**Gráficos de controle CUSUM para
monitoramento de dados de sobrevivência**

Jocelânio Wesley de Oliveira

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM ESTATÍSTICA

Programa: Pós-graduação em Estatística
Orientador: Prof. Dr. Antonio Carlos Pedroso de Lima

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES/CNPq

São Paulo, julho de 2018

Gráficos de controle CUSUM para monitoramento de dados de sobrevivência

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 18/05/2018. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Antonio Carlos Pedroso de Lima (Orientador) - IME-USP
- Prof. Dr^a. Linda Lee Ho - EP-USP
- Prof. Dr^a. Dione Maria Valença - UFRN
- Prof. Dr. Roberto da Costa Quinino - UFMG
- Prof. Dr^a. Airlane Pereira Alencar - IME-USP

Agradecimentos

Ao meu *orientador-pai* Antonio Carlos, que despendeu o máximo de atenção para incentivar o meu desenvolvimento neste trabalho e também em aspectos pessoais, e por isso destaco o substantivo composto que escolhi para caracterizá-lo.

Aos professores do departamento de Estatística do IME pelos conhecimentos fornecidos, em especial a Gisela e Lane que me apoiaram também em questões pessoais.

Aos professores da comissão julgadora pelas várias considerações que levaram a um texto melhor.

Aos muitos colegas estudantes do IME que me auxiliaram em várias disciplinas difíceis.

Aos funcionários das secretarias do IME que me atenderam paciente e eficientemente em diversos assuntos burocráticos, em especial a Regiane que me ajudou muito principalmente na fase de conclusão do curso.

Aos meus familiares e amigos de longa data do Rio Grande do Norte, que só podia reencontrar nos curtos períodos de férias, mas que sempre me foram acolhedores. Em especial a Vanessa e Fabrício, por me receber generosamente em sua casa para férias em Natal, e a Kalogênia e Natanael, que moram em São Paulo e me ajudaram muito aqui em período de dificuldades.

Aos meus amigos conhecidos em São Paulo, os quais me apresentaram e fizeram aproveitar melhor a cidade. Em especial ao grupo de amigos que mantém contato frequente pelo aplicativo Telegram, com destaque também para Carlos Goering, que com sua experiência em programação me sugeriu dicas importantes para as simulações computacionais realizadas no estudo.

Aos guias espirituais que possam ter me inspirado serenidade e sapiência quando foi necessário para enfrentar os diversos desafios.

Às agências CAPES/CNPq pelo apoio financeiro.

Resumo

OLIVEIRA, J. W. **Gráficos de controle CUSUM para monitoramento de dados de sobrevivência**. 2018. 124 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2018.

Neste trabalho propomos gráficos de controle tipo CUSUM para monitoramento de tempos de sobrevivência. Nossa proposta é desenvolver diferentes estatísticas para o escore do gráfico CUSUM de forma prospectiva. Inicialmente propomos um gráfico CUSUM não paramétrico para monitoramento de populações homogêneas que avalia a variação na estatística log-rank como forma de identificar se há uma mudança significativa no risco de falha ao longo do tempo. Algumas abordagens diferentes foram consideradas e em destaque colocamos o gráfico ZDiff CUSUM, que tem como escore o incremento na estatística Z do teste log-rank em relação à inspeção anterior. Foi constatado, via simulação, que este método é eficiente. Posteriormente investigamos abordagens que levam em conta heterogeneidade na população por meio do modelo de Cox, considerando medidas baseadas na razão de verossimilhanças e em resíduos martingal e deviance. Através de simulações, verificou-se que o método com base na razão de verossimilhanças se mostrou ágil para detectar alteração na taxa de falha, quando se conhece a intensidade da mudança e este valor é informado na construção do teste. Por outro lado, os gráficos CUSUM com base em resíduos são mais simples e se mostraram eficazes para identificar aumentos no padrão da sobrevivência. Estes três métodos e o ZDiff CUSUM foram aplicados a dados de um estudo conduzido no Instituto do Coração (InCor) envolvendo pacientes com insuficiência cardíaca. Foi detectado que ao longo do tempo estes pacientes apresentam sobrevida maior, o que pode estar ligado à melhoria no tratamento e procedimentos realizados no hospital. Como conclusão, sugerimos que os gráficos tipo CUSUM com resíduos do modelo de Cox e o método não paramétrico com teste log-rank podem ser alternativas para utilização na prática em monitoramento de dados de sobrevivência.

Palavras-chave: Análise de sobrevivência, gráfico CUSUM, teste log-rank, modelo de Cox.

Abstract

OLIVEIRA, J. W. **CUSUM control charts to monitor survival data**. 2018. 124 f. Tese (Doutorado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2018.

In this work we propose CUSUM control charts to monitor survival times. Our proposal is to develop different statistics for the CUSUM chart score in a prospective way, to take into account SA approaches. We initially consider a non-parametric approach to monitor homogeneous populations. This CUSUM evaluates the variation on the log-rank test statistics as a way to identify significant changes in the risk of failure. Some different expressions for this have been considered and, in particular, we propose a ZDiff CUSUM chart computed as the increment on the log-rank test statistics Z at each inspection point in relation to the previous one. Based on simulation studies it was found that this method is efficient. Subsequently we investigated approaches that take into account heterogeneity in the population through the Cox model, considering measures based on the likelihood ratio and on martingal and deviance residuals. Through simulations, it was verified that the method based on the likelihood ratio was agile to detect a change in the hazard rate, when the intensity of the change is known and this value is informed in the construction of the test. On the other hand, CUSUM methods based on residuals are simpler and have been shown to be effective in identifying increases in survival pattern. These three methods and the ZDiff CUSUM were applied to a dataset from a study conducted at the Heart Institute (InCor) on patients with heart failure. It has been found that, over time, these patients have greater survival, which may be linked to improved treatment and procedures performed at the hospital. As a conclusion, we suggest that the CUSUM methods based on Cox model residuals and the nonparametric method on the log-rank test may be alternatives for practice in monitoring survival data.

Keywords: Survival analysis, CUSUM chart, log-rank test, Cox model.

Sumário

1	Introdução	1
2	Revisão da Literatura	5
2.1	Esquema usual de inspeção e monitoramento	5
2.2	Monitoramento com heterogeneidade	9
2.3	CUSUM ajustado ao risco com base no modelo de Cox	14
2.4	Monitoramento prospectivo ajustado ao risco de tempos até a ocorrência de um evento	18
3	Monitoramento da Sobrevivência de Populações Homogêneas	21
3.1	Gráficos log-rank CUSUM	21
3.2	Aplicação a dados do InCor e proposta de novos gráficos	25
3.3	Comparação das abordagens em dados simulados	31
3.4	Obtenção de limites de controle e desempenho do gráfico ZDiff CUSUM	40
3.5	Estudo de simulação	46
3.6	Modelo de ponto de mudança	49
4	Monitoramento da Sobrevivência de Populações Heterogêneas	59
4.1	Monitoramento com verossimilhança completa do modelo de Cox	59
4.2	Simulação: Cox LHR CUSUM	61
4.3	Monitoramento com resíduo martingal e deviance	69
4.4	Simulação: Cox Martingal CUSUM e Cox Deviance CUSUM	70
5	Simulação Comparativa	77
6	Aplicação dos Métodos nos Dados do InCor	93
7	Conclusões e Perspectivas	105
A	Processos Pontuais	107
B	Fórmula do CUSUM	109
C	Cálculos de Razão de Chances	111
	Referências Bibliográficas	113

Capítulo 1

Introdução

Ferramentas de controle estatístico de processos (CEP), como gráficos de controle, têm sido propostas como forma de detectar mudanças significativas no comportamento de parâmetros importantes para a área de saúde, prontamente propiciando intervenções em procedimentos subjacentes. Woodall (2006) faz uma revisão sobre diversos métodos presentes na literatura. Por exemplo, é de interesse detectar se, com o passar do tempo, um procedimento cirúrgico realizado por alguma equipe médica passa a apresentar um risco significativamente maior de ocorrer falha nos pacientes, o que sugeriria a necessidade de rever a eficácia do método e/ou procedimentos da equipe, possivelmente levando intervenções para melhorar o desempenho.

O gráfico de controle CUSUM (*Cumulative Sum*) é em geral preferido quando se deseja detectar pequenos desvios em um determinado processo e é, portanto, adequado para aplicações na área médica, em que é importante atentar para mudanças de pequena intensidade na qualidade. Este gráfico se configura por acumular pequenas diferenças entre o que se observa nas características do processo monitorado e o perfil tomado como referência, quando tal processo está sob controle. Além disso, é um gráfico que considera a informação sobre todas as observações de forma geral, e desta forma avalia o processo como um todo, ao contrário do que ocorre nos gráficos de controle de Shewhart, em que a decisão é tomada com a informação mais recente disponível (Shewhart, 1931). Basicamente, este gráfico é construído a partir de um escore que representa a diferença entre os dados observados e o que se espera encontrar quando o processo está sob controle.

Diante da importância de se efetuar um controle da qualidade dos serviços realizados em hospitais ou no âmbito da saúde em geral, faz-se necessária a utilização de métodos para monitorar e detectar possíveis desvios nos resultados que são observados no contexto médico. Para desenvolvimento deste trabalho, temos como motivação um estudo conduzido no Instituto do Coração (InCor) da Faculdade de Medicina na Universidade de São Paulo. O conjunto de dados refere-se a duas coortes de pacientes com insuficiência cardíaca, que foram atendidos a partir do ano 1991 até junho de 2003 (coorte 1) e de julho de 2003 até 2007 (coorte 2).

De acordo com os pesquisadores responsáveis pelo estudo, as duas coortes representam circunstâncias operacionais bastante diferentes quanto à dinâmica do tratamento. A primeira coorte (1991) teve o objetivo de selecionar pacientes para o programa de transplante cardíaco e incluiu indivíduos em condições de saúde mais graves; a colaboração da Unidade Clínica com esse programa subsistiu até junho de 2003. Uma característica desta coorte é que incluiu pacientes muito graves e com número grande de óbitos (eventos) para alguns grupos.

A partir de então (2 de julho do ano 2003) foi iniciada uma nova coorte, na qual inclui-se todos os

pacientes atendidos na Unidade Clínica de Ambulatório Geral (todos os pacientes ambulatoriais do SUS que eram encaminhados para o InCor) e que receberam o diagnóstico de insuficiência cardíaca. Espera-se que nesse grupo ocorram tanto pacientes mais graves quanto menos graves. Análises prévias sugerem que há uma melhoria na sobrevida dos pacientes da coorte 2 em comparação com os indivíduos atendidos e acompanhados na coorte 1, mas é um ponto de interesse também detectar a intensidade desta mudança e o momento em que foi estabelecida.

Para a análise realizada, foram considerados os dados referentes a 3378 pacientes, sendo 1274 presentes na Coorte 1 e 2104 da Coorte 2. A Figura 1.1 ilustra as curvas de sobrevivência estimadas para os grupos de dados considerados. No gráfico à esquerda, cada uma das duas coortes é considerada em sua totalidade e mostra que a curva da coorte 2 se mantém superior ao padrão que é observado para a coorte 1 até aproximadamente 5 anos. No da direita, são considerados como dados de *referência* somente os primeiros 975 pacientes, os quais deram entrada na unidade até 1998, que foi o ano indicado pelos pesquisadores de mudança relevante nos procedimentos de tratamento realizados. Os pacientes remanescentes se referem aos dados de *monitoramento*. Esta segunda representação sugere que pacientes com início de acompanhamento a partir de 1999 apresentam um perfil de sobrevivência superior aos demais.

Figura 1.1: Comparação de curvas de Kaplan-Meier para os dados considerando (a) coorte 1 versus coorte 2 e (b) dados de treinamento com os primeiros 975 pacientes da coorte 1, que deram entrada até 1998 versus dados de monitoramento (i.e. todos os pacientes seguintes).

No entanto, deve-se atentar para a influência das distintas características individuais de cada paciente, antes de se afirmar que essa mudança no padrão da sobrevivência representa, de fato, uma melhoria na qualidade do atendimento. No exemplo considerado foi verificado que algumas variáveis como idade e alguns resultados de exames clínicos, como a fração de ejeção e o diâmetro diastólico do ventrículo esquerdo, são capazes de influenciar o risco de óbito (Marçula et al., 2011). Estes dados se caracterizam pela presença de muitas medidas omissas (NA's).

Na Figura 1.2 são apresentadas algumas comparações entre as coortes, considerando as diferentes classes funcionais, que representam avaliações do estado de saúde dos pacientes (em ordem crescente de gravidade). Os 4 níveis de classe funcional foram agrupados, resultando em duas clas-

(a) FUNCIONAL NA versus informada

(b) FUNCIONAL I ou II versus III ou IV

Figura 1.2: Comparação de curvas de Kaplan-Meier para os dados considerando indivíduos com (a) classe funcional NA versus classe funcional conhecida (b) classe funcional I ou II versus classe funcional III ou IV.

ses: I/II (com 566 indivíduos na Coorte 1 e 1083 na Coorte 2) e III/IV (535 na Coorte 1 e 863 na Coorte 2), de modo a contornar o problema da ocorrência de poucos pacientes em algum dos grupos. Observa-se a superioridade da sobrevivência para os pacientes da coorte 2, nos dois grupos de classe funcional considerados.

Diante destas discussões, entende-se que é de interesse desenvolver uma metodologia que consiga identificar, ao longo do tempo, os momentos em que ocorrem alterações no padrão de sobrevida, com o devido controle do efeito destas informações por meio das covariáveis, que pode afetar a variável de interesse avaliada. Dentro deste contexto, há na literatura diversos trabalhos propostos que consideram modelagens paramétricas para os dados, identificando algum parâmetro associado com o controle da qualidade do processo. Desta forma, testes envolvendo hipóteses simples são estabelecidos, com base em gráficos de controle ajustados, sendo que o valor especificado para esse parâmetro na hipótese alternativa representa o processo fora de controle. O instante exato de ocorrência da mudança não é conhecido; os gráficos fornecem apenas indicações de que uma alteração foi estabelecida.

Ao contrário da abordagem paramétrica usualmente aplicada em controle, modelos semiparamétricos podem proporcionar uma análise sem suposições sobre a distribuição dos dados. Por exemplo, para o modelo de Cox a única suposição necessária é a proporcionalidade de riscos, que é em geral simples de ser verificada.

O principal objetivo deste trabalho é revisar e destacar alguns métodos presentes na literatura sobre monitoramento no contexto de vigilância e desenvolver novas técnicas capazes de identificar, prospectivamente, possíveis alterações no padrão de sobrevivência, que consideram a heterogeneidade dos dados com gráficos de controle construídos por meio de uma modelagem semiparamétrica do risco de falha com o modelo de Cox. Também um método não paramétrico com a estatística de teste log-rank é explorado. Temos como foco a aplicação de gráficos de controle CUSUM para detecção de mudanças no tempo de sobrevivência com discussão sobre a obtenção de limites de

controle ótimos.

