de 7.3. Um novo modelo foi ajustado apenas com a variável protocolo e, o valor obtido para o CIC foi de 2.53, por isso ele foi considerado modelo final. O teste para a bondade do ajuste resultou não significativo (p-valor=0.08).

Para a fase FAC nenhuma variável foi significativa para os modelos univariados. Na fase Docetaxel, a única variável significativa para os modelos univariados foi polimorfismos em ABCB1, o modelo final apresentou CIC=1.45 e p-valor associado de 0.1.

### 6.3.6 Modelos ajustados para diarreia

Os resultados para os modelos marginais ajustados para diarreia estão apresentados na tabela 6.14. Os modelos univariados para todos os ciclos apresentaram como significativas as variáveis protocolo, interação entre protocolo e idade, e interação entre idade e polimorfismos em CBR3(rs8133052). Dessas variáveis apenas protocolo

Tabela 6.13: Estimativas para os parâmetros dos modelos marginais para mucosite

	Modelo 1*		Modelo 3 ***	
	OR	p-valor	OR	p-valor
Protocolo FAC	Referente			
Protocolo Docetaxel	2.88	0.03		
ABCB1(rs1128503) (0)	Referente			
ABCB1(rs1128503) (1 ou 2)			4.1	0.04

\*modelo incluindo todos os ciclos

\*\*\* modelo incluindo os ciclos da fase Docetaxel

Tabela 6.14: Estimativas para os parâmetros dos modelos marginais para diarreia

	Modelo 1*		Modelo 3 ***	
	OR	p-valor	OR	p-valor
Protocolo FAC	Referente			
Protocolo Docetaxel	7.5	0.001		
Protocolo FAC e Idade ( $j=55$ )	Referente			
Protocolo Docetaxel ou Idade ( $j=55$ )	0.91	0.007		
Dislipidemia (não) e CBR3 (rs8133052) (0)	Referente			
Dislipidemia (sim) ou CBR3 (rs8133052)(1 ou 2)			0.04	0.001

\*modelo incluindo todos os ciclos

\*\*\* modelo incluindo os ciclos da fase Docetaxel

e interação entre protocolo e idade permaneceram significativas no modelo final, que apresentou CIC=2.56. Um novo modelo foi ajustado apenas com protocolo e interação entre protocolo e idade, o valor para CIC observado foi de 1.09 e por isso ele foi considerado o modelo final. O teste para a bondade do ajuste resultou não significativo (p-valor=0.23).

Para a fase FAC nenhuma variável foi significativa nos modelos univariados. Na fase Docetaxel polimorfismos em ABCB1 (rs1128503) e interação entre dislipidemia e polimorfismos CBR3 (rs8133052) foram as variáveis significativas para os modelos univariados, dessas variáveis apenas a interação entre dislipidemia e polimorfismos em CBR3 (rs8133052) permaneceu significativa no modelo final que teve o valor de CIC igual a 4.25. Um novo modelo foi ajustado apenas com essa interação e o valor para CIC observado foi 1.36. O teste para a bondade do ajuste foi não significativo (p-valor=0.08).

## 6.4 Resultados para os modelos lineares generalizados mistos

Os três modelos mistos ajustados para cada reação podem ser representados por

$$\log\left[\frac{P(Y_{it} \leq k | b_i)}{1 - P(Y_{it} \leq k | b_i)}\right] = \beta_{0k} - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_3 - \dots - \beta_n x_n - b_i \quad k = 1, 2 \quad (6.3)$$

onde  $k$  assume os valores 1 e 2, relativos aos níveis das reações e as covariáveis  $x_n$  são selecionadas dos modelos univariados. Foi necessário considerar apenas o intercepto aleatório  $b_i$  para o ajuste do modelo em virtude da convergência ser afetada pela quantidade de parâmetros a serem estimados. O primeiro ajuste considerou os seis ciclos de tratamento, o segundo e terceiro modelo foram ajustados para a fase FAC e Docetaxel, respectivamente. O propósito de ajustar modelos para as duas fases foi verificar o efeito de mudança de protocolo para a significância das variáveis, assim como nos modelos marginais. As variáveis dependentes (reações), são as mesmas para os três modelos. As variáveis independentes para o segundo e terceiro modelo foram as variáveis e interações entre polimorfismos e variáveis clínicas. Para o primeiro modelo, além dessas variáveis, foi incluída uma variável indicativa do protocolo quimioterápico e as interações relacionadas a ela. Somente o nível da reação e protocolo foram consideradas variáveis tempo dependente. Modelos univariados foram ajustados inicialmente e, variáveis e interações que apresentaram  $p$ -valor  $< 0.1$  foram incluídas no modelo final. Um novo ajuste foi realizado nos casos em que havia covariáveis não significativas nos modelos finais  $p$ -valor  $> 0.05$ . A comparação dos modelos e o teste para a significância do intercepto aleatório foi realizada utilizando o teste da razão de verossimilhanças.

O valores do desvio padrão para os interceptos aleatórios dos primeiros modelos foram 1.51, 2.12, 2.05, 2.44, 1.92 e 1.60, para as variáveis desfecho astenia, mialgia, artralgia, dor abdominal mucosite e diarreia respectivamente. Para os modelos mistos para a fase FAC os valores foram: 2.1, 3.44, 2.45, 4.24, 5.34, 2.02, 1.85 para as variáveis desfecho astenia, mialgia, artralgia, dor abdominal, mucosite e diarreia respectivamente. As estimativas para o desvio padrão do intercepto aleatório dos modelos ajustados para a fase Docetaxel foram: 1.89, 2.55, 2.8, 2.78, 2.18, 3.13, 1.07 para as variáveis desfecho mialgia, artralgia, dor abdominal, mucosite e diarreia respectivamente.

### 6.4.1 Modelos ajustados para astenia

As estimativas para os modelos mistos ajustados para astenia são apresentadas na tabela 6.15. Os modelos univariados considerando todos os ciclos apresentaram

como significativas as variáveis protocolo, IMC e polimorfismos em CYP2B6. Dessas variáveis, apenas protocolo foi significativa para o modelo final. Um novo modelo foi ajustado considerando apenas o protocolo como covariável, o teste da razão de verossimilhanças para a comparação desses dois modelos resultou não significativo (p-valor= 0.19) e por isso o modelo com única covariável medicamento foi considerado como modelo final. O teste da razão de verossimilhanças para o intercepto aleatório resultou significativo (p-valor <0.05).

Os modelos univariados para a fase FAC apresentou como significativa apenas polimorfismos em GSTP1(rs1695) (p-valor=0.02). O teste para o intercepto aleatório foi significativo, confirmando a importância do intercepto para o modelo (p-valor<0.05).

Para a fase Docetaxel, as variáveis IMC e interação entre IMC e polimorfismos em CBR3 foram significativas para os modelos univariados e permaneceram significativas no modelo final. O teste da razão de verossimilhanças para o intercepto aleatório resultou significativo (p-valor<0.05).

#### **6.4.2 Modelos ajustados para mialgia**

Os resultados para os modelos mistos ajustados para mialgia são apresentados na tabela 6.16. Os modelos univariados para todos os ciclos apresentaram como variáveis significativas protocolo e interação entre menopausa e polimorfismos no gene ABCB1 (rs1128503), dessas variáveis apenas protocolo foi significativa para o modelo final. Um novo modelo foi ajustado apenas com essa variável como variável independente. O teste da razão de verossimilhança entre esses dois modelos foi não significativo (p-valor=0.98) e por isso o modelo com apenas a covariável protocolo foi considerado como modelo final. O teste da razão de verossimilhanças para o intercepto foi significativo (p-valor<0.05) o que indica necessidade do intercepto aleatório.

Na fase FAC nenhuma variável foi significativa para os modelos univariados. Na fase Docetaxel apenas polimorfismos no gene ABCB1(rs1128503) foi significativo. O teste de verossimilhança para o intercepto aleatório foi significativo (p-valor=0.02).

#### **6.4.3 Modelos ajustados para artralgia**

Os resultados para os modelos mistos para artralgia são apresentados na tabela 6.17. Os modelos univariados para todos os ciclos apresentaram como significativas as variáveis protocolo, hipertensão e interação entre IMC e polimorfismos em ABCB1(rs1045642) que permaneceram significativas no modelo final. O teste de razão de verossimilhança para o intercepto resultou significativo, p-valor<0.05, indicando a necessidade do intercepto aleatório.

Tabela 6.15: Estimativas para os parâmetros dos modelos mistos para astenia

	Modelo 1*		Modelo 2**		Modelo 3***	
Protocolo FAC	OR	p-valor	OR	p-valor	OR	p-valor
Protocol Docetaxel	Referente					
GSTP1(rs1695) (0)	7.3	(;0.05)				
GSTP1(rs1695) (1 or 2)			Referente			
IMC(<30)			7.03	0.02		
IMC (>=30)					Referente	
CBR3(rs8133052) (0) ou IMC (<30)					8.2	0.005
CBR3(rs8133052) (1 or 2) e IMC (>=30)					Referente	
					0.04	0.03

\*modelo incluindo todos os ciclos

\*\* modelo incluindo os ciclos da fase FAC

\*\*\* modelo incluindo os ciclos da fase Docetaxel

Tabela 6.16: Estimativas para os parâmetros dos modelos mistos para mialgia

	Modelo 1*		Model 3 ***	
	OR	p-valor	OR	p-valor
Protocolo FAC	Referente			
Protocolo Docetaxel	12.1	¡0.05		
ABCB1 (rs1128503) (0)	Referente			
ABCB1 (rs1128503) (1 or 2)			0.14	0.02

\*modelo incluindo todos os ciclos

\*\*\* modelo incluindo os ciclos da fase Docetaxel

Tabela 6.17: Estimativas para os parâmetros dos modelos mistos para artralgia

	Model 1*		Model 2 **	
	OR	p-valor	OR	p-valor
Protocolo FAC	Referente			
Protocolo Docetaxel	8.1	¡0.05		
ABCB1 (rs1045642) (0) ou IMC (<30)	Referente			
ABCB1 (rs1045642) (1 or 2) e IMC (>=30)	3.29	0.04		
Hipertensão (não)	Referente		Referente	
Hipertensão (sim)	0.24	0.04	0.08	0.04

\*modelo incluindo todos os ciclos

\*\* modelo incluindo os ciclos da fase FAC

\*\*\* modelo incluindo os ciclos da fase Docetaxel

Os modelos univariados para a fase FAC apresentou como significativas hipertensão, IMC e polimorfismos em ABCB1(rs1045642), dessas variáveis, apenas hipertensão foi significativa no modelo final. Um novo modelo foi ajustado apenas com hipertensão como covariável e o teste de para comparação dos modelos resultou não significativo e o último modelo foi adotado como modelo final. O teste para a significancia do intercepto resultou significativo (p-valor<0.05). Para a fase Docetaxel nenhuma variável foi significativa para os modelos univariados.

#### 6.4.4 Modelos ajustados para dor abdominal

Os resultados para os modelos mistos ajustados estão apresentados na tabela 6.18. Os resultados para os modelos univariados ajustados para todo o tratamento apresentaram como significativas as variáveis protocolo e interação entre IMC e polimorfismos em CBR3 que permaneceram significativas no modelo final. O teste da razão de verossimilhanças para o intercepto resultou significativo (p-valor<0.05).

Os modelos univariados para a fase FAC não apresentaram nenhuma variável



Tabela 6.18: Estimativas para os parâmetros dos modelos mistos para dor abdominal

	Modelo 1*		Modelo 3 ***	
	OR	p-valor	OR	p-valor
Protocolo FAC	Referente			
Protocolo Docetaxel	6.15	0.05		
CBR3(rs8133052) (0) e IMC(<30)	Referente			
CBR3(rs8133052) (1 ou 2) ou (IMC≥30)	0.05	0.04		
Hipertensão (não)	Referente			
Hipertensão (sim)			0.12	0.03
CYP2B6(rs3745274) (0)	Referente			
CYP2B6(rs3745274) (1 ou 2)			0.11	0.04
CYP2B6(rs3745274) (0) ou Hipertensão (não)	Referente			
CYP2B6(rs3745274) (1 ou 2) e Hipertensão (sim)			10.7	0.04

\*modelo incluindo todos os ciclos

\*\*\* modelo incluindo os ciclos da fase Docetaxel

significativa. Para a fase Docetaxel, os modelos univariados apresentaram como significativas as variáveis hipertensão, polimorfismos em CYP2B6(rs3745274) e interação entre hipertensão e polimorfismos em CYP2B6(rs3745274), que permaneceram significativas no modelo final. O teste para o intercepto

### 6.4.5 Modelos ajustados para mucosite

Os resultados para os modelos mistos ajustados para mucosite estão apresentados na tabela 6.19. Os modelos univariados para todo o tratamento apresentaram como significativas protocolo e polimorfismos no gene ABCB1, dessas variáveis, apenas protocolo foi significativa no modelo final. Um novo modelo foi ajustado para essa covariável e o teste da verossimilhança para a comparação entre esses dois modelos resultou não significativo (p-valor=0.7) e por isso o modelo apenas com a covariável protocolo foi considerado como modelo final. O teste para o intercepto aleatório resultou significativo (p-valor<0.05).

Os modelos univariados para a fase FAC apresentaram como significativas as variáveis diabetes e polimorfismos no gene ABCB1, dessas variáveis apenas diabetes foi significativa para o modelo final. O modelo apresentando apenas diabetes com covariável foi considerado o melhor modelo e assumido como modelo final (teste para comparação entre os dois modelos resultou não significativo p-valor=0.31). O teste para a significância do intercepto aleatório resultou significativo (p-valor=0.001). Na fase Docetaxel nenhuma variável foi significativa para os modelos univariados.

Tabela 6.19: Estimativas para os parâmetros dos modelos mistos para mucosite

	Modelo 1*		Modelo 2 **	
	OR	p-valor	OR	p-valor
Protocolo FAC	Referente			
Protocolo Docetaxel	4.85	0.007		
Diabetes (não)	Referente			
Diabetes (sim)			7.2	0.04

\*modelo incluindo todos os ciclos

\*\*modelo incluindo os ciclos da fase FAC

Tabela 6.20: Estimativas para os parâmetros dos modelos mistos para diarreia

	Modelo 1*		Modelo 3 ***	
	OR	p-valor	OR	p-valor
Protocolo FAC	Referente			
Protocolo Docetaxel	7.53	0.0001		
Menopausa (não)	Referente			
Menopausa (sim)			0.07	0.005
ABCB1(rs8133052) (0)	Referente			
ABCB1(rs8133052) (1 ou 2))			0.14	0.03
Menopausa (não) e ABCB1(rs8133052) (0)	Referente			
Menopausa(sim) ou ABCB1(rs8133052)(1 ou 2)			12.4	0.01

\*modelo incluindo todos os ciclos

\*\*\* modelo incluindo os ciclos da fase Docetaxel

### 6.4.6 Modelos ajustados para diarreia

Os resultados para os modelos mistos ajustados para diarreia são apresentados na tabela 6.20. Os modelos univariados para todo o tratamento apresentaram como significativa a penas a variável protocolo. O teste para a significância do intercepto resultou significativo (p-valor<0.05).

Para a fase FAC nenhuma variável foi significativa para os modelos univariados. Para a fase Docetaxel, os modelos univariados apresentaram como significativas as interações entre menopausa e polimorfismos em ABCB1 e dislipidemia e polimorfismos em CBR3, dessas interações, apenas a interação entre menopausa a ABCB1 foi significativa para o modelo final. O novo modelo ajustado apenas com essa interação não apresentou diferença significativa para o modelo anterior (p-valor=0.39) e, por isso foi considerado como modelo final. O intercepto aleatório foi significativo (p-valor=0.02).

## 6.5 Análise de resíduos

Para comparar o ajuste nos modelos marginais e mistos os resíduos padronizados foram calculados em ambos modelos segundo as referências [24, 95, 97]

Os gráficos dos resíduos dos modelos encontram-se no capítulo Apêndice A.

# Capítulo 7

## Discussão

Fatores clínicos, genéticos e protocolos quimioterápicos podem estar relacionados a gravidade de reações relatadas por pacientes durante o tratamento quimioterápico para o câncer de mama [15–18],[35, 38]. A identificação de fatores associados às reações pode auxiliar na condução de terapêuticas mais eficazes, que promovam uma melhor qualidade de vida para os pacientes.

