

RAE – CEA – 18P06

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO: “CLINICAL AND FUNCTIONAL DATA, GENETIC VARIABILITY AND PHARMACOLOGICAL THERAPIES RECEIVED BY GROUP OF BRAZILIAN PATIENTS WITH DUCHENNE MUSCULAR DYSTROPHY”.

Carlos Alberto de Bragança Pereira

Lucas Ramalho Anderson

Zheng Zhangzhe

São Paulo, Maio de 2018

CENTRO DE ESTATÍSTICA APLICADA – CEA – USP**RELATÓRIO DE ANÁLISE ESTATÍSTICA**

TÍTULO: “Relatório de análise estatística sobre o projeto: Clinical and functional data, genetic variability and pharmacological therapies received by group of Brazilian patients with Duchenne muscular dystrophy”

PESQUISADORAS: Gabriela de Souza Barbosa

Natália Guerra

ORIENTADOR: Marco Antônio de Albuquerque

INSTITUIÇÃO: Hospital Universitário da USP – SP

FINALIDADE DO PROJETO: Publicação

RESPONSÁVEIS PELA ANÁLISE: Carlos Alberto de Bragança Pereira

Lucas Ramalho Anderson

Zheng Zhangzhe

REFERÊNCIA DESTE TRABALHO:

Pereira, C.A.B.; Anderson, L.R.; Zheng, Z. Z. **Relatório de análise estatística sobre o projeto: “Clinical and functional data, genetic variability and pharmacological therapies received by group of Brazilian patients with Duchenne muscular dystrophy”**. São Paulo, IME – USP, 2018. (RAE – CEA – 18P06).

FICHA TÉCNICA

REFERÊNCIAS:

BUSSAB, W.O. e MORETTIN, P.A. **Estatística Básica. 9ª ed.** Saraiva, São Paulo: 2017

COLOSIMO, E.A. e GIOLO, S.R. **Análise de Sobrevivência Aplicada. 1ª ed.** Edgard Blücher: 2006

CONOVER, W.J. **Practical Nonparametric Statistics. 3ª ed.** John Wiley & Sons, New York: 1999.

FONSECA, A.T.Q.S.M., *et al.* *Doenças neuromusculares*. In: **Tratado de neurologia infantil**. Atheneu, São Paulo: 2017.

GEYER, J. C. **Nonparametric Tests and Confidence Intervals**. University of Minnesota. Disponível em: www.stat.umn.edu/geyer/old03/5102/notes/rank.pdf. Acesso: 25 jun. 18.

KIM, J.S. e PROSCHAN. F. *Piecewise exponential estimator of the survivor function*. In: **IEEE Transactions on Reliability**, vol. 40, pp. 134 – 139, 1991.

KLEIN, J.P. e MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. Springer, New York: 1997.

MARDIA, K. V. *et al.* **Multivariate Analysis. 10ª ed.** Academic Press, London: 1995

STONE. K. *et al.* **Occupational therapy and Duchenne Muscular Dystrophy. 1ª ed.** Wiley: 2007

WANG, D. N *et al.* (2017). *Clinical and mutational characteristics of Duchenne muscular dystrophy patients based on a comprehensive database in South China*. In **Neuromuscular Disorders** 27, 715–722, 2017

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Microsoft Office Excel for Windows (versão 2007)

Microsoft Office Excel for Windows (versão 2016)

Microsoft Office Word for Windows (versão 2007)

Microsoft Office Word for Windows (versão 2016)

R (Versão 3.5.0)

TÉCNICAS ESTATÍSTICAS UTILIZADAS:

Análise Descritiva Unidimensional (03:010)

Análise Descritiva Multidimensional (03:020)

Análise de Sobrevida (13:070)

Estimação Não Paramétrica (04:080)

Estimador Piecewise Exponencial da Função de Sobrevida (Outros – 04:990)

Testes de Hipóteses Não Paramétricas (05:070)

ÁREA DE APLICAÇÃO:

Medicina Epidemiologia (14:040)

SUMÁRIO

Resumo	5
1. Introdução	7
2. Objetivos	8
3. Descrição do estudo	10
4. Descrição das variáveis	11
5. Análise descritiva	13
5.1. Questões relativas às mutações	13
5.2. Questões relativas aos pacientes	16
6. Análise inferencial	20
6.1 Intervalos de confiança	20
6.2 Modelo de Cox	20
7. Conclusão	22
Apêndice A	23
Apêndice B	34
Apêndice C	49

Resumo

A distrofia muscular de Duchenne é uma doença causada por mutações associadas ao cromossomo X. As consequências da moléstia no portador são a degeneração dos músculos, levando a uma progressiva perda de funções do organismo (perda de marcha, problemas cardíacos e respiratórios), que acarretam em morte precoce. Ao compararmos o estudo de referência¹ com o que foi observado em clínica, podemos ver que os cenários são parecidos, apesar de terem ocorrido alguns tipos de mutações diferentes em ambos os estudos.

Com o intuito de procurar associar os tipos de mutações com as demais variáveis (sintomas iniciais e deficiência mental), concluímos que a variável de interesse (tipo de mutação) é independente dos sintomas iniciais e de deficiência mental. Ao aplicarmos o teste exato de Fisher para se testar independência entre o tipo de mutação e as variáveis citadas, os valores-p associados são dados por 45,17% e 10,69%, respectivamente. Além disto, averiguamos que não há evidências estatísticas para dizer que existem diferenças entre as médias de CPK, tempos de início de sintomas e tempo entre diagnóstico e sintomas iniciais por tipo de mutação. Aplicando o teste exato de Wilcoxon para avaliar se existe diferença entre as distribuições da variável CPK por tipo de mutação, pudemos observar um valor alto para o valor-p (51,9%). Para comparar os os tempos de início de sintomas, comparamos as três curvas de sobrevivência associadas aos tipos de mutação. Aplicando o teste log-rank, concluímos que não existe diferença entre as curvas (valor-p dado por 55,4%). A respeito do tempo entre idade de diagnóstico e início de sintomas, pudemos observar que os boxplot para a variável são praticamente os mesmos por tipo de mutação.

Como os indivíduos que não tomam corticoides ainda não pararam de andar (a única informação que utilizamos para indicar a piora dos pacientes), não podemos avaliar se a utilização de corticoides é diferente de sua não utilização para a evolução

¹ WANG, D. N *et al.* (2017). *Clinical and mutational characteristics of Duchenne muscular dystrophy patients based on a comprehensive database in South China*. In **Neuromuscular Disorders** 27, 715–722, 2017

da doença. Porém, pudemos realizar a comparação entre os corticoides, pois, das pessoas que perderam a marcha, todas utilizavam algum dos corticoides (prednisona ou deflasacort). Concluímos que não existe diferença entre os corticoides ao compararmos suas curvas de sobrevivência (utilizando como tempo de referência a distância entre a idade de perda de marcha e a idade em que começaram a tomar corticoides). Ao aplicarmos o teste log-rank, obtivemos o valor-p dado por 27,6%.

Propusemos o modelo semi-paramétrico de riscos proporcionais de Cox para avaliar o risco de perda de marcha dos pacientes, onde a única variável explicativa estatisticamente significativa foi a idade de início de corticoide (quanto mais tarde o paciente tomar, maior o risco de se perder a marcha). Calculamos também intervalos de confiança (95%) para a média e para a mediana do tempo necessário para se diagnosticar a doença a partir de seus sintomas iniciais, ambos dados por [8,11] .

1. Introdução

A distrofia muscular de Duchenne é uma doença genética recessiva, relacionada ao cromossomo X. Por este motivo, sua ocorrência é muito mais frequente na população do sexo masculino do que no feminino, e normalmente a doença se manifesta antes dos 5 anos em meninos².

O fenômeno ocorre quando um gene defeituoso no cromossomo leva a uma deficiência na produção de distrofina (proteína que mantém a integridade das paredes celulares dos músculos). Sem a proteína, diversos eventos fisiológicos ocorrem, levando à destruição irreversível das células musculares³.

Com o tempo, além do indivíduo perder a movimentação dos membros, começam a surgir complicações cardíacas e pulmonares, sendo estas as causas mais comuns de morte dos pacientes. As mutações genéticas que podem levar a esta condição são diversas, levando a graus variados de severidade. Dependendo da gravidade da doença, o indivíduo pode possuir algum tipo de deficiência mental⁴.

Quando existe a desconfiança de que o indivíduo é portador da doença, é indicado um exame para se determinar a concentração de creatinofosfoquinase (CPK) no sangue. Quanto maior a concentração desta enzima, maior o indicativo da existência de alguma doença neuromuscular. Caso se confirme que a concentração de CPK esteja alta, passa-se para o exame de confirmação da doença. Pode-se fazer o teste molecular, considerado o diagnóstico padrão-ouro, que permite aconselhamento genético, diagnóstico pré-natal, planejamento do tratamento paliativo, entre outros. Porém, é um teste que é pouco disponível na rede pública. Logo, utiliza-se mais a

² STONE. K. *et al.* **Occupational therapy and Duchenne Muscular Dystrophy**. 1^a ed. Wiley: 2007
p.1-3

³ Idem

⁴ Ibidem

biópsia muscular, que apesar de ser um método mais invasivo, é o método mais utilizado na rede pública brasileira⁵.

A doença não possui cura, mas existem diversos tratamentos para abrandar os sintomas e prolongar a vida dos afetados pela moléstia. Dentre estas medidas, as mais comuns são uso de corticoides e cardioprotetores. Existem tratamentos alternativos, como fisioterapia e hidroterapia.

O presente estudo visa compreender melhor o perfil dos pacientes atendidos no ambulatório de Neurologia Infantil do Hospital das Clínicas e como a evolução da doença ocorre.

2. Objetivos

⁵ FONSECA, A.T.Q.S.M., et al. *Doenças neuromusculares*. In: **Tratado de neurologia infantil**. Atheneu, São Paulo: 2017.

Os objetivos deste relatório de análise estatística são:

1. Descrever as frequências obtidas para as variáveis de interesse que possivelmente estão associadas à doença de Duchenne (tipos de mutações genéticas, início dos sintomas, idade de diagnóstico, localização das mutações) e compará-las com os dados da literatura⁵.
2. Descrever a associação entre os diferentes tipos de mutações genéticas encontradas e compará-las com a gravidade clínica;
3. Averiguar se há benefícios do tratamento com corticoides precocemente;
4. Descrever o tempo para se diagnosticar a doença a partir do diagnóstico.

Para o presente estudo, utilizaremos como referência o artigo “Clinical and mutational characteristics of Duchenne muscular dystrophy patients based on a comprehensive database in South China”⁶, apresentado nas referências.

⁶ WANG, D.N. et al. Op.cit.

3. Descrição do estudo

A base de dados considerada é constituída por todos os pacientes diagnosticados com a doença no Hospital das Clínicas entre agosto de 2014 e agosto de 2017, totalizando 73 observações. Trata-se de um estudo observacional e prospectivo, pois o paciente é acompanhado a partir do momento em que a doença é diagnosticada, e anotam-se seus dados.

4. Descrição das variáveis

- **Idade DX:** Idade (em anos) que o paciente foi diagnosticado com a doença
- **CPK:** Concentração da enzima creatinofosfoquinase (U/L), calculada após o diagnóstico (tempo indeterminado) – taxas elevadas podem indicar doenças relacionadas aos músculos ou nervos (a referência para homens é entre 22U/L e 334 U/L).
- **História familiar:** Se a família apresenta histórico da doença ou não
 - **0** – Sem informação
 - **1** – Não
 - **2** – Sim
- **Quadro inicial:** Como a doença se manifestou no paciente no início
 - **0** – Assintomático
 - **1** – Atraso na aquisição da marcha
 - **2** – Alteração no padrão da marcha previamente adquirida e dentro da normalidade
 - **3** – Perda de habilidades motoras previamente adquiridas
 - **4** – Dor nos membros inferiores
- **Início dos sintomas:** Idade (em anos) em que começou o quadro inicial
- **Genética:** Tipo de mutação que o paciente carrega e em quais exons ela ocorre (somente para as mutações do tipo deleção e duplicação, para as demais não faz sentido falar em posições de exons). Por haver uma grande quantidade de mutações, mantivemos as mutações do tipo deleção e duplicação em separado e agregamos as demais mutações (nonsense, missense, intrônica, pontual) na categoria “outras”, por não haver quantidade suficiente das mesmas.

- **Idade atual:** Idade (em anos) que o paciente possui nos dias de hoje (2018)
- **Corticoide atual:** Qual corticoide o paciente utiliza (se utiliza) hoje em dia (2018). Os pacientes só podem tomar um tipo de corticoide por período.
 - **0** – Não utiliza
 - **1** – Deflasacort
 - **2** – Prednisona
- **Idade de início do corticoide:** Idade (em anos) que o paciente começou a tomar corticoides (independente de qual tenha sido prescrito)
- **Tempo de uso:** Tempo de uso (em anos) de corticoides do paciente
- **Fisioterapia:** Número de vezes na semana que o paciente pratica fisioterapia
- **Hidroterapia:** Número de vezes na semana que o paciente pratica hidroterapia
- **Cardioprotetor:** Qual cardioprotetor o paciente utiliza (se utiliza)
 - **0** – Não usa
 - **1** – Enalapril
 - **2** – Enalapril e carvedilol
- **Idade de perda de marcha:** Idade (em anos) que o paciente perdeu a marcha (se perdeu)
- **Deficiência intelectual:** Indicador de pacientes com deficiência mental
 - **0** – Não
 - **1** – Sim

5. Análise descritiva

Para a análise descritiva, procuramos descrever ao máximo as informações do banco de dados, explorando tanto as variáveis que podem estar relacionadas diretamente com as mutações gênicas como as variáveis relacionadas aos doentes. Para uma análise univariada utilizamos box plots, histogramas e gráficos de barras⁷. Para uma análise bivariada, a análise de correspondência foi a escolhida⁸. Também foram construídos gráficos de Kaplan-Meier⁹ (e sua suavização exponencial¹⁰) para a análise de sobrevivência.