Este trabalho está organizado da seguinte forma: O Capítulo 2 contém uma revisão de métodos, com detalhamento de alguns resultados, e apresenta algumas abordagens de monitoramento discutidas na literatura para controlar a qualidade em estudos na área médica, em que se faz necessário levar em conta a heterogeneidade das características dos indivíduos, com diferentes riscos de óbito associados, e desta forma consideram uma modelagem semiparamétrica para dados de sobrevivência. Estas propõem um gráfico CUSUM ajustado ao risco, que leva em conta particularidades de cada indivíduo, por meio de modelos de regressão. São apresentados também alguns estudos que fazem uso de modelos paramétricos para dados de sobrevivência, e que monitoram através do gráfico CUSUM uma mudança no tempo esperado de sobrevivência.

No Capítulo 3, propomos um desenvolvimento não paramétrico do CUSUM para monitoramento de populações homogêneas, com base na estatística de teste obtida na família G^p de testes proposta por Harrington & Fleming (1982), que tem como caso particular o teste log-rank usual ($\rho = 0$) que é enfatizado. Ainda dentro do contexto de populações homogêneas, apresentamos um modelo paramétrico de ponto de mudança para identificar variação na sobrevida esperada, com uma abordagem que utiliza máxima verossimilhança para obter estimativas dos parâmetros. Consideramos também um procedimento em duas etapas, uma baseada na estatística log-rank para identificação do ponto de mudança para, em seguida, estimar os parâmetros da distribuição por máxima verossimilhança.

No Capítulo 4 propomos gráficos CUSUM que abordam o modelo semiparamétrico de Cox para monitorar a sobrevivência de populações heterogêneas. São exploradas abordagens que utilizam a razão de verossimilhanças e resíduos para o modelo de Cox usual. No Capítulo 5 os métodos propostos com base no modelo de Cox são comparados via estudo de simulação. Em seguida, os gráficos desenvolvidos são aplicados aos dados do estudo de motivação no Capítulo 6. Finalmente, são feitas conclusões e discussões sobre os métodos elaborados no Capítulo 7.

Capítulo 2

Revisão da Literatura

Este capítulo apresenta uma revisão detalhada sobre métodos de monitoramento com uma visão geral sobre inspeção e com ênfase em gráficos de controle CUSUM, estes apresentados também em sua extensão para acomodar características específicas de indivíduos em estudos no contexto de vigilância. Tais trabalhos são esclarecidos a fim de determinar uma abordagem adequada de monitoramento para dados de sobrevivência, que requerem a utilização de um modelo para os dados que deve ser levado em conta no gráfico de controle.

2.1 Esquema usual de inspeção e monitoramento

Discutimos nessa seção a estrutura básica de inspeção e monitoramento elaborada por Page (1954). Quando observações de um determinado processo são tomadas de forma consecutiva e ordenada, podemos considerar divisões em subconjuntos, de forma que cada um deles pode ser considerado como uma amostra aleatória de uma distribuição comum, porém com diferentes valores de um ou mais parâmetros da distribuição original. Por exemplo, se itens de algum processo possuem uma característica com distribuição normal de variância constante, pode-se considerar que a distribuição de um grupo de itens possua média μ_1 e outro grupo tenha uma média diferente μ_2 . Os problemas a serem abordados neste estudo consideram a identificação das subamostras, detectando-se alterações no valor do(s) parâmetro(s).

Tais problemas podem surgir em várias áreas de aplicação, e na área industrial um dos principais enfoques diz respeito à detecção de mudanças na qualidade dos itens em um processo de produção contínuo. Alguns de tais processos mantêm uma qualidade aproximadamente constante do produto por períodos consideráveis; ocasionalmente, devido provavelmente a uma falha em algum ponto do processo, a qualidade piora e uma grande proporção dos itens torna-se inaceitável. A qualidade do produto pode ser avaliada por alguma característica mensurável (por exemplo, quando o comprimento de peças é normalmente distribuído com variância constante, o comprimento médio pode ser usado como uma indicação de qualidade), ou pela fração de itens que falham a atender uma ou mais especificações. Em geral, é possível atribuir um número θ , denominado *medida de qualidade*, para o produto que pode ser representado por um parâmetro da distribuição. O interesse é então em detectar mudanças em θ .

Um esquema amplamente utilizado consiste em examinar amostras de tamanho fixo em intervalos regulares de tempo; uma estatística da amostra (por exemplo, média, amplitude, ou número de defeitos) é plotada em um gráfico de controle e uma ação corretiva é tomada se o seu valor

ultrapassa limites de controle preestabelecidos (ver Shewhart, 1931). Como se observa na prática, há uma probabilidade α de que algum ponto eventualmente caia fora dos limites e uma ação será tomada mesmo quando não ocorre mudança na medida de qualidade θ .

Se a fração de itens que é amostrada permanece constante, a quantidade de itens produzidos antes de uma ação ser tomada é proporcional ao número de itens inspecionados. Desta forma, estabelece-se como medida de desempenho do controle de qualidade o *número médio de amostras inspecionadas* até uma interrupção do processo sinalizada pelo método de avaliação. Esta medida é definida, em inglês, como *average run length (ARL)*. Quando a qualidade permanece constante, o ARL de um processo é o número esperado de itens amostrados antes de uma ação ser tomada. O ARL é claramente uma função de θ .

Quando a qualidade do produto é satisfatória, o ARL reflete o gasto relacionado à interrupção desnecessária da produção, quando ocorrem alarmes falsos, e está associado à probabilidade de parar o processo quando este está sob controle, ou seja, com uma variabilidade natural (erro do Tipo I). É desejado, então, que o valor do ARL seja o maior possível nessa situação com o processo sob controle. Por outro lado, se acontece uma deterioração na qualidade o ARL mede o atraso em sua detecção. Portanto, a quantidade de itens não conformes produzidos antes de uma ação reparadora ser tomada está relacionada com a probabilidade de manter o processo em andamento, quando este está irregular (erro do Tipo II). Nesta situação, deseja-se um ARL igual a 1, ou com o menor valor possível.

Para o esquema descrito, assumindo que o processo tem qualidade constante, e que as observações inspecionadas são estatisticamente independentes, o número r de amostras com n itens examinadas até a ocorrência de um alerta pode ser analisado como uma variável aleatória com distribuição geométrica, isto é,

$$P(r = k) = p^{k-1}(1 - p), \quad k = 1, 2, \dots,$$

em que $p = p(\theta)$ é a probabilidade de que um determinado ponto amostral esteja entre os limites de controle, ou seja, o gráfico não detecta uma mudança. Consequentemente, $E(r) = (1 - p)^{-1}$ e a função *ARL*, que define o número médio de inspeções até à ocorrência de um alarme, é dada por

$$ARL(\theta) = 1/(1 - p).$$

O esquema de inspeção descrito até agora é formulado com base ou em um único ponto registrado no gráfico de controle ou em uma quantidade fixa dos pontos registrados mais recentes. Portanto, essas regras ignoram toda a informação disponível no gráfico com relação à amostragem anterior. A fim de fazer uso de toda a informação contida nas inspeções já realizadas, Page (1954) sugere um *esquema sequencial de inspeção de processo*, que consiste em regras suplementares que utilizam a informação a respeito de todas as observações inspecionadas (ou algumas delas) desde a última ação tomada e que são adequadas para detectar qualquer magnitude de mudança no parâmetro monitorado.

Consideramos primeiro um exemplo simples de uma regra para controlar a fração de itens defeituosos produzidos por um processo industrial, sob um esquema *unilateral*, que é um tipo de inspeção designada para detectar desvios de θ em uma única direção. Suponha que amostras de 20 itens são tomadas a cada hora, e seja d_j o número de itens defeituosos na j -ésima amostra após a

última ação tomada. Considere a regra:

$$\begin{aligned}
 &\text{tomar uma ação se } d_j \geq 3, \\
 &\text{ou } d_j + d_{j-1} \geq 4, \\
 &\text{ou } d_j + d_{j-1} + d_{j-2} \geq 5, \\
 &\text{etc.} \\
 &\text{i.e. } \sum_{i=0}^r d_{j-i} \geq 3 + r \text{ para pelo menos um } r \quad (0 \leq r \leq j-1).
 \end{aligned} \tag{2.1}$$

Com essa regra, a decisão de tomar ou não uma ação é feita com base em cada amostra e todas as anteriores. É possível especificar o critério de ação (2.1) de forma diferente. Considere que um escore +19 é atribuído a cada item defeituoso na amostra e -1 a cada item não defeituoso. Seja w_j o escore total observado na j -ésima amostra inspecionada e $C_j = \sum_{i=1}^j w_i$, com $C_0 = 0$. Então, (2.1) é equivalente a

$$\text{Tomar uma ação se } C_j - C_{j-r} \geq 40 \text{ para todo } r (1 \leq r \leq j),$$

ou seja,

$$\text{Tomar uma ação se } C_j - \min_{0 \leq i \leq j-1} C_i \geq 40.$$

Supomos agora que, em um esquema sequencial, observações são inspecionadas individualmente em intervalos regulares, e que a decisão de tomar ou não uma ação devido a uma suspeita de deterioração no item é feita depois de cada observação avaliada. É possível então considerar uma regra semelhante à definida acima, e de forma geral atribuir um escore $w_k \in \mathbb{R}$ à k -ésima observação e plotar o escore acumulado $C_j = \sum_{k=1}^j w_k$ em um gráfico. Então deve-se tomar uma ação se

$$C_j - \min_{0 \leq i \leq j-1} C_i \geq h, \tag{2.2}$$

com $h > 0$ representando o limite superior de controle.

Intuitivamente, entende-se que um aumento expressivo no valor do escore acumulado é um indício de que a produção está gerando uma quantidade indevida de itens defeituosos. Sendo assim, tomar a diferença entre o escore acumulado até à amostra atual e o menor escore acumulado dentre cada uma das amostragens anteriores, que dá ideia da amplitude dessa medida, é uma forma de controlar o processo. Caso esta amplitude seja muito grande, a ponto de ultrapassar algum valor máximo tolerável, deve-se produzir um alerta para que haja uma interferência no processo.

Esse sistema de escore é escolhido de tal forma que o trajeto da média dos escores w_j é decrescente se a qualidade é satisfatória, ou seja, de gradiente negativo, e é crescente quando a qualidade não é satisfatória. Para avaliar o ARL, defina

$$C'_j = \max(C'_{j-1} + w_j, 0), \quad j = 1, 2, \dots,$$

com $C'_0 = 0$. Então, $C'_j = 0$ quando $C_j < \min_{0 \leq i \leq j-1} C_i$. A quantidade C'_j definida é a base para o procedimento elaborado por Page (1954) que define um gráfico de somas acumuladas (CUSUM), e, conforme demonstrado no Apêndice B, (2.2) é equivalente a

$$\text{Tomar uma ação se } C'_j \geq h. \tag{2.3}$$

A construção desse escore considera uma estrutura sequencial, em que o cálculo do escore atual envolve as informações de todas as amostragens anteriores. No entanto, é possível que, devido a uma sequência numerosa de itens produzidos sob controle, o valor do escore seja reduzido a ponto de retornar a zero, anulando a contribuição de todas as amostras inspecionadas até então. Escores sequencialmente acumulados que ultrapassam o limite h levam à interrupção do processo; por outro lado o retorno destes ao valor zero indica que a produção está funcionando de forma desejada. Consequentemente, a contribuição das amostras inspecionadas na tomada de decisão encerra-se no limite de controle superior (h) ou no limite inferior zero, quando a produção continua normalmente.

O erro do tipo I, mencionado anteriormente, acontece justamente quando o processo está sob controle e ainda assim o escore acumulado ultrapassa o limite superior h . Do mesmo modo, o escore acumulado pode atingir o limite inferior zero em uma situação que o processo está irregular, o que leva a uma reaplicação do método com uma produção fora de controle continuada (erro do tipo II).

Devido à natureza aleatória da produção, na prática a variação nos resultados é inevitável, mas pode ser avaliada probabilisticamente. Isso remete a uma metodologia conhecida como *teste sequencial de Wald*. Este teste é definido considerando-se limites (a, b) e escores $w_1, w_2, \dots$ associados às amostras $i = 1, 2, \dots$ observadas. Define-se então $C_i = C_{i-1} + w_i$, com $C_0 = w_0$, em que w_0 é um escore inicial, e tem-se um procedimento de testes no qual observações são tomadas sequencialmente enquanto $a < C_i < b$ (Wald, 1947). Pode ser visto que o CUSUM é uma sequência de testes de Wald com limites em $(0, h)$ e escore inicial zero. Mais detalhes sobre esse tipo de teste sequencial são apresentados adiante.

De forma geral, considere que amostras de tamanhos possivelmente diferentes são testadas sequencialmente por este método. Seja então $P(C_0)$ a probabilidade de que um teste de Wald com limites em $(0, h)$ e escore inicial C_0 termine no limite inferior (ou seja, $C_i < 0$) e sejam $N(C_0), N_1(C_0)$ e $N_2(C_0)$ o número médio de itens amostrados, respectivamente, incondicional, condicionado ao teste terminando no limite inferior e condicionado ao teste terminando no limite superior. Um teste que termina no limite inferior é chamado de “teste de aceitação”, do contrário temos um “teste de rejeição”.

Para o teste proposto com escore inicial zero ($C_0 = 0$) e limites em $(0, h)$, seja $P(0)$ a probabilidade de que o teste seja de “aceitação”. Com a suposição de independência entre as observações, tem-se que a probabilidade de que haja r testes de aceitação antes da ocorrência de um teste de rejeição é $\{P(0)\}^r \{1 - P(0)\}$, de acordo com a distribuição geométrica (com $r = 0, 1, \dots$), e segue que o número esperado de tais testes é então

$$\sum_{r=0}^{\infty} r \{P(0)\}^r \{1 - P(0)\} = \frac{P(0)}{1 - P(0)}. \quad (2.4)$$

Note que o número médio de itens amostrados, incondicional, pode ser calculado como $N(0) = N_1(0)P(0) + N_2(0)(1 - P(0))$. Segue que o número esperado de itens amostrados antes de uma ação ser tomada é

$$\begin{aligned} ARL &= \frac{P(0)}{1 - P(0)} N_1(0) + N_2(0) \\ &= \frac{N(0)}{1 - P(0)}. \end{aligned} \quad (2.5)$$

O gráfico CUSUM tem por vantagem a detecção mais rápida de mudanças pequenas em um

parâmetro, em comparação com métodos do controle de inspeção não sequencial. Além disso, há diversas formas de se construir o gráfico por intermédio de modificações no escore atribuído às observações inspecionadas, o que permite uma versatilidade para a técnica. É possível, por exemplo, reformular o escore w_k de modo a acomodar características específicas a dados de sobrevivência, e com isso aplicar esse método em estudos desta natureza. Mais informações a respeito deste gráfico podem ser encontradas em Hawkins & Olwell (1998) e Montgomery (2007).

2.2 Monitoramento com heterogeneidade

Em geral, aplicações industriais apresentam homogeneidade dos itens no processo de produção, enquanto que a heterogeneidade está presente em estudos relacionados à saúde. Desta forma, é importante considerar um ajuste do gráfico que leve em conta características específicas das observações.

Woodall (2006) descreve diversos trabalhos que utilizam gráficos de controle no contexto de vigilância. Dentre essas aplicações, considera-se a avaliação da qualidade de hospitais, e como variáveis de interesse pode-se monitorar taxas de infecção, taxas de quedas de pacientes, ou tempos de espera dos mais variados tipos. Frequentemente adota-se um modelo de regressão logística que associa variáveis relacionadas à saúde com taxas de mortalidade para pacientes.