O efeito da mudança de protocolo no conjunto de características individuais associadas ao risco para o desenvolvimento de uma reação ainda é pouco conhecido na literatura, possivelmente devido ao fato de grande parte dos estudos serem relacionados a um tipo de protocolo específico ou a uma fase de um tratamento quimioterápico. A utilização pouco frequente de ferramentas estatísticas para dados categóricos longitudinais em estudos epidemiológicos também pode ser um componente que contribui para a existência de poucos estudos nesse sentido.

Neste trabalho, a utilização de modelos marginais e modelos lineares generalizados mistos foi de extrema importância para a identificação de fatores associados a ocorrência de reações, e na confirmação da hipótese de que mudanças de protocolos podem estar associadas a mudanças nas características individuais associadas às reações, além da mudança na gravidade das reações.

Os resultados mostraram que o perfil das 166 mulheres participantes do estudo foi caracterizado por mulheres com idade menor 55 anos, mulheres que não estavam em menopausa, e mulheres que apresentaram pelo menos um polimorfismo nos genes descritos, excetuando-se os polimorfismos no gene *SLCO1B3*.

A evolução das proporções dos níveis das reações mostrou um aumento nas proporções de níveis graves para as reações durante os ciclos da fase Docetaxel quando comparados aos ciclos da fase FAC. Esse resultado vai ao encontro de alguns estudos que relatam maior toxicidade tratamentos com Docetaxel que em outros protocolos [35, 38]. Em particular, a relação entre Docetaxel e níveis maiores de fadiga/astenia está presente em alguns trabalhos [38] e [35].

Os testes Qui-quadrado mostraram que na fase FAC e Docetaxel, mialgia e ar-

tralgia estiveram associadas, o que pode estar relacionado ao fato de serem reações musculares. Na fase Docetaxel, as reações gastrointestinais associadas foram muco-site e dor abdominal. Além disso, a associação foi significativa também para astenia e dor abdominal. Como astenia está relacionada a fraqueza ou condição delibitadora, ela pode estar associada a dores ou outras reações, inclusive gastrointestinais.

Os modelos para astenia mostraram que, em todo o tratamento, a única variável significativa foi protocolo, tanto para os modelos marginais quanto para os modelos mistos, com o protocolo Docetaxel associado a maiores chances para níveis graves da reação. Isso sugere que o protocolo Docetaxel agrava a reação quando comparado ao protocolo FAC. O teste para o ajuste do modelo marginal indicou um bom ajuste ( $p$ -valor=0.18), apesar disso, os resíduos para o modelo misto apresentaram menor variação e o teste de razão de verossimilhança apontou a necessidade para o intercepto aleatório para o melhor ajuste do modelo. Isso indica que o modelo misto melhor se adequou aos dados.

Na fase FAC, o modelo de regressão logística apresentou como variáveis relacionadas a maiores chances para níveis graves de astenia polimorfismos em CYP2B6 (rs3211371), ABCB1(rs3211371), GSTP1(rs1695) e IMC ( $\geq 30$ ). Polimorfismos em GSTP1(rs1695) também estavam associados a maiores chances para níveis graves de astenia nos modelos marginais e mistos. O teste para o ajuste do modelo marginal foi não significativo indicando um bom ajuste aos dados. Apesar do teste de razão de verossimilhanças para o modelo misto indicar a necessidade do intercepto aleatório para o modelo, os resíduos para o modelo marginal apresentaram menor variação, o que sugere que o melhor ajuste é dado pelo modelo marginal.

Na fase Docetaxel a única variável significativa para os modelos de regressão logística foi polimorfismos em CYP2B6(rs3745274), associado a menores chances para níveis graves da reação. Os modelos marginal e misto para essa fase apresentaram como variáveis significativas IMC ( $\geq 30$ ), associado a maiores chances para níveis graves de astenia e interação entre IMC ( $\geq 30$ ) e polimorfismos em CBR3 (rs8133052), associada a menores chances para níveis graves. O teste para o ajuste do modelo marginal resultou não significativo ( $p$ -valor=0.1), o que indica bom ajuste. Apesar do teste de razão de verossimilhanças no modelo misto resultar significativo, indicando a necessidade de um intercepto aleatório, os resíduos para o modelo marginal apresentaram menor dispersão, o que sugere a melhor adequação aos dados pelo modelo marginal.

Os modelos para mialgia para todo o tratamento apresentaram o protocolo Docetaxel como associado a maiores chances para níveis graves da reação, tanto para o modelo marginal como para o misto, agravando as reações quando comparado ao protocolo FAC. O modelo marginal também apresentou como significativa a interação entre polimorfismos em ABCB1(rs1128503) e menopausa, associada a meno-

res chances para níveis graves da reação. O teste para ajuste do modelo marginal resultou significativo, e o teste de razão de verossimilhanças para o intercepto aleatório também resultou significativo, o que indica a necessidade do intercepto para o ajuste aos dados. Apesar disso, os resíduos para o modelo marginal apresentaram menor dispersão o que sugere um melhor ajuste pelo modelo marginal.

Para a fase FAC, o modelo de regressão logística apresentou como variáveis associadas a maiores chances para níveis graves de mialgia polimorfismos em ABCB1 (rs1128503), polimorfismos em ABCB1(rs1045642) e hipertensão. Para os modelos marginais e mistos, nenhuma variável foi significativa na fase FAC.

Na fase Docetaxel, o modelo de regressão logística apresentou como variáveis associadas a maiores chances para níveis graves de mialgia polimorfismos em CYP2B6 (rs3745274) e diabetes. Tanto o modelo marginal como o misto apresentaram como significativa a variável polimorfismos em ABCB1(rs1128503), associada a menores chances para níveis graves de mialgia. O teste para ajuste do modelo marginal indicou um bom ajuste e os resíduos apresentaram menor variação que os resíduos para o modelo misto, apesar do teste da razão de verossimilhanças indicar a necessidade de um intercepto aleatório para o modelo.

Os modelos marginais e mistos para artralgia para todo o tratamento mostraram que o protocolo Docetaxel estava associado a maiores chances para níveis graves de artralgia. No modelo marginal, a interação entre polimorfismos em ABCB1 (rs1045642) e hipertensão foi significativa, associada a menores chances para níveis graves. Nos modelos mistos, esse mesmo polimorfismo foi significativo na interação com IMC, além dessa interação, hipertensão também estava associada a menores chances para níveis graves da reação. O teste da razão de verossimilhanças para os modelos mistos resultou significativo, apesar disso, o teste para a bondade do ajuste do modelo marginal foi não significativo, e os resíduos apresentaram menor variação, o que indica o modelo marginal como modelo que melhor se ajusta aos dados.

Para a fase FAC, o modelo de regressão logística apresentou IMC ( $\geq 30$ ) como associado a maiores chances para níveis graves de artralgia. O modelo marginal também apresentou IMC ( $\geq 30$ ) como associado a maiores chances para níveis graves de artralgia e hipertensão associada a menores chances. Para o modelo misto, hipertensão também foi significativa e associada a menores chances para níveis graves. O teste para a bondade do ajuste para o modelo marginal resultou não significativo, os gráficos para os resíduos também indicaram um melhor ajuste pelo o modelo marginal.

Na fase Docetaxel, nenhuma variável foi significativa para o modelo misto. O modelo de regressão logística indicou como significativas diabetes, associada a menores chances para níveis graves e hipertensão e idade, associados a menores chances para níveis graves. O modelo marginal não apresentou um bom ajuste, o teste para

qualidade do ajuste foi significativo. A interação entre polimorfismos em SLCO1B3 e idade foi significativa para esse modelo, associada a menores chances para níveis graves.

Os modelos para todo o tratamento para dor abdominal indicaram o protocolo como variável associada a maiores chances para níveis graves de dor abdominal tanto para o modelo marginal quanto para o misto. Para o modelo misto, a interação entre polimorfismos em CBR3(rs8133052) e IMC( $\geq 30$ ) foi significativa, associada a menores chances para níveis graves. O teste para o ajuste do modelo marginal foi não significativo e os resíduos apresentaram menor dispersão que o modelo misto, o que indica um melhor ajuste aos dados.

Para a fase FAC, o modelo de regressão logística apresentou como variáveis associadas a maiores chances níveis graves de dor abdominal polimorfismos em CBR3(rs8133052) e menopausa. Os modelos marginal e misto não apresentaram nenhuma variável significativa.

Para a fase Docetaxel, o modelo de regressão logística apresentou diabetes como variável associada a maiores chances para níveis mais graves. O modelo marginal apresentou um bom ajuste e as variáveis associadas a maiores chances para níveis graves foram polimorfismos em CBR3(rs8133052) e interação entre polimorfismos em CBR3(rs8133052) e hipertensão. Para o modelo misto, a interação entre polimorfismos em CYP2B6 e hipertensão foi significativa e associada a maiores chances para níveis graves das reações. A avaliação dos resíduos indicou um melhor ajuste pelo modelo marginal, apesar do teste da razão de verossimilhanças apontar o intercepto aleatório como significativo.

Os modelos marginal e misto para mucosite utilizando todos os ciclos apresentaram como significativa a variável protocolo. O protocolo Docetaxel estava associado a maiores chances para níveis graves de mucosite. O teste para o ajuste do modelo marginal, resultou não significativo, indicando um bom ajuste aos dados. Os resíduos para o modelo marginal também indicaram um melhor ajuste quando comparados aos resíduos do modelo misto.

Para a fase FAC, o modelo de regressão logística apresentou como variáveis associadas a maiores chances para níveis graves polimorfismos em ABCB1(rs1128503) e diabetes. Para os modelos marginais nenhuma variável foi significativa. Para o modelo misto diabetes também estava associada a maiores chances para níveis graves de mucosite, o teste de razão de verossimilhanças resultou significativo, apontando a necessidade para o intercepto aleatório.

Para a fase Docetaxel, o modelo de regressão logística apresentou como variáveis associadas a maiores chances para níveis graves de mucosite polimorfismos em SLCO1 B3 e diabetes. O modelo marginal apresentou um bom ajuste para essa fase, com única variável significativa polimorfismos em ABCB1(rs1128503), associ-

ado a maiores chances para níveis graves. O modelo misto não apresentou nenhuma variável significativa.

Os modelos para Diarreia, ajustados utilizando todos os ciclos, apresentaram como significativa a variável protocolo, tanto para o modelo marginal quanto para o misto. O protocolo Docetaxel estava associado a maiores chances para níveis graves de diarreia. O modelo marginal apresentou um bom ajuste aos dados de acordo com o teste para a bondade do ajuste. Os resíduos para o modelo marginal apresentaram menor variação e sugerem um melhor ajuste quando comparado ao modelo misto.

Para a fase FAC, o modelo de regressão logística apresentou como variável significativa  $IMC(\geq 30)$ , associada a maiores chances para níveis graves da reação. Os modelos marginal e misto não apresentaram nenhuma variável significativa.

Na fase Docetaxel, o modelo de regressão logística apresentou  $IMC (\geq 30)$  associado a maiores chances para níveis graves de diarreia. O modelo marginal para essa fase apresentou um bom ajuste com a única variável significativa interação entre polimorfismos em CBR3 e dislipidemia, associada a menores chances para níveis graves. O modelo misto apresentou como variável significativa a interação entre polimorfismos em ABCB1 e menopausa, associada a maiores chances para níveis graves. Os resíduos para o modelo marginal apresentaram menor dispersão, o que sugere um melhor ajuste quando comparado ao modelo misto.

Apesar do teste de McNemar indicar a não alteração das proporções dentro de cada fase, a suposição de independência entre as respostas dentro de cada fase não é razoável em nosso estudo. Os modelos longitudinais ajustados para cada fase foram modelos mais adequados para descrever a relação entre os ciclos de cada fase que os modelos de regressão logística ordinal, uma justificativa pode ser dada a partir da significância para a maioria dos parâmetros de correlação nos modelos marginais, e para o intercepto aleatório nos modelos mistos. Além disso, a introdução de uma estrutura de correlação entre as respostas foi importante, como indicado pelo conjunto de variáveis e valores de estimativas similares obtidos nos modelos marginais e mistos.

Para a maioria das modelagens realizadas, a análise dos resíduos mostrou que os modelos marginais se ajustaram melhor aos dados que os modelos mistos. Uma possível causa é que a quantidade de parâmetros ajustados no modelo misto é maior que no modelo marginal e o método de estimação por verossimilhança, requer hipóteses sobre os efeitos aleatórios que nem sempre são suposições razoáveis para variáveis ordinais, como por exemplo a hipótese de distribuição normal para os interceptos aleatórios. Por outro lado, o modelo marginal, utiliza o método de equações de estimativas generalizadas, que por não possuir as hipóteses de máxima verossimilhança, se adapta melhor a variáveis categóricas no sentido de que nenhuma suposição sobre o tipo de distribuição precisa ser estabelecida e os estimado-



res possuem boas propriedades independente da estrutura de correlação escolhida. Então, mesmo que a estrutura de correlação tenha um parâmetro apenas, as boas propriedades, como convergência e existência de uma distribuição assintótica, são garantidas. Isso permite que o modelo marginal, mesmo com menos parâmetros e sem incluir a aleatoriedade no intercepto, possua um ajuste melhor. Infelizmente ainda não há na literatura, uma maneira de testar diferenças entre modelos marginais e mistos, pelo fato que um utiliza a quase-verossimilhança para o processo de estimação, e o outro utiliza a verossimilhança.

Alguns dos fatores encontrados neste estudo já foram descritos na literatura como por exemplo a associação encontrada entre polimorfismos ABCB1(rs1128503) e fadiga/astenia em tratamentos quimioterápicos com Docetaxel, descrita em um estudo de 2009 [48] além disso, IMC( $\geq 30$ ) também é citado na literatura como relacionado a maiores chances para níveis mais graves de fadiga/astenia [33, 34]. Comorbidades e menopausa foram variáveis relacionadas a artralgia em um estudo de 2013 [36], em nosso estudo encontramos associação entre diabetes, hipertensão e obesidade e a ocorrência de níveis mais graves de artralgia, que não identificamos nos estudos incluídos na revisão de literatura. A associação entre idade e diarreia foi descrita em um estudo de 2015, [30] e o modelo marginal que ajustamos para essa reação apresentou a interação entre protocolo Docetaxel e idade como variável significativa. Apesar de não encontrados registros das demais associações, algumas podem ter apoio biológico como o fato de polimorfismos na família CYP3A estar associado a tumores mais agressivos e possivelmente a reações mais graves ao tratamento.

A natureza de múltiplas causas associadas ao desenvolvimento de reações adversas ao tratamento quimioterápico para o câncer de mama, requer um estudo integrado de fatores, clínicos, ambientais e genéticos, como uma maneira de melhor compreender o processo de ocorrência de reações e melhorar a qualidade de vida dos pacientes. Nos modelos que ajustamos, foram incluídas variáveis genéticas, clínicas e protocolos utilizados durante o tratamento FAC-D, e as interações relacionadas. A inclusão representou um ganho com relação aos resultados existentes na literatura pois integra variáveis em um único perfil de susceptibilidade.

As interações entre variáveis encontradas para os modelos para astenia na fase Docetaxel indicaram IMC( $\geq 30$ ) como associado a maiores chances para níveis mais graves da reação, mas para as mulheres possuíam polimorfismos no gene CBR3 (rs8133052), o efeito mudava, ou seja, mulheres que além de apresentarem IMC( $\geq 30$ ), portassem polimorfismos em CBR3 (rs8133052) tinham menores chances para desenvolver níveis graves de astenia. Mulheres em menopausa, portadoras de polimorfismos em ABCB1(rs1128503) apresentaram menores chances para desenvolver níveis graves de mialgia ao longo de todo o tratamento FAC-D. Para

artralgia, as chances para níveis graves da reação ao longo de todo tratamento foram menores para mulheres hipertensas que possuíam polimorfismos em ABCB1 (rs1045642) e mulheres com IMC( $\geq 30$ ) que possuíam esse mesmo polimorfismo. Para dor abdominal, os modelos para todo o tratamento indicaram que mulheres com IMC( $\geq 30$ ) e polimorfismos no gene CBR3 (rs8133052), apresentaram menores chances para níveis graves. Na fase Docetaxel, as mulheres portadoras desse polimorfismo e hipertensas tinham maiores chances para níveis graves de dor abdominal. Para diarreia na fase Docetaxel, o perfil de susceptibilidade para níveis mais graves da reação foi mulheres em menopausa portadoras de polimorfismos em ABCB1.