5.1. Questões relativas às mutações

Note-se primeiramente (Tabela A.1) que, dos 73 pacientes originais, 51 carregam mutações relativas à deleção de exons, 10 apresentam duplicação de exons (foram observadas 6 grandes duplicações, 2 pequenas duplicações e 2 mutações do tipo duplicação sem posição definida), 9 tiveram mutações que consideramos como “outras” (mutações pontuais, nonsense, missense e intrônica) e 3 ainda aguardam resultado. Como não se conhece o tipo de mutação que estes últimos pacientes carregam, procuramos relacionar as variáveis que podem estar associadas à mutação apenas com diagnósticos já conhecidos.

Com respeito aos exons afetados, criamos gráficos de barras para observar a região que mais é afetada apenas para as mutações do tipo deleção e duplicação, pois para as demais mutações não faz sentido falarmos de posições de exons.

⁷ BUSSAB, W.O. e MORETTIN, P.A. **Estatística Básica**. 6ª ed. Saraiva, São Paulo: 2009.

⁸ MARDIA, K. V. *et al.* **Multivariate Analysis**. 10ª ed. Academic Press, London: 1995

⁹ KLEIN, J.P. e MOESCHBERGER, M. L. **Survival analysis: techniques for censored and truncated data**. Springer, New York: 1997

¹⁰ KIM, J.S. e PROSCHAN, F. *Piecewise exponential estimator of the survivor function*. In: **IEEE Transactions on Reliability**, vol. 40, pp. 134 – 139, 1991.

Assim como no artigo de referência, podemos notar (Figuras B.1 e B.2) que há um pico de deleções entre os exons 48 e 52, e que a região à esquerda deste pico é muito mais densa do que a região à direita. Há apenas a divergência entre o que observamos e o artigo de que a região à esquerda observada é muito mais densa relativamente do que o estudo chinês.

Quanto aos genes afetados pela duplicação, não conseguimos efetivamente realizar uma comparação, pelo número pequeno de pacientes que sofrem deste tipo de mutação (foram observados apenas 6 pacientes com grandes duplicações). Porém, aparentemente, a mutação ocorre de maneira uniforme entre os exons, o que também ocorre no estudo de comparação (Figuras B.3 e B.4).

A respeito dos tipos mais frequentes de mutação, o que foi observado (Tabela A.1) é diferente do que foi registrado no estudo de referência (Tabela A.2). Pode-se observar que aproximadamente 77% dos pacientes do estudo chinês sofrem de mutações do tipo deleção, enquanto que isto só ocorre em 57% da amostra do estudo. Além disto, existem mutações que não existem na referência e que foram observadas (como é o caso das mutações do tipo intrônica e pontual) e mutações existentes no estudo chinês que não foram observadas no presente estudo (pontos de splice). Apesar disso, as frequências relativas de mutações do tipo duplicação são semelhantes (9,09% e 10,9% no observado e na comparação, respectivamente).

Quanto ao “Quadro Inicial”, realizamos uma análise de correspondência (ver Figura B.5), onde pode-se notar que o tipo duplicação está mais associado à falta de sintomas iniciais, que o grupo de outras mutações está tão associado à perda de habilidades motoras previamente adquiridas quanto a deleção e que aparentemente deleção está associada aos demais sintomas, a saber: atraso na aquisição da marcha, alteração no padrão da marcha previamente já adquirida e dor nos membros inferiores. Porém há a ressalva de que há poucas observações tanto para o grupo de duplicação quanto para o grupo de outras mutações (10 e 9, respectivamente). Como não foram observados pacientes com dor nos membros inferiores para as mutações do tipo duplicação e outras e só observamos indivíduos assintomáticos para as mutações do tipo deleção e duplicação (Tabela A.3), temos que a análise fica

distorcida. Aplicando o teste exato de Fisher (valor-p dado por 45,17%), não rejeitamos a hipótese de que há independência entre o tipo de mutação e o quadro inicial.

Para avaliar se existe diferença entre o início dos sintomas da doença para cada tipo de mutação, construímos um gráfico de Kaplan-Meier, em que o tempo de interesse é o início dos sintomas (Figura B.6). Podemos observar que de fato, não parece haver relação entre as duas variáveis, isto é, o tipo de mutação não parece influenciar na idade de início dos sintomas (fazendo o teste log-rank, temos que o valor-p é dado por 0,554, isto é, não rejeitamos a hipótese de que as curvas de sobrevivência são iguais).

Com relação a algum tipo de deficiência mental, temos que 47 pacientes não possuem, 22 possuem alguma e 3 não se conhece a informação. Desta forma, trabalhamos apenas com a informação conhecida. Podemos observar (ver Figura B.7 e Tabela A.4) que aparentemente existe alguma relação com o tipo de mutação, pois temos que apenas 25% dos pacientes do grupo que possui mutações do tipo deleção possui algum tipo de deficiência, enquanto que relativamente este número aumenta para 60% no grupo que possui a mutação do tipo duplicação. Podemos observar que aparentemente os outros tipos de mutação seguem um comportamento próximo ao grupo de deleções, em que apenas 33% dos pacientes possuem algum tipo de deficiência. Aplicando o teste exato de Fisher (valor-p dado por 10,69%), não temos evidência para rejeitar que o tipo de mutação é independente da deficiência mental.

Construímos também o *box plot* das concentrações de CPK por tipo de mutação (ver Figura B.8 e medidas resumo na Tabela A.5). Podemos observar que aparentemente não há diferença entre as concentrações da enzima para os diferentes tipos de mutação (a média é dada por 12187,04 U/L e a mediana é de 10892 U/L, desconsiderando os tipos de mutação). Pelo teste de Wilcoxon (valor-p dado por 51,9%), não rejeitamos a hipótese de que as médias de CPK são iguais para os três tipos de mutação.

Construímos um *box plot* por tipo de mutação, para a variável diferença entre a idade de diagnóstico do paciente e a idade inicial dos sintomas (Figura B.9). Pode-se

observar que aparentemente a distribuição deste tempo é muito próxima para todos os tipos de mutação. Como é comum este tempo ser maior do que dois anos, categorizamos os mesmos em “tempo menor ou igual a dois anos” e “tempo maior do que dois anos”, para podermos entender melhor se o tipo de mutação genética pode ter relação com uma demora maior ou menor para o diagnóstico, a partir de seus primeiros sintomas (ver Figura B.10 e Tabela A.6). Novamente nota-se que aparentemente não há diferença entre as mutações para o tempo considerado. O tempo entre diagnóstico e idade inicial de sintomas maior do que 2 anos ocorre para 56%, 67% e 44% para as mutações deleção, duplicação e outras, respectivamente. Deve-se lembrar que existem apenas 9 observações para os últimos dois tipos de mutações citados.

5.2. Questões relativas aos pacientes

A base de dados compreende 73 jovens com idades entre 3 e 18 anos atualmente. Podemos observar que a distribuição de idades é simétrica, sendo a média igual a 9,01 anos e a mediana igual a 9 anos (ver Tabela A.7 e Figura B.11). Para avaliar se existem pacientes jovens demais na amostra (que potencialmente desenvolveram pouco a doença), construímos dois *box plots* que podem ser observados na Figura B.12. No primeiro *box plot* podemos observar a distribuição das idades atuais dos jovens que não tomam corticoides. Já no segundo, podemos observar a distribuição das idades em que os pacientes, que tomam corticoides hoje em dia, começaram a tomar os mesmos. Da comparação de distribuições, podemos averiguar que aparentemente não há diferença entre as mesmas. Aplicando o teste de Wilcoxon não pareado, obtemos um valor-p de 69,21%, isto é, não rejeitamos a hipótese de que as distribuições são iguais.

Um de nossos objetivos é avaliar se a utilização de corticoides está desacelerando o progresso da doença. Porém, como nenhum dos pacientes que não toma corticoide teve perda de marcha, não é possível avaliar esta diferença. No entanto, podemos comparar os corticoides entre si (ver Tabela A.8). Para os grupos que usam corticoides,

construímos o gráfico de Kaplan-Meier, em que o tempo medido é a diferença entre a idade em que o paciente perde a marcha e a idade que começou a tomar o corticoide atual (Figura B.13). No gráfico, cada “escada” representa a função de sobrevivência relativa aos grupos de pacientes que tomam corticoides distintos (totalizando 61 pacientes, sendo 38 pertencentes ao grupo que usa deflasacort e 23 ao grupo que utiliza prednisona, ver Tabela A.8). Como a função de sobrevivência é contínua, utilizamos a técnica PEXE (Piexewise Exponential) para suavizar o gráfico e melhor comparar as duas curvas (Figura B.14 e Fórmula C.1). Observando ambos os gráficos, temos que aparentemente o tempo até a perda da marcha é aproximadamente o mesmo para os dois grupos (existem apenas 11 falhas observadas, sendo que 3 pertencem ao grupo que tomou prednisona e 8 ao grupo dos pacientes que tomaram deflasacort). Ao aplicarmos o teste log-rank, obtemos um valor-p de 27,6%, isto é, não rejeitamos a hipótese de que as curvas de sobrevivência são iguais. Como alguns pacientes tomaram corticoides no mesmo ano em que perderam a marcha para os indivíduos que tomaram deflasacort (ver Tabela A.9), a suavização da curva de sobrevivência ficou deformada. Além disto, não utilizamos os tempos negativos, justamente pelo fato da suposição de que o tempo deve ser positivo.

Construímos também dois *box plots* (Figura B.15) para a idade inicial em que cada paciente começou a tomar corticoides. Podemos observar que as distribuições de idades são diferentes, em que a mediana da idade inicial para as pessoas que tomam prednisona é de 5,5 anos, enquanto que a mesma medida é dada por 8 anos para os pacientes que tomam deflasacort (Tabela A.10). Realizando o teste de Wilcoxon (valor-p < 5%), ou seja, rejeitamos a hipótese de que as distribuições são iguais.

Com o conhecimento do histórico familiar da doença, construímos um *box plot* em que a variável de interesse é o tempo entre o diagnóstico dos pacientes e o início dos sintomas. Observando a Figura B.16 podemos ver que, aparentemente, não há diferença entre os dois grupos para esta diferença de tempo. Como no caso anterior, categorizamos os intervalos de tempo em “tempo menor ou igual a dois anos” e “tempo maior do que dois anos”, para podermos entender melhor como se dá essa relação entre histórico familiar e o tempo mencionado (ver Tabela A.11). Podemos observar no

gráfico de barras (Figura B.17) que, aparentemente, a proporção de tempo “menor do que 2 anos” é maior no grupo que possui histórico familiar (o tempo menor do que 2 anos ocorre para 53% e 42% para os pacientes com histórico familiar e sem histórico familiar, respectivamente).

Por se tratar de uma doença rara, é comum haver uma grande distância entre a idade de diagnóstico e a idade inicial dos sintomas. Podemos observar que a média e a mediana da idade de diagnóstico são dadas por 7,05 e 7 anos, respectivamente, enquanto que as mesmas medidas para a idade de sintomas são dadas por 3,58 e 3,5 (Tabela A. 12 e A. 13). Desta forma, construímos um gráfico de barras contendo as duas informações (Figura B.18) para comparar com o gráfico gerado pelo estudo de referência (Figura B.19). Podemos observar que o comportamento das distribuições dos dois tempos é similar para ambos. Construímos então o *box plot* da diferença entre as idades citadas (Figura B.20 e Figura B.21) e pode-se notar que de fato a distância entre estes períodos é evidente (podemos ver na Tabela A.14 que a média desta diferença é 3,44 e a mediana é 3, ambas medidas em anos), que concorda com o artigo de referência (existe um “gap” muito grande entre os sintomas e o diagnóstico para a criança começar a se tratar).

O uso de cardioprotetores é um indicativo de que a doença está se agravando. Como não possuímos o tempo desta “piora”, isto é, não conhecemos o momento em que os meninos começaram a utilizar os remédios, não podemos afirmar quais variáveis estão associadas ao progresso mais acelerado da doença. Da mesma forma, não podemos também averiguar se a prática de terapias (como a fisioterapia e a hidroterapia) estão associadas a uma redução do ritmo da moléstia pois não possuímos a informação de quanto tempo o indivíduo fez terapia. Sendo assim, podemos apenas avaliar as condições dos meninos observados, sem construir relações de causalidade com as demais variáveis.

A respeito da procura de terapias para o tratamento, podemos ver (Tabela A.15 e Figura B.22) que aproximadamente 20% dos pacientes não faz fisioterapia, e que aproximadamente 50% dos pacientes faz a mesma apenas uma vez por semana. Já a

hidroterapia é feita por apenas aproximadamente 30% dos meninos (Tabela A.16 e Figura B.23).