Diversas modificações em gráficos de controle comumente aplicados têm sido propostas para ajustar o risco individual de falha. Steiner et al. (2000) e Steiner et al. (2001) desenvolveram gráficos CUSUM ajustados ao risco que consideram a abordagem de Page (1954). Basicamente, é aplicado um modelo de regressão logística, que envolve o risco de falha para pacientes que realizaram uma cirurgia, e considera-se alterações na taxa de falha como forma de avaliar a qualidade do procedimento. Trata-se de uma variável binária indicando se ocorre falha para um determinado indivíduo em até 30 dias subsequentes ao procedimento cirúrgico, o que reflete um desempenho insatisfatório.

Assim, considera-se a chance de falha como uma medida da qualidade, e é de interesse ser detectado com rapidez um aumento expressivo nessa chance, para que se possa tomar providências em relação à qualidade do atendimento. O gráfico CUSUM pode ser feito com base em razões de funções de verossimilhança e estas relações são afetadas pelas variações nas probabilidades sob controle de mortalidade para os pacientes. Como é usual, o CUSUM definido considera um *teste de razão de probabilidades sequencial* (TRPS), como introduzido por Wald (1947), cuja ideia inicial é brevemente descrita a seguir.

Seja X uma variável aleatória com densidade $f(x, \theta)$ envolvendo um parâmetro desconhecido θ . Suponha que deseja-se testar as hipóteses $H_0: f(x, \theta = \theta_0)$ contra $H_1: f(x, \theta = \theta_1)$. Com base em observações $(x_1, x_2, \dots)$, tomadas sequencialmente, o teste deve ser tal que, se θ_0 é o verdadeiro valor do parâmetro deseja-se decidir por θ_0 com probabilidade pelo menos $1 - \alpha$, e se θ_1 é o verdadeiro valor do parâmetro deseja-se decidir por θ_1 com probabilidade pelo menos $1 - \beta$. A solução ótima para este problema é obtida pelo Lema de Neyman-Pearson, em que, para um dado n , o teste mais poderoso, isto é, com menor β , depende da razão de funções de verossimilhança

$$L_n = \prod_{i=1}^n \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} \quad (2.6)$$

e o teste decide a favor (ou contra) θ_0 se L_n é menor que (ou maior) que uma constante. O escore CUSUM pode ser definido em geral como o logaritmo da razão de funções de verossimilhança (2.7).

$$w_i = \log L_n = \sum_{i=1}^n \log f(x_i, \theta_1) - \log f(x_i, \theta_0), \quad (2.7)$$

em que L_n é definido em (2.6).

O teste da razão de probabilidades sequencial é análogo a este procedimento e tem propriedades ótimas análogas. Funciona da seguinte forma: para cada observação obtida, calcula-se L_n (em que n é o total de observações coletadas até o momento). Se

$$\begin{aligned} B < L_n < A, & \quad \text{continuar a amostragem;} \\ L_n < B, & \quad \text{parar a amostragem e decidir por } \theta_0; \\ L_n > A, & \quad \text{parar a amostragem e decidir por } \theta_1. \end{aligned}$$

Equivalentemente, é mais usual na prática considerar logaritmos, tomando

$$\log B < l_n = \log L_n < \log A$$

para a região de continuação e, de forma análoga, definindo regiões de aceitação e rejeição de θ_0 . Uma das propriedades do teste é que $A \simeq (1 - \beta)/\alpha$ e $B \simeq \beta/(1 - \alpha)$ (ver Wald, 1947).

Para o exemplo de monitoramento binário discutido por Steiner et al. (2000) e Steiner et al. (2001), considere a variável y_i para o paciente i , e admita que $y_i = 1$ se ocorrem complicações devido à falha da cirurgia para este paciente dentro de 30 dias após sua realização e, caso contrário, $y_i = 0$. Deste modo, $f(y_i|\theta) = p(\theta)^{y_i}[1 - p(\theta)]^{1-y_i}$, e considere que $p(\theta_0) = c_0$ é a probabilidade de complicações sob a hipótese nula (estado sob controle) e $p(\theta_A) = c_A$ é a probabilidade de complicações sob a hipótese alternativa (situação fora de controle), que reflete uma mudança a ser detectada pelo método (esta pode representar uma deterioração ou melhoria da qualidade do procedimento). O gráfico CUSUM proposto permite, então, comparar sequencialmente as hipóteses $H_0: p(\theta) = c_0$ versus $H_1: p(\theta) = c_A$.

Para a proposta de monitoramento binário descrita, tem-se como resultado do logaritmo da razão de verossimilhanças sob H_1 e H_0 que os escores w_i em (2.7) a serem atribuídos para o gráfico CUSUM são:

$$w_i = \log \left\{ \left[\frac{c_A}{c_0} \right]^{y_i} \left[\frac{1 - c_A}{1 - c_0} \right]^{1-y_i} \right\}.$$

A heterogeneidade presente nas características observadas para os indivíduos, que é representada em um vetor de covariáveis $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$, é acomodada com o ajuste de um modelo que define uma relação do tipo $p_i(\theta) = g(\theta, \mathbf{x}_i)$. Pode-se, por exemplo, considerar a função de ligação g como os parâmetros estabelecidos em um modelo de regressão logística, que podem ser estimados antes da operação. Como cada paciente tem um risco pré-operativo de referência específico, redefine-se as hipóteses testadas em termos da razão entre a chance de ocorrer complicação no período de 30 dias após o procedimento cirúrgico com relação à chance de complicação pré-operativa.

Sejam R_0 e R_A as razões de chances de ocorrer complicação (pós-operativa/pré-operativa) sob a hipótese nula e alternativa, respectivamente, e considere $R_A > R_0$, isto é, há interesse em detectar

aumentos na chance de complicações após a cirurgia. Dada uma probabilidade de complicação pré-operativa p_i , a chance pré-operativa de ocorrer complicações para o paciente i é então $p_i/(1 - p_i)$. Portanto, como pode ser conferido no Apêndice C, para o paciente i a chance de complicação após a realização da cirurgia é $R_0 p_i/(1 - p_i)$ sob H_0 , enquanto que a chance de complicação pós-operativa é $R_A p_i/(1 - p_i)$ sob H_A , que corresponde a uma probabilidade de complicação $R_A p_i/(1 - p_i + R_A p_i)$ sob H_A . Para esse teste sequencial, as hipóteses confrontadas são: H_0 : razão de chances = R_0 versus H_A : razão de chances = R_A . Assim, os escores w_i em (2.7) para o gráfico CUSUM, calculados como o logaritmo da razão de verossimilhanças, são expressos por:

$$w_i = \log \left[\frac{1 - p_i + R_0 p_i}{1 - p_i + R_A p_i} \left(\frac{R_A}{R_0} \right)^{y_i} \right]. \quad (2.8)$$

A razão de verossimilhanças é maior que 1 quando há maior evidência para a hipótese alternativa, o que retorna um escore positivo, e assim ao acumular vários desses escores o gráfico pode ultrapassar o limite superior de controle h estabelecido, o que sinaliza uma mudança na qualidade. Já para a situação em que a evidência é maior para a hipótese nula, a razão de verossimilhanças é menor que 1, e os escores são negativos, e isto possibilita que os escores acumulados no CUSUM retornem ao ponto inicial, além de provocar um retardo nos alarmes falsos.

Na motivação apresentada, é razoável assumir que pacientes de diversas idades e condições de saúde podem apresentar diferentes probabilidades de sobrevivência $S(t|\mathbf{x}_i) = P(T > t|\mathbf{x}_i)$. Desta forma deve-se considerar que as funções de sobrevivência são condicionadas ao perfil do indivíduo, isto é, sob controle espera-se que para um paciente i em que é observado um conjunto de p características $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$, há uma função $S(t|\mathbf{x}_i)$ associada ao risco individual de este paciente apresentar óbito. As estimativas dessas probabilidades, que estão associadas a um risco específico de cada indivíduo apresentar o evento em estudo, são então utilizadas na construção dos gráficos de controle, incorporando a heterogeneidade inerente a este tipo de dados, por exemplo com um modelo de sobrevivência.

Desta forma, entende-se que o parâmetro de qualidade monitorado, θ , pode estabelecer uma relação do tipo $\theta \equiv \theta^* + \mathbf{X}\beta$, em que $\mathbf{X}\beta$ é um preditor linear com a matriz de covariáveis $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$, e β é o parâmetro que mede o efeito dessas características do indivíduo no tempo de sobrevivência. Assim, θ^* é visto como uma quantidade de referência que pode ter um valor estabelecido sob controle e que pode ser monitorado. Com isso, a medida de qualidade fica representada com a inclusão do efeito das diferentes covariáveis através do modelo de regressão, e tem-se então um gráfico de controle ajustado para dados de sobrevivência.

Sego et al. (2009) consideraram uma abordagem paramétrica com base no modelo de regressão de tempo de falha acelerado (MTFA) (Lawless, 2003), que assume como medida de qualidade o parâmetro de escala da distribuição do tempo de falha, θ (na notação utiliza neste trabalho). Os autores desenvolvem esse método para as distribuições Weibull e log-logística.

Para melhor entendimento desta abordagem, considere um cenário em que uma variável aleatória T , que define o tempo até a ocorrência de um evento em estudo, com certa distribuição de probabilidade, é de interesse. Tem-se ainda que a observação deste evento na prática está sujeita a uma outra variável aleatória C , que define tempos de censura desses eventos, e com isso, os dados contém tempos observados $t_1, t_2, \dots, t_n$, realizações de uma variável aleatória definida como $T^* = \min\{T, C\}$ e indicadores de falha $\delta_1, \delta_2, \dots, \delta_n$, com $\delta_i = I(T_i \leq C_i), i = 1, 2, \dots, n$. Também se

considera um conjunto de covariáveis $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$, $i = 1, 2, \dots, n$, que possibilita a aplicação do modelo de tempo de falha acelerado. Esta estrutura de dados define um conjunto de informações $\mathcal{D} = (n, T^*, \boldsymbol{\delta}, \mathbf{X})$, em que $\mathbf{X}_{n \times p}$ é a matriz com as covariáveis $\mathbf{x}_i$ para cada indivíduo e $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_n)'$. A relação estabelecida no MTFA é dada por

$$\log(T_i^*) = \mu + \mathbf{x}_i' \boldsymbol{\beta} + \sigma \varepsilon_i, \quad (2.9)$$

em que ε_i são erros aleatórios com distribuição padrão (com média 0 e variância 1), μ é o parâmetro de locação e σ é o parâmetro de escala da distribuição do logaritmo de T_i^* , e $\boldsymbol{\beta}$ é um vetor de parâmetros de regressão.

Admitindo independência entre as censuras C e os tempos de falha T , a função de verossimilhança para este modelo é dada por:

$$L(\boldsymbol{\eta}; \mathcal{D}) = \prod_{i=1}^n f(t_i; \boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{x}_i)^{\delta_i} S(t_i; \boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{x}_i)^{1-\delta_i}, \quad (2.10)$$

em que $\boldsymbol{\eta}$ são os parâmetros associados à distribuição do tempo de falha T (incluindo o parâmetro de qualidade $\boldsymbol{\theta}$), $f(t_i; \boldsymbol{\eta}, \mathbf{x}_i)$ é a função de densidade de probabilidade desta variável e $S(t_i; \boldsymbol{\eta}, \mathbf{x}_i)$ é a correspondente função de sobrevivência ajustada às características particulares do indivíduo i . Se por exemplo a variável aleatória T possui distribuição Weibull, sua função de densidade de probabilidade é dada por:

$$f(t) = \frac{\gamma}{\theta} \left(\frac{t}{\theta} \right)^{\gamma-1} \exp \left[- \left(\frac{t}{\theta} \right)^{\gamma} \right], \quad \text{para } t \geq 0, \quad (2.11)$$

sendo $\gamma > 0$ e $\theta > 0$ os parâmetros de forma e de escala, respectivamente. Para esta variável aleatória T , o valor esperado é $\theta \Gamma(1 + 1/\gamma)$, com função de sobrevivência $S(t) = \exp \left[- \left(\frac{t}{\theta} \right)^{\gamma} \right]$.

Com isso, foi definido também um gráfico CUSUM que se baseia na razão de verossimilhanças para este modelo, que testa basicamente hipóteses proporcionais do tipo $H_0: \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1: \boldsymbol{\theta} = \boldsymbol{\theta}_1 = \xi_1 \boldsymbol{\theta}_0$, com $\xi_1 \in \mathbb{R}_+^* - \{1\}$. Sego et al. (2009) atentam para o fato de que o parâmetro $\boldsymbol{\theta}$ está diretamente ligado ao tempo médio (ou mediano) de sobrevivência, e incluir covariáveis neste modelo para explicar esse parâmetro e monitorá-lo permite avaliar o desempenho da sobrevivência dos pacientes após um procedimento cirúrgico. O gráfico intitulado RAST CUSUM (Risk-adjusted survival time CUSUM) foi comparado com a proposta de regressão logística apresentada em (2.8), e verificou-se vantagem para o RAST CUSUM na detecção de deterioração da performance (ver Sego et al., 2009), proporcionando um ARL_1 menor.

Determinando o logaritmo da razão de verossimilhanças utilizado na construção desse método CUSUM, obtém-se que, para o modelo Weibull, o escore CUSUM é definido por:

$$w_i = (1 - \xi_1^{-\gamma}) \left(\frac{t_i}{\theta_0 \exp(\mathbf{x}_i' \boldsymbol{\beta})} \right)^{\gamma} - \delta_i \gamma \log \xi_1, \quad (2.12)$$

em que $\boldsymbol{\beta}$ é um vetor com parâmetros de regressão que representam os efeitos das covariáveis na medida de qualidade, e θ_0 é a quantidade de referência, sob controle, para esta medida, que é estipulada previamente.

De um modo análogo, Oliveira et al. (2016) propõem um gráfico CUSUM com base em um

modelo de regressão que considera a existência de uma proporção dos indivíduos na população com sobrevivência longa (Chen et al., 1999). Maller & Zhou (1996) descrevem tais modelos e enfatizam a importância dessa ideia da fração de cura, que é a possibilidade de que alguns indivíduos estejam “curados”. Eles, então, admitem um modelo de mistura para representar essa parcela dos indivíduos que não deve apresentar o evento em observação.

Chen et al. (1999) propuseram uma modelagem alternativa para a função de sobrevivência na presença de indivíduos curados, ao considerar a existência de uma variável aleatória M que representa a quantidade de causas que levam uma determinada observação a falhar. Para os indivíduos em que $M = 0$, o evento não ocorrerá mais, e estes são considerados curados; assim, a proporção de curados na população é dada pela probabilidade de M valer zero. Para os indivíduos em que $M = m > 0$, supõe-se então que para cada causa dentre as m há um tempo de sobrevivência associado, e finalmente o tempo referente a estes indivíduos é o menor destes. Rodrigues et al. (2009) buscaram uma unificação deste modelo, a qual é brevemente explicada a seguir.