O fato de a maioria dos trabalhos presentes na literatura estarem relacionados a identificação de fatores associados a reações para um determinado quimioterápico dificulta o reconhecimento do efeito de mudanças de quimioterápicos para os fatores de risco e para a gravidade das reações ao longo de um tratamento. Neste trabalho, por meio dos modelos longitudinais foi possível reconhecer, no caso do tratamento FAC-D, o impacto da mudança de protocolo quimioterápico.

De uma forma geral, o protocolo Docetaxel estava associado a maiores chances para níveis graves de todas as reações quando comparado ao protocolo FAC, o que indica que o Docetaxel aumenta a gravidade das reações ao longo do tratamento. Isso pode afetar a identificação de perfis de susceptibilidade para todo o tratamento FAC-D já que o protocolo pode ocultar outras associações em modelos com várias covariáveis. Por isso, o estudo das fases é importante para identificação do conjunto de variáveis e interações relacionadas às reações quando as mulheres estão sob o efeito do mesmo medicamento.

Os resultados mostraram que além da mudança de protocolo aumentar gravidade das reações no tratamento FAC-D, para protocolos diferentes, o conjunto de variáveis relacionadas às reações também é diferente e não foi possível encontrar perfis comuns de susceptibilidade para as duas fases.

Os fatores identificados neste estudo podem ajudar a caracterizar perfis de pacientes suscetíveis ao desenvolvimento de reações adversas durante o tratamento FAC-D. Esses perfis, compreendidos como um conjunto integrado de variáveis clínicas, genéticas podem contribuir de maneira mais efetiva para o monitoramento e a escolha de terapêuticas que visem minimizar a ocorrência de reações adversas nos pacientes.

# Capítulo 8

## Conclusões finais

Reação adversa a medicamento é uma resposta não intencional, que ocorre após a administração de um medicamento em dose normalmente usada em seres humanos. No caso de tratamentos quimioterápicos, reações são esperadas e podem estar associadas a eficácia do tratamento. Por isso, é possível que haja subvalorização de efeitos adversos o que dificulta mapear fatores de risco. A notificação de reações em tratamentos quimioterápicos é uma maneira de identificar fatores relacionados a reações adversas e auxiliar políticas públicas relacionadas a segurança de medicamentos.

No caso do tratamento quimioterápico FAC-D, algumas reações também são esperadas, mas elas podem ser agravadas por um determinado perfil clínico, genético, pela própria mudança de protocolo e pela maneira com que cada protocolo interage com um determinado perfil clínico/genético do paciente.

Existem lacunas na literatura nesse sentido, relacionadas a um estudo integrado das múltiplas causas associadas às reações adversas e a efeitos da mudança de protocolo ao longo de um tratamento, no conjunto de fatores de risco e na gravidade das reações. Uma das possíveis causas pode ser atribuída ao fato de grande parte dos estudos serem realizados para um tipo de quimioterápico, e agravada pela utilização pouco frequente de ferramentas estatísticas para dados categóricos longitudinais.

Em nosso estudo do tratamento FAC-D, os resultados apontam as equações de estimativas generalizadas como a metodologia mais apropriada para identificação de características individuais relacionadas às reações, e indicam que a gravidade das reações adversas pode ser afetada pelo modo como as variáveis se relacionam, já que algumas interações entre perfis clínico e genético foram significativas para os modelos ajustados. Dessa forma, um estudo integrado das múltiplas causas pode auxiliar, de uma maneira mais eficiente, a identificação de perfis de risco e a condução de terapêuticas individualizadas.

O protocolo foi principal fator associado às reações adversas ao longo de todo tratamento, sendo o Docetaxel o quimioterápico que aumentou a gravidade das reações. Por meio de estudo das fases nós confirmamos a hipótese de que mudanças

de protocolo causam mudanças no conjunto de variáveis significativas, o que indica que diferentes protocolos implicam diferentes perfis de susceptibilidade, e no caso do tratamento FAC-D não foram encontrados perfis comuns para os dois protocolos.

Os resultados sugerem o monitoramento de reações adversas em tratamentos FAC-D, em especial durante a fase Docetaxel, em que foram encontrados níveis mais graves para as reações. Nessa fase, as reações musculares têm como perfis associados a maiores chances para níveis mais graves, IMC ( $\geq 30$ ), diabetes e polimorfismos em CYP2B6(rs3745274). Para as reações gastrointestinais, polimorfismos em CBR3(rs8133052), interação entre polimorfismos em CBR3(rs8133052) e hipertensão, polimorfismos em ABCB1(rs1128503), IMC ( $\geq 30$ ) e interação entre polimorfismos em ABCB1(rs1128503) e menopausa foram variáveis associadas a maiores chances para níveis mais graves.

### **Perspectivas futuras**

Como proposta futura, pretendemos aplicar métodos Bayesianos para o ajuste dos GLMM com intuito de realizar comparações com as estimativas para os parâmetros encontradas utilizando metodologia clássica.

A definição Bayesiana para modelos mistos é idêntica à abordagem frequentista. Os modelos mistos são uma das conexões mais próximas que podemos realizar entre métodos clássicos e Bayesianos. A grande diferença é a estimação dos parâmetros, que é feita utilizando máxima verossimilhança nos métodos clássicos, e MCMC nos métodos Bayesianos.

# Referências Bibliográficas

- [1] A. Jemal and et al., “Global cancer statistics. *ca cancer j clin*, 2011. 61(2): p. 69-90.” 1
- [2] INCA, “Instituto nacional do cancer <http://www.inca.gov.br/estimativa/2014/index.asp?id=5>, acessado em 06-2014.” 1
- [3] NCCN, “National comprehensive cancer network <http://www.nccn.org>, acessado em 03-2013.” 1
- [4] OMS, “International drug monitoring: the role of national centres. world health organization. technical report series, n. 498, 1972.” 1
- [5] E. W. Eichelbaum M, Ingelman-Sundberg M, “Pharmacogenomics and individualized drug therapy. *annu rev med*, p. 119:137, 2006.” 1, 4.2
- [6] J. Oates, “A ciência da farmacoterapia. in: Brunton, l. l. lazo, j. s. parker, k. l. eds goodmann gilman. *as bases farmacológica da terapêutica*, 11. ed. rio de janeiro: Mcgraw hill interamericana do brasil, p. 107:125, 2006.” 1
- [7] B. Chabner and et al., “Antineoplásicos. in: Goodman gilman, *as bases farmacológicas da terapêutica* (10<sup>a</sup> ed). rio de janeiro: Editora: Mcgraw-hill, 2005. p. 1041-1096.” 1
- [8] G. Wilkinson, “Farmacocinética: Dinâmica da absorção, da distribuição e da eliminação dos fármacos. in: Goodman gilman, *as bases farmacológicas da terapêutica* (10<sup>a</sup> ed). rio de janeiro: Editora mcgraw-hill, 2005. p. 3-24.” 1
- [9] N. Network, “Clinical practice guidelines in oncology - breast cancer, nccn, v.i., 2010.” 1
- [10] I. Henderson, “Improved outcomes from adding sequential paclitaxel but not from escalating doxorubicin dose in adjuvant chemotherapy regimen for patients with node-positive primary breast cancer. *j clin oncol*, 2003. 21(6).976-83.” 1

- [11] E. Mamounas, “Paclitaxel after doxorubicin plus cyclophosphamide as adjuvant chemotherapy for node-positive breast cancer: results from nsabp b-28. *j clin oncol*, 2005.23(16).3686-96.”
- [12] M. Martin, “Adjuvant docetaxel for node-positive breast cancer. *n engl j med*, 2005.352(22).2302-13.” 1
- [13] D. Solimando, “Drug information handbook for oncology, 7 ed. 2008. hudson:lexicoms drug reference handbooks.” 1, 3
- [14] E. Chu and V. Devita, “Physician’s cancer chemotherapy drug manual. 1 ed. 2009.” 1, 3
- [15] E. Davies and et.all., “Adverse drug reactions in hospitals: a narrative review. *curr drug saf*, 2007. 2(1). p 79-87.” 1, 3, 7
- [16] J. Wooten, “Adverse drug reactions:part. *south med j*. 103(10). p 1025-8.”
- [17] S. Menon, “Reações adversas a medicamentos (rams).saúde em revista. 2005. 7(16). p 71-9.” 1
- [18] K. Beard and A. Lee, “Reações adversas a medicamentos. 2009. p 21.” 1, 3, 7
- [19] A. NegaoDias and etall, “Reações alérgicas a medicamentos. *jornal de pediatria*. 2004. p 259-66.” 1
- [20] D. T. Relling MV, “Pharmacogenetics and cancer therapy. *nature reviews cancer*; n. 1, p. 99:108, 2001.” 1
- [21] V. Ananth and G. Kleinbaum, “Regression models for ordinal responses: A review of methods and applications. *international journal of epidemiology*. 1997.” 1
- [22] C. Liu, T. Cripe, and M. Kim, “Statistical issues in longitudinal data analysis for treatment efficacy studies in the biomedical sciences. *mol ther*. 2010 sep;18(9):1724-30.” 1
- [23] R. Twisk, “Applied longitudinal data analysis for epidemiology. cambridge university press. 2003.” 1
- [24] A. Agresti, “Categorical data analysis. 2 ed. wiley series in probability and statistics. 2002.” 1, 3, 4.3, 4.4.4, 4.5, 4.6.1, 4.6.3, 4.7, 6.5
- [25] J. A. McCullagh, P. Nelder. 1, 4.6.2

- [26] K. Liang and S. Zeger, “Longitudinal data analysis using generalized linear models (1986). *biometrika*. 73. 13:22.” 1, 4.5, 4.6, 4.6.3
- [27] P. Diggle, P. Heagerty, K. Y. Liang, and S. Zeger, “Analysis of longitudinal data.(2001). *statistical science series. oxford university press,oxford.*” 1, 4.6
- [28] E. Abreu and S. Koifman, “Fatores prognósticos para cancer de mama feminino. *revista brasileira de cancerologia*.48(1):p.113-31.” 1
- [29] O. Minton, S. Alexander, and P. Stone, “Breast cancer research and treatment. 2012. volume 136, issue 2, pp 513-520. identification of factors associated with cancer related fatigue syndrome in disease-free breast cancer patients after completing primary treatment.” 1, 3, 3
- [30] B. Stein, N. Petrelli, H. Douglass, D. Driscoll, G. Arcangeli, and N. Meropol, “Age and sex are independent predictors of 5-fluorouracil toxicity. analysis of a large scale phase iii trial. *cancer*. 1995.75.11.17.” 3, 7
- [31] M. McCarthy, D. Awde, H. Ghandi, and I. Kocha, “Risk factors associated with mucositis in cancer patients receiving 5-fluorouracil , *oral oncology*.34 (6).484 - 490.1998.” 3
- [32] A. Maroun, L. Anthony, N. Blais, R. Burkes, S. Dowden, G. Dranitsaris, B. Samson, A. Shah, M. Thirlwell, M. Vincent, and R. Wong, “Prevention and management of chemotherapy-induced diarrhea in patients with colorectal cancer: a consensus statement by the canadian working group on chemotherapy-induced diarrhea. *curr oncol*. 2007 february. 14(1). 13.20.” 1, 3, 3
- [33] A. K. Donovan, J. B. Small, A. M. Andrykowski, P. Munster, and B. P. . Jacobsen, “Utility of a cognitive-behavioral model to predict fatigue following breast cancer treatment. *health psychology*. vol.26. 4. 464-472.” 1, 3, 3, 7
- [34] N. Jong, A. Courtens, H. Abu-Saad, and et al.(2002), “Fatigue in patients with breast cancer receiving adjuvant chemotherapy : a review of the literature.*cancer nurs*.25; 283-297.” 1, 3, 3, 7
- [35] H. Burstein, J. Manola, J. Younger, L. M. Parker, C. A. Bunnell, R. Scheib, U. Matulonis, J. Garber, K. Clarke, L. Shulman, and E. . Winer, “J clin oncol. 2000 mar;18(6):1212-9.docetaxel administered on a weekly basis for metastatic breast cancer.” 1, 3, 7

- [36] L. Castel, K. Hartmann, I. Mayer, B. Saville, J. Alvarez, C. Boomershine, V. Abramson, A. Chakravarthy, D. Friedman, and D. Cella, "Time course of arthralgia among women initiating aromatase inhibitor therapy and a postmenopausal comparison group in a prospective cohort. *cancer*. 2013 jul 1;119(13):2375-82." 1, 3, 7
- [37] J. Dobson, "An introduction to generalized linear models. *champan and hall*. 2002." 1, 4.3, 4.5
- [38] J. . Hainsworth, "Practical aspects of weekly docetaxel administration schedules. *.oncologist review*, 2004;9(5):538-45." 1, 7
- [39] F. Moore, S. Vanter, C. Boyden, J. Lokich, and R. Wilson, "Adrenalectomy with chemotherapy in the treatment of advanced breast cancer: objective and subjective response rates, duration and quality of life. *surgery* 1974, 76:376-390." 3
- [40] T. Priestman and M. Baum, "Evaluation of quality of life in patients receiving treatment for advanced breast cancer. *lancet* 1976, 1:899-900." 3
- [41] A. Montazeri, "Health related quality of life in breast cancer patients: A bibliographic review of the literature from 1974 to 2007. *journal of experimental and clinical cancer research* 2008, 27 32." 3
- [42] R. M. Huang RS, "Pharmacogenetics and pharmacogenomics of anticancer agents. *ca cancer j. clin*, p. 42:55, 2009." 3, 4.2, 4.2
- [43] F. Marre, G. Sanderink, and G. Sousa, "Hepatic biotransformation of docetaxel (taxotere®)in vitro: Involvement of the cyp3a subfamily in humans. *cancer research* n.56, p.1296-1302, 1996." 3, 4.2
- [44] M. Shou, M. Martinet, K. Korzekwa, and et al., "Role of human cytochrome p450 3a4 and 3a5 in the metabolism of taxotere and its derivatives: enzyme specificity, interindividual distribution and metabolic contribution in human liver. *pharmacogenetics*, n.8, p. 391-401, 1998." 3, 4.2
- [45] Y. Miyoshi, T. Taguchi, and S. Kim, "Prediction of response to docetaxel by immunohistochemical analysis of cyp3a4 expression in human breast cancers. *breast cancer*, v. 12, n. 1, p. 11-15, 2005." 3, 4.2
- [46] V. J. Baker DS, Sparreboom A, "Clinical pharmacokinetics of docetaxel. *clin pharmacokinet*, n. 45, v. 3, p. 235:252, 2006." 3, 4.1.1, 4.2