Quanto ao uso de cardioprotetores, podemos ver na Tabela A.17 e na Figura B.24 que 61,64% dos meninos não utiliza nenhum tipo de cardioprotetor, 34,24% utiliza enalapril e os demais utilizam enalapril e carvedilol. Isto é, até o dado momento, a maior parte dos pacientes não apresenta um dos principais indicadores de piora da doença. Já sobre os corticoides, podemos observar na Tabela A.8 e Figura B.25 que 16,44% (12 pacientes) não tomam, e que aproximadamente 52% (38 jovens) dos mesmos toma deflasacort. Os demais tomam prednisona (31,51%, equivalente a 23 pacientes).

6. Análise inferencial

6.1 Intervalos de confiança

Para responder à questão sobre o tempo de demora para se diagnosticar a doença a partir dos sintomas iniciais, propusemos métodos não paramétricos para estimar sua média¹¹ e mediana¹². A decisão foi tomada pois, apesar de haver 70 observações (Tabela A.14), não desejamos assumir uma distribuição para o tempo citado. Assim, o intervalo de confiança (95%) atribuído à mediana deste tempo é dado por [2,4] (em anos), e para a média, temos que o intervalo é dado por [3,4] (em anos). Isto é, espera-se que a média e a mediana reais da população estejam 95 vezes entre 100 ensaios repetidos dentro dos intervalos calculados.

Há interesse também em avaliar a idade de perda de marcha para os pacientes que tomam corticoides. Como existem apenas 17 pacientes que pararam de andar (Tabela A.8), estimamos a idade média¹³ e sua mediana¹⁴ por métodos não paramétricos. Sendo assim, os intervalos de confiança (95%) para a mediana e a média são dados por [8,11] e [8,11], respectivamente. A interpretação dos intervalos calculados é similar ao caso anterior.

6.2 Modelo de Cox

Para solucionar um dos questionamentos propostos pelas pesquisadoras, optou-se por adotar um modelo de Cox¹⁵ (Fórmula C.2) para avaliar o risco da perda de marcha dos pacientes. Para tal utilizou-se as variáveis explicativas: “Genética”,

¹¹ GEYER, J. C. **Nonparametric Tests and Confidence Intervals**. University of Minnesota. Disponível em: www.stat.umn.edu/geyer/old03/5102/notes/rank.pdf. Acesso: 25 jun. 18.

¹² CONOVER, W.J. **Practical Nonparametric Statistics**. 3ª ed. John Wiley & Sons, New York: 1999.

¹³ GEYER, J. C. Op.cit.

¹⁴ CONOVER, W.J. Op.cit.

¹⁵ Colosimo, E.A. Giolo, S.R. **Análise de Sobrevida Aplicada**. 1ª ed. Edgard Blücher: 2006

“Corticoide atual” e “Idade de início do corticoide” (anos). No total, utilizamos 50 observações para o ajuste do modelo.

Ao nível de significância de 5%, observa-se que apenas a variável idade de início do corticoide é significativa (valor-p menor que 5%). A administração de prednisona ou deflasacort pelo paciente não apresenta diferenças significantes no risco da perda de marcha. No caso de pacientes que possuem mutação genética do tipo deleção, duplicação ou outras, também não há diferenças significantes (Tabela A.18).

Reajustamos o modelo semiparamétrico de Cox somente com a variável preditora idade de início de corticoide e as estimativas estão apresentadas na Tabela A.19. Pelo modelo, podemos interpretar que, para cada ano que o paciente tardar em tomar o corticoide, seu risco para perder a marcha aumenta em 34,13%, com um intervalo de confiança (95%) de [6,7% , 68,6%].

Para o modelo final, efetuamos o diagnóstico do ajuste por meio dos resíduos de Schoenfeld¹⁶ (Figura B.26), em que se conclui que a hipótese de riscos proporcionais é adequada aos dados. Em outras palavras, os riscos de perda da marcha dos pacientes não dependem do tempo ($\lambda_0(t)$, isto é, a taxa de risco basal é igual para todos os pacientes). Vale ressaltar que dispomos de poucas observações, uma vez que o evento de interesse (perda de marcha) foi observado em um reduzido número de pacientes (9 de 50 pacientes). Apesar disso, vemos comparando resíduos de Cox-Snell com a curva de Kaplan-Meier (Figura B.27), que o modelo parece adequado.

¹⁶ KLEIN, J.P. Op.cit.

7. Conclusão

Apesar de haver poucas observações para mutações do tipo duplicação e “outras”, existe grande concordância entre os resultados encontrados neste estudo brasileiro com os do estudo de referência com pacientes chineses. A análise descritiva do presente estudo indicou que não existe relação de dependência entre o tipo de mutação genética com sintomas iniciais e deficiência mental. Também concluímos que não há evidências estatísticas (considerando $\alpha = 5\%$ como nível descritivo) para diferenças entre as médias de CPK, tempos de início de sintomas, tempo entre diagnóstico e sintomas iniciais por tipo de mutação (deleção, duplicação e “outras”).

Utilizamos a variável perda de marcha para comparar os corticoides quanto à evolução da doença nos indivíduos. Porém, como nenhum paciente que não toma corticoides parou de andar ainda, não podemos comparar a utilização de corticoides com sua não utilização. Logo, ao se comparar prednisona com deflasacort, concluímos que não existe diferença entre ambos (considerando $\alpha = 5\%$ como nível descritivo).

Propusemos o modelo semi-paramétrico de riscos proporcionais de Cox, em que a única variável explicativa estatisticamente significativa foi a idade de início de tratamento com corticoide (quanto mais tarde o paciente tomar, maior o risco de se perder a marcha). Cabe lembrar que todos os pacientes estão piorando. Desta forma, tomar corticoides mais cedo (por exemplo), não fará com que o paciente deixe de perder a marcha futuramente. Podemos apenas interpretar que espera-se que o paciente demore mais tempo para parar de andar.

Calculamos também intervalos de 95% confiança para a média bem como para a mediana do tempo necessário para se diagnosticar a doença a partir de seus sintomas iniciais, obtendo-se [3,4] e [2,4] (em anos), respectivamente.

Calculamos também os intervalos de confiança (95%) para a média e a mediana das idades de perda de marcha dos pacientes. Ambos os intervalos são dados por [8,11]. Isto é, espera-se que a média e a mediana reais da população estejam 95 vezes entre 100 ensaios repetidos dentro do intervalo calculado.

Apêndice A

Tabela A.1 Distribuição da variável Genética

Grandes mutações	Total	Proporção
Grandes deleções	51	69,86%
Grandes duplicações	6	8,22%
Pequenas mutações		
Pequenas duplicações	2	2,74%
Mutações pontuais	4	5,48%
Nonsense	2	2,74%
Missense	1	1,37%
Outras mutações		
Duplicação (sem posição)	2	2,74%
Mutação intrônica	2	2,74%
Não tem resultados		
Missing	1	1,37%
Aguarda resultado	2	2,74%
Total	73	100%

Tabela A.2 Distribuição da variável Genética - artigo de comparação

Grandes mutações	Total	Proporção
Grandes deleções	74	57,8%
Grandes duplicações	14	10,9%
Pequenas mutações		
Pequenas deleções	8	6,2%
Pequenas duplicações	10	7,8%
Splice sites	4	3,1%
Mutações pontuais		
Nonsense	17	13,3%
Missense	1	0,8%
Total	128	100%

Tabela A.3 Frequência de Tipo de mutação por Quadro inicial

Tipo de mut.\Quadro inicial	Deleção	Duplicação	Outras	Total	Proporção
Assintomático	1	1	0	2	2,90%
Atraso na aquisição da marcha	20	6	2	28	40,58%
Alteração no padrão da marcha previamente já adquirida e dentro da normalidade	7	1	1	9	13,04%
Perda de habilidades motoras previamente adquiridas	18	2	6	26	37,68%
Dor em mmii	4	0	0	4	5,80%
Total	50	10	9	69	
Proporção	72,47%	14,49%	13,04%		100%

Tabela A.4 Frequência de Tipo de mutação por Deficiência mental

Tipo de mutação\Deficiência mental	Não	Sim	Total	Proporção
Deleção	36	12	48	72%
Duplicação	4	6	10	15%
Outros	6	3	9	13%
Total	46	21	67	
Proporção	69%	31%		100%

Tabela A.5 Medidas resumo de CPK por Tipo de mutação

Tipo de mutação	n	Mínimo	Mediana	Média	Máximo	Desvio Padrão
Deleção	47	668	10982	12906	44606	8012,428
Duplicação	10	3696	8715	9922	19756	5158,854
Outros	8	2656	14931	12679	21000	6256,43
Total	65					

Tabela A.6 Frequência do Tipo de mutação por Tempo de referência

Tempo de referência\Tipo de mutação	Deleção	Duplicação	Outras	Total	Proporção
Maior que 2 anos	28	6	4	38	56%
Menor ou igual a 2 anos	22	3	5	30	44%
Total	50	9	9	68	
Proporção	74%	13%	13%		100%

Tabela A.7 Medidas resumo da Idade atual

	n	Mínimo	Máximo	Primeiro Quantil	Mediana	Terceiro quantil	Média
Idade atual (anos)	73	3	18	7	9	11	9,01

Tabela A.8 ¹⁷ Distribuição do Uso de Corticoides por Perda da marcha

Corticoide atual \ Perda da marcha	Não	Sim	Total	Proporção
Não usa corticoide	12	0	12	16,44%
Deflasacort	26	12	38	52,05%
Prednisona	18	5	23	31,51%
Total	56	17	73	
Proporção	76,71%	23,29%		100%

Tabela A.9 Medidas resumo do Tempo entre início de corticoide e perda da marcha

	n	Mínimo	Máximo	Primeiro Quantil	Mediana	Terceiro quantil	Média
Tempo geral (anos)	11	-1	7	0	2	2	1,64
Deflasacort (anos)	7	0	4	0,5	2	2	1,57
Prednisona(anos)	4	-1	7	-1	0,5	3,25	1,75

Tabela A.10 Medidas resumo da Idade do início de corticoides

¹⁷ Prednisona: um paciente não tem idade de início do corticoide e tempo de uso e 2 pacientes têm a perda de marcha antes de tomar corticoide

	n	Mínimo	Máximo	Primeiro Quantil	Mediana	Terceiro quantil	Média	NA's
Idade do início de corticoides(anos)	60	1	16	6	8	9	7,42	1
Idade do início de deflasacort	38	2	16	7	8	9	8,29	0
Idade do início de prednisona	22	1	14	4	5,5	8	5,91	1

Tabela A.11 Frequência da variável Histórico familiar por Tempo de referência

Tempo de referência\Histórico familiar	Sem histórico	Com histórico	Total	Proporção
Maior que 2 anos	29	7	36	55%
Menor ou igual a 2 anos	21	8	29	45%
Total	50	15	65	
Proporção	77%	23%		100%

Tabela A.12 Medidas resumo da Idade de diagnóstico

	n	Mínimo	Máximo	Primeiro Quartil	Mediana	Terceiro quartil	Média	NA's
Idade de diagnóstico (anos)	70	1	15	5	7	9	7,05	3

Tabela A.13 Medidas resumo da Idade de sintomas

	n	Mínimo	Máximo	Primeiro Quartil	Mediana	Terceiro quartil	Média	NA's
Idade de sintomas (anos)	72	0	9	2	3,5	5	3,58	1

Tabela A.14 Medidas resumo de diferença entre idade de diagnóstico e idade de sintomas

	n	Mínimo	Máximo	Primeiro Quartil	Mediana	Terceiro quartil	Média	NA's
Diferença das idades (anos)	70	-5	11	2	3	5	3,44	3

Tabela A.15 Distribuição de Frequência de Variável Fisioterapia

Fisioterapia (fre. semanal)	0	1	2	3	Total
Número de pacientes	14	37	20	1	72
Proporção	19,44%	51,39%	27,78%	1,39%	100%

Tabela A.16 Distribuição de Frequência de variável Hidroterapia

Hidroterapia (freq. semanal)	0	1	total
Número de pacientes	49	23	72
Proporção	68,06%	31,94%	100%

Tabela A.17 Distribuição de frequências da variável Cardioprotetor

Cardioprotetor	não usa	usa enalapril	usa enalapril e carvedilol	Total
Número de pacientes	45	25	3	73
Proporção	61,64%	34,24%	4,12%	100%

Tabela A.18 Coeficientes do modelo de Cox

	Coeficiente	exp(coeficiente)	D.P	Valor-p
Corticoide atual (Prednisona)	-0,677	0, 508	1,140	0,552
Idade de início do corticoide	0, 260	1,296	0,117	0,027
Génética (Duplicação)	-19,100	<0,001	11020	0,999
Génética (Outras)	-0, 545	0, 580	1,105	0,622

Tabela A.19 Coeficientes do modelo final de Cox

	Coeficiente	exp(coeficiente)	D.P	Valor- p	Lim.Inf 95%	Lim.Sup 95%
Idade de início do corticoide	0,2936	1,3413	0,1168	<0,02	1,067	1,686