Considerando uma amostra de tamanho n e que a cada indivíduo estão associadas as seguintes variáveis aleatórias:

- $M \equiv$ quantidade de causas que podem levar ao acontecimento do evento estudado; com distribuição de probabilidade $p_m = P_v(M = m)$ conforme um parâmetro v ;
- $R_j \equiv$ tempo até acontecer uma falha devido à causa j . Para $M = m$ supomos que $R_1, R_2, \dots, R_m$ são independentes e identicamente distribuídas (*iid*), independentes de M , com função de sobrevivência $S(t)$;
- $T \equiv$ tempo até ocorrência do evento, obtido como

$$T = \min\{R_0, R_1, \dots, R_m\},$$

com $P(R_0 = \infty) = 1$, uma vez que $M = 0$ implica em nunca observar o evento estudado, o que reflete a ideia de um “tempo infinito”.

Rodrigues et al. (2009) consideram a função de sobrevivência incondicional (populacional) dada por

$$S_p(t) = p_0 + \sum_{m=1}^{\infty} p_m S(t)^m, \quad (2.13)$$

sendo $p_0 = P_v(M = 0)$ a fração de cura. É desenvolvido o modelo de tempo de promoção, em que a variável aleatória M segue uma distribuição Poisson com parâmetro v . Neste caso, obtém-se a função de sobrevivência (imprópria) de longa duração:

$$S_p(t) = \exp\{-v[1 - S(t)]\}, \quad (2.14)$$

sendo $S(t)$ uma função de sobrevivência própria associada às variáveis $R_j, j = 1, \dots, M$. Logo, quando t tende a infinito, a função S_p tende a $\exp(-v)$ que é a fração de cura.

No modelo de tempo de promoção a relação entre v e as covariáveis pode ser descrita como

$$v_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}) \quad (2.15)$$

e a fração de cura é dada por

$$p_0(\mathbf{v}_i) = \exp[-\exp(\mathbf{x}_i'\boldsymbol{\beta})]. \quad (2.16)$$

Dada a possibilidade de censura, define-se uma variável $Y_i = \min\{T_i, C_i\}$ que é de fato observável, com o indicador δ_i análogo à modelagem sem fração de cura, e com isso os dados completos são $\mathcal{D}_c = (n, \mathbf{Y}, \boldsymbol{\delta}, \mathbf{M}, \mathbf{X})$, em que $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ e $\mathbf{M} = (M_1, M_2, \dots, M_n)$. A função de verossimilhança completa para esta nova abordagem envolve a variável latente $\mathbf{M}$ (na prática não se conhece o total de causas possíveis que provocam uma falha). Considera-se a função verossimilhança marginal, dada por

$$L^*(\boldsymbol{\phi}; \mathcal{D}) = \prod_{i=1}^n f_p(y_i | \boldsymbol{\eta}, \mathbf{v}_i)^{\delta_i} S_p(y_i | \boldsymbol{\eta}, \mathbf{v}_i)^{1-\delta_i}, \quad (2.17)$$

em que $\mathcal{D}$ são os dados observáveis apenas, sem a variável latente $\mathbf{M}$ e $\boldsymbol{\phi} = (\boldsymbol{\eta}, \mathbf{v})$. Note que $\mathbf{v}_i$ envolve os parâmetros de regressão $\boldsymbol{\beta}$ e as covariáveis $\mathbf{x}_i$ de cada indivíduo i e $\boldsymbol{\eta}$ são parâmetros associados à distribuição de sobrevivência dos indivíduos não curados. Por exemplo, para o caso Weibull definido são $\boldsymbol{\gamma}$ e $\boldsymbol{\theta}$ (o parâmetro monitorado).

De modo análogo à abordagem paramétrica usual de monitoramento apresentada por Sego et al. (2009), o gráfico de controle proposto por Oliveira et al. (2016) com modelo de fração cura investiga o parâmetro de escala ($\boldsymbol{\theta}$) da distribuição do tempo de falha associada aos indivíduos não curados. Assim, o gráfico CUSUM envolve o logaritmo da razão de verossimilhanças para este modelo em um teste de hipóteses para este parâmetro. Este gráfico é denotado por RACUF CUSUM (Risk-adjusted with cure fraction CUSUM).

É feita uma verificação por meio de simulações que esta proposta com um modelo que contempla a possibilidade da existência de indivíduos curados nos dados tem em geral um desempenho melhor que o RAST CUSUM apresentado (Oliveira et al., 2016).

Para o modelo Weibull, por exemplo, o escore CUSUM apresentado nessa proposta é dado por:

$$w_i = \delta_i \left[-\gamma \log \xi_1 + \left(\frac{y_i}{\theta_0} \right)^\gamma (1 - \xi_1^{-\gamma}) \right] - e^{\mathbf{x}_i'\boldsymbol{\beta}} \left\{ \exp \left[-\left(\frac{y_i}{\theta_0} \right)^\gamma \right] - \exp \left[-\left(\frac{y_i}{\xi_1 \theta_0} \right)^\gamma \right] \right\}, \quad (2.18)$$

em que $\boldsymbol{\beta}$ é um vetor de parâmetros do modelo de tempo de promoção que associa as covariáveis observadas à fração de cura.

Grigg & Farewell (2004) descrevem alguns gráficos de controle ajustados ao risco, no contexto geral da área de saúde, e dão ênfase para a eficácia dos métodos CUSUM. Para estas abordagens com o gráfico CUSUM, a eficiência do método é demonstrada em termos do ARL, em vez de simplesmente considerar as probabilidades de erro do tipo I e do tipo II como é feito em algumas propostas que consideram também um teste de hipóteses sequenciais. Mas há uma analogia entre estes dois conceitos, como foi discutido anteriormente.

Nas seções seguintes, são apresentadas abordagens semiparamétricas para construção dos gráficos de controle ajustados ao risco, com ênfase para o gráfico CUSUM e com a utilização do modelo de Cox para estimar probabilidades de sobrevivência.

2.3 CUSUM ajustado ao risco com base no modelo de Cox

Biswas & Kalbfleisch (2008) propõem um gráfico de controle CUSUM para monitoramento da sobrevivência com base no modelo de Cox. O estudo em discussão é realizado com informações de

todos os 59650 transplantes de rim realizados em 258 centros de transplante nos Estados Unidos, de janeiro de 1997 até dezembro de 2001. O intuito é monitorar ao longo do tempo a qualidade desses procedimentos para os pacientes que chegam, atentando para uma alteração expressiva na taxa de falha dos transplantes (falha do enxerto ou morte do paciente no período de um ano após o transplante), tendo como base um desempenho médio nacional que leva em conta o perfil de todos os centros combinados.

Considere uma instituição específica e suponha que transplantes ocorrem ao longo do tempo iniciando em um tempo de origem $t = 0$, que se refere ao momento em que se dá início ao estudo/acompanhamento. Sejam os sucessivos instantes (no calendário) de transplante $B_1, B_2, \dots$, para os pacientes $1, 2, \dots$ e Y_i o tempo, em anos, decorrido do transplante até uma falha do enxerto ou morte, e assuma que Y_i e B_i são independentes, dadas as covariáveis X_i medidas no momento do transplante. O tempo de falha *cronológico*, i.e. com relação ao instante de origem do estudo $t = 0$, é $T_i = B_i + Y_i$. Supomos que, condicionado às covariáveis X_i , há uma distribuição sob H_0 de interesse conhecida, que é definida pela função taxa de falha $\alpha_i(y)$. Esta distribuição sob a hipótese nula é considerada como a referência, que indica uma situação de qualidade sob controle. Nas aplicações, admite-se que esta distribuição é estimada das experiências de todos os centros de transplante combinados.

Biswas & Kalbfleisch (2008) consideram um modelo de Cox, com as covariáveis X_i medidas, que é da forma $\alpha_i(y) = \alpha_0(y)e^{X_i'\beta}$, para $y > 0$. Assim, $\alpha_i(y)$ representa a taxa de falha média nacional para um indivíduo com covariável X_i e $\alpha_0(y)$ é a função taxa de falha *baseline*. É relativamente comum utilizar como resposta o status do paciente um ano após o transplante. Desta forma, define-se como uma *falha qualificativa* da cirurgia o paciente que morre ou o enxerto que falha no período de um ano após a data do transplante.

Considere o tempo cronológico t , e defina um processo de contagem $N_i^D(t)$ correspondente ao i -ésimo transplante. Especificamente, seja $N_i^D(t) = 1$ para todo $t \geq T_i$ se o tempo de falha do transplante $T_i \in (B_i, B_i + 1]$ e $N_i^D(t) = 0$ caso contrário. Seja $dN_i^D(t) = N_i^D(t + dt) - N_i^D(t)$, em que dt é uma quantidade infinitesimal. Como o interesse é avaliar se há alteração na função taxa de falha, ou seja, se a partir de um determinado instante o processo muda de $\alpha(y)$ para $\xi \alpha(y)$, com $\xi > 0$, adota-se a notação $\xi \alpha(y)$ para a função de risco, a qual permite construir o teste de hipóteses $H_0: \xi = 1$ versus $H_1: \xi = \xi_1$, com $\xi_1 \neq 1$, em que $\xi_1 > 1$ indica aumento no risco de falha e $\xi_1 \in (0, 1)$ sugere redução. Em geral, para garantir a não negatividade da função taxa de falha, considera-se $\xi = e^\mu$, $\mu \in \mathbb{R}$, e as hipóteses são, então, $\mu = 0$ versus $\mu = \mu_1 \neq 0$. Desta forma, o processo $dN_i^D(t)$ pode ser caracterizado como

$$P(dN_i^D(t) = 1 | T_i \geq t, B_i, Z_i) = \begin{cases} e^\mu \alpha_i(t - B_i) dt & \text{se } 0 \leq t - B_i \leq 1, T_i \geq t \\ 0 & \text{caso contrário} \end{cases}. \quad (2.19)$$

Denota-se o lado direito da equação (2.19) por $e^\mu d\Lambda_i(t)$ e note que isto representa o risco instantâneo de uma falha qualificativa de transplante. Aqui e^μ é o fator pelo qual o risco na instituição em estudo difere da taxa nacional $d\Lambda_i(t)$, ou taxa nula sob $\mu = 0$. Note que há duas diferentes escalas de tempo de interesse: o processo $N_i^D(t)$ é definido em tempo cronológico enquanto que o tempo de falha Y_i é medido a partir do tempo de transplante B_i .

Para construção do monitoramento em tempo contínuo, seja $N^D(t) = \sum_{i \geq 1} N_i^D(t)$ o número total observado de falhas de transplante qualificativas que ocorreram até o instante t . Considera-se como

resposta o processo $N^D(t)$, $t > 0$, de forma que a contribuição na função de verossimilhança em qualquer intervalo pequeno $(t, t + dt]$ leva em conta a informação sobre todos os transplantes em risco de uma falha qualificativa no instante t .

Para uma constante $\theta > 0$ selecionada, considere a estatística da razão de funções de verossimilhança correspondente a um teste de $\mu = 0$ versus $\mu = \theta$. É utilizada a função de verossimilhança com base em $dN_i^D(t)$, a resposta no intervalo $(t, t + dt]$, condicionada à informação disponível até o instante t . A razão de funções de verossimilhança, com base nos dados $dN_i^D(t)$, com intensidade $e^\theta d\Lambda_i(t)$ versus $d\Lambda_i(t)$, em que consideramos (2.19) e um modelo de probabilidade Bernoulli para o incremento diferencial $dN_i^D(t)$, é

$$\begin{aligned} \frac{L_1(\theta)}{L_0(\theta)} &= \prod_{i \geq 1} \left\{ \frac{[e^\theta d\Lambda_i(t)]^{dN_i^D(t)} [1 - e^\theta d\Lambda_i(t)]^{1-dN_i^D(t)}}{[d\Lambda_i(t)]^{dN_i^D(t)} [1 - d\Lambda_i(t)]^{1-dN_i^D(t)}} \right\} \\ &= \prod_{i \geq 1} \left[\frac{e^{\theta dN_i^D(t)} [1 - e^\theta d\Lambda_i(t)]^{1-dN_i^D(t)}}{[1 - d\Lambda_i(t)]^{1-dN_i^D(t)}} \right]. \end{aligned}$$

Por um argumento de condicionamento repetido, e fazendo $dt \rightarrow 0$ no limite, a função de verossimilhança com base na informação disponível até o instante t é a integral produto dos componentes da verossimilhança em $(t, t + dt]$ (ver Kalbfleisch & Prentice, 2002). Tem-se então o seguinte:

$$\pi \prod_{s \in (0, t]} \prod_{i \geq 1} \left[\frac{e^{\theta dN_i^D(s)} [1 - e^\theta d\Lambda_i(s)]^{1-dN_i^D(s)}}{[1 - d\Lambda_i(s)]^{1-dN_i^D(s)}} \right].$$

Usando propriedades da integral produto (Bashirov et al., 2008) e do produtório, podemos calcular o numerador e denominador da expressão separadamente, e temos que

$$\begin{aligned} \pi \prod_{s \in (0, t]} \prod_{i \geq 1} e^{\theta dN_i^D(s)} [1 - e^\theta d\Lambda_i(s)]^{1-dN_i^D(s)} &= \pi \prod_{s \in (0, t]} \prod_{i \geq 1} e^{\theta dN_i^D(s)} \pi \prod_{s \in (0, t]} \prod_{i \geq 1} [1 - e^\theta d\Lambda_i(s)]^{1-dN_i^D(s)} \\ &\approx \pi \prod_{s \in (0, t]} e^{\sum_{i \geq 1} \theta dN_i^D(s)} \pi \prod_{s \in (0, t]} \prod_{i \geq 1} [e^{-e^\theta d\Lambda_i(s)}]^{1-dN_i^D(s)} \\ &= \pi \prod_{s \in (0, t]} e^{\theta dN^D(s)} \pi \prod_{s \in (0, t]} e^{-\sum_{i \geq 1} [1-dN_i^D(s)] e^\theta d\Lambda_i(s)} \\ &= e^{\int_0^t \theta dN^D(s)} e^{-\sum_{i \geq 1} \int_0^t [1-dN_i^D(s)] e^\theta d\Lambda_i(s)} \\ &= e^{\int_0^t \theta dN^D(s)} e^{-e^\theta \sum_{i \geq 1} \int_0^t d\Lambda_i(s)} \\ &= e^{\int_0^t \theta dN^D(s)} e^{-e^\theta \int_0^t \sum_{i \geq 1} d\Lambda_i(s)} \\ &= e^{\theta \int_0^t dN^D(s)} e^{-e^\theta \int_0^t d\bar{\Lambda}(s)} \\ &= e^{\theta N^D(t)} e^{-e^\theta \bar{\Lambda}(t)} \end{aligned}$$

é o numerador da expressão, em que $\bar{\Lambda}(t) = \sum_{i \geq 1} \Lambda_i(t)$, que representa a taxa de falha acumulada.