- [47] B. Goh, S. Lee, L. Wang, and et al., “Explaining interindividual variability of docetaxel pharmacokinetics and pharmacodynamics in asians through phenotyping and genotyping strategies. *journal of clinical oncology*, v. 20, n. 17, p.3683-3690, 2002.” 3, 4.2
- [48] S. M. Tsai, C. Y. Lin, S. H. Wu, L. A. Hou, H. Ma, L. Y. Tsai, and M. F. . Hou, “Side effects after docetaxel treatment in taiwanese breast cancer patients with *cyp3a4*, *cyp3a5*, and *abcb1* gene polymorphisms. *clin chim acta*. 404, 160-165.” 3, 4.2, 7
- [49] I. Cascorbi, “P-glycoprotein: Tissue distribution, substrates, and functional consequences of genetic variations. *drug transportres, handbook of experimental pharmacology*, 201, p. 261-283, 2011.” 3, 4.2
- [50] B. L. Zhang, T. Sun, B. N. Zhang, S. Zheng, N. Lü, X. Xu, B. H.and Wang, G. J. Chen, D. K. Yu, and D. X. . Lin, “Polymorphisms of *gstp1* is associated with differences of chemotherapy response and toxicity in breast cancer. *chin med j (engl)*. 124, 199-204.” 3
- [51] V. Rocha, R. Porcher, J. Fernandes, and et al., “Association of drug metabolism gene polymorphisms with toxicities, graft-versus-host disease and survival after hla-identical sibling hematopoietic stem cell transplantation for patients with leukemia. *leukemia*, n. 23, v. 3, p.545-556, 2009.” 3, 4.2
- [52] H. Kunitoh, N. Saijo, K. Furuse, and et al, “Neuromuscular toxicities of paclitaxel 210 mg m(-2) by 3-hour infusion. *br j cancer* 77:1686- 1688, 1998.” 3
- [53] J. Garrison, J. McCune, R. Livingston, H. Linden, J. Gralow, G. Ellis, and H. West, “Myalgias and arthralgias associated with paclitaxel. *oncology (williston park)*. 2003 feb;17(2):271-7.” 3
- [54] J. Mao, H. Su, R. Feng, M. Donelson, R. Aplenc, R. Rebbeck, F. Stanczyk, and A. Demichele, “Association of functional polymorphisms in *cyp19a1* with aromatase inhibitor associated arthralgia in breast cancer survivors. *breast cancer research* 2011, 13:r8.” 3
- [55] A. N. Tucker, K. A. Tkaczuk, L. M. Lewis, D. Tomic, C. K. Lim, and J. A. . Flaws, “Polymorphisms in cytochrome *p4503a5* (*cyp3a5*) may be associated with race and tumor characteristics, but not metabolism and side effects of tamoxifen in breast cancer patients. *cancer lett*. 217, 61-72.” 3
- [56] M. V. Imyanitov EN, “Molecular based choice of cancer therapy: Realities and expectations. *clinica chimica acta*, n, 379, p. 1:13, 2007.” 4.1

- [57] A. R. e. a. Fisher B, Ravdin RG, “Surgical adjuvant chemotherapy in cancer of the breast: Results of a decade of cooperative investigation. *annals of surgery*, p.337:356,1968.” 4.1
- [58] E. S. Fisher B, Carbone P, “1 phenylalanine mustard (l pam) in the management of primary breast cancer. a report of early findings. *n engl j med*, p 117:122, 1975.” 4.1
- [59] V. P. e. a. Bonadonna G, Brusamolino E, “Combination chemotherapy as an adjuvant treatment in operable breast cancer. *n engl j med*, n. 294, p. 405:410, 1976 apud hassan msu, ansari j, spooner d, hussain sa. chemotherapy for breast cancer (review). *oncology reports*, p. 1121:1131. 2010.” 4.1
- [60] S. D. H. S. Hassan MSU, Ansari J, “Chemotherapy for breast cancer (review). *oncology reports*, p. 1121:1131, 2010.” 4.1
- [61] C. C. e. a. Laporte S, Jones S, “Consistency of effect of docetaxel containing adjuvant chemotherapy in patients with early stage breast cancer independent of nodal status: Metanalysis of 12 randomized clinical trials. presented at the 2009 san antonio breast cancer symposium, san antonio, texas, december 11, 2009.” 4.1
- [62] e. Klasco RK, “Drugdex system, base de dados na internet. greenwood village (colorado): Thomson micromedex; 1974:2012. citado em 13 de fevereiro de 2012.” 4.1.1, 4.1.1
- [63] D. M. Pinto N, Ludeman SM, “Pharmacogenetic studies related to cyclophosphamide based therapy. *pharmacogenomics*, p. 1897:1903, 2009.” 4.1.1
- [64] v. d. B. J. e. a. Engels FK, Loos WJ, “Therapeutic drug monitoring for the individualization of docetaxel dosing: a randomized pharmacokinetic study. *clin cancer res*, v.17, n. 2, p. 353:362, 2011.” 4.1.1
- [65] A. J. e. a. Tran A, Jullien V, “Pharmacokinetics and toxicity of docetaxel: Role of pharmacokinetics and toxicity of docetaxel: Role of cyp3a, mdr1 and gst polymorphisms. *pharmacogenetics and genomics*, n. 79, v. 6, p. 570:580, 2006.” 4.1.1, 4.1.1, 4.2
- [66] O. E. Paulusma CC, “The canalicular multispecific organic anion transporter and conjugated hyperbilirubinemia in rat and man. *j mol med (berl)* v. 75, n. 6, p. 420:428, 1997.” 4.1.1

- [67] K. M. e. a. Kiyotani K, Mushiroda T, “Association of genetic polymorphisms in *slco1b3* and *abcc2* with docetaxel induced leukopenia. *cancer science*, n. 99, v.5, p. 967:972, 2008.” 4.1.1
- [68] 4.2
- [69] J. L. Haller C, “Adverse drug reactions: Moving from chance to science. *clinical pharmacology & therapeutics*, p. 761:764, 2011.” 4.2
- [70] V. S. RHN., “Cancer treatment and pharmacogenetics of cytochrome p450 enzymes. *invest. new drugs*, n. 23, p. 513:522, 2005.” 4.2
- [71] S. E. T. K. Lamba JK, Lin YS, “Genetic contribution to variable human *cyp3a* mediated metabolism. *advanced drug delivery reviews*, p. 1271:1294, 2002.” 4.2
- [72] T. Lang, K. Klein, J. Fisher, and et al., “Extensive genetic polymorphism in the human *cyp2b6* gene with impact on expression and function in human liver. *pharmacogenetics*, n.11, p.399-415, 2001.” 4.2
- [73] L. Fan, B. Goh, C. Wong, and et al., “Genotype of human carbonyl reductase *cbr3* correlates with doxorubicin disposition and toxicity. *pharmacogenetics and genomics*, n. 18, p. 621-629, 2008.” 4.2
- [74] B. Mannervik, Y. Awasthi, P. Board, and et al., “Nomenclature for human glutathione transferases. *biochem j*, n. 282, p. 305-306, 1992.” 4.2
- [75] J. Seidegard and G. Eketrom, “The role of human glutathione transferases and epoxide hydrolases in the metabolism of xenobiotics. *environ health perspect*, n. 105, p. 791-9, 1997.” 4.2
- [76] H. Autrup, “Genetic polymorphism in human xenobiotic metabolizing enzymes as susceptibility factors in toxic response. *mutat. res.* n. 464, p. 65-76, 2000.” 4.2
- [77] J. Abraham, H. Earl, P. Pharoah, and C. Caldas, “Pharmacogenetics of cancer chemotherapy. *biochimica et biophysica acta*, n.1766, p. 168-183, 2006.” 4.2
- [78] M. Watson, L. Meyer, A. Thomson, and etal, “Psychological factors predicting nausea and vomiting in breast cancer patients on chemotherapy. *eur j cancer* 34:831-837, 1998.” 4.2

- [79] S. Zhong, M. Huang, X. Yang, S. Liang, and et al., “Relationship of glutathione s-transferase genotypes with side-effects of pulsed cyclophosphamide therapy in patients with systemic lupus erythematosus. *br. j. clin. pharmacol*, n. 62, p. 457-472, 2006.” 4.2
- [80] S. C. Kimchi, J. Oh, I. Kim, and et al., “A ”silent” polymorphism in the *mdr1* gene changes substrates specificity. *science*, n. 315, p. 525-528, 2007.” 4.2
- [81] Z. Sauna, S. C. Kimchi, S. Ambudkar, and et al., “The sounds of silence: Synonymous mutations affect function. *pharmacogenomics*,: p. 527-532, 2007.” 4.2
- [82] S. Lal, A. Mahajan, W. Chen, and et al., “Pharmacogenetics of target genes across doxorubicin disposition pathway: A review. *current drug metabolism*, n. 11, p. 115-128, 2011.” 4.2
- [83] K. J. Fahrmayr C, Fromm MF, “Hepatic *oatp* and *oct* uptake transporters: their role for drug drug interactions and pharmacogenetic aspects. *drug metab rev*. 2010;42:380:401.” 4.2
- [84] J. Nelder and R. Wedderburn, “(1972). *generalized linear models*. *j.r. statist. soc.* 135.370-84.” 4.4
- [85] J. L. Singer and D. F. Andrade, “Análise de dados longitudinais (1986). *são paulo: Associação brasileira de estatística*.” 4.5
- [86] D. A. Harville, “1997.*maximum likelihood approaches to variance component estimation and to related problems*. *journal of the american statistical association* 72 (358) 320 338.” 4.5
- [87] N. M. Laird and W. J.H., “1982.*random effects models for longitudinal data*. *biometrics*. 1982 dec 38(4)963 74.” 4.5
- [88] K. C. B. D. B. M. Ziegler, A., “Familial associations of lipid profiles: a generalized estimating equations approach. (2000) *statistics in medicine*, 19, 3345:3357.” 4.6
- [89] T. D. Wickens, “Categorical data analysis. (1998). *annu. rev. psychol.* 48, 537:558.” 4.6
- [90] R. W. M. Wedderburn, “Quasi likelihood functions, generalized linear models, and the gauss newton method.(1974). *biometrika*, 61, 439:447.” 4.6.2
- [91] G. Kauermann and R. Carrol, “A note of the efficiency of sandwich covariance matrix estimation. 2001. *j. amer. statist.* 96.1387-1397.” 4.6.3

- [92] P. Heagerty and S. Zeger, “Marginal regression models for clustered ordinal measurements. 1996. *journal of american statistical association*. 91.1024-1036.” 4.6.3, 4.6.4
- [93] S. Lipsitz, M. Laird, and P. Harrington, “Generalized estimated equations for correlated binary data: using the odds ratio as a measure of association.(1991). *biometrika*. 78. 153-160.” 4.6.3
- [94] W. Pan, “Akaike information criterion in generalized estimating equations. *biometrics*, 57(1), 120:125. 2001.” 4.6.4
- [95] W. Y. Hin, L.Y., “Working correlation structure identification in generalized estimating equations. *statistics in medicine*, 28(4), 642:658.(2009).” 4.6.4, 6.3, 6.5
- [96] S. Oh, K. C. Carriere, and T. Park, “Model diagnostic plots for repeated measures data using the generalized estimating equations approach, *comput. stat. data anal*. 2008. 53.1,2008.” 4.6.4
- [97] K. Lin, “Goodness of fit tests for modeling longitudinal ordinal data. *computational statistics and data analysis*, 54, 1972:1880. (2010).” 4.6.4, 6.3, 6.5
- [98] W. Pan, “Goodness of fit tests for gee with correlated binary data, *scandinavian journal of statistics*, 29, 101:110. 2002.” 4.6.4
- [99] P. H. Schutzenmeister A., “Residual analysis of linear mixed models using simulation approach, *computational statistics and data analysis*, 56, 1405:1416, 2012.” 4.7.2
- [100] P. L. Christensen, R., “Case deletion diagnostics for mixed models. *technometrics*. 34.38:45,1992.” 4.7.2
- [101] A. Trotti and . etal., “Development of a comprehensive grading system for adverse effects of cancer treatment. *semin radiat oncol*.13(3):p.176-81.ctcae v3.0.” 5.2
- [102] E. Saad and etal. (2002).48., “Critérios comuns de toxicidade do instituto nacional do câncer dos estados unidos. *revista brasileira de cancerologia*.” 5.2