Apêndice B

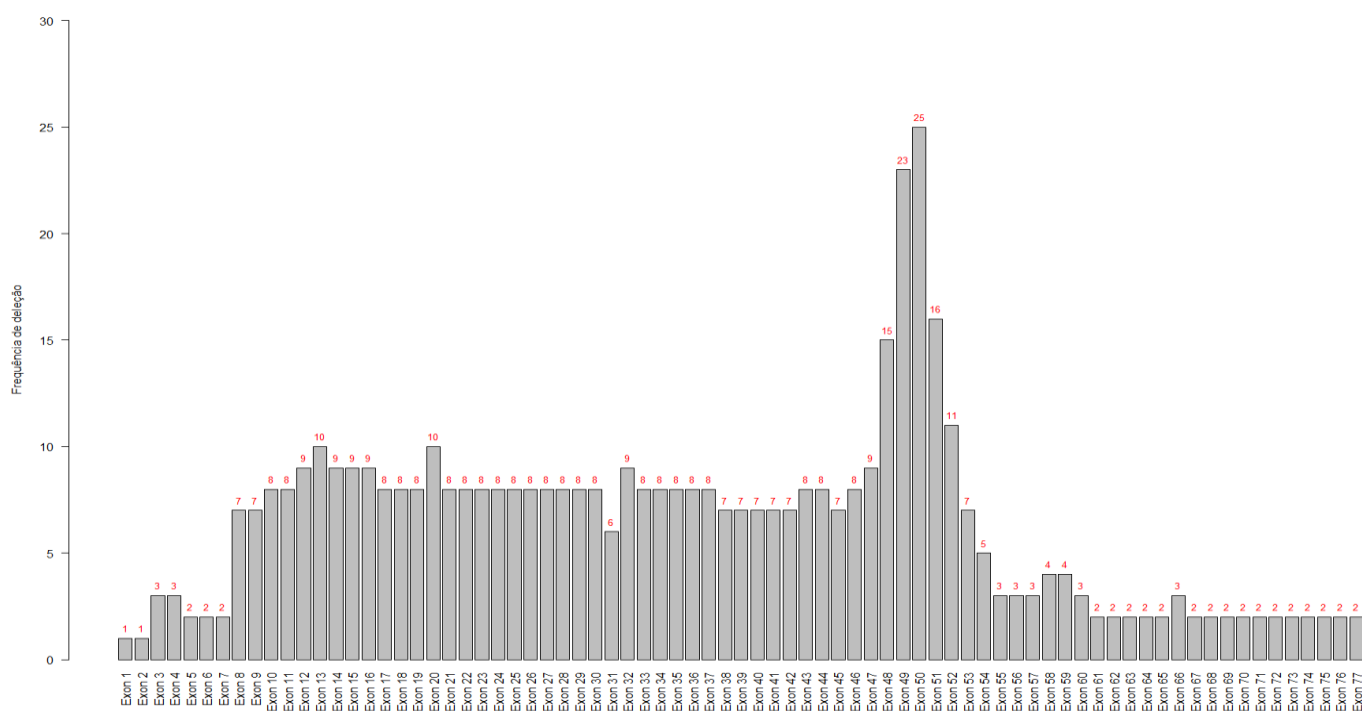


Figura B.1 Mutações do tipo deleção por exon dos pacientes observados

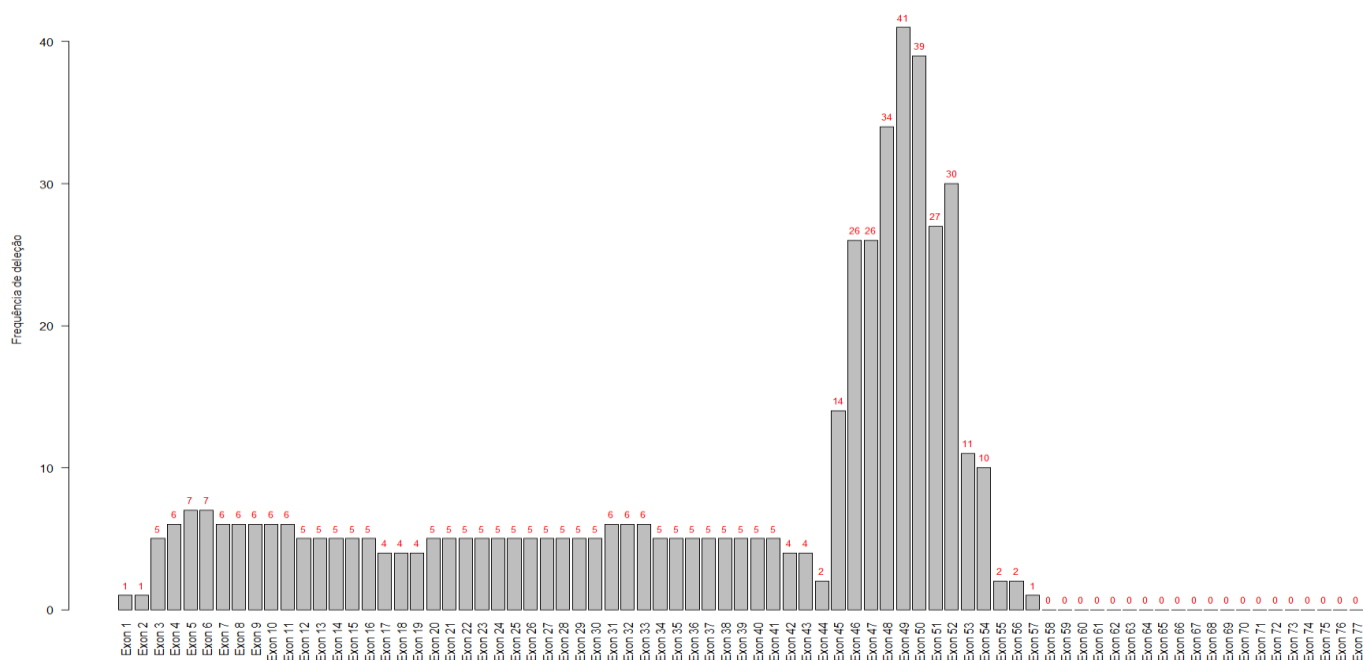


Figura B.2 Mutações do tipo deleção por exon do estudo de comparação

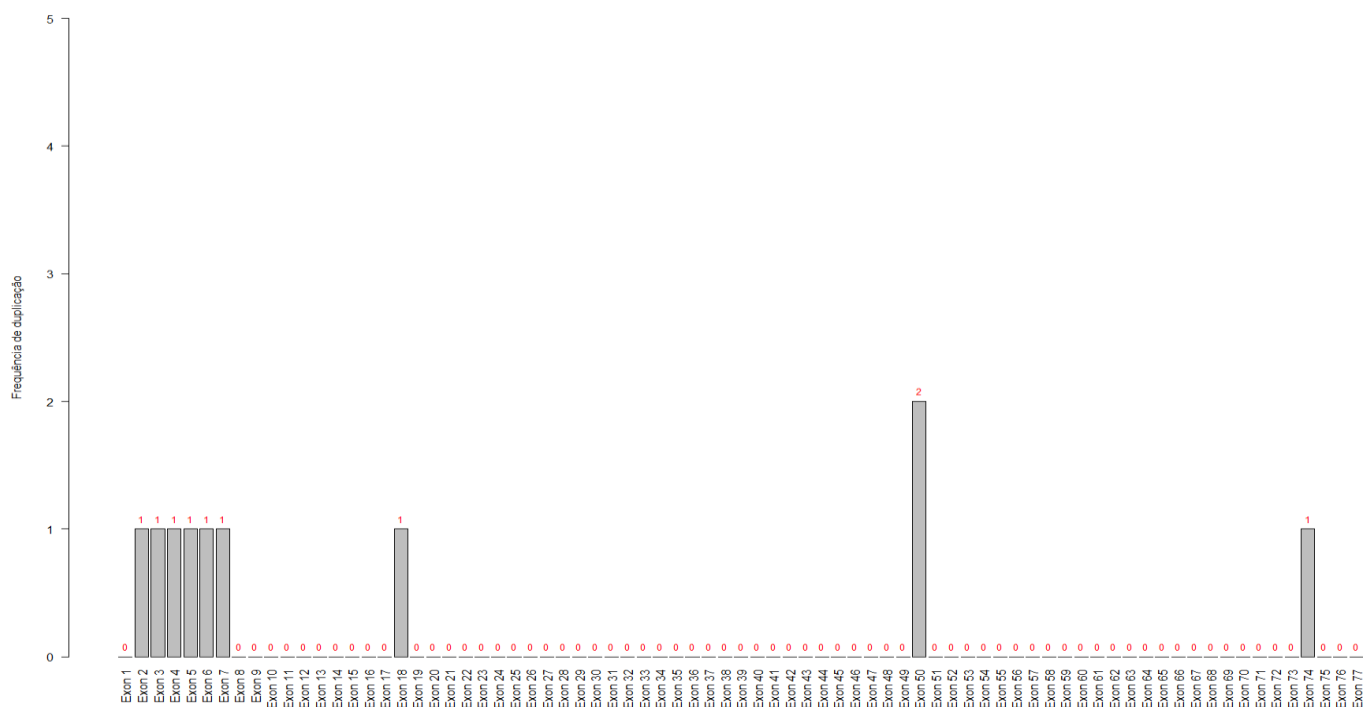


Figura B.3 Mutações do tipo duplicação por exon dos pacientes observados

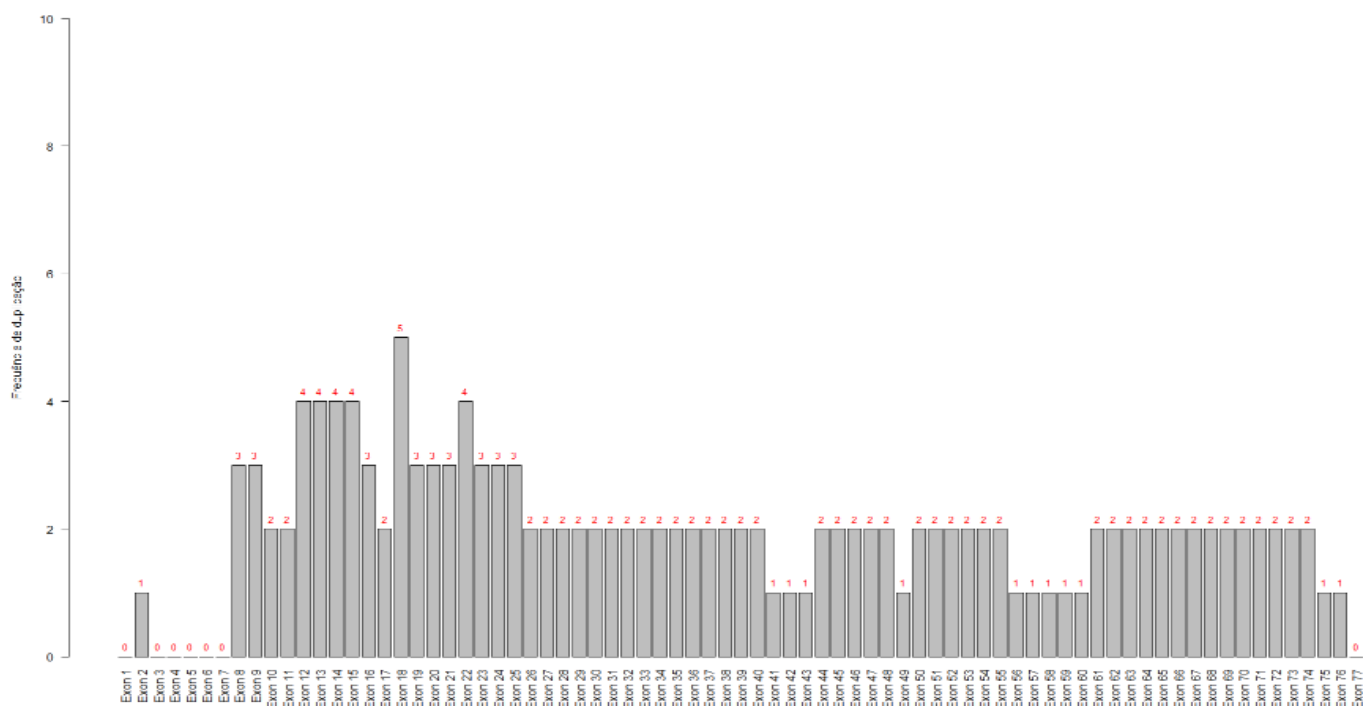


Figura B.4 Mutações do tipo duplicação por exon do estudo de comparação

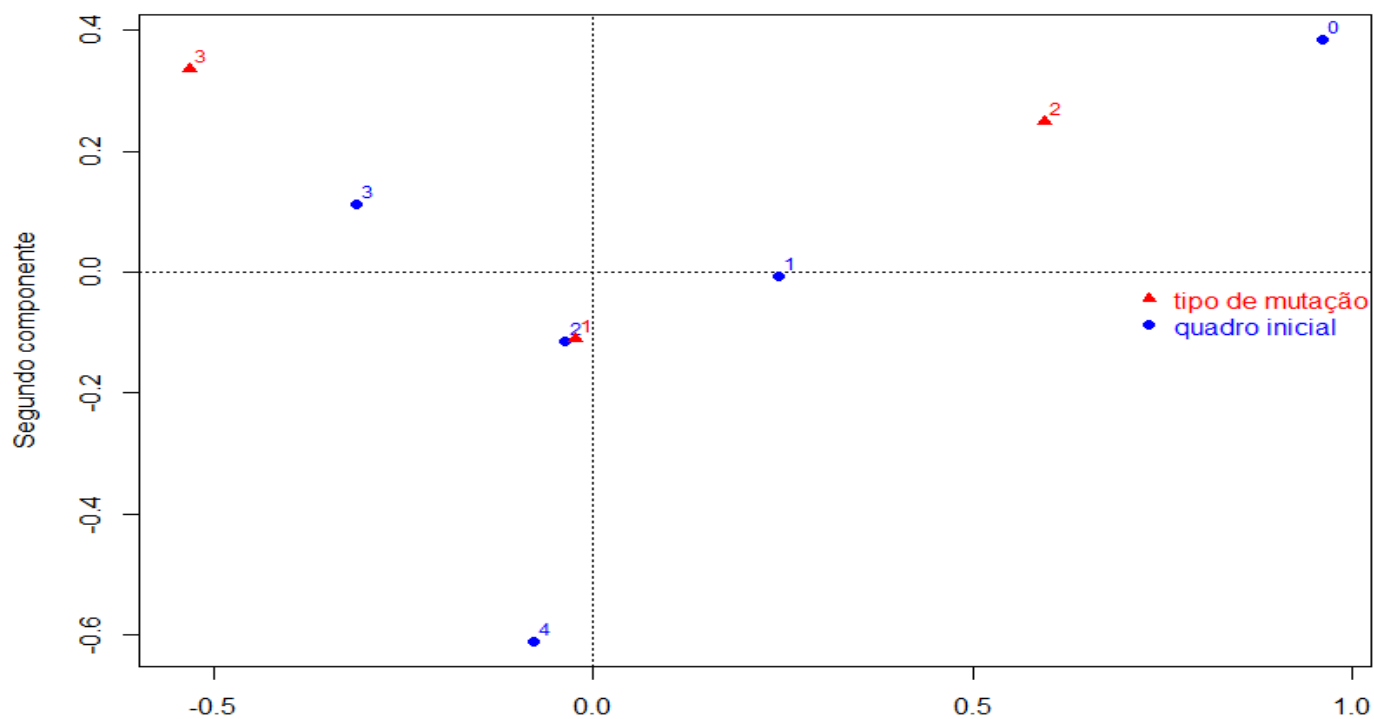


Figura B.5 Análise de correspondência entre quadro inicial e tipo de mutação¹

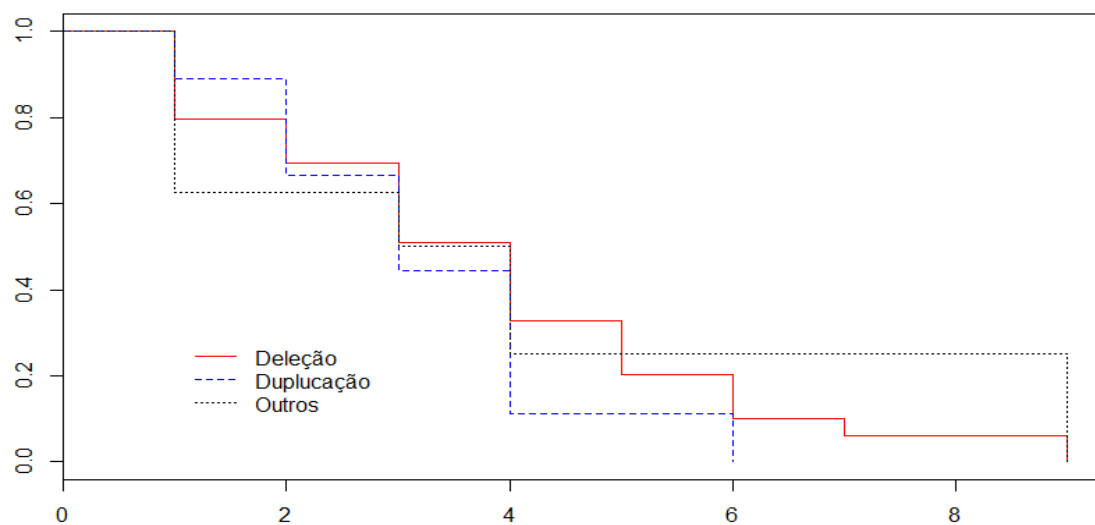