O denominador da expressão é dado por

$$\begin{aligned}
\prod_{s \in (0, t]} \prod_{i \geq 1} [1 - d\Lambda_i(s)]^{1 - dN_i^D(s)} &\approx \prod_{s \in (0, t]} \prod_{i \geq 1} [e^{-d\Lambda_i(s)}]^{1 - dN_i^D(s)} \\
&= \prod_{s \in (0, t]} e^{-\sum_{i \geq 1} [1 - dN_i^D(s)] d\Lambda_i(s)} \\
&= e^{-\sum_{i \geq 1} \int_0^t [1 - dN_i^D(s)] d\Lambda_i(s)} \\
&= e^{-\sum_{i \geq 1} \int_0^t d\Lambda_i(s)} \\
&= e^{-\int_0^t \sum_{i \geq 1} d\Lambda_i(s)} \\
&= e^{-\int_0^t d\bar{\Lambda}(s)} \\
&= e^{-\bar{\Lambda}(t)}.
\end{aligned}$$

Segue que a razão de verossimilhanças pode ser escrita como:

$$\frac{e^{\theta N^D(t)} e^{-e^\theta \bar{\Lambda}(t)}}{e^{-\bar{\Lambda}(t)}} = e^{\theta N^D(t)} e^{(1 - e^\theta) \bar{\Lambda}(t)} = e^{\theta N^D(t)} e^{-(e^\theta - 1) \bar{\Lambda}(t)}, \quad (2.20)$$

e o logaritmo da razão de funções de verossimilhança até o instante t é dado pelo processo

$$U_t = \theta N^D(t) - (e^\theta - 1) \bar{\Lambda}(t), \quad t > 0. \quad (2.21)$$

De forma a atualizar o CUSUM continuamente em cada instante t , trabalhamos em termos dos incrementos dU_t de U_t

$$dU_t = \theta dN^D(t) - (e^\theta - 1) d\bar{\Lambda}(t). \quad (2.22)$$

Utiliza-se aqui notação de derivada para t contínuo, mas na prática quando o tempo é discreto, pode-se adaptar para utilizar este método. Se uma alternativa bilateral for de interesse, calcula-se o CUSUM como $U_t = \int_0^t dU_s$ e tendências fortes, acréscimo ou decréscimo, indicariam uma mudança no processo, ultrapassando o limite no gráfico de controle. Esta proposta é uma extensão para tempo contínuo do CUSUM usual em tempo discreto introduzido em Page (1954). O CUSUM para a alternativa unilateral é $\{G_t\}$, em que $G_0 = 0$ e

$$G_{t+dt} = \max(0, G_t + dU_t) \text{ para } t > 0.$$

Pode ser visto, de forma análoga ao CUSUM em tempo discreto apresentado no Apêndice B, que isso é equivalente a

$$G_t = U_t - \min_{0 \leq s \leq t} U_s, \quad t > 0.$$

É fácil ver que U_t tem saltos de tamanho θ somente quando há uma falha qualificativa (só nestes instantes $dN^D(t) \neq 0$); caso contrário ele tem tendência decrescente. Note também que $\min_{0 \leq s \leq t} U_s$ é não crescente e reduz o valor somente quando $G_t = 0$. Em (2.21), $(e^\theta - 1)$ pode ser interpretado como o excesso de risco sob a hipótese alternativa, enquanto que $\bar{\Lambda}(t)$ pode ser interpretado como a taxa de falha acumulada esperada sob a hipótese nula.

O processo U_t pode também ser visto como um extensão contínua do passeio aleatório em escala de tempo discreta (por exemplo, ver Cox & Miller, 1965). Em tempo discreto, o desvio é

definido como o tamanho médio do passo. De forma análoga, o parâmetro de desvio é definido no caso contínuo em termos da taxa média de mudança do processo U_t . É de algum interesse notar que (2.22) é exatamente igual ao incremento do logaritmo da razão de verossimilhanças para um problema de teste em que $N^D(t)$ é um processo de Poisson com função média $\bar{\Lambda}(t)$.

Para implementar o procedimento CUSUM, monitora-se G_t em tempo contínuo até que $G_t > h$, em que $h > 0$ é um limite superior fixado, no qual um sinal é dado. O valor h é escolhido de forma a garantir algumas propriedades para o esquema de monitoramento. Seja ζ_h o tempo até que o limite em h é atingido, dado por

$$\zeta_h = \inf\{t > 0 : G_t \geq h\}.$$

Note que $E(\zeta_h)$ é o *average run length (ARL)* e desempenha um papel importante para determinar os valores dos parâmetros para o CUSUM. Este CUSUM foi obtido considerando um teste de razão de probabilidades sequencial (TRPS) em que a intensidade para $dN_i^D(t)$ é $e^\theta d\Lambda_i(t)$ *versus* $d\Lambda_i(t)$.

2.4 Monitoramento prospectivo ajustado ao risco de tempos até a ocorrência de um evento

Gandy et al. (2010) propõem um gráfico CUSUM que monitora a função taxa de falha, de modo similar a Biswas & Kalbfleisch (2008), mas concentram-se mais na estrutura prospectiva de monitoramento, conforme descrito a seguir. Suponha que indivíduos chegam nos instantes $B_1, B_2, \dots$ com vetores de covariáveis $X_1, X_2, \dots$ e experienciam um evento depois de $T_1, T_2, \dots$ unidades de tempo. Um indivíduo pode ser censurado depois de $C_1, C_2, \dots$ unidades de tempo. Assumimos que para cada indivíduo T_i e C_i são condicionalmente independentes dados B_i e X_i . A única adição à configuração da análise de sobrevivência padrão são os tempos de chegada B_i .

No tempo de calendário $t \geq B_i$ observamos para o i -ésimo indivíduo o vetor de covariáveis X_i , o atual tempo em risco

$$Y_i(t) = \min\{T_i, C_i, t - B_i\}$$

e o indicador do evento $\delta_i(t) = I\{T_i = Y_i(t)\}$, em que I denota a função indicadora.

Suponha que, sob controle, a função taxa de falha da variável T_i condicional às covariáveis X_i é $\alpha_{0i}(\cdot) = \alpha_0(\cdot, X_i)$. Esta taxa de falha, intuitivamente, corresponde a um estado de referência, por exemplo, com base no desempenho passado. Assuma que depois de um tempo desconhecido $\tau \in [0, \infty]$ a taxa de falha muda para um valor fora de controle $\alpha_{1i}(\cdot) = \alpha_1(\cdot, X_i)$. Assim, para um indivíduo que chega no instante $B_i \leq \tau$, a taxa de falha da variável T_i é $\alpha_{0i}(s)$ para $s \leq \tau - B_i$ e $\alpha_{1i}(s)$ para $s > \tau - B_i$. Denotamos as correspondentes taxas de falha integradas por $A_{ji}(t) = \int_0^t \alpha_{ji}(s) ds$, $j = 0, 1$.

Definimos um gráfico CUSUM com base na razão de verossimilhanças. As funções de verossimilhança sob controle ($j = 0$) e fora de controle ($j = 1$) até o tempo de calendário t são

$$L_j(t) = \prod_{i: B_i \leq t} \alpha_{ji}(T_i)^{\delta_i(t)} \exp[-A_{ji}\{Y_i(t)\}] \quad j = 0, 1.$$

Conforme Gandy et al. (2010), a correspondente estatística de teste do logaritmo da razão de verossimilhanças para o estado sob controle *versus* fora de controle é

$$R(t) = \sum_{i: B_i \leq t} \delta_i(t) \log \left\{ \frac{\alpha_{1i}(T_i)}{\alpha_{0i}(T_i)} \right\} - \sum_{i: B_i \leq t} [A_{1i}\{Y_i(t)\} - A_{0i}\{Y_i(t)\}].$$

Com base nisso, é definido um gráfico de controle CUSUM $Z(t) = R(t) - \min_{s \leq t} R(s)$ que sinaliza no tempo $\zeta_h = \inf\{t: Z(t) > h\}$ para algum limite $h > 0$. Vários critérios podem ser considerados na escolha do limite h . Uma abordagem comum é determinar um ARL sob controle no tempo de calendário, que é o tempo esperado até que o limite seja ultrapassado sem que haja alguma mudança, isto é, $E(\zeta_h | \tau = \infty)$. Outra abordagem é determinar uma probabilidade de alarme falso em uma janela de tempo dada, por exemplo, $P(\zeta_h \leq U | \tau = \infty) = \alpha$ para algum $U > 0$ e $0 < \alpha < 1$.

As metodologias discutidas até então assumem que a informação para cada indivíduo está disponível todo tempo enquanto ele está em risco. Esse pode não ser o caso. Por exemplo, a informação sobre um indivíduo pode estar disponível apenas depois que o evento ou censura tiver ocorrido. A única mudança resultante na abordagem é permitir que cada indivíduo contribua na verossimilhança somente depois de sair, ao redefinir $Y_i(t)$ como

$$Y_i(t) = \min(T_i, C_i) I\{\min(T_i, C_i) + B_i \leq t\}.$$

O gráfico CUSUM resultante $Z(t)$ consiste apenas de saltos, ele não possui mais um desvio contínuo. Para determinar o limite de controle nesta situação deve-se recorrer a simulações.

Gandy et al. (2010) comentam ainda sobre dois tipos de hipóteses que podem ser testadas, proporcionais e gerais, como é apresentado a seguir. Considere uma hipótese alternativa proporcional, em que $\alpha_{1i}(s) = \xi_1 \alpha_{0i}(s)$ para algum $0 < \xi_1 < \infty$. Biswas & Kalbfleisch (2008) estudaram o caso $\xi_1 > 1$. O logaritmo da razão de verossimilhanças simplifica para

$$R(t) = \log(\xi_1)N(t) - (\xi_1 - 1)\Lambda(t), t \geq 0,$$

em que $N(t) = \sum_i \delta_i(t)$ é o número de eventos até o instante t e $\Lambda(t) = \sum_i A_{0i}\{Y_i(t)\}$. Proporcionalidade aqui refere-se apenas a relação entre α_{0i} e α_{1i} . A taxa de falha sob controle α_{0i} pode ser estimada com base no desempenho passado por meio de qualquer tipo de modelo de tempo até falha, não somente o modelo de riscos proporcionais. Se $\xi_1 > 1$, então os saltos do gráfico $Z(t)$ são para cima e a parte contínua tem tendência decrescente. Se $\xi_1 < 1$ então os saltos são para baixo e a tendência na parte contínua é crescente.

Hipóteses alternativas gerais podem levar a um CUSUM com saltos tanto para cima como para baixo. Além disso, entre os saltos, o gráfico CUSUM não precisa ser monótono. Uma alternativa plausível ao método com alternativas proporcionais é considerar mudanças no tempo, isto é, $A_1(s) = A_0\{f(s)\}$ para alguma função não negativa monotonicamente crescente f . A taxa de falha alternativa é então $\alpha_1(s) = \alpha_0\{f(s)\}f'(s)$. Como extensão, pode-se deixar a função f depender de covariáveis.

Um caso especial tem sido empregado por Sego et al. (2009). Eles consideram certos modelos paramétricos de tempo de falha acelerado como o modelo sob controle e estudam alterações no parâmetro de escala, o que acarreta uma função de tempo linear, isto é, $f(s) = as$, para algum $a \neq 1$.

Uma alternativa de mudança no tempo pode levar a casos não proporcionais, como o exemplo simples (2.23) demonstra. Suponha que $\alpha_0(t) = 1$ para $t \leq 1$ e $\alpha_0(t) = 1/2$ para $t > 1$. Para a alternativa $A_1(t) = A_0(2t)$, obtemos

$$\frac{\alpha_1(t)}{\alpha_0(t)} = \begin{cases} 1, & \text{se } \frac{1}{2} < t \leq 1 \\ 2, & \text{se } t \leq \frac{1}{2} \text{ ou } t > 1. \end{cases} \quad (2.23)$$

Quando usamos abordagens não paramétricas ou semiparamétricas fundamentadas no estimador de Nelson-Aalen ou no modelo de Cox para modelar α_{0i} e α_{1i} , computar o processo da razão de verossimilhanças $R(t)$ e, conseqüentemente, o gráfico CUSUM $Z(t)$ é complicado pela necessidade de calcular a razão entre α_{0i} e α_{1i} . A estimação da taxa de falha geralmente envolve suavização de uma estimativa da taxa de falha integrada via suavizadores Kernel, por exemplo. Em contraste com o caso de alternativas proporcionais, transformações no tempo levam a um processo com várias alturas de salto.

Capítulo 3

Monitoramento da Sobrevivência de Populações Homogêneas

Os métodos apresentados no capítulo anterior são uma revisão do que se tem na literatura a respeito do monitoramento de variáveis que estão ligadas a especificidades das observações, o que deve ser feito por gráficos de controle ajustados. Tais métodos concentram-se, na maior parte, em uma estrutura paramétrica, que sugere a adoção de um modelo e distribuição para os dados. Neste capítulo apresentamos algumas propostas para construir gráficos de controle em uma estrutura não paramétrica com o teste log-rank.

Na literatura, alguns autores propõem a utilização da estatística log-rank para identificar diferenças no padrão do tempo de sobrevida, quando consideram-se populações homogêneas. Nas subseções seguintes, discutimos brevemente essas ideias de monitoramento e desenvolvemos novas propostas para detectar mudanças no perfil da sobrevivência de forma prospectiva, como a construção de um gráfico CUSUM que faz uso da estatística de teste log-rank.

3.1 Gráficos log-rank CUSUM

Primeiramente desenvolvemos um método não paramétrico para avaliação de mudanças no perfil da sobrevivência com a utilização da estatística linear de postos do teste de comparação entre curvas de sobrevida com dois grupos (Peto & Peto, 1972) para construção de um escore para monitoramento com o gráfico de controle CUSUM. Assumimos o desenvolvimento proposto em Harrington & Fleming (1982), que define a família G^p de testes.

Para comparação entre curvas de sobrevivência, a utilização desta estatística de teste, que tem como caso particular o teste log-rank, foi considerada em Li et al. (2012) para construir um gráfico de controle como uma forma de avaliar se há diferença significativa entre duas amostras com dados de sobrevivência. Considera-se que há uma amostra de treinamento para servir de referência tendo o processo sob controle, e tem-se então subamostras de monitoramento que serão comparadas a fim de detectar alguma alteração na sobrevivência. Li & Kong (2015) propuseram uma generalização para essa abordagem não paramétrica, em que há um teste unilateral comparando a função de risco estimada para a amostra de monitoramento com uma estimativa precisa da função de risco sob controle, obtida a partir de uma amostra com dados históricos sob controle.