# Apêndice A

## Resíduos dos modelos marginais e mistos:

### A.0.1 Primeiro modelo marginal e misto para astenia:

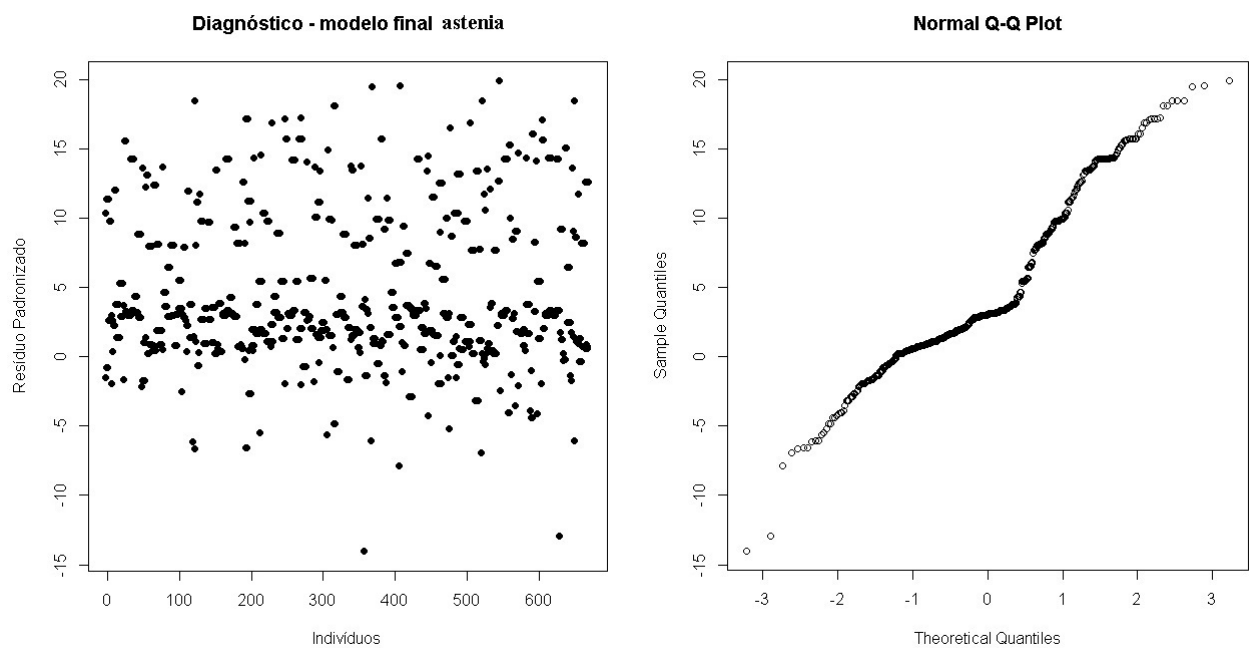


Figura A.1: Resíduos do primeiro modelo marginal para astenia

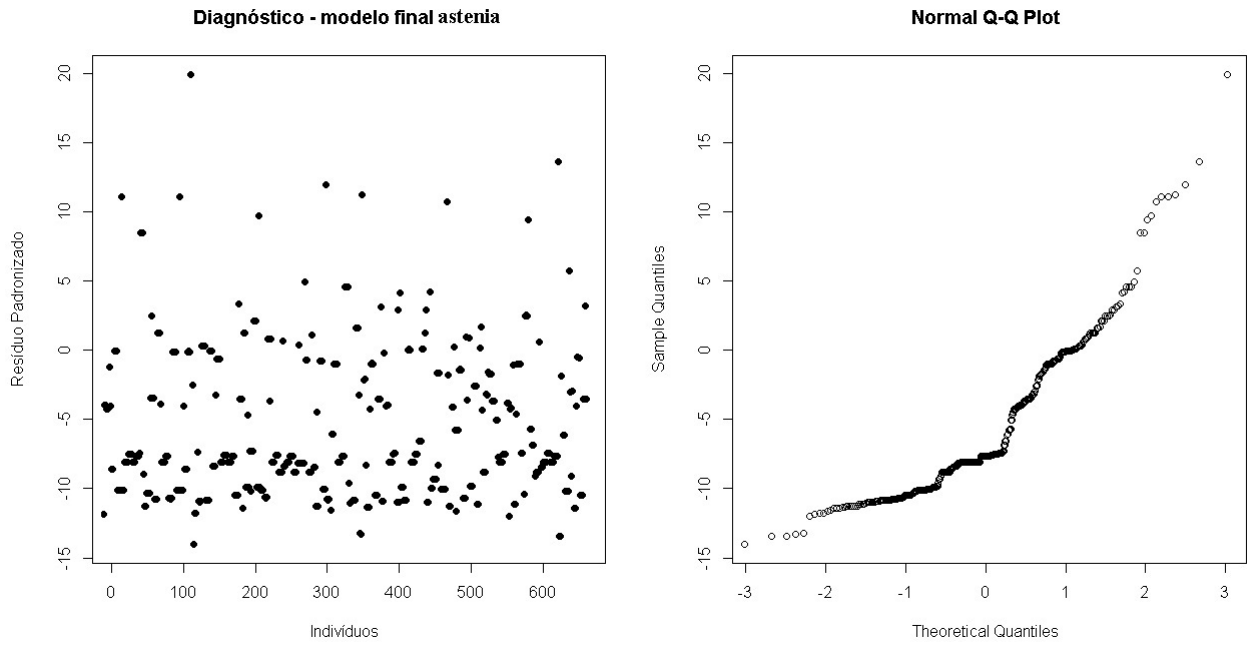


Figura A.2: Resíduos do primeiro modelo misto para astenia

### A.0.2 Primeiro modelo marginal e misto para mialgia:

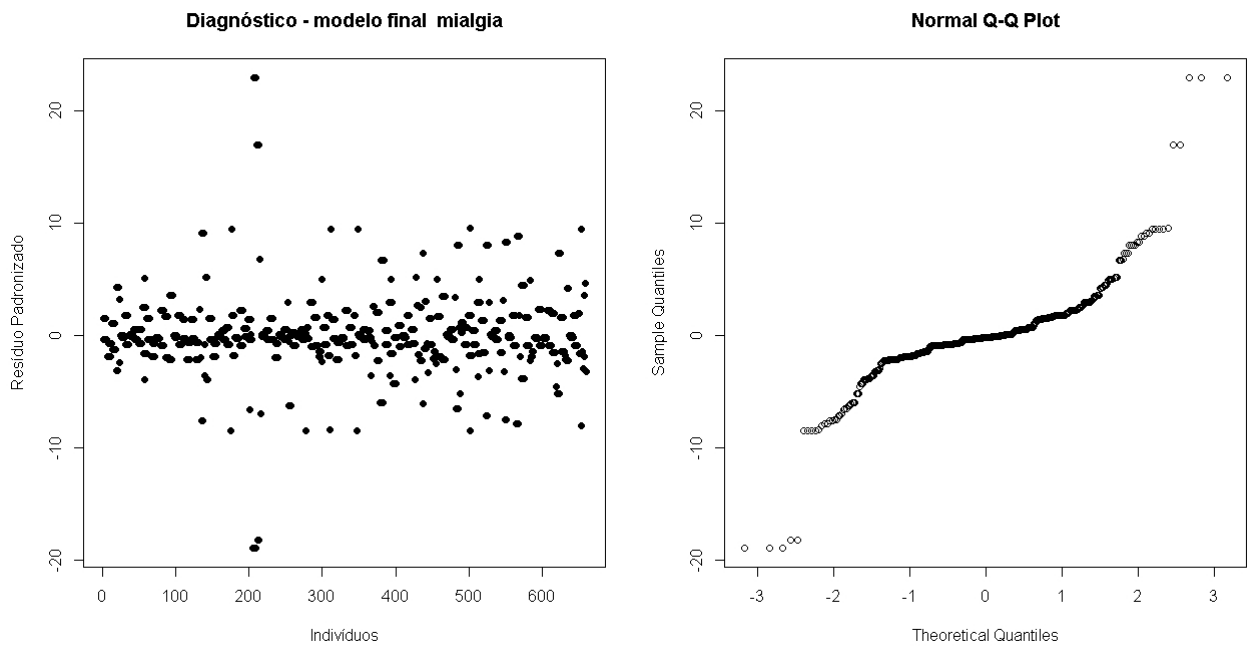


Figura A.3: Resíduos do primeiro modelo marginal para mialgia

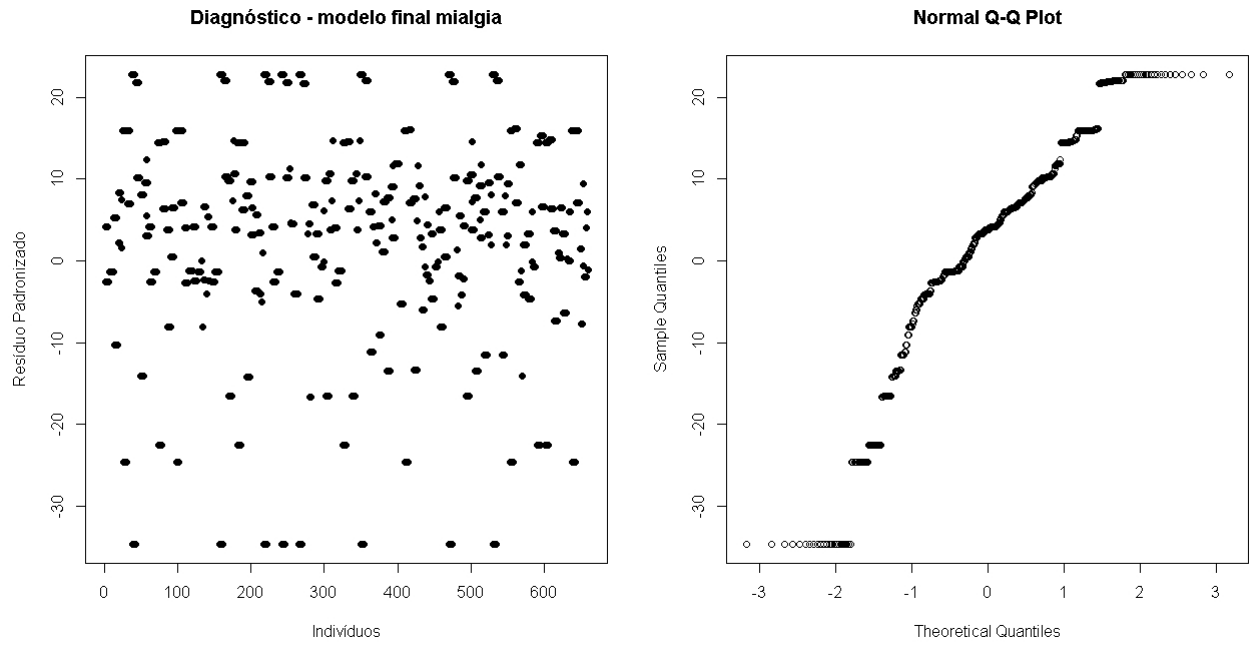


Figura A.4: Resíduos do primeiro modelo misto para mialgia

### A.0.3 Primeiro modelo marginal e misto para artralgia:

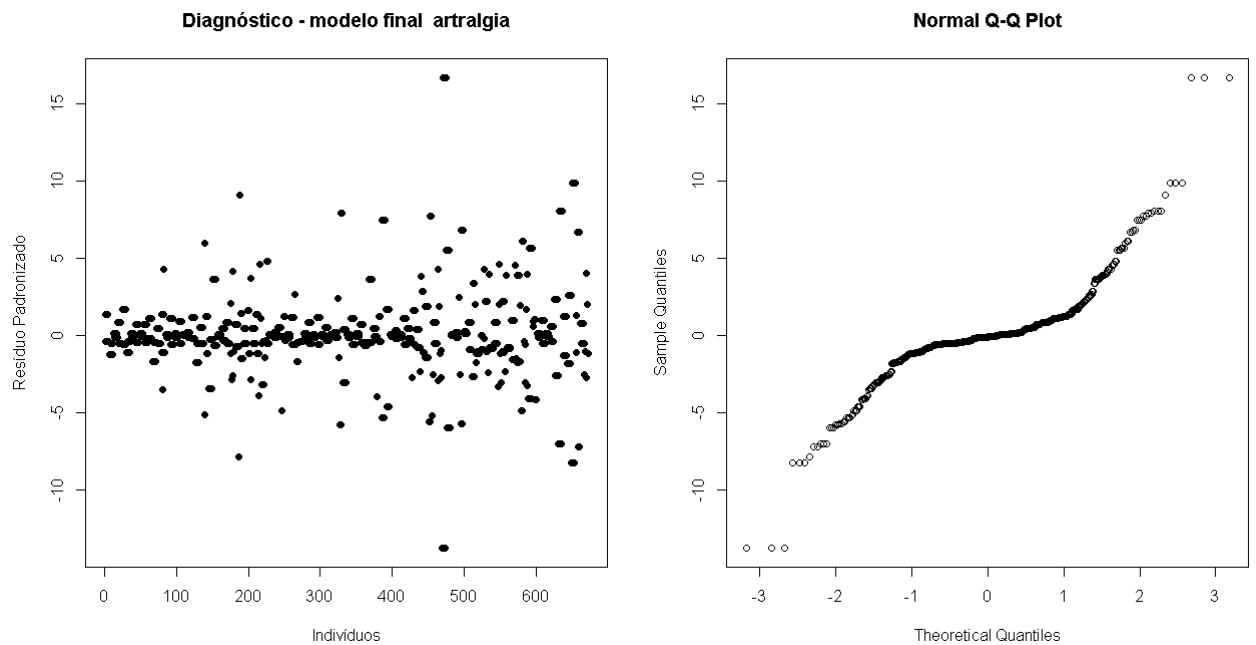


Figura A.5: Resíduos do primeiro modelo marginal para artralgia



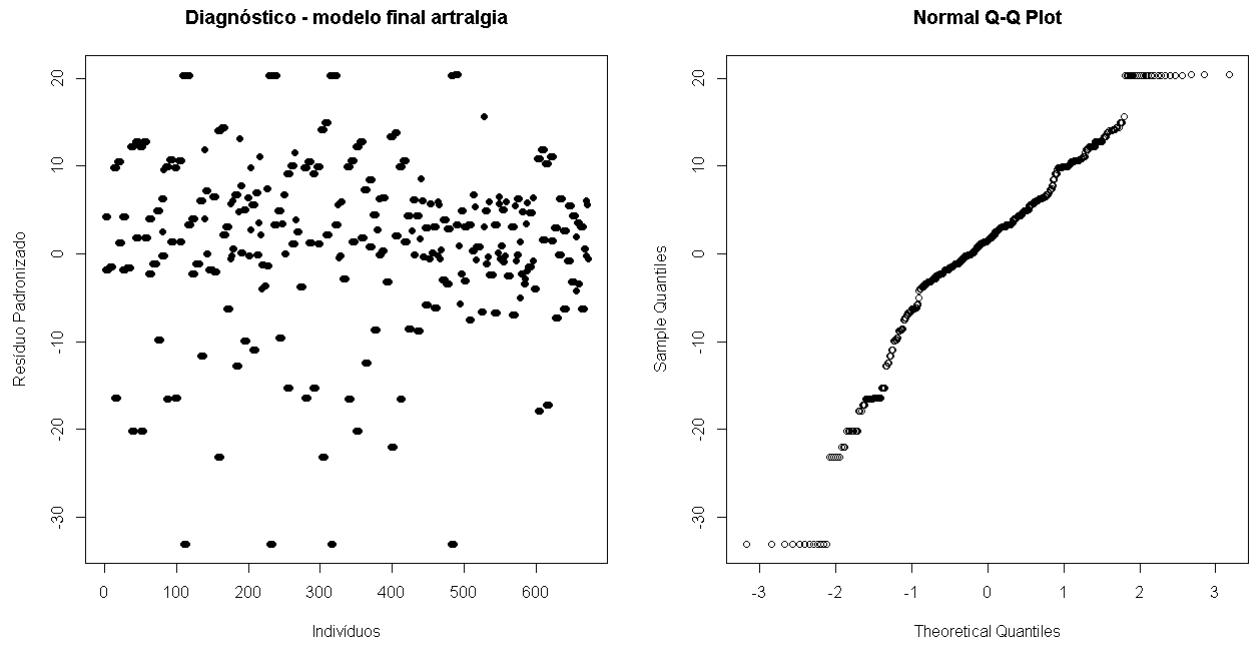


Figura A.6: Resíduos do primeiro modelo misto para artralgia