Figura B.6 Kaplan-Meier para a idade de início de sintomas dos pacientes por tipo de mutação

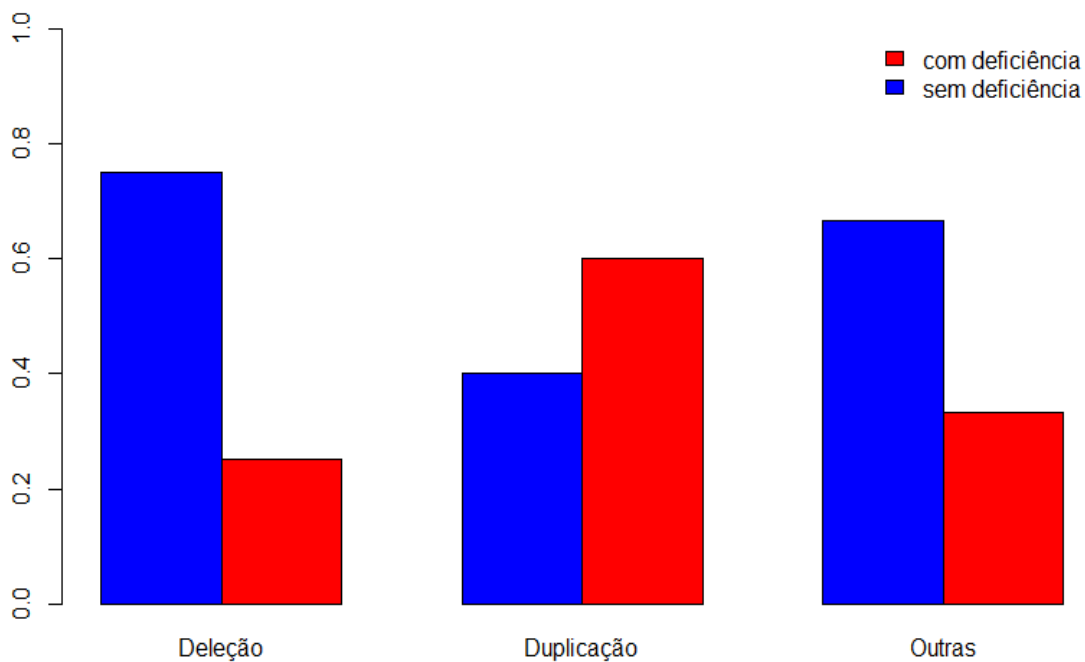


Figura B.7 Proporção de pacientes com algum tipo de deficiência mental por tipo de mutação

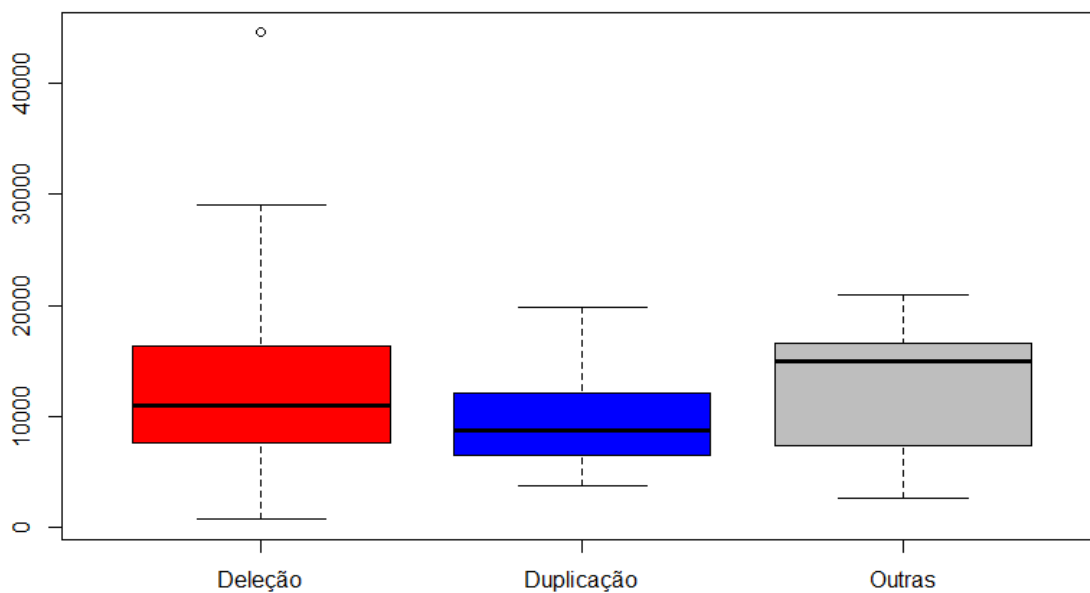


Figura B.8 Boxplots de concentração de CPK por tipo de mutação

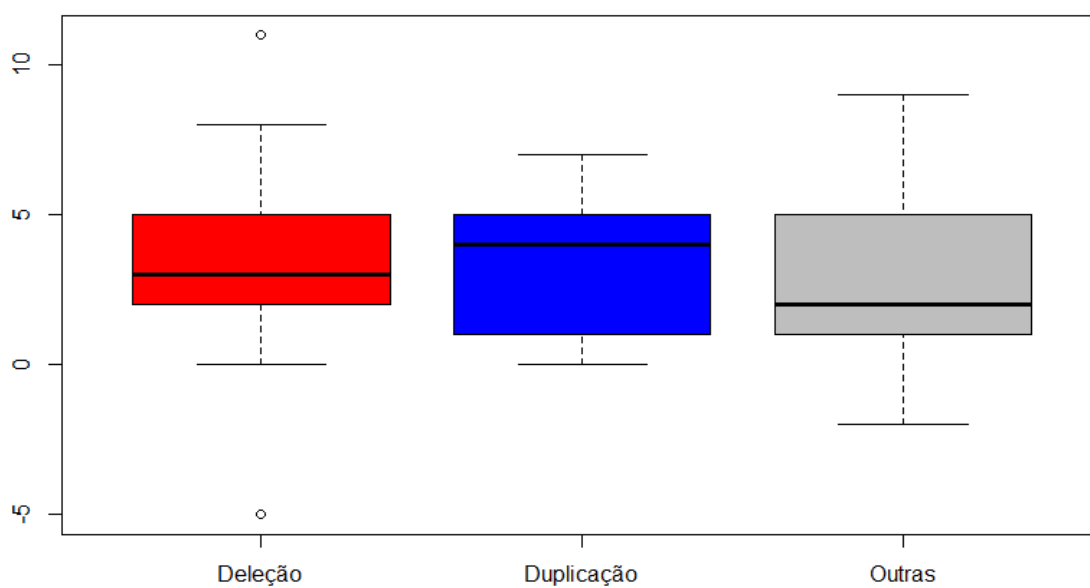


Figura B.9 Boxplots de tempo entre diagnóstico e sintomas iniciais por tipo de mutação

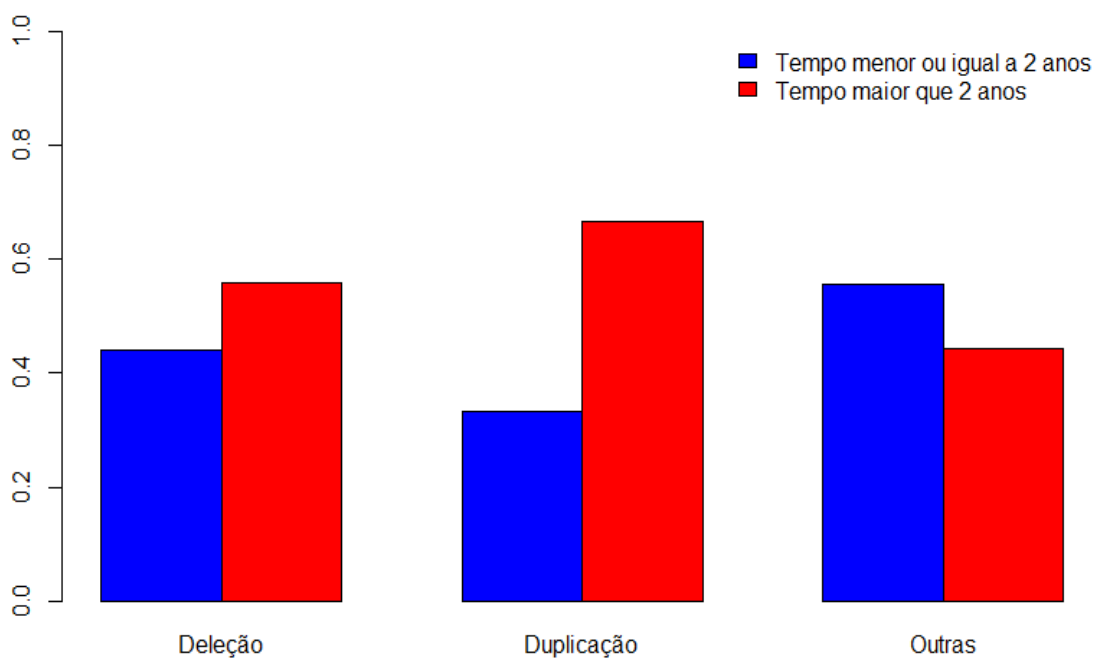


Figura B.10 Proporção de tempo de demora entre diagnóstico e início de sintomas por tipo de mutação