Considerando notação parecida com a adotada em Li et al. (2012) para comparação entre duas

amostras ($j = 1, 2$), sejam então $t_1 < t_2 < \dots < t_L$ os tempos de falha distintos na amostra completa (com combinação dos dados de cada grupo avaliado). Desta forma, as falhas ocorrem em L instantes diferentes. No instante t_i , são observadas d_{ij} falhas no j -ésimo conjunto de dados, de um total de r_{ij} indivíduos em risco, cujo tempo observado não é menor que t_i , em que $i = 1, 2, \dots, L$ é o índice do tempo de ocorrência e $j = 1, 2$ é o índice do conjunto de dados. Sejam $d_{\cdot}(t_i) = d_{\cdot i} = \sum_{j=1}^2 d_{ij}$ e $r_{\cdot}(t_i) = r_{\cdot i} = \sum_{j=1}^2 r_{ij}$, respectivamente, o total de falhas no instante $t_i, i = 1, 2, \dots, L$ e o total de indivíduos em risco neste instante na amostra combinada. O objetivo é testar as hipóteses:

$$\begin{aligned} H_0 : h_1(t) &= h_2(t), \text{ para todo } t \leq \tau \\ H_1 : h_1(t) &\neq h_2(t), \text{ para algum } t \leq \tau \end{aligned}$$

em que τ pode ser selecionado como o menor valor dentre os maiores tempos observados, um de cada grupo de dados. Essa condição se faz necessária devido à impossibilidade de obter estimativas não paramétricas da função de risco para cada grupo após esse instante τ . A estatística de teste não paramétrica para duas amostras, com base no estimador de Nelson-Aalen para a função de risco acumulada é dada por:

$$Z_j(\tau) = \frac{\sum_{i=1}^L W(t_i) \left\{ d_{ij} - r_{ij} \frac{d_{\cdot i}}{r_{\cdot i}} \right\}}{\sqrt{\sum_{i=1}^L W(t_i)^2 \frac{r_{ij}}{r_{\cdot i}} \left(1 - \frac{r_{ij}}{r_{\cdot i}} \right) \left\{ \frac{r_{\cdot i} - d_{\cdot i}}{r_{\cdot i} - 1} \right\} d_{\cdot i}}}, \quad (3.1)$$

em que $W(t_i)$ é uma função peso não negativa e $W(t_i) = 0$ quando $r_{\cdot i} = 0$ (Klein & Moeschberger, 2005, p. 206). A estatística tem, sob a hipótese nula, distribuição normal padrão para amostras grandes (Schoenfeld, 1981), e nota-se que $Z_1(\tau) = -Z_2(\tau)$. A função $W(t_i)$ pode variar definindo diferentes famílias de testes. A proposta discutida em Harrington & Fleming (1982) considera $W(t_i) = [\hat{S}(t_i)]^\rho$, em que $\hat{S}(t_i)$ é o estimador de Kaplan-Meier para a função de sobrevivência no instante t_i , e ρ define a ponderação (importância) das diferenças observadas ao longo do tempo (por exemplo, $\rho = 1$ sugere que há um peso maior para a informação observada nos menores tempos de falha). Desta forma, admitindo a distribuição normal padrão desta estatística de teste, os limites de controle para um gráfico com esses valores podem ser facilmente obtidos ao se fixar um nível de significância para o teste.

No entanto, a aplicação desse gráfico de controle ocorre diretamente a dados com subgrupos, ao comparar cada novo subgrupo formado com uma amostra de referência fixada e plotar um valor da estatística, o que difere de um monitoramento feito diariamente para comparar um grupo de controle, que se atualiza dinamicamente, com um segundo grupo de novos indivíduos que chegam, e também têm informação atualizada a cada dia. Porém, é possível fazer uma adaptação deste método para construir uma inspeção prospectiva em um conjunto de dados, em que o grupo de controle é pré-definido, e novos subgrupos podem ser a informação atualizada sobre um grupo de monitoramento.

Pressupõe-se que os dados avaliados podem ser reestruturados de modo a construir um monitoramento prospectivo, com a atualização feita de forma dinâmica diariamente com a informação disponível e os indivíduos presentes nos dados. Desta forma, considere um cenário em que pacientes dão entrada no protocolo de atendimento, similarmente ao acompanhamento feito nos dados do InCor, e que os dados registrados são organizados na ordem destas datas de referência para cada

indivíduo. Com isso, tem-se a origem da contagem de tempo a partir da primeira destas datas.

Admitimos que até um determinado ponto estão situados indivíduos que configuram um grupo de controle para o estudo, e a partir desta data se dá início a uma investigação que compara a sobrevivência desse segundo grupo, que contém pacientes novos, com este grupo de controle. O interesse é avaliar então de forma prospectiva se há diferença entre o perfil da sobrevivência dos novos pacientes (grupo monitorado) e a dos indivíduos que formam o grupo controle. Com esta finalidade, a abordagem de Harrington & Fleming (1982) define um teste para comparação entre duas curvas de sobrevivência com base na informação sobre a quantidade de falhas e indivíduos em risco que é avaliada nos instantes de falha observados, em cada um dos dois grupos.

Propomos então um CUSUM que considera o valor da estatística de teste (3.1) ao quadrado, $Q_j(\tau) = [Z_j(\tau)]^2$, que tem assintoticamente distribuição qui-quadrado e deve ser referida simplesmente como qui-quadrado posteriormente. Definimos um escore para cada instante em que se aplica o teste comparativo dos dois grupos de interesse, com a informação dos indivíduos disponibilizada de forma prospectiva. Por exemplo, os dados podem ser atualizados diariamente, e com isso é possível monitorar prospectivamente. Também desta forma adaptamos a proposta de monitoramento para subgrupos ao considerar que a cada dia é feita a comparação do grupo de monitoramento atualizado com o de referência, e é feita a comparação com esta abordagem.

Para construir o monitoramento prospectivo, com a notação adotada, considere que a cada dia no tempo cronológico de acompanhamento tem-se a informação mais atual sobre cada indivíduo presente nos dados. Admitindo que cada paciente dá entrada no estudo em um tempo de calendário possivelmente diferente, seja B_i o instante de início do acompanhamento referente ao indivíduo i . Assim, o conjunto de informações disponível no momento t do calendário contém $Y_i(t) = \min\{T_i, C_i, t - B_i\}$ que representa o tempo observado para o paciente i até o momento e $\delta_i(t)$ o correspondente indicador se esse tempo é de falha. Neste caso, T_i é o verdadeiro tempo de falha e C_i o tempo de censura associado ao indivíduo e temos que $\delta_i(t) = I\{Y_i(t) = T_i\}$.

Como exemplo da estrutura de dados que é construída, considere um exemplo em que os 5 primeiros indivíduos dão entrada no estudo em dias subsequentes (ou seja, seus tempos de entrada são $B_1 = 0, B_2 = 1, B_3 = 2, B_4 = 3, B_5 = 4$), e que dentre estes somente os indivíduos 2 e 4 apresentam falha, respectivamente, 28,7 e 29,3 dias depois dos instantes em que entraram. Considere estes 5 primeiros indivíduos como o grupo de controle e que após estes se dá início à formação do grupo de monitoramento, constituído por mais 5 indivíduos com datas de entrada também sequenciais ($B_6 = 5, B_7 = 6, B_8 = 7, B_9 = 8, B_{10} = 9$). Para este grupo de monitoramento, admita que somente os indivíduos 6 e 10 falham, respectivamente, 28,7 e 23,1 dias depois de darem entrada. Então, o formato dos dados nos dias de acompanhamento $t = 29$ a $t = 34$ após a entrada do indivíduo 1 fica conforme mostrado na Tabela 3.1.

Nota-se que as informações em relação aos indivíduos em risco e aos tempos de falha observados podem variar, o que modifica diariamente as quantidades definidas para construção da estatística de teste, que é um somatório para distintos tempos de falha. O indivíduo está em risco em determinado tempo de falha t_i se ele não apresentou falha nem foi censurado até o instante imediatamente anterior, ou seja, se seu tempo de sobrevivência observado (falha ou censura) é maior ou igual que t_i . Por exemplo, no dia 30 há apenas um tempo de falha observado nos dados, e neste instante de falha conta-se apenas com os 2 primeiros indivíduos em risco no total, já que todos os outros têm tempo observado inferior a 28,7.

	Dia 29				Dia 30		Dia 31		Dia 32		Dia 33		Dia 34	
i	B_i	Y_i	δ_i		Y_i	δ_i	Y_i	δ_i	Y_i	δ_i	Y_i	δ_i	Y_i	δ_i
1	0	29	0		30	0	31	0	32	0	33	0	34	0
2	1	28	0		28,7	1	28,7	1	28,7	1	28,7	1	28,7	1
3	2	27	0		28	0	29	0	30	0	31	0	32	0
4	3	26	0		27	0	28	0	29	0	29,3	1	29,3	1
5	4	25	0		26	0	27	0	28	0	29	0	30	0
6	5	24	0		25	0	26	0	27	0	28	0	28,7	1
7	6	23	0		24	0	25	0	26	0	27	0	28	0
8	7	22	0		23	0	24	0	25	0	26	0	27	0
9	8	21	0		22	0	23	0	24	0	25	0	26	0
10	9	20	0		21	0	22	0	23	0	23,1	1	23,1	1

Tabela 3.1: Estruturação dos dados de forma prospectiva.

Já no dia 31, há apenas o tempo de falha 28,7 para construir novamente a tabela de contingência, e embora o tempo de falha seja o mesmo, nesta configuração dos dados temos que o total de indivíduos em risco é 3, pois além dos dois primeiros, o indivíduo 3 da amostra também teve tempo de observação maior ou igual a 28,7. O mesmo ocorre no dia 32, em que o indivíduo 4 passa a ser mais um em risco neste único tempo de falha observado.

Em sequência, no dia 33 de acompanhamento, há uma nova falha do indivíduo 4, e nesta tabela adicional há apenas 4 indivíduos em risco no total, todos inclusos no grupo de controle. Também observa-se uma nova falha para o indivíduo 10, pertencente ao grupo 2, o que acrescenta mais um tabela de contingência no instante de falha 23,1, ocasião em que há um total de 10 indivíduos em risco.

Para o dia 34 continua a mesma quantidade dos tempos de falha para construção de tabelas, porém há uma mudança no total de falhas para o instante 28,7, que conta com mais uma no grupo de monitoramento (referente ao indivíduo 6). Percebe-se, com isso, que o valor da estatística de teste é alterado constantemente a cada dia de monitoramento, devido à entrada de novos indivíduos, e também pela maior informação que se tem a respeito de toda a amostra.

Como escore para definir um gráfico CUSUM levamos em conta possíveis fórmulas que ponderam de alguma forma a evolução do valor da estatística de teste qui-quadrado ao longo do tempo. Conforme discutido anteriormente, seja τ_t o tempo máximo para comparação dos grupos no instante de calendário t , e defina $Q_t = Q_j(\tau_t) = Z_j^2(\tau_t)$, o valor dessa estatística de comparação dos grupos nesse momento, em que j é o grupo definido para cálculo (que consideramos ser o grupo de controle). Uma primeira formulação considera a razão entre os valores de Q_t dada por

$$w_t = I\{Q_{t-1} \neq 0\} a_t \log \left(b_t \frac{Q_t}{Q_{t-1}} \right), \quad (3.2)$$

em que $a_t \geq 0$ e $b_t > 0$ são possíveis pesos para ponderar o escore a cada dia de inspeção. Com isso, as somas acumuladas de w_t podem ser utilizadas na formulação de um gráfico CUSUM como discutido. Denotamos este gráfico por *logRatio CUSUM*. Para simplificar, apresentamos ilustrações para os casos em que os pesos b_t são constantes iguais a 1. Este método se caracteriza por identificar, em um monitoramento prospectivo, aumentos expressivos no valor da estatística de teste que podem ocorrer de um dia para o outro, e indicar que a diferença entre os grupos comparados é cada vez mais provável.

3.2 Aplicação a dados do InCor e proposta de novos gráficos

Nesta seção exploramos a utilização do gráfico logRatio CUSUM proposto com a aplicação do método nos dados do InCor com $\rho = 0$ ($W(t_i) = 1, \forall t_i$), que configura o usual teste log-rank, que é o teste mais poderoso quando a suposição de riscos proporcionais se verifica nos dados. Com isso, são elaboradas diferentes propostas de gráficos CUSUM com utilização da estatística de teste log-rank.

Vamos concentrar este estudo para o caso do teste log-rank, mas entendemos que é válido investigar em estudos futuros a possível eficiência do teste de Wilcoxon ($\rho = 1$) que é em geral preferível quando esta suposição não é válida. No entanto, para os dados do InCor foi averiguada uma similaridade no comportamento de ambos os testes. Nesta ilustração, aplicamos o monitoramento após a entrada dos primeiros 975 pacientes da Coorte 1 (que representam o grupo de controle), e a partir do primeiro dia de acompanhamento do paciente seguinte, o que ocorre aproximadamente no começo de 1999.

Figura 3.1: Linhas de tempo dos pacientes presentes nos dados (a) no primeiro dia de monitoramento e (b) no último dia de monitoramento.

A Figura 3.1 exibe as linhas de acompanhamento para cada indivíduo presente nos dados em dois tempos de calendário distintos, a saber, no primeiro dia de monitoramento (em que há apenas 1 indivíduo no grupo 2, com tempo observado igual a 1) e no término do estudo. Nota-se que o grupo 1 (controle) é estruturado com uma taxa de entrada dos pacientes bem menor que a do grupo 2, como pode ser observado na subida mais rápida presente na Figura 3.1b. Percebe-se ainda que há muitas censuras (e também algumas falhas) com valores grandes, o que motiva a utilização dos modelos de longa duração.

Na Figura 3.2 são ilustradas as comparações das curvas de sobrevida estimadas para os grupos de controle e monitoramento em diferentes datas. Vê-se que pelo menos 1 ano é necessário para se notar graficamente a diferença entre as duas curvas. Assim, a avaliação com métodos de controle para este tipo de dado está sujeita a um atraso que é inerente à necessidade de se observar mais tempo para se ter uma melhor comparação, o que pode naturalmente atrasar a detecção de mudanças, e traz um desafio na elaboração dos métodos de controle. De fato, para investigar se um novo tratamento proporciona maior sobrevida aos pacientes, há que se observar os indivíduos por um

Figura 3.2: Kaplan-Meier comparativo dos grupos controle e monitoramento (a) na data da primeira falha no grupo 2 (b) no primeiro ano de monitoramento (c) no segundo ano de monitoramento e (d) no quinto ano de monitoramento.

tempo razoável para que um possível efeito de melhora se manifeste e possa ser averiguado.

É importante, ainda, que esta melhora seja confirmada o quanto antes após esse período de espera necessário, para que seja possível avaliar se vale investir na mudança de tratamento, e principalmente, é essencial identificar rapidamente se há uma piora dos pacientes para que providências sejam tomadas a tempo de reverter uma situação grave. Os gráficos apresentados na Figura 3.2 informam no título a quantidade total de pacientes no momento, $n(t)$, o total de falhas observadas até o momento, $D(t)$ e quantos tempos distintos de falha foram observados ao todo, $L(t)$. Na legenda essas mesmas informações são apresentadas para cada grupo (i.e. $n_j(t)$ é o total de pacientes, $D_j(t)$ é o total de falhas e $L_j(t)$ é o total de falhas distintas no grupo j).

A Figura 3.3 mostra o logRatio CUSUM com dois possíveis pesos a_t . No eixo x está informado o dia no tempo de calendário, contado a partir da data de entrada do primeiro indivíduo pertencente ao grupo de controle. Nota-se que o valor do CUSUM cresce rapidamente nos primeiros dias do

monitoramento, o que pode alertar para uma diferença entre os riscos de falha. No entanto, é importante discutir os problemas de instabilidade que podem ocorrer devido ao valor da estatística (com distribuição qui-quadrado assintótica) no início de acompanhamento que é calculado com poucas observações no grupo monitorado. Daí a importância de se considerar pesos que levem em conta essas dificuldades iniciais. Deve-se discutir então na prática a partir de que dia é confiável fazer uma comparação entre dois grupos com o teste log-rank, o que está relacionado com a quantidade de observações e falhas nos grupos.