#### A.0.4 Primeiro modelo marginal e misto para dor abdominal:

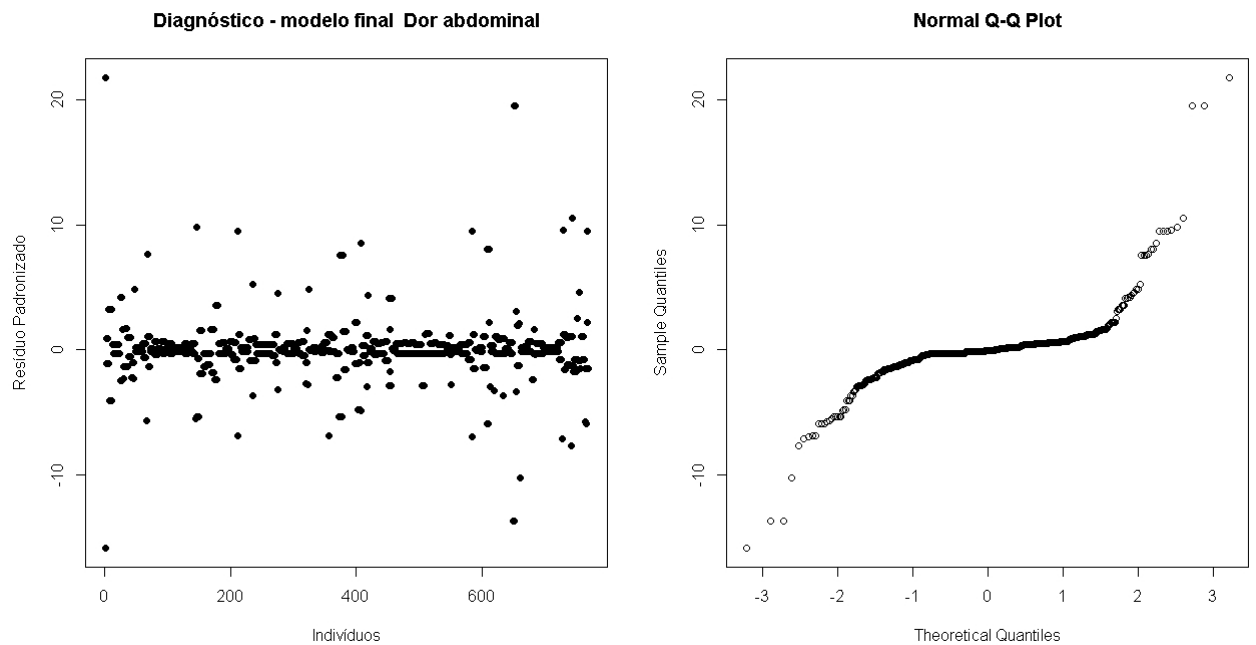


Figura A.7: Resíduos do primeiro modelo marginal para dor abdominal

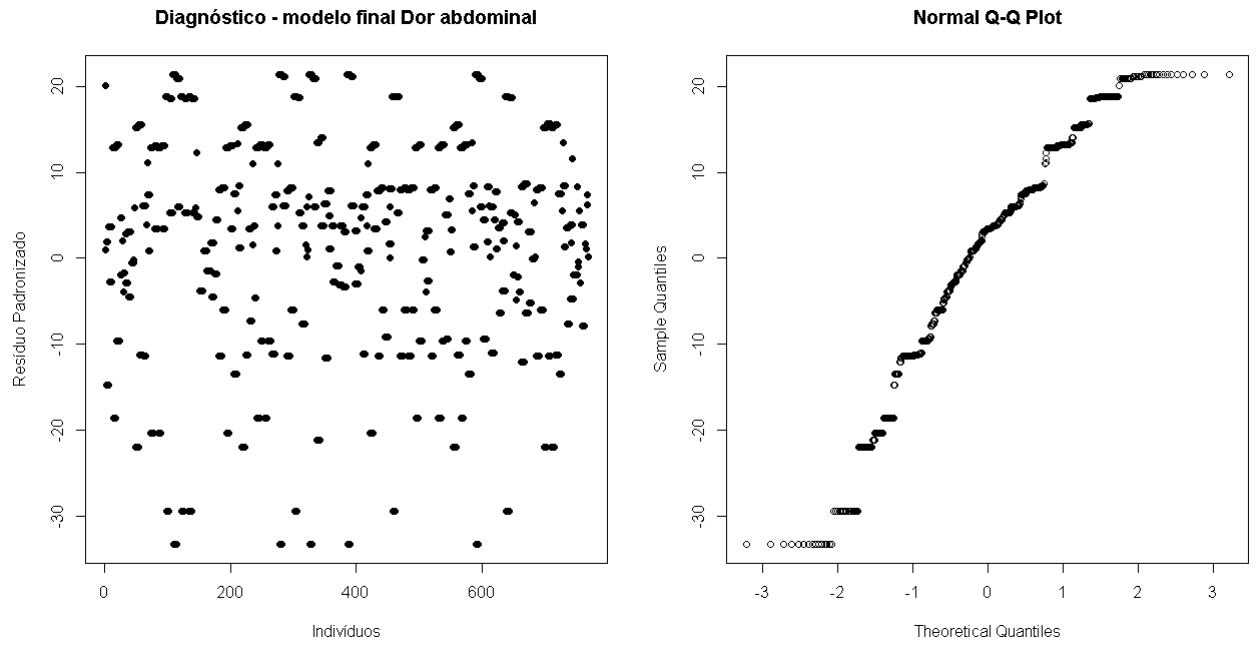


Figura A.8: Resíduos do primeiro modelo misto para dor abdominal

### A.0.5 Primeiro modelo marginal e misto para mucosite:

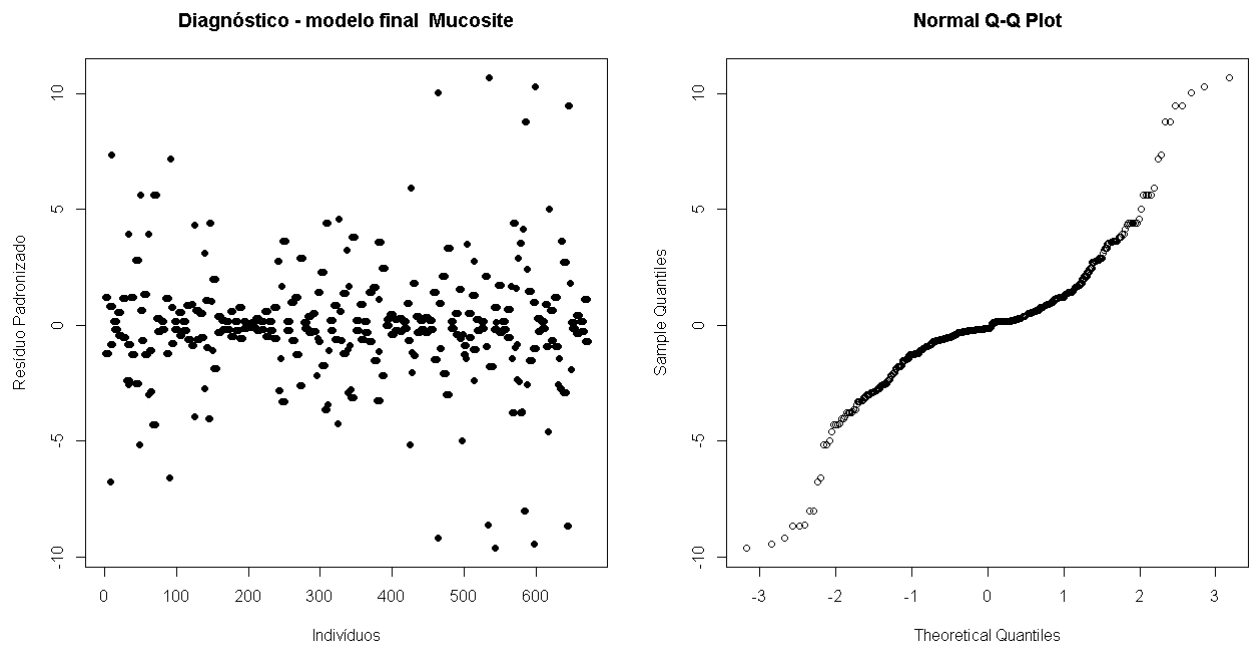


Figura A.9: Resíduos do primeiro modelo marginal para mucosite

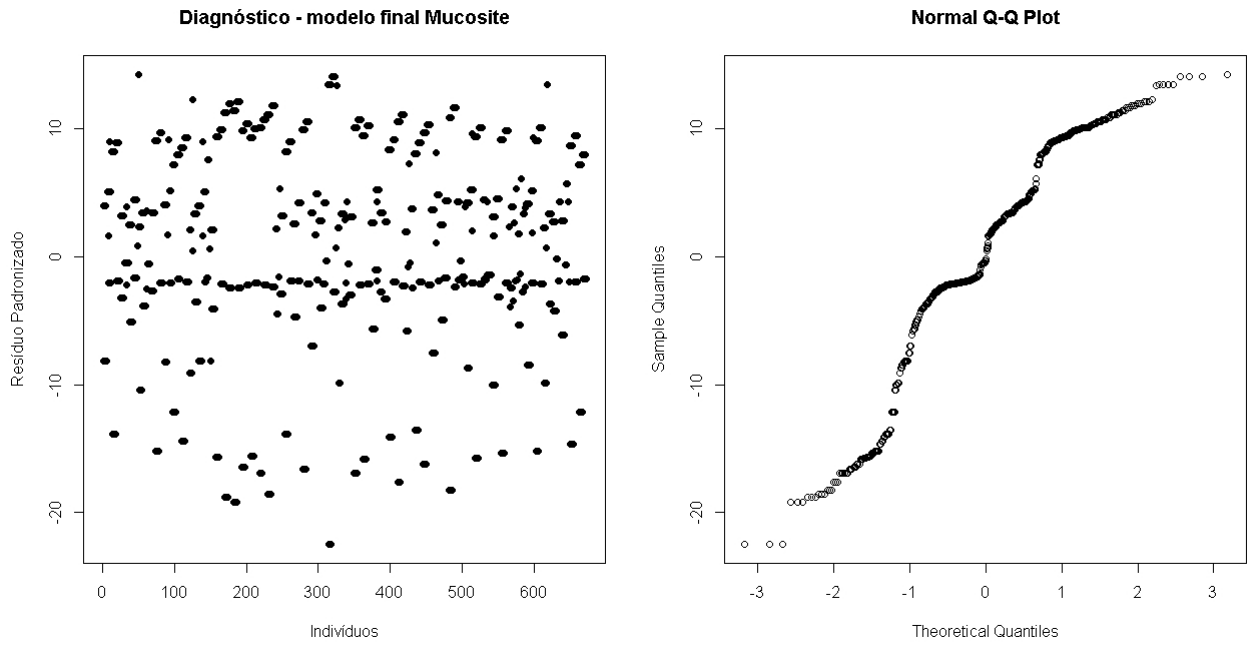


Figura A.10: Resíduos do primeiro modelo misto para mucosite

### A.0.6 Primeiro modelo marginal e misto para diarreia:

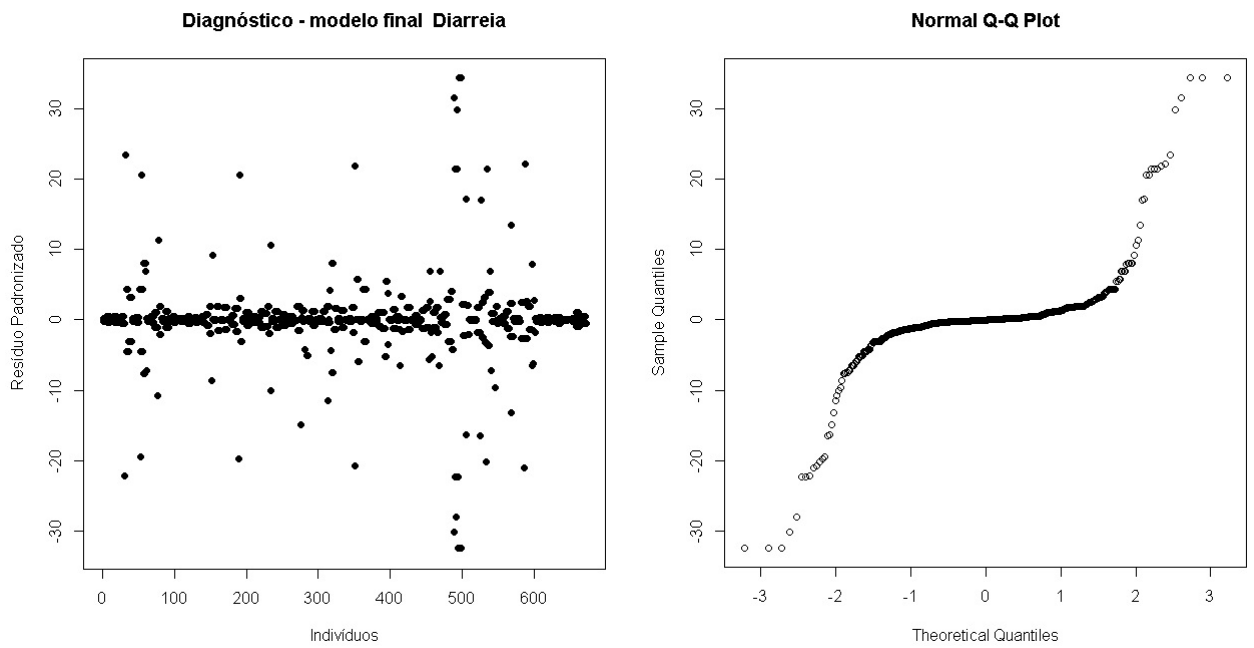


Figura A.11: Resíduos do primeiro modelo marginal para diarreia

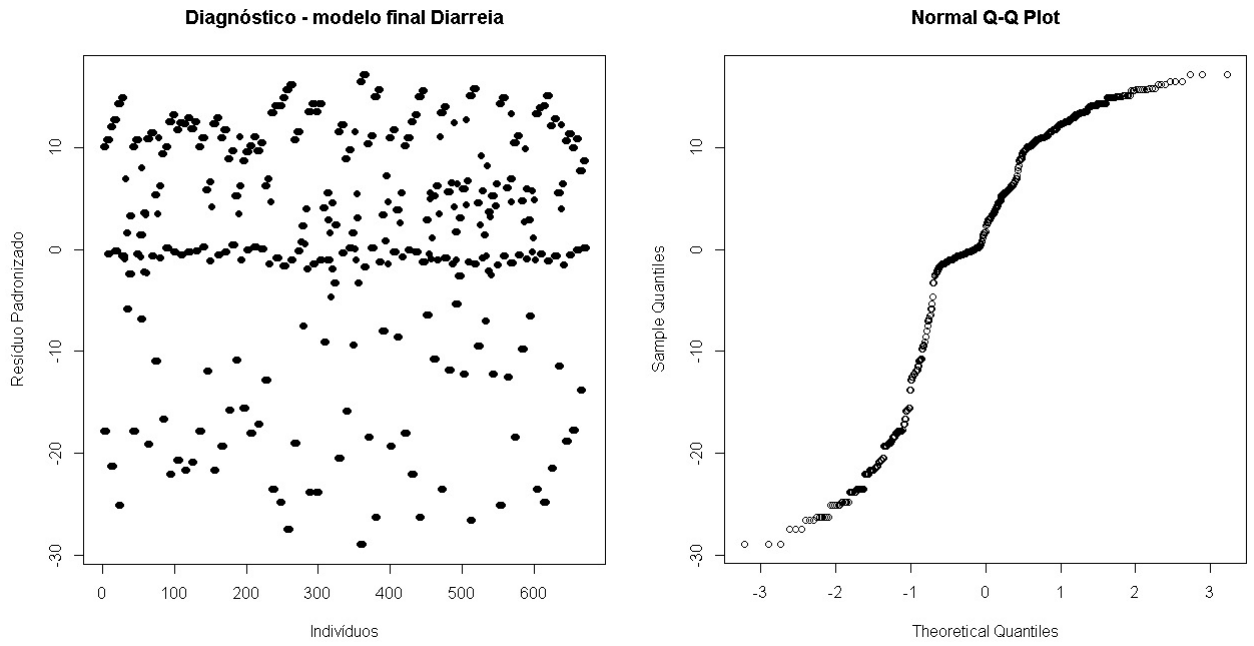


Figura A.12: Resíduos do primeiro modelo misto para diarreia

### A.0.7 Demais modelos para astenia:

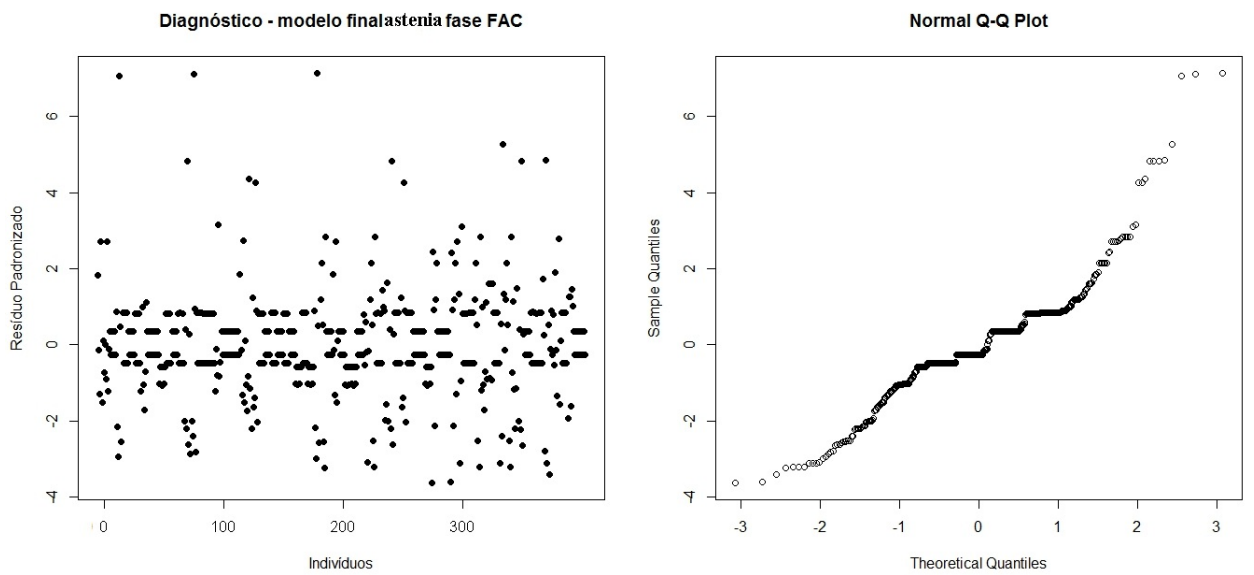


Figura A.13: Resíduos do segundo modelo marginal para astenia- fase FAC