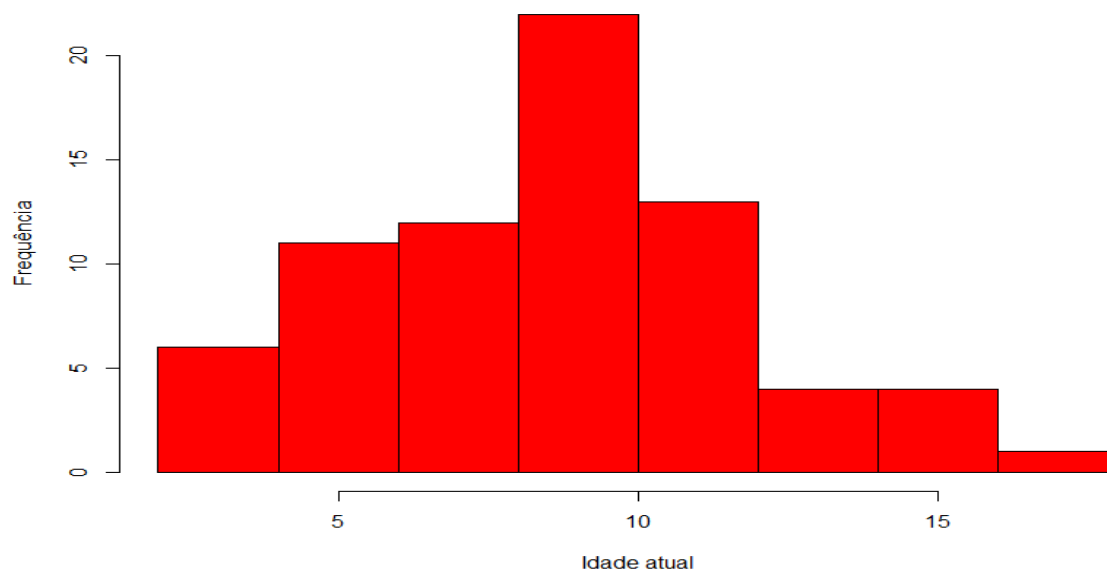


Figura B.11 Histograma de Frequência de idade atual dos pacientes

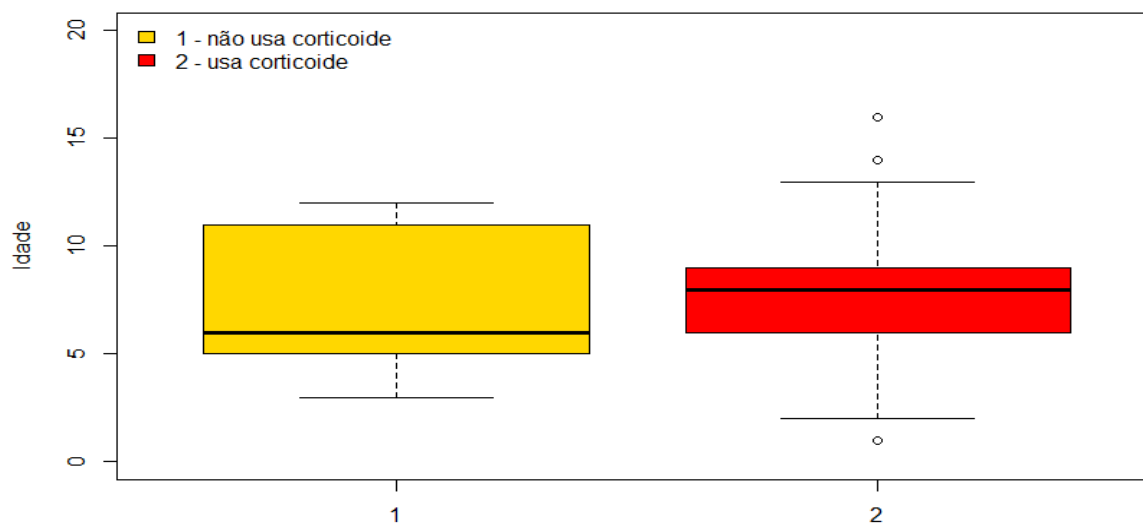


Figura B.12 Box-plot de comparação de idade atual com idade inicial de corticoide

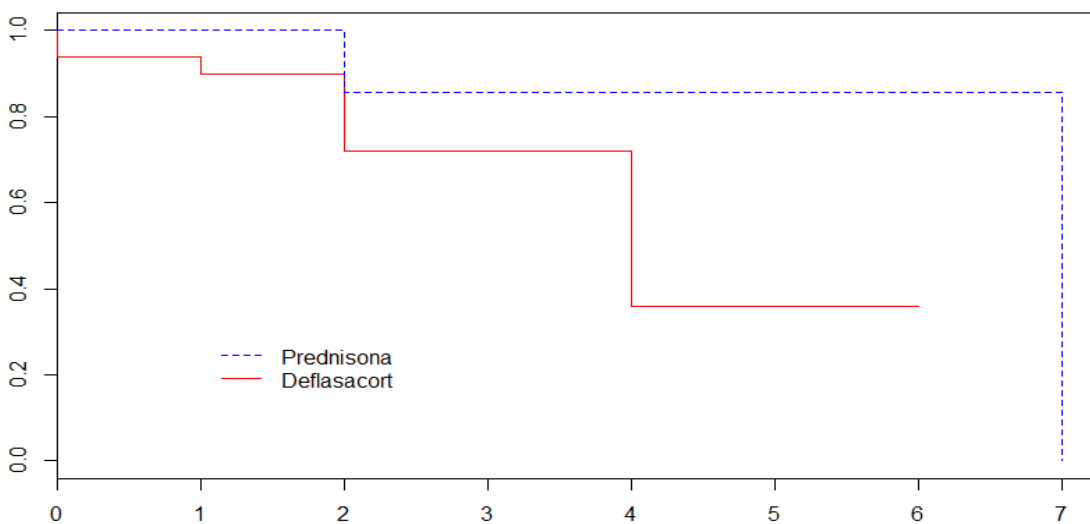


Figura B.13 Kaplan-Meier para o intervalo de tempo entre perda de marcha e início de corticóide

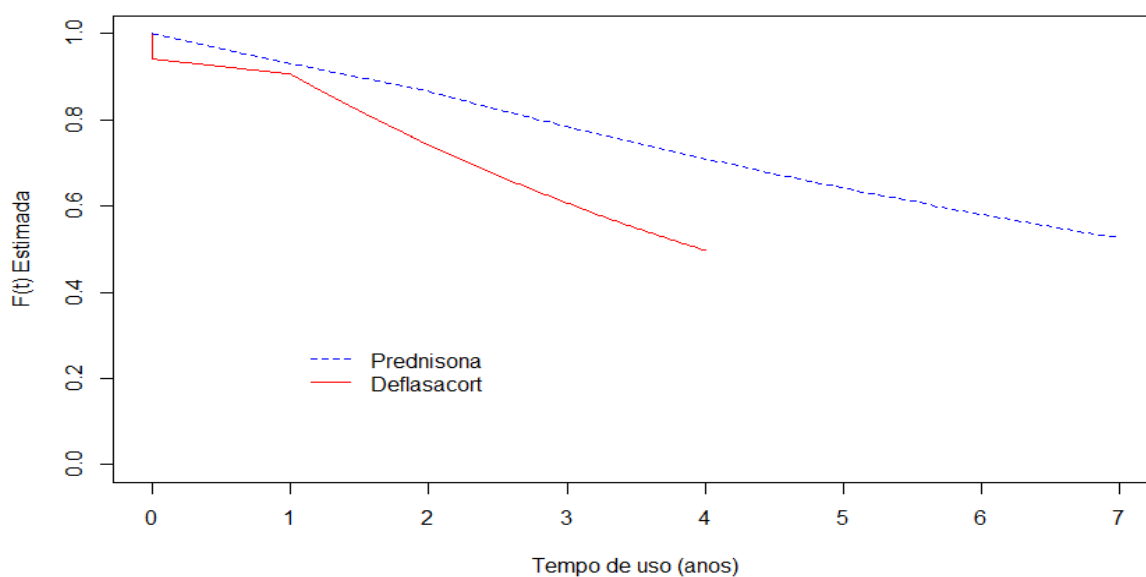


Figura B.14 Piecewise exponencial estimator – PEXE para o intervalo de tempo entre perda de marcha e início de corticóide

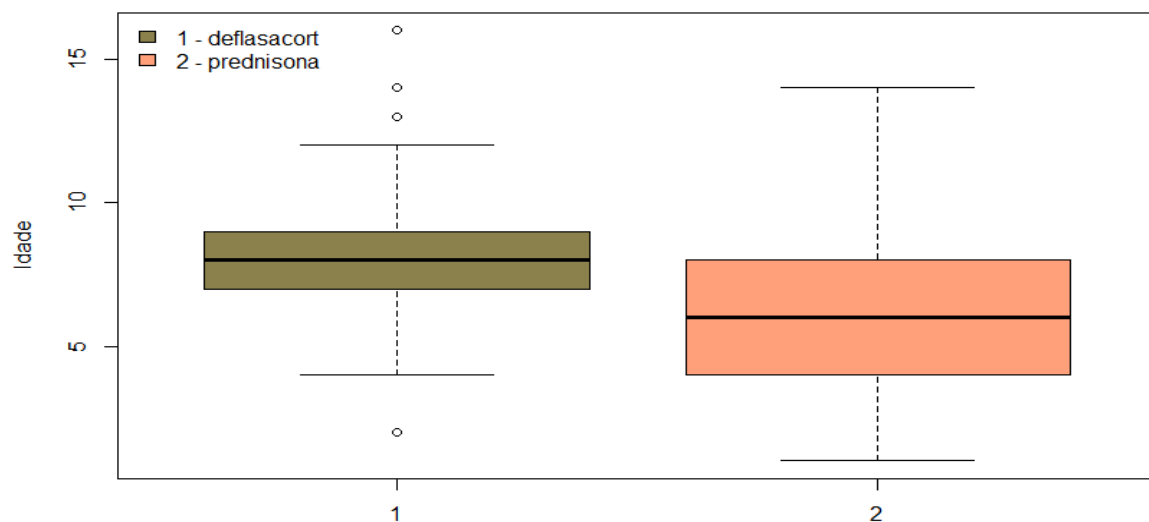


Figura B.15 Box-plot de comparação de idade inicial de corticoides

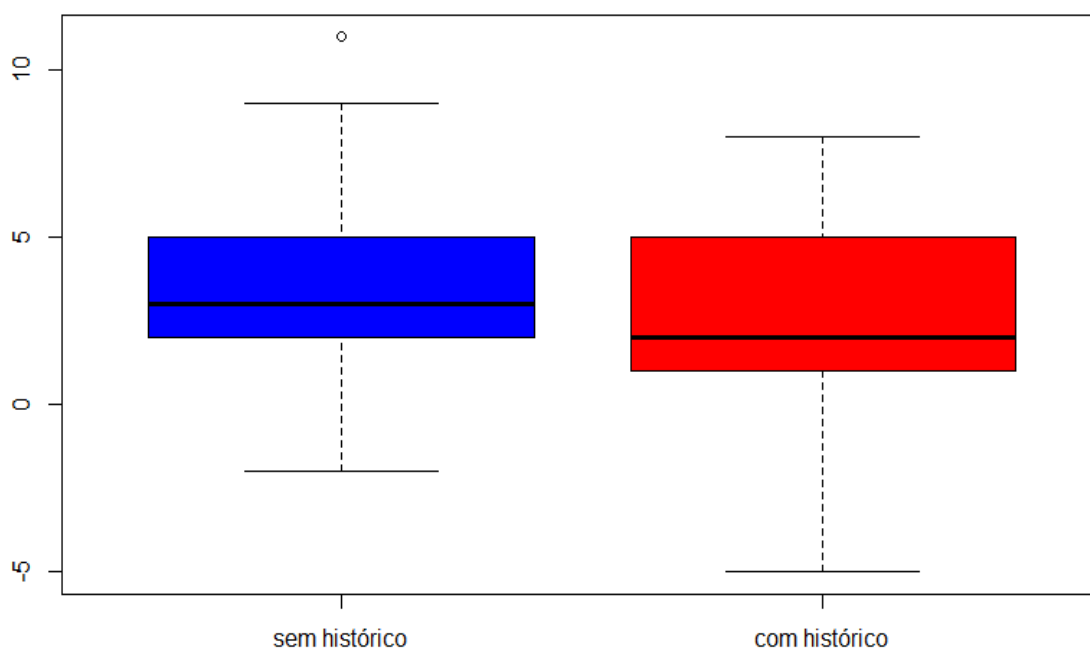


Figura B.16 Boxplots de distância de tempo entre diagnóstico e início dos sintomas por histórico familiar