(a) *logRatio CUSUM versus Qui-Quadrado (Q_t)* (b) *logRatio CUSUM versus Qui-Quadrado (Q_t)*

Figura 3.3: Monitoramento do *logRatio CUSUM* (a) com $\rho = 0$, $a_t = 1$ e $b_t = 1$ e (b) com $\rho = 0$, $a_t = \sqrt{Q_t/Q_{t-1}}$ e $b_t = 1$.

Contudo, os escores construídos desta forma estão atrelados à variação da estatística de forma individual (vê apenas a mudança de um dia para o seguinte). Com isso, outra formulação com ideia semelhante à que foi proposta anteriormente considera a razão entre os valores médios de Q_t , como mostrado em (3.3).

$$w_t = I\{\bar{Q}_{t-1} \neq 0\} a_t \log \left(b_t \frac{\bar{Q}_t}{\bar{Q}_{t-1}} \right), \quad (3.3)$$

em que $a_t \geq 0$ e $b_t > 0$ são pesos para ponderar o escore a cada dia de inspeção. Denota-se esse gráfico por *logMeanRatio CUSUM*. Nesta fórmula, $\bar{Q}_t$ representa a média de todos os valores obtidos para a estatística qui-quadrado até o momento t . Desta forma, avalia-se a cada dia um aumento na média de todo o histórico desse valor de teste como forma de alertar para uma evidência de diferença entre os grupos comparados.

Uma outra formulação de CUSUM considera a diferença entre os valores médios de Q_t , como mostrado em (3.4).

$$w_t = a_t (\bar{Q}_t - \bar{Q}_{t-1}), \quad (3.4)$$

com $a_t > 0$, que simboliza uma ponderação a ser atribuída para cada escore. Denota-se esse gráfico por *MeanDiff CUSUM*. De forma similar à proposta anterior, tem-se que um aumento significativo no valor médio da estatística de teste deve produzir um alerta pelo gráfico CUSUM que se baseia em acúmulos destes escores.

É possível de forma geral definir duas quantidades de interesse a serem comparadas no momento atual e no passado, $K_t = g_1(Q_1, Q_2, \dots, Q_t)$ e $K'_t = g_2(Q_1, Q_2, \dots, Q_t)$, em que $g_1(\cdot) \geq 0$ e $g_2(\cdot) \geq 0$ são funções das estatísticas de teste Q_t observadas até o momento t , e espera-se que a diferença entre K_t e K'_t aumente conforme a hipótese alternativa seja mais evidente. Com isso, define-se o gráfico logKRatio CUSUM da forma:

$$w_t = I\{K'_t \neq 0\} a_t \log \left(b_t \frac{K_t}{K'_t} \right) \quad (3.5)$$

e o *KDiff CUSUM*:

$$w_t = a_t (K_t - K'_t), \quad (3.6)$$

Pode-se considerar por exemplo $K_t = \max_{1 \leq i \leq t} Q_i$ e $K'_t = \max_{1 \leq i \leq t-1} Q_i$ para avaliar se o máximo das estatísticas de teste aumenta significativamente conforme o tempo (ou utilizar $K_t = Q_t$, para permitir que o gráfico decresça caso novos valores da Q_t comecem a ser inferiores ao máximo observado); ou um gráfico menos rigoroso pode adotar $K'_t = \text{mediana}(Q_1, Q_2, \dots, Q_{t-1})$ contra $K_t = \text{mediana}(Q_1, Q_2, \dots, Q_t)$ ou $K_t = Q_t$.

(a) *logMeanRatio CUSUM versus Qui-Quadrado* (b) *MeanDiff CUSUM versus Qui-Quadrado (Q_t)*

Figura 3.4: Monitoramento do CUSUM (a) *logMeanRatio* com $\rho = 0$ e $a_t = 1$ e $b_t = 1$ e (b) *MeanDiff* com $\rho = 0$ e $a_t = 1$.

A Figura 3.4 apresenta aplicações destes gráficos com os dados do InCor. Observa-se que o gráfico MeanDiff CUSUM tende a seguir o padrão de crescimento que se observa nos valores da qui-quadrado, enquanto que o logMeanRatio CUSUM já sobe ligeiramente nos primeiros dias de monitoramento. Isto traz discussões novamente com respeito à razoabilidade do teste nos primeiros dias, quando há poucas observações e pouco tempo de acompanhamento para o grupo de monitoramento, o que resulta em pequena quantidade de falhas e poucos indivíduos em risco provindos deste grupo. Entende-se assim a importância de se considerar diferentes pesos (por exemplo, dar menor peso a estes primeiros dias de monitoramento no gráfico) para evitar a produção de alarmes falsos pelo CUSUM já no início.

Devido aos valores inferiores a 1 que ocorrem logo no começo dos cálculos, é possível observar

por exemplo escores do tipo $\log\left(\frac{0,1}{0,01}\right) \simeq 2,3$, que claramente não deviam produzir um alerta no gráfico, mas resultam em um valor alto para acumular no CUSUM e acarretam esse crescimento irregular no início dos gráficos que utilizam a razão de qui-quadrados na fórmula. Dar pesos de valor zero (ou quase) para essas situações seria uma forma de contornar este tipo de problema.

Outra maneira de construir um escore CUSUM é considerar incrementos na própria estatística de teste $Z_t = Z_1(\tau)$, em vez da qui-quadrado. Ou seja, temos:

$$w_t = a_t(Z_t - Z_{t-1}), \quad (3.7)$$

em que $a_t \in \mathbb{R}$. Com isso, é possível ter um gráfico que detecta aumento na sobrevivência (com $a_t > 0$) e também um que pode detectar redução ($a_t < 0$). Denotamos este gráfico por *ZDiff CUSUM* e uma ilustração do comportamento deste é apresentada na Figura 3.5.

(a) CUSUM com 975 pacientes no grupo controle (b) CUSUM com 833 pacientes no grupo controle

Figura 3.5: Monitoramento do ZDiff CUSUM com $\rho = 0$ e $a_t = 1$ x $a_t = -1$ versus $Z_1(\tau)$ considerando como grupo de controle (a) os 975 primeiros pacientes e (b) os 833 primeiros pacientes.

Na Figura 3.5a são plotados dois gráficos que simultaneamente monitoram aumento ou redução na sobrevivência, acompanhados do gráfico da estatística de teste $Z_1(\tau)$, feito com uma escala diferente (exibida no lado direito) na mesma região gráfica. O gráfico desenhado acima do eixo horizontal zero (referente à escala do CUSUM) considera $a_t = 1$, e então verifica se há uma melhoria na sobrevida dos pacientes do segundo grupo em relação aos primeiros 975 (grupo controle), enquanto que o gráfico plotado abaixo (que é o CUSUM calculado com $a_t = -1$ plotado com o sinal trocado) monitora uma redução na sobrevivência dos pacientes ao longo do tempo. O gráfico apresentado na Figura 3.5b é feito de forma semelhante, mas considera como grupo controle as primeiras 833 observações da coorte 1, o que representa dados até o ano de 1997, 1 semestre antes da mudança sugerida pelos pesquisadores (que equivale a considerar dados até o dia 2295 do acompanhamento). Com isso, marca-se no gráfico em uma linha vertical o ponto que corresponde a este instante de mudança (o último momento em que um paciente dá entrada no protocolo antes da mudança), e espera-se que um método de controle venha a detectar alteração na sobrevivência somente após este ponto.

De fato o valor do CUSUM superior se mantém baixo antes desta linha, e começa a crescer após este ponto, com aumento expressivo na faixa dos 5000 dias. Para o CUSUM inferior, observa-se que

se preserva estável (próximo de zero) na maior parte do tempo, apresentando uma queda por volta dos 6000 dias, mas que depois se estabiliza. Tal acontecimento pode ser devido à proximidade do momento final deste estudo, em que há poucos indivíduos em risco, e novamente ocorre um possível prejuízo no cálculo da estatística de teste. Contudo, é notável que o método se mostra eficaz na detecção do aumento na sobrevida dos pacientes, algo que é esperado pelos pesquisadores com a mudança de tratamento que foi proposta.

Ilustrações semelhantes a esta considerando diferentes tamanhos de amostra para os grupos foram produzidas e estão disponíveis em <<https://www.ime.usp.br/~acarlos/CUSUM>>. Por exemplo, o arquivo nomeado como **InCor1_LR_CUSUM_n0.833_d6905.pdf** contém diversos gráficos CUSUM construídos com essas formulações para monitoramento que considera o grupo de controle com os 833 primeiros pacientes.

São apresentadas diferentes possibilidades para o peso considerado no escore CUSUM, e motivados pela necessidade de aguardar mais tempo para se conseguir uma comparação melhor entre as curvas de sobrevida temos ilustrações com diferentes descartes de inspeção (uma certa quantidade de dias é ignorada, e os cálculos são feitos só após isso). Outro exemplo considera no controle 881 observações (isto é, somente até o final do mês junho de 1998) e permite também investigar como o gráfico se comporta antes e depois do ponto da mudança, mas neste caso um pouco mais próximo desta data. Em todos é notável uma estabilidade melhor para o gráfico que utiliza diferenças entre estatísticas $Z_1(\tau)$.

É necessário discutir ainda que valor se deve atribuir para o limite de controle nestes gráficos, de modo a saber em que momento eles alertariam alguma mudança. Para o gráfico da estatística de teste $Z_1(\tau_i)$ adaptado ao monitoramento prospectivo, os quantis da normal referentes aos níveis de significância desejados podem ser uma alternativa, no entanto, deve-se levar em conta que os testes feitos de forma sequencial podem comprometer o nível de confiança global do teste, e daí os limites deviam ser bem mais amplos que o usual valor 1,96 de modo a evitar alarmes falsos, mas não tão grandes para não retardar também a detecção de uma alteração verdadeira na qualidade. Há então que se desenvolver métodos para obtenção de valores ótimos para essas aplicações do gráfico na prática.

Para este mesmo banco de dados do InCor, e com o grupo de controle definido como os primeiros 975 pacientes, uma aplicação paramétrica de monitoramento foi utilizada por Oliveira et al. (2016) e uma detecção de melhoria na sobrevida dos pacientes foi apontada pelo gráfico RACUF CUSUM. Ao considerar um limite de controle 5 (que é superior ao valor estipulado nesta aplicação), isto ocorre na observação 1214 dos dados completos, que corresponde a um indivíduo da segunda coorte (com data de entrada em 28/12/2001). Com isso, o gráfico percebe a alteração no perfil da sobrevida em torno de 3 anos após o ponto sugerido pelos médicos.

No entanto, o valor deste limite de controle deve atender medidas de desempenho ótimas, e a importância disso é discutida nessa aplicação. Oliveira et al. (2016) propõem uma forma de obter um valor ótimo o limite, e para essa aplicação conseguem que o CUSUM detecte uma mudança na observação 1052 (indivíduo que dá entrada no início do mês outubro de 1999 (correspondente ao dia 3129 do acompanhamento), que é um pouco menos que 1 ano após o início do monitoramento). Contudo, deve-se ainda se considerar que esta análise paramétrica é apresentada de forma retrospectiva, e assim tem a desvantagem de não fornecer um sinal de mudança durante a transição dos procedimentos. Uma adaptação deste método para essa estrutura de monitoramento prospectivo

Figura 3.6: monitoramento do RACUF CUSUM com $\xi_1 = 1.1$ e $\xi_1 = 0.9$ (a) na data 19/09/2003 e (b) em 01/01/2004.

foi construída para fazer esta comparação, e foi constatado que o gráfico só sinaliza uma mudança na data 19/09/2003 (já bem mais tardia), como pode ser visto na Figura 3.6a. Neste caso o primeiro sinal é observado em particular na observação 1432, que é um indivíduo com entrada em 19/08/2003, ou seja, 1 mês antes do momento em que o gráfico consegue detectar pela primeira vez.

Neste gráfico, são representadas duas intensidades de mudança, um aumento de 10% no tempo médio de sobrevivência (isto é, $\xi_1 = 1.1$) na região superior e uma redução de 10% ($\xi_1 = 0.9$, representado como valores negativos do CUSUM na região inferior). O gráfico da direita mostra um momento em torno de 3 meses depois, quando é mais visível o comportamento de detecção do CUSUM para estes dados.

Nesse aspecto, o método semiparamétrico que propomos do ZDiff CUSUM já parece ser mais ágil ao detectar diferença entre os grupos monitorados, pois com o mesmo limite de controle 5 apresenta um sinal no gráfico que monitora aumento da sobrevida dos pacientes por volta de 4000 dias no tempo de calendário (que corresponde ao início do ano 2002), pouco mais que 3 anos após o momento em que se acredita ocorrer uma melhora. Assim, ele pode ser uma alternativa para monitorar sem necessidade de definir uma distribuição para os tempos de falha, algo necessário para construir os gráficos ajustados ao risco que foram discutidos.

3.3 Comparação das abordagens em dados simulados

Para avaliar o comportamento deste gráfico de controle em diferentes situações, realizamos algumas aplicações com dados simulados. Os tempos de sobrevivência são gerados conforme o modelo de tempo de promoção exponencial com duas covariáveis do tipo fator de dois níveis e parâmetros adequados para compor 4 cenários distintos em relação aos percentuais de censura e cura: com fração de cura 20% (e duas possibilidades de censura em relação à amostra completa, 40% e 20%) ou sem fração de cura (e percentuais de censura 20% ou 0%). Para cada um dos cenários considera-se ainda duas possíveis mudanças no perfil da sobrevida ($\xi_1 = 2$ ou $\xi_1 = 0.5$), que representam respectivamente uma duplicação no tempo médio de sobrevivência ou uma redução

à metade, e a situação em que não há mudança ($\xi_1 = 1$). Para o grupo 1, sob controle, são geradas $n_0 = 1000$ observações, e considera-se $n_1 = 2000$ indivíduos para o grupo 2.

(a) Linhas de tempos dos dados

(b) Kaplan-Meier comparativo dos grupos, com $n_0 = 1000$ versus $n_1 = 2000$

(c) Kaplan-Meier comparativo dos grupos remane- jados, com $n_0 = 900$ versus $n_1 = 2100$

(d) Kaplan-Meier comparativo dos grupos remane- jados, com $n_0 = 500$ versus $n_1 = 2500$

Figura 3.7: Avaliação do perfil dos dados gerados com mudança de intensidade $\xi_1 = 2$, para o cenário com 40% de censura e 20% de cura.

Para os dados com mudança $\xi_1 = 2$ apresentamos na Figura 3.7 o comportamento das linhas de tempo dos indivíduos no calendário e a comparação das curvas de sobrevida estimadas para 3 diferentes situações de divisão da amostra. Primeiro considera-se os grupos da forma que foram gerados, depois compara-se os 900 primeiros indivíduos do grupo controle com um grupo de monitoramento formado pelos 100 seguintes em conjunto com o grupo 2 (fora de controle), o que ilustra uma situação em que o monitoramento começa um pouco antes do ponto de mudança. E a terceira situação leva em conta apenas os 500 primeiros indivíduos do grupo controle em comparação com os demais, o que faz uma mistura maior de indivíduos do grupo controle com os do grupo 2.