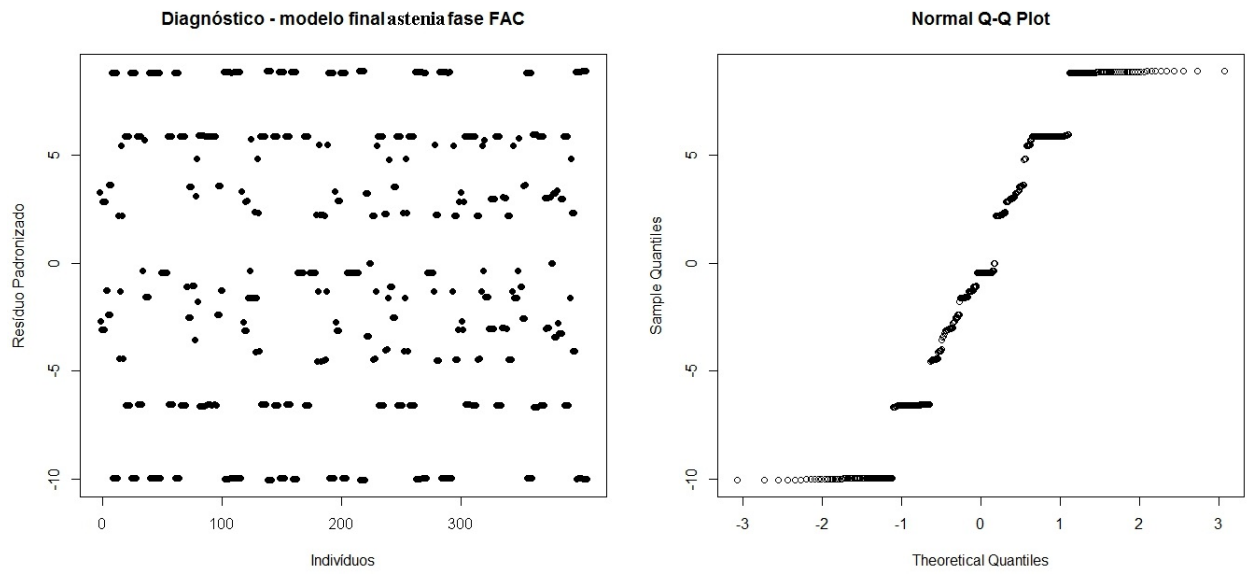


Figura A.14: Resíduos do segundo modelo misto para astenia - fase FAC

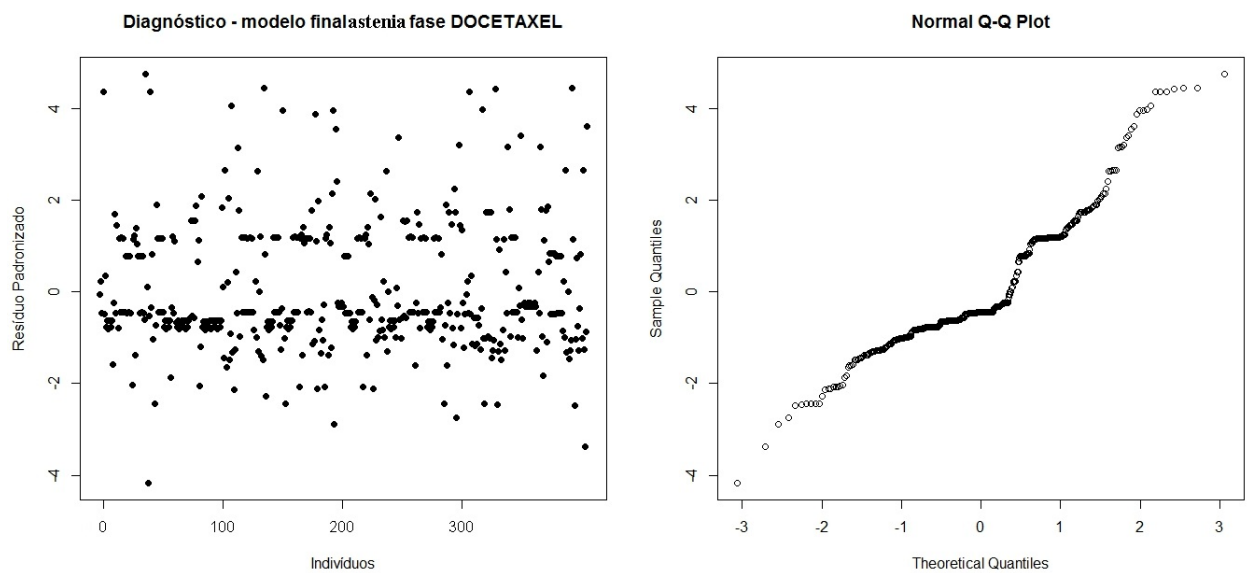


Figura A.15: Resíduos do terceiro modelo marginal para astenia - fase Docetaxel

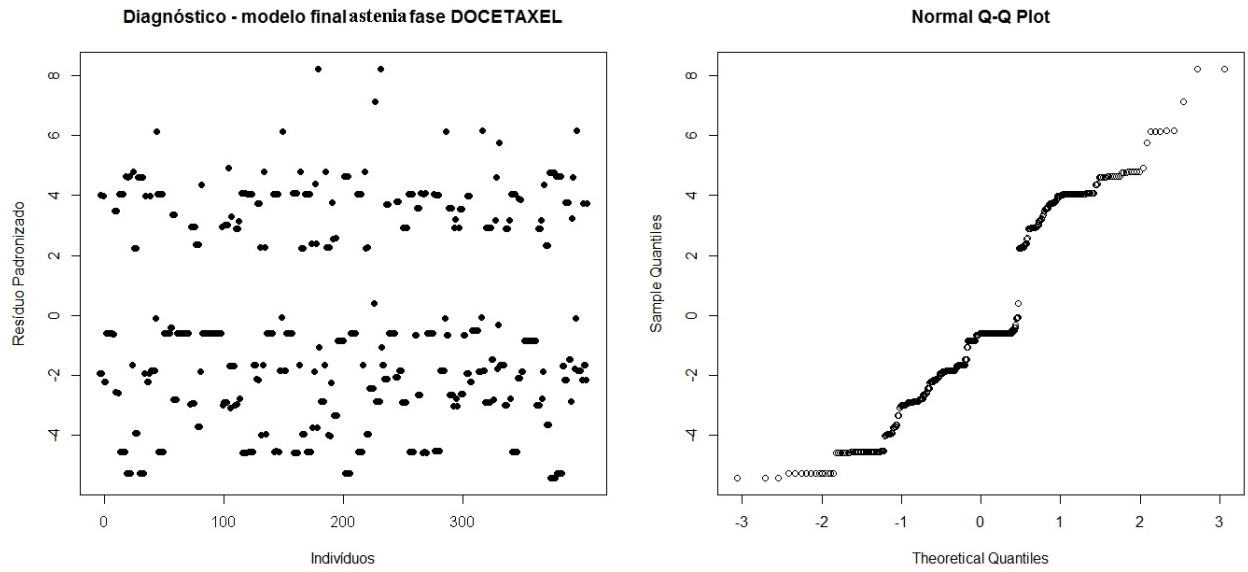


Figura A.16: Resíduos do terceiro modelo misto para astenia - fase Docetaxel

### A.0.8 Demais modelos para mialgia:

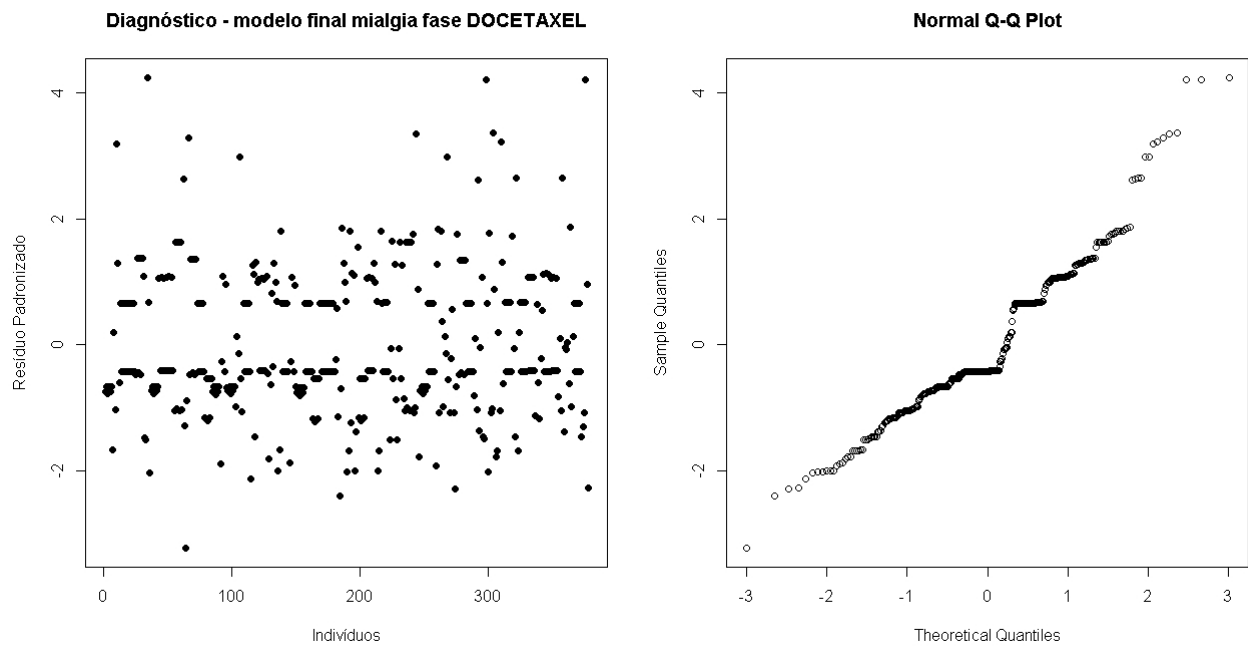


Figura A.17: Resíduos do terceiro modelo marginal para mialgia - fase Docetaxel

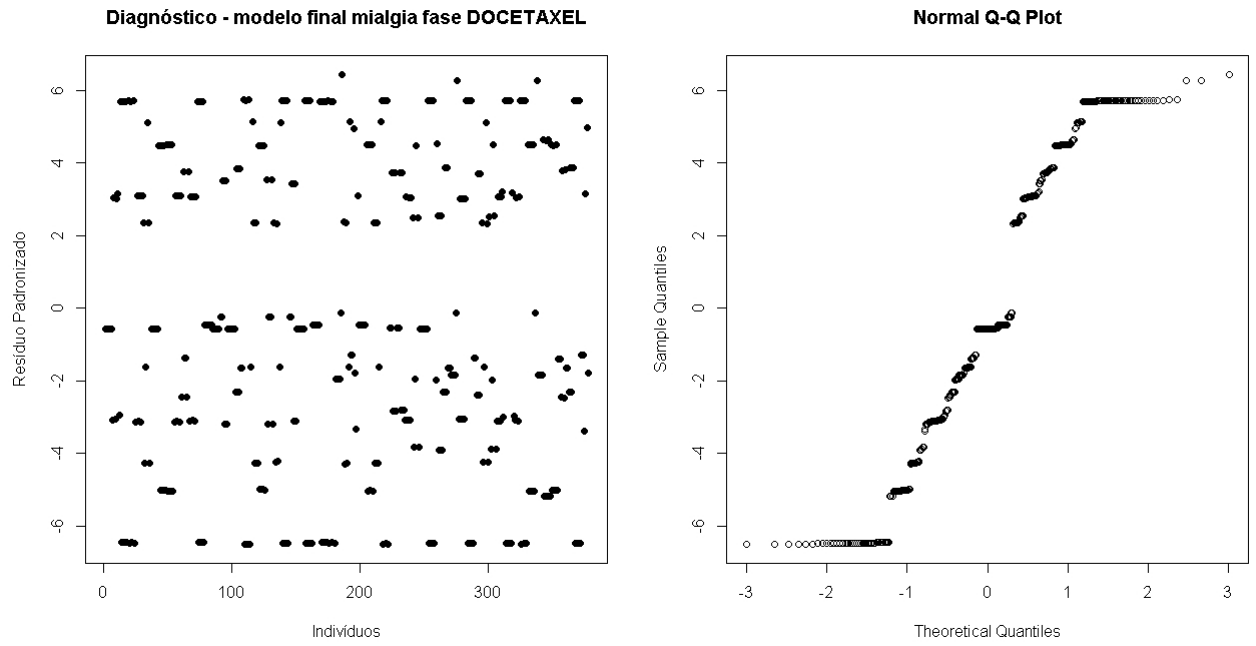


Figura A.18: Resíduos do terceiro modelo misto para mialgia - fase Docetaxel

### A.0.9 Demais modelos para artralgia:

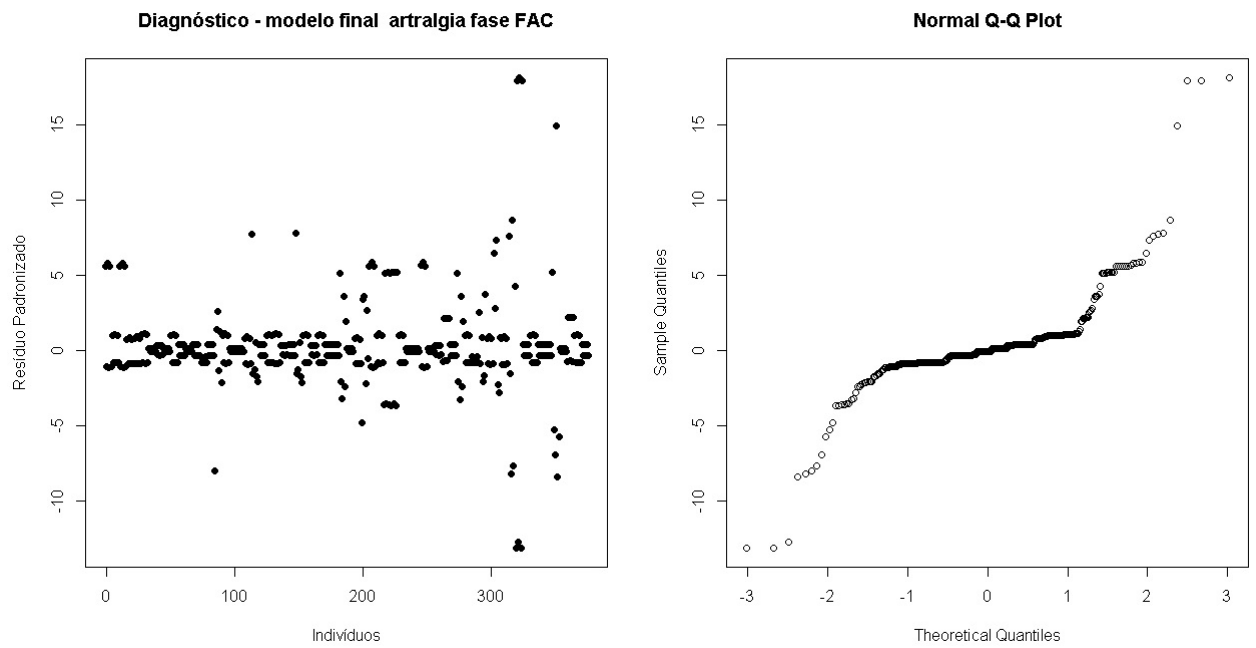


Figura A.19: Resíduos do segundo modelo marginal para artralgia - fase FAC

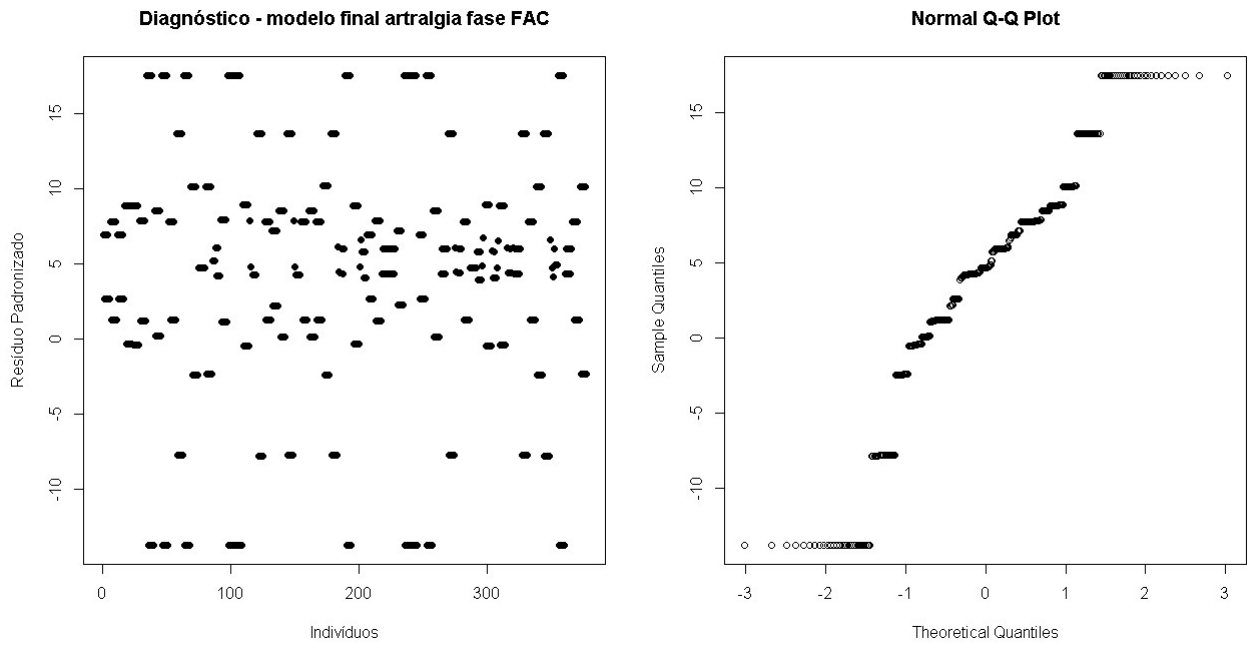


Figura A.20: Resíduos do segundo modelo misto para artralgia - fase FAC