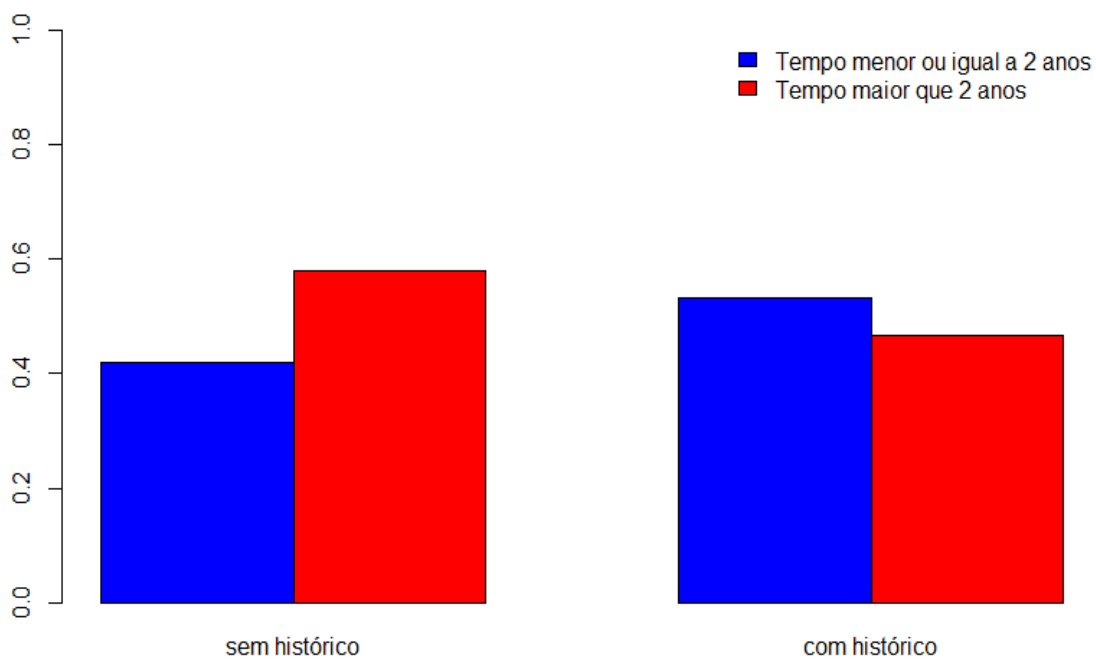


Figura B.17 Proporção de tempo de demora entre diagnóstico e início de sintomas por histórico familiar

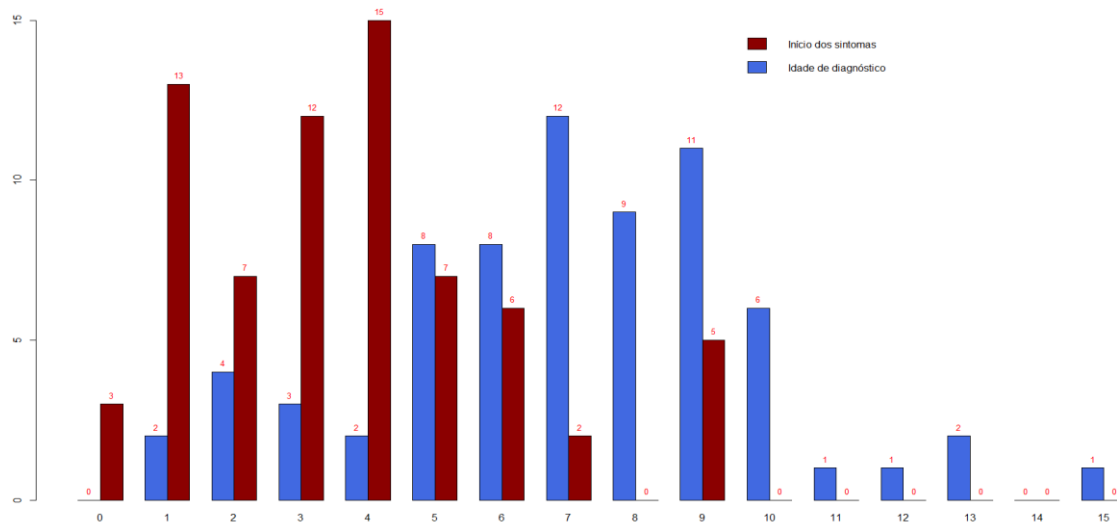


Figura B.18 Gráfico de barra de idades de início de sintomas e de diagnóstico

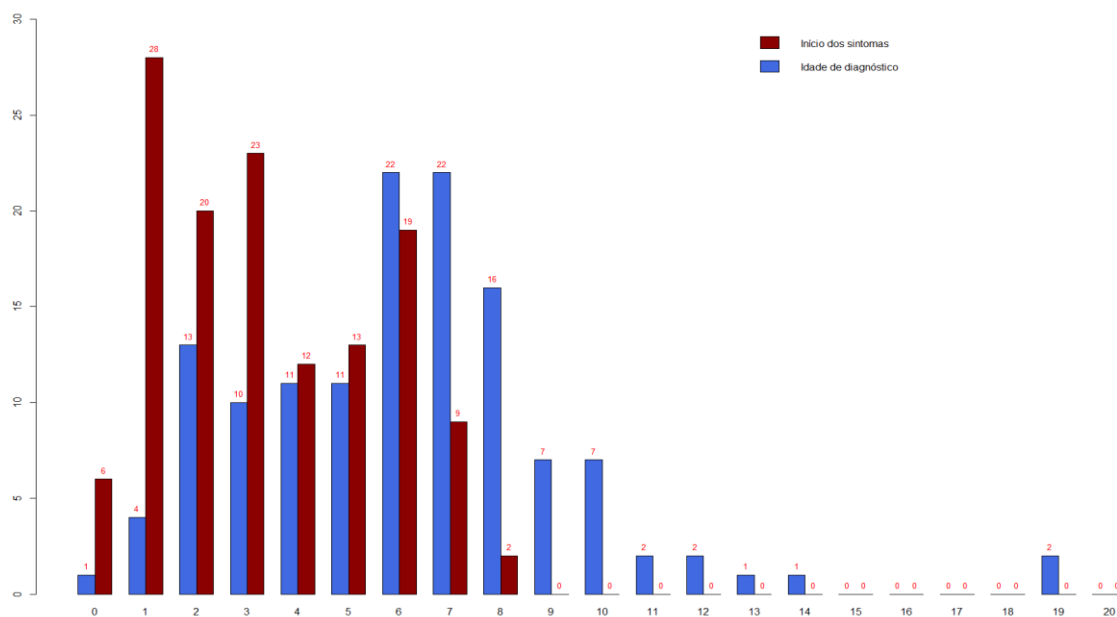


Figura B.19 Gráfico de barra de idades de início de sintomas e de diagnóstico do estudo de comparação

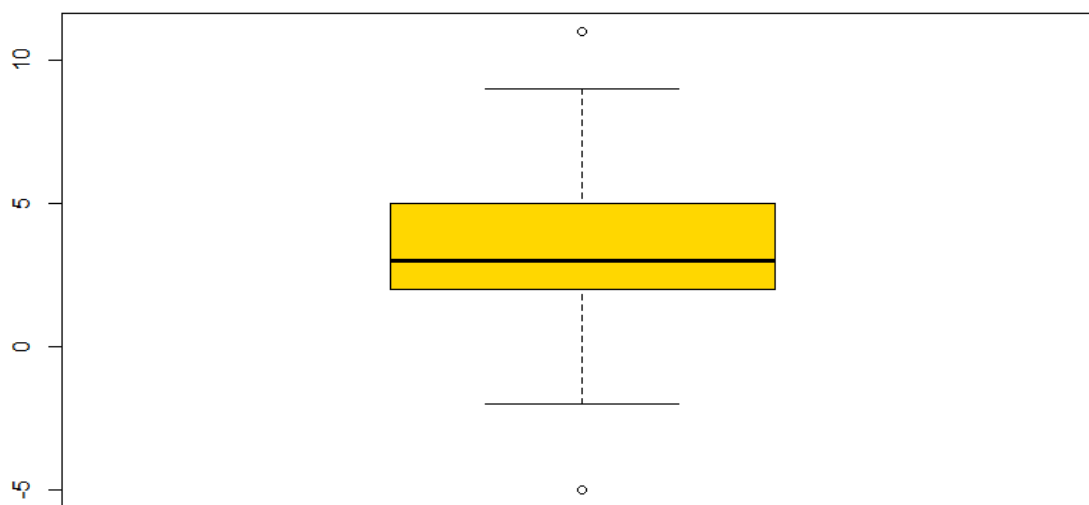


Figura B.20 Boxplot da diferença entre a idade de diagnóstico e a idade de início dos sintomas

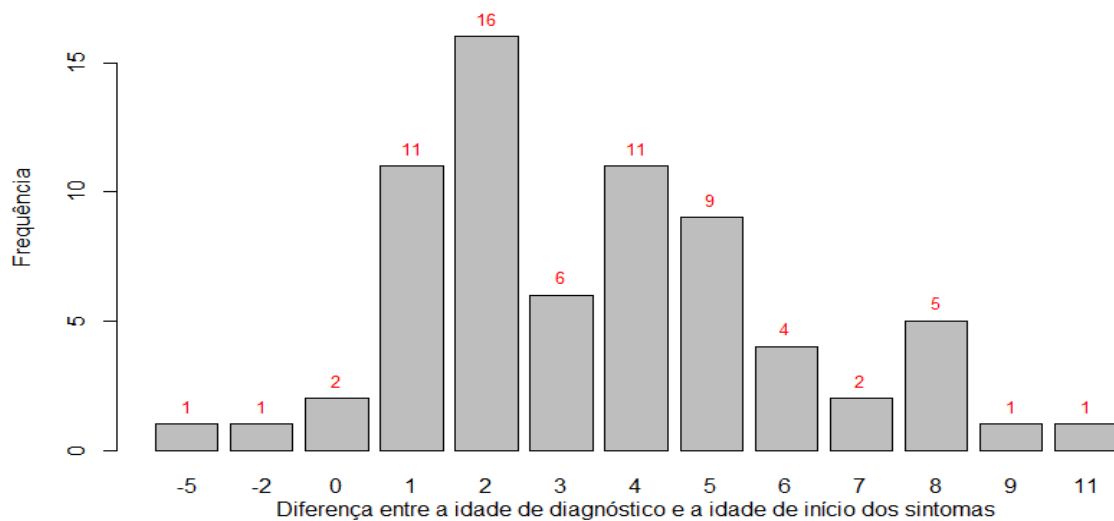


Figura B.21 Frequência da diferença entre a idade de diagnóstico e a idade de início dos sintomas

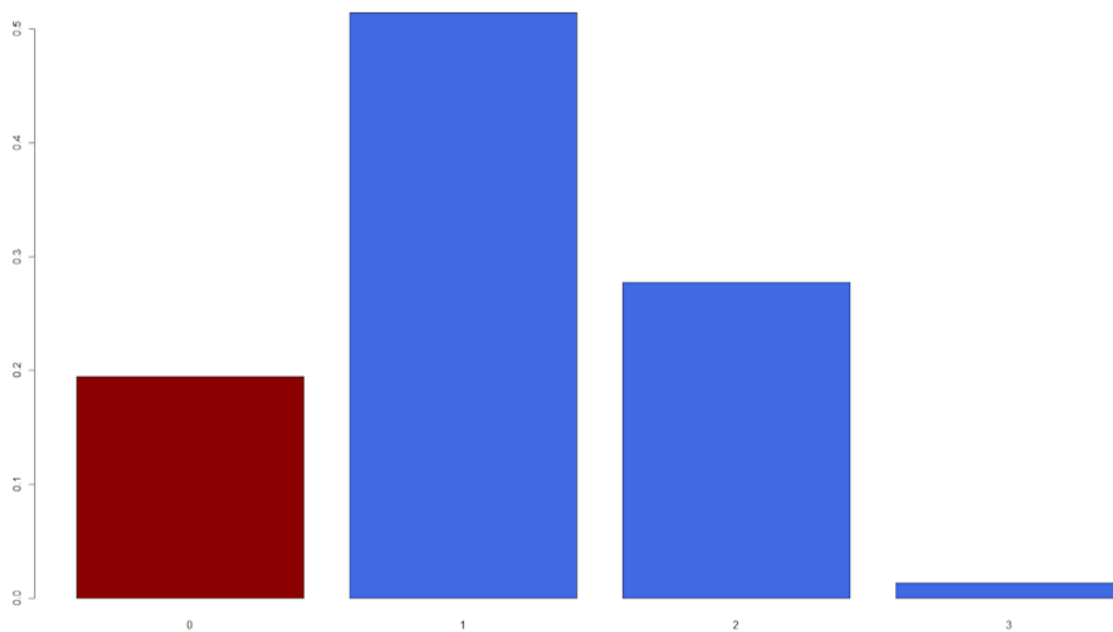


Figura B.22 Proporção de quantidade de vezes que os pacientes fazem fisioterapia por semana

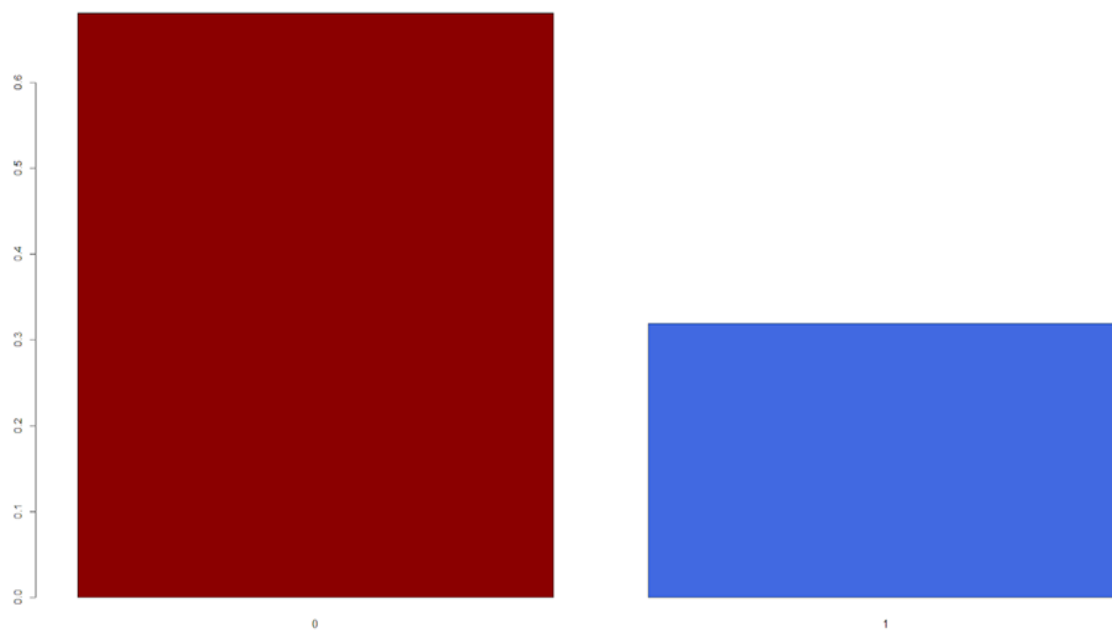


Figura B.23 Proporção de quantidade de vezes que os pacientes fazem hidroterapia por semana

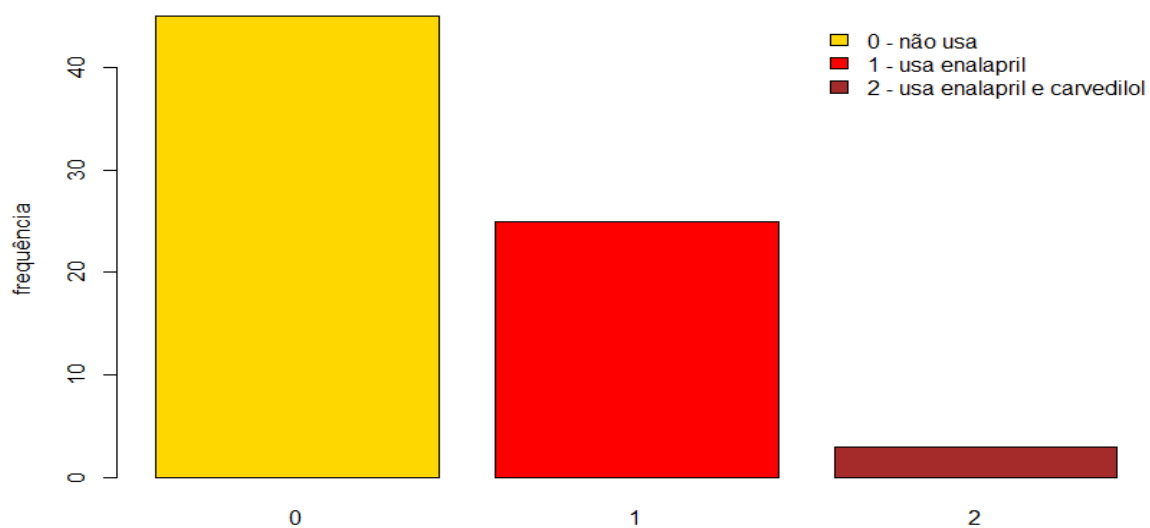


Figura B.24 Frequência de quantidade de pacientes uso de cardioprotetor

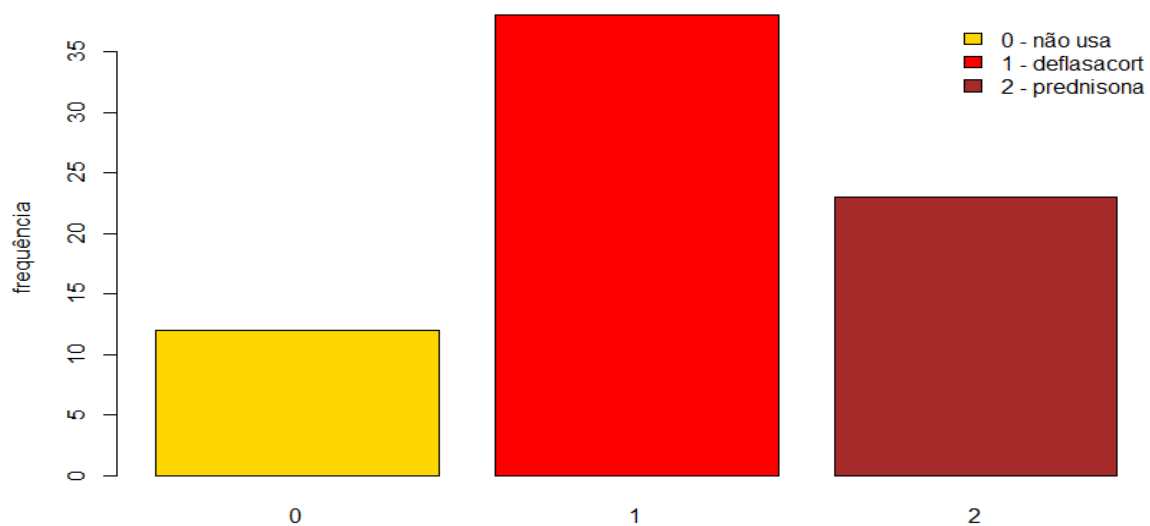


Figura B.25 Frequência de quantidade de pacientes uso de Corticoide atual

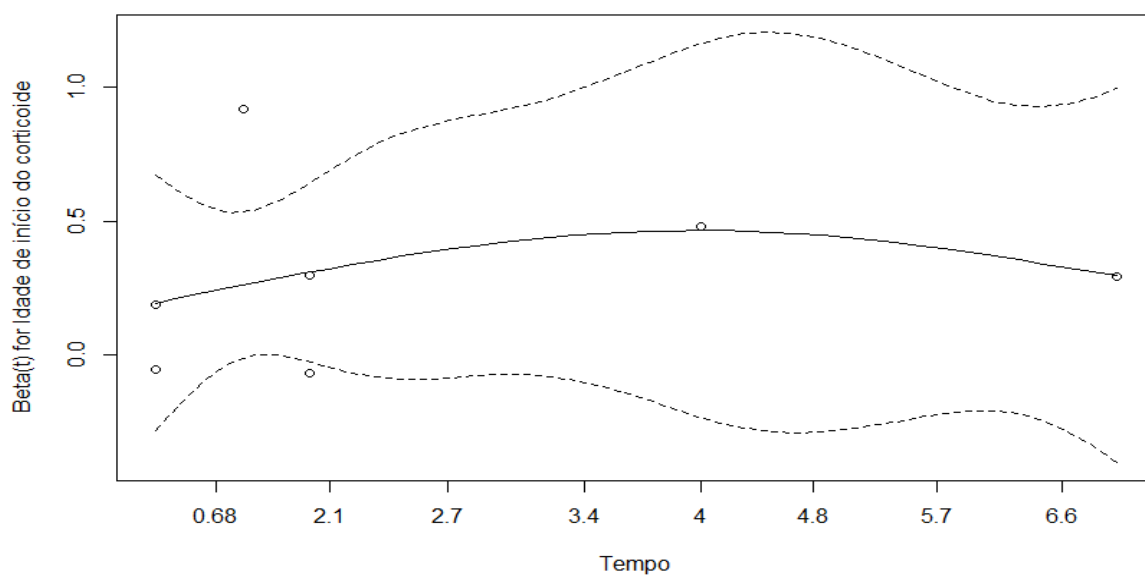


Figura B.26 Resíduos de Schoenfeld

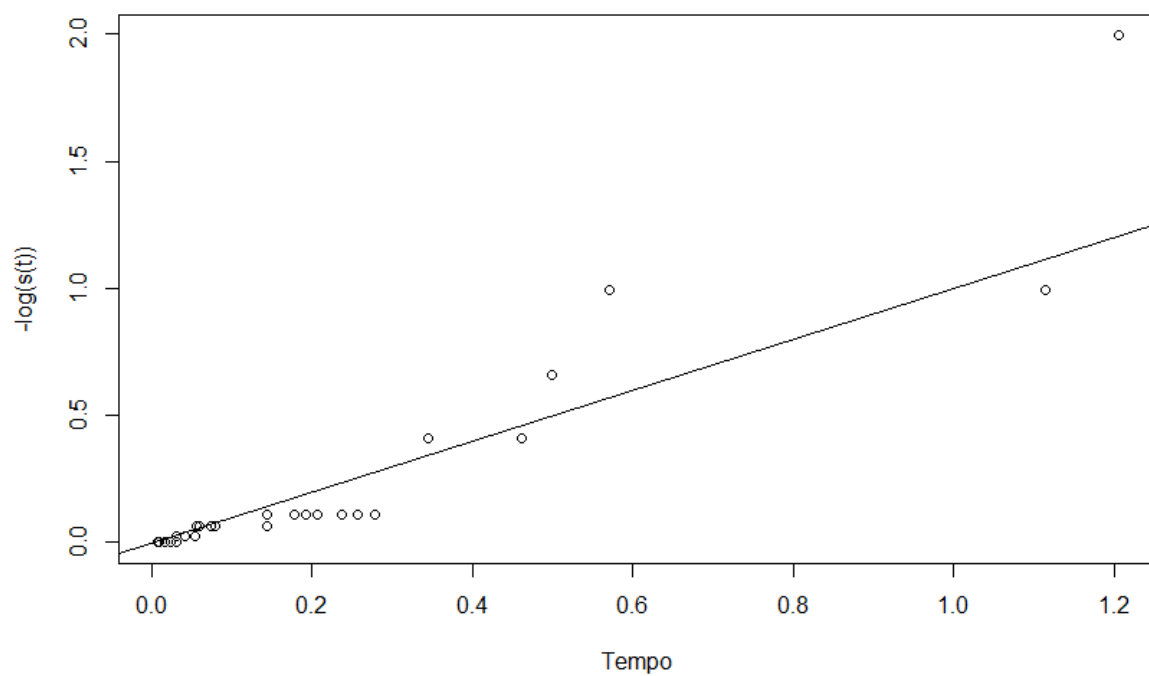


Figura B.27 Resíduos de Cox-Snell Comparação com Kaplan-Meier

Apêndice C

Fórmula C.1 Estimativa de Piexewise Exponential

1. Deflasacort¹⁸

$$\widehat{S(t)} = \begin{cases} e^{-60,60t}, & \text{se } 0 \leq t < 0,001 \\ e^{-0,0606-0,0376*(t-0,001)}, & \text{se } 0,001 \leq t < 1 \\ e^{-0,0982-0,2*(t-1)}, & \text{se } 1 \leq t < 2 \\ e^{-0,2982-0,2*(t-2)}, & \text{se } 2 \leq t < 4 \\ \text{não podemos estimar} & \text{se } t \geq 4 \end{cases}$$

2. Prednisona

$$\widehat{S(t)} = \begin{cases} e^{-0,0714t}, & \text{se } 0 \leq t < 2 \\ e^{-0,1428-0,1*(t-2)}, & \text{se } 2 \leq t < 7 \\ \text{não podemos estimar} & \text{se } t \geq 7 \end{cases}$$

¹⁸ Ajustamos os valores da variável “tempo de uso” de 0 para 0,001

Fórmula C.2 Modelo de Cox

O modelo de Cox é um modelo semi-paramétrico¹⁹, cuja função de risco possui a seguinte forma:

$$\lambda(t) = \lambda_0(t) \exp \{X' \beta\}$$

Em que,

X é um vetor com as covariáveis $X = (x_1, \dots, x_p)$ e β é o vetor de parâmetros associado a X , $\lambda(t) = \lambda_0(t)$ quando $X = 0$.

Suposição do modelo:

A razão de risco não depende do tempo, isto é,

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t) \exp \{x_i' \beta\}}{\lambda_0(t) \exp \{x_j' \beta\}} = \exp \{x_i' \beta - x_j' \beta\}$$

Considerando:

$$\lambda(t) = \lambda_0(t) \exp \{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4\}$$

em que,

$\lambda(t)$: variável resposta (risco)

$\lambda_0(t)$: uma função não-negativa do tempo

x_1 : $\begin{cases} 1, \text{ se o paciente usa prednisona} \\ 0, \text{ se o paciente usa deflasacort} \end{cases}$

x_2 : Idade de início do corticoide

x_3 : $\begin{cases} 1, \text{ se o paciente carrega mutação do tipo duplicação} \\ 0, \text{ caso contrário} \end{cases}$

¹⁹ KLEIN, J.P. Op.cit.

$$x_4: \begin{cases} 1, \text{ se o paciente } \textit{carrega mutação do tipo "outras"} \\ 0, \text{ caso contrário} \end{cases}$$

β_1 : o log da variação esperada no risco de se perder a marcha atribuído à utilização de prednisona em relação ao deflasacort;

β_2 : o log da variação esperada no risco de se perder a marcha atribuído à variação da idade de início de corticoide;

β_3 : o log da variação esperada no risco de se perder a marcha atribuído aos indivíduos portadores da mutação duplicação em relação às demais mutações;

β_4 : o log da variação esperada no risco de se perder a marcha atribuído aos indivíduos portadores do grupo de mutações “outras” em relação às demais mutações.