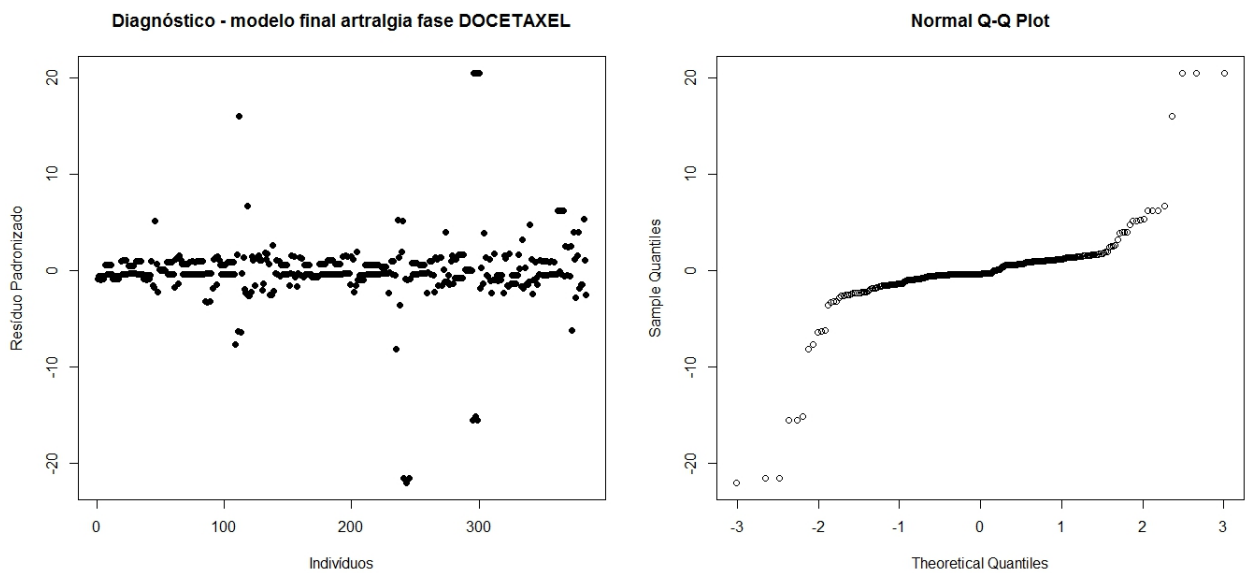


Figura A.21: Resíduos do terceiro modelo marginal para artralgia - fase Docetaxel



## A.0.10 Demais modelos para dor abdominal:

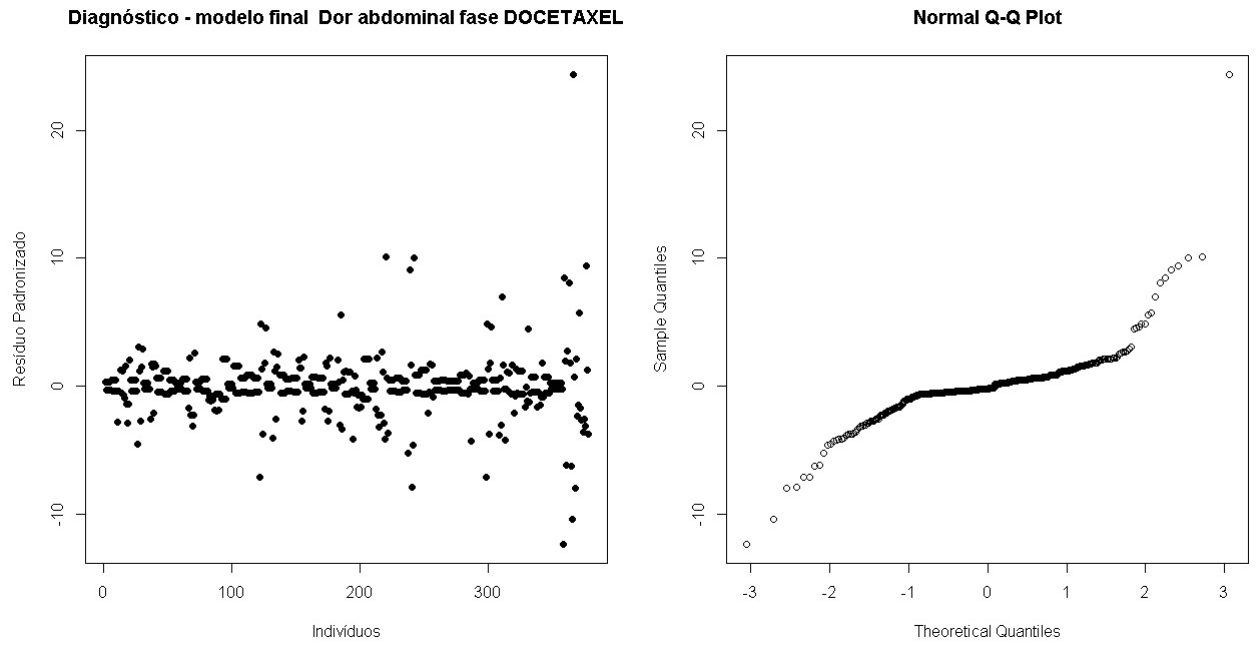


Figura A.22: Resíduos do terceiro modelo marginal para dor abdominal - fase Docetaxel

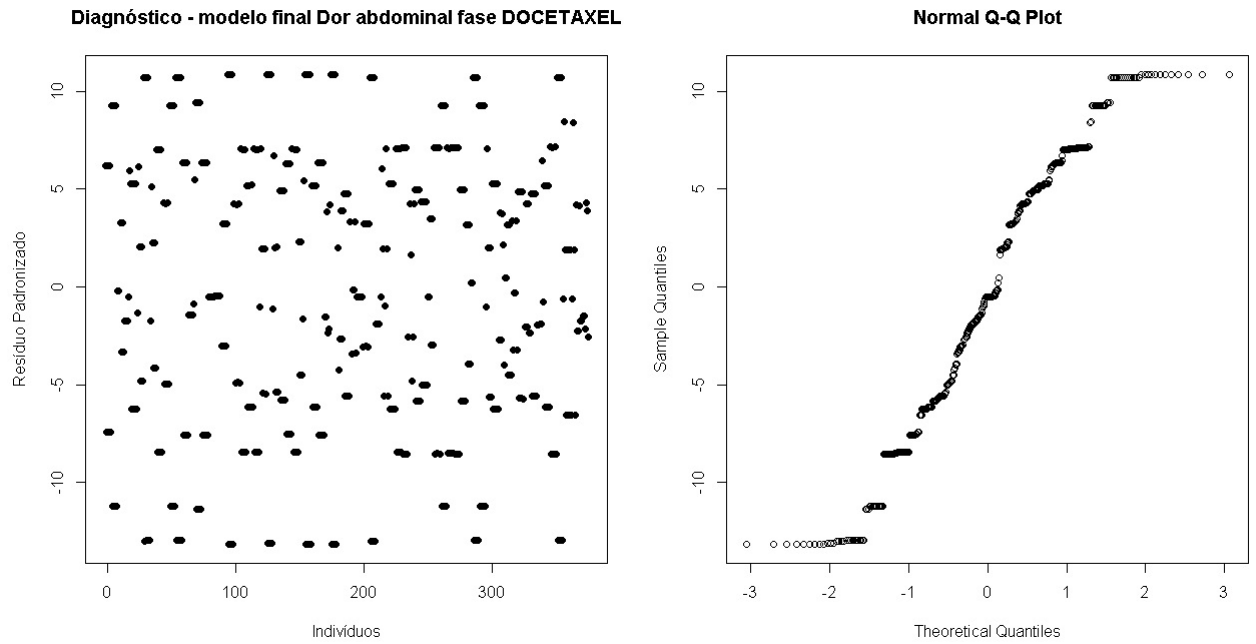


Figura A.23: Resíduos do terceiro modelo misto para dor abdominal - fase Docetaxel

### A.0.11 Demais modelos para mucosite:

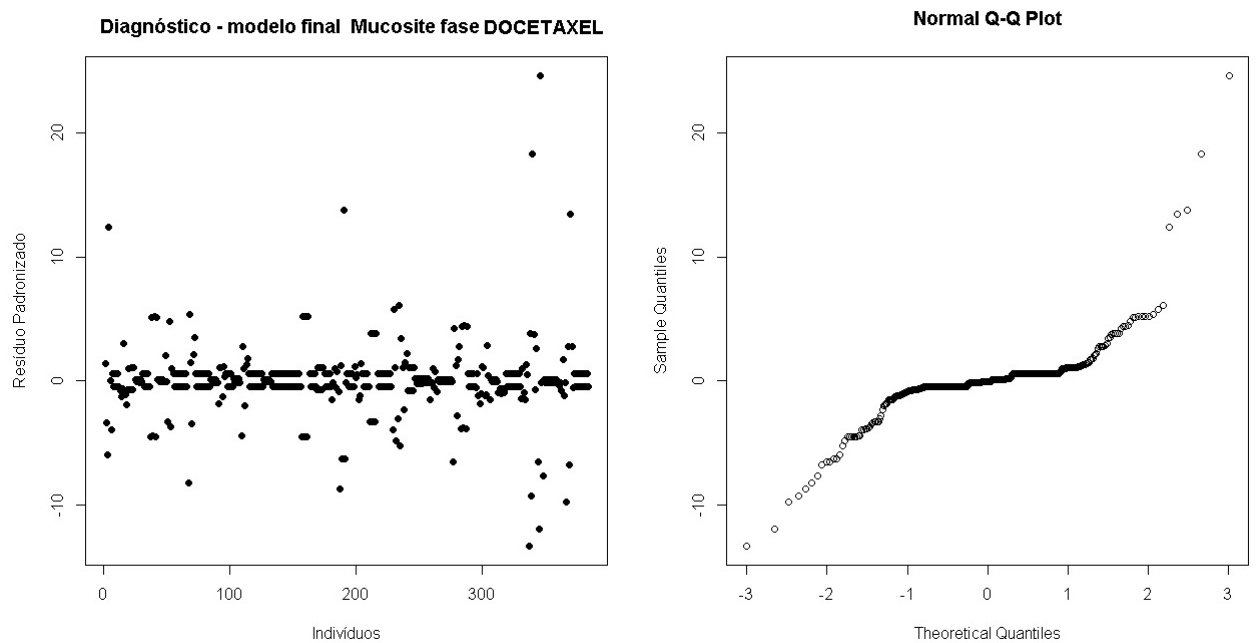


Figura A.24: Resíduos para o terceiro modelo marginal para mucosite - fase Docetaxel

## A.0.12 Demais modelos para Diarreia:

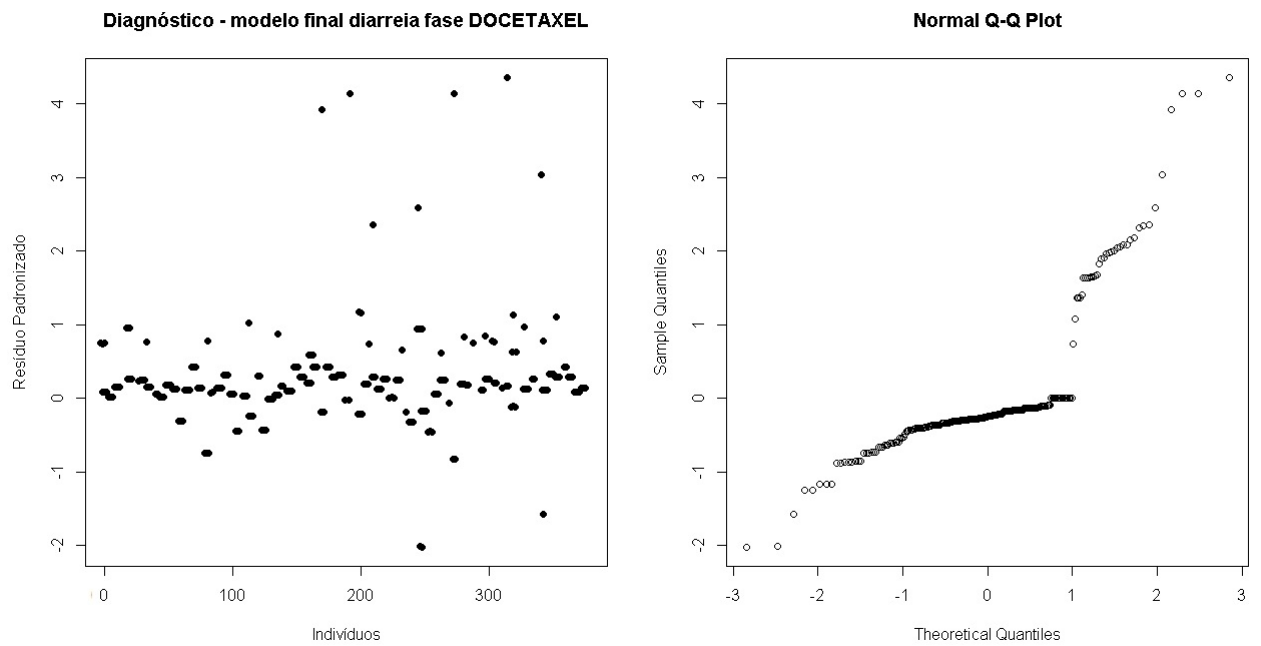


Figura A.25: Resíduos para o terceiro modelo marginal para diarreia - fase Docetaxel

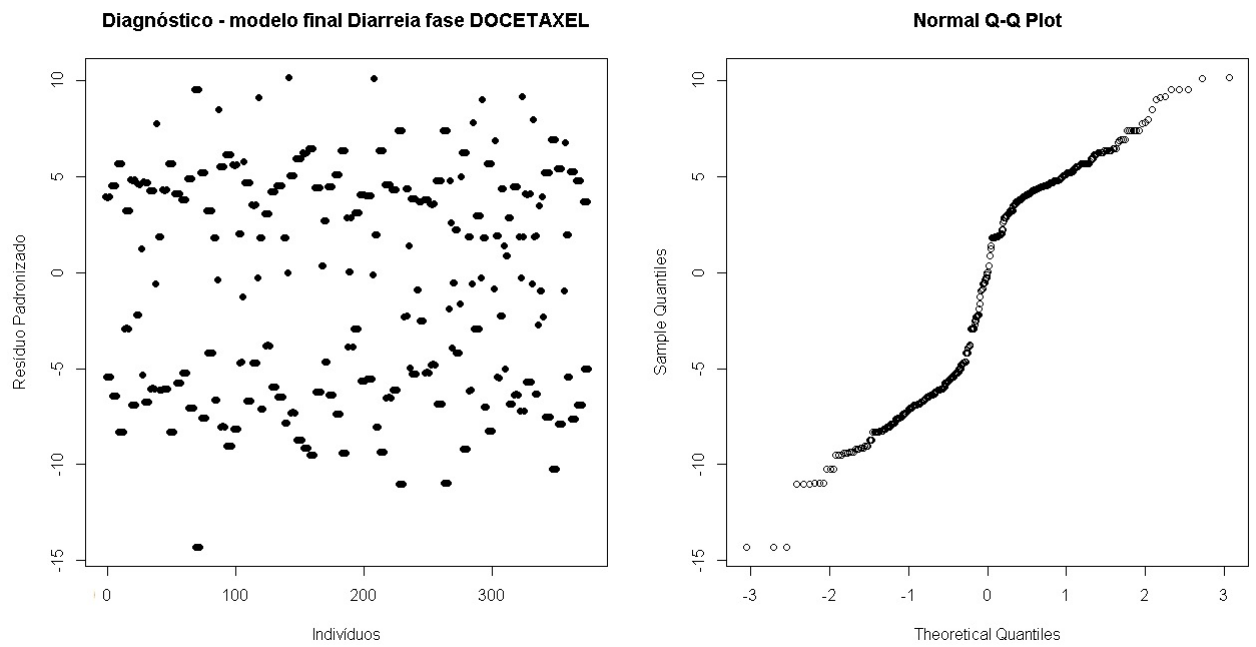


Figura A.26: Resíduos para o terceiro modelo misto para diarreia - fase Docetaxel