Nota-se a dificuldade em concluir visualmente pelo gráfico das linhas de acompanhamento que houve uma melhora na sobrevivência após as primeiras 1000 observações, e isso enfatiza a importância da construção dos métodos capazes de identificar essas alterações no perfil da sobrevida

de forma objetiva. As curvas de sobrevivência estimadas se mostram diferentes, com a do grupo monitorado sempre superior, nas três situações comparadas, mas no terceiro caso elas já ficam bem próximas devido à diferente subdivisão da amostra completa que junta parte do grupo controle com o segundo, e neste caso deve ser mais difícil que um gráfico de controle perceba ao longo do tempo a melhora na sobrevida dos indivíduos. Na prática é então importante definir adequadamente a amostra sob controle a ser tomada como referência para que o monitoramento seja eficiente.

Figura 3.8: Monitoramento dos gráficos CUSUM com teste log-rank em dados com mudança de intensidade $\xi_1 = 2$, para o cenário com 40% de censura e 20% de cura.

A Figura 3.8 mostra o comportamento dos diferentes gráficos CUSUM considerados na situação em que todo o grupo controle é tomado como amostra de referência para comparação. Percebe-se que o MeanDiff CUSUM tende a seguir uma trajetória parecida com a da estatística qui-quadrado. Já os gráficos que utilizam a razão de qui-quadrados têm uma trajetória diferente, porém apresentam escores acumulados muito grandes logo no começo, o que reflete a instabilidade na comparação de grupos nos primeiros dias de monitoramento e também o problema de razões entre valores inferiores a 1. O gráfico ZDiff, que difere dos outros métodos por permitir detecção de diferenças entre os grupos nas duas direções, parece ser o mais consistente nesta situação, e consegue acumular no CUSUM superior que detecta a melhoria da sobrevivência. Devemos então discutir como determi-

nar a melhor utilização dos gráficos CUSUM para detectar diferenças entre grupos com base na estatística qui-quadrado.

Figura 3.9: Monitoramento dos gráficos CUSUM com teste log-rank em dados com mudança de intensidade $\xi_1 = 2$, para o cenário com 40% de censura e 20% de cura, com remanejamento de metade da amostra sob controle para o grupo monitorado.

Novamente, observa-se uma maior estabilidade do ZDiff CUSUM no monitoramento feito com a amostra de treinamento que leva em conta apenas metade do grupo controle, conforme exibido na Figura 3.9. Para esta aplicação é notada incapacidade do MeanDiff e logMeanRatio em acumular escores para detectar a diferença entre os grupos (possivelmente pelo efeito de suavização da média), e deve-se então rever a construção destes escores formulados nos valores médios da estatística de teste. Já para o logRatio há um crescimento que se inicia no CUSUM entre os dias 1500 e 2000, que apesar da instabilidade no começo deste período (provavelmente devido a razões por valores menores que 1) pode competir na detecção com um devido ajuste por pesos nos escores diários. Tanto para o logRatio quanto para o ZDiff não ocorre um acúmulo preocupante no CUSUM antes da linha vertical (que se refere ao último momento em que indivíduos chegam da amostra de treinamento), o que configura um alarme falso, e para o ZDiff principalmente, percebe-se um crescimento permanente do CUSUM após esse instante, que sugere eficácia do método para advertir sobre uma alteração na sobrevida.

Para outros cenários de percentual da censura e cura no dados, e também para dados com mudança de intensidade $\xi_1 = 0.5$ (redução no padrão de sobrevida), foi observado um comportamento semelhante desses gráficos, com maior estabilidade do ZDiff CUSUM. Estes resultados podem ser consultados em arquivos disponibilizados na pasta compartilhada informada anteriormente. Discutimos a seguir como eles funcionam em dados cujo grupo monitorado é gerado da mesma distribuição sob controle, ou seja, não há mudança no padrão da sobrevida ($\xi_1 = 1$).

(a) Linhas de tempos dos dados

(b) Kaplan-Meier comparativo dos grupos, com $n_0 = 1000$ versus $n_1 = 2000$

(c) Kaplan-Meier comparativo dos grupos remane- jados, com $n_0 = 900$ versus $n_1 = 2100$

(d) Kaplan-Meier comparativo dos grupos remane- jados, com $n_0 = 500$ versus $n_1 = 2500$

Figura 3.10: Avaliação do perfil dos dados gerados sem mudança ($\xi_1 = 1$), para o cenário com 40% de censura e 20% de cura.

Como exibido na Figura 3.10, as curvas de sobrevivência estimadas são praticamente idênticas nas três divisões da amostra completa. Logo, espera-se que o CUSUM não acumule e permaneça oscilando entre valores pequenos no gráfico, como deve ser o comportamento da própria estatística de teste qui-quadrado. A Figura 3.11 mostra este monitoramento na comparação dos dois grupos na formação em que foram gerados, e exceto pelo logRatio, todos se mantêm oscilando entre valores absolutos menores que 3, sem apresentar um perfil de crescimento (ou decréscimo no caso do ZDiff) persistente ao final do monitoramento. Novamente nota-se que o gráfico logRatio é

inconsistente, por apresentar valores no CUSUM superiores a 10.

Figura 3.11: Monitoramento dos gráficos CUSUM com teste log-rank em dados sem mudança ($\xi_1 = 1$), para o cenário com 40% de censura e 20% de cura.

Para o monitoramento que considera somente metade do grupo 1 gerado como amostra de referência para comparação, segue um comportamento parecido dos gráficos, como mostrado na Figura 3.12. Antes da linha vertical, os gráficos *logMeanRatio* e *MeanDiff* apresentam o CUSUM distribuído entre valores pequenos, e após a linha vermelha há um crescimento notável nos escores acumulados, mas que ainda se mantém em um valor baixo. Neste caso, a linha demarca o momento em que a última observação do grupo 1 originalmente gerado entra no monitoramento, mas todo o conjunto de dados é formado por indivíduos sob controle, e então espera-se que não haja crescimento do CUSUM, o que configura alarme falso nesta situação.

O gráfico *ZDiff CUSUM* tem por diferencial a possibilidade de reconhecer se a possível diferença no padrão da sobrevida entre dois grupos comparados se dá em favor ao grupo monitorado ou ao grupo de referência, uma vez que pode testar se há alteração neste padrão nos dois sentidos. Devemos enfatizar este método, mas também consideramos que o CUSUM para monitorar diferenças entre os grupos com base no comportamento da estatística qui-quadrado é interessante, e na prática caso ocorra uma detecção de diferença, o sentido desta pode ser verificado com métodos

Figura 3.12: Monitoramento dos gráficos CUSUM com teste log-rank em dados sem mudança ($\xi_1 = 1$), para o cenário com 40% de censura e 20% de cura, com remanejamento de metade da amostra sob controle para o grupo monitorado.

alternativos, por exemplo ao observar o simples confronto entre curvas de sobrevivência estimadas para cada grupo é possível saber que a diferença favorece o grupo monitorado ou alerta para uma piora.

A seguir destacamos o comportamento apenas do ZDiff CUSUM para o cenário de mudança $\xi_1 = 0.5$. Como pode ser observado na Figura 3.13, a curva estimada para o grupo monitorado fica inferior à do grupo controle, o que descreve uma situação de redução no tempo médio de sobrevivência. Embora seja visível a diferença entre esses grupos, na aplicação foi observada uma dificuldade para o monitoramento CUSUM, como pode ser visto na Figura 3.14. Possivelmente devido ao alto percentual de censura nos dados (e também o fato de haver uma fração de cura) a estatística de teste normal decresce mas não o suficiente para acumular muitos escores no CUSUM, o que retarda a detecção da queda na sobrevida. Contudo, o CUSUM que monitora redução ainda consegue ficar abaixo de -4 em todos os gráficos.

Entretanto, aparece na Figura 3.14b um crescimento alto antes da linha vertical (que configura um alarme falso), e notamos que isso ocorre pela alteração de um valor negativo para um positivo da estatística normal, algo que acumula muito no CUSUM superior e está ligado também com a

instabilidade inerente ao método nos primeiros dias de monitoramento.

(a) Linhas de tempos dos dados

(b) Kaplan-Meier comparativo dos grupos, com $n_0 = 1000$ versus $n_1 = 2000$

(c) Kaplan-Meier comparativo dos grupos remane- jados, com $n_0 = 900$ versus $n_1 = 2100$

(d) Kaplan-Meier comparativo dos grupos remane- jados, com $n_0 = 500$ versus $n_1 = 2500$

Figura 3.13: Avaliação do perfil dos dados gerados com mudança ($\xi_1 = 0.5$), para o cenário com 40% de censura e 20% de cura.

Isso enfatiza a importância de considerar pesos menores para esses escores ou descartá-los (peso zero), de modo a evitar produção de alarmes falsos pelo gráfico logo no começo. Por exemplo, a Figura 3.14d mostra este mesmo CUSUM com o descarte dos primeiros 30 dias do monitoramento, e nota-se a remoção total do erro de detecção do CUSUM superior antes da linha vertical, e ele se comporta como deve ser quando não há evidências de aumento no padrão de sobrevivência, mas sem com isso prejudicar a detecção correta da redução na sobrevida pelo CUSUM inferior. Neste caso poucos dias são desperdiçados, e deve-se discutir então possíveis critérios para se definir um ponto ótimo de partida para monitorar, que seja confiável e o mais breve possível.

Outras simulações não exibidas aqui, que estão disponíveis na pasta compartilhada, confirmam que em geral é mais fácil identificar diferenças na sobrevida dos grupos quando há menor percentual de censuras. Ainda assim, no cenário discutido aqui que tem 40% de censura e ainda apresenta um percentual de cura 20% foi possível detectar com o CUSUM a diferença entre os grupos. A

Figura 3.14: Monitoramento dos gráficos ZDiff CUSUM com teste log-rank em dados com mudança de intensidade $\xi_1 = 0.5$, para o cenário com 40% de censura e 20% de cura.

dificuldade maior parece ser em algumas das situações quando as curvas Kaplan-Meier estimadas estão muito próximas ou se cruzam, mas como se trata de uma simulação ilustrativa isso pode ser atribuído à semente utilizada para gerar o conjunto de dados específico, e com isso mesmo configurados para gerar um grupo fora de controle com tempo médio de falha 2 vezes maior ou menor os dados se mostram semelhantes na comparação das curvas de sobrevivência estimadas.

Além disso, no cenário com mudança de intensidade $\xi_1 = 0.5$ exibido anteriormente, parece ser mais demorada a detecção, mesmo com curvas de sobrevivência visualmente diferentes, algo que deve estar associado ao fato de ter poucos tempos de falha para compor a estatística de teste normal, que neste caso atingiu um valor mínimo em torno de -5, algo bem menos expressivo do que se pode constatar em outras situações ilustradas. Também para alguns casos de intensidade $\xi_1 = 2$, em que foi feito o remanejamento de indivíduos do grupo controle para o monitoramento, uma dificuldade é imposta para o método devido à mistura de observações geradas sob distribuições diferentes. Neste caso a estatística de teste normal chega a no máximo 2, mas ainda assim o CUSUM consegue acumular escores e mostrar crescimento que leva à detecção. Como há ainda o aspecto prospectivo que interfere na espera para obtenção das informações, que é um desafio já

mencionado e que deve trazer discussões mais profundas, acreditamos que o método já parece se mostrar capacitado. Entende-se então que há vários fatores que precisam ser levados em conta na avaliação do desempenho destes métodos, e um estudo mais extensivo de simulação deve ser feito posteriormente, com a comparação do comportamento destes métodos em confronto com alguns desenvolvidos na literatura.

3.4 Obtenção de limites de controle e desempenho do gráfico ZDiff CUSUM

Uma das dificuldades da aplicação do teste log-rank está na característica da composição dos grupos comparados, que deve atender a alguma quantidade mínima de indivíduos ou eventos observados para que a estatística de teste descrita retorne um resultado confiável. Diversos autores discutem a respeito do tamanho amostral ideal para realizar a comparação de dois grupos com o log-rank. No entanto, tais discussões referem-se a um propósito de estudo controlado, em que se deseja realizar apenas um teste comparativo depois de concluído o acompanhamento de indivíduos pertencentes a dois grupos de interesse.

Contudo, nosso interesse está em realizar testes sequenciais, a partir do momento em que se tem indivíduos de um grupo a ser monitorado, para ser confrontado a um grupo de referência já definido e acompanhado previamente. Com isso, é provável que no começo do monitoramento haja poucos indivíduos (e conseqüentemente poucas falhas observadas) no segundo grupo, o que torna a estatística log-rank mais instável para fazer a comparação. O teste sequencial tem por vantagem justamente fornecer análises úteis sobre um experimento no decorrer de sua realização, neste caso, poder notar uma melhora ou piora na sobrevida de pacientes o quanto antes, sem a necessidade de esperar todo o acompanhamento previsto pelos pesquisadores para chegar a essa conclusão.

Na proposta em questão, que avalia o comportamento da estatística de teste log-rank ao longo de tempo, deve-se então atentar para a necessidade de que haja um período mínimo de acompanhamento do segundo grupo antes de se iniciar a construção do gráfico de controle. A instabilidade que pode ocorrer, principalmente nos primeiros dias de estudo, pode estar associada com a variabilidade do número de falhas, que é muito pequena no começo e faz com que a estatística de teste possa retornar um valor muito grande (já que o denominador é a raiz quadrada dessa variância), sem que necessariamente isso seja uma evidência de diferença entre as curvas de sobrevida dos dois grupos, o que pode então provocar alarmes falsos no monitoramento.

Outra característica desse método é que todas as falhas observadas (ou uma grande parcela) para construção da estatística de teste são provavelmente provenientes do grupo de referência no período inicial de monitoramento. Desta forma, há uma tendência de que o valor de $Z_1(\tau)$ seja positivo por certo período (dando evidência à hipótese alternativa), até que se comece a observar falhas no grupo 2, permitindo que esta possa então assumir um valor negativo. Aguardar um período antes de se iniciar a construção do gráfico de controle é portanto importante para evitar alarmes falsos logo no início.

Dentre as propostas de gráficos CUSUM com base no teste log-rank, notamos previamente um desempenho mais eficiente para o ZDiff CUSUM, então iremos concentrar as próximas análises neste método específico. Na literatura é comum o uso de um valor de referência na construção do escore CUSUM, e acreditamos que isto pode ser útil nesse método para também evitar alarmes

falsos que possam acontecer no decorrer do monitoramento, especialmente nos períodos iniciais que apresentam maior instabilidade, como já discutido. O CUSUM com valor de referência é da forma

$$C_t = \max(0, C_{t-1} + w_t - k), \quad (3.8)$$

em que $k > 0$ é o valor de referência.

Uma característica observada na estatística log-rank sequencial para dois grupos determinados é permanecer constante por alguns intervalos de tempo, devido ao fato de não haver novas informações sobre falhas ou indivíduos em risco para compor o cálculo. Desta forma, definimos um escore CUSUM que subtrai um valor de referência somente quando o escore obtido com o incremento no log-rank é positivo, de modo a evitar que o gráfico CUSUM decresça nos períodos de constância no valor $Z_1(\tau)$. Assim, o gráfico ZDiff CUSUM fica determinado pelo escore:

$$w_t = a_t[Z_t - Z_{t-1}] - kI\{a_t[Z_t - Z_{t-1}] > 0\}, \quad (3.9)$$

em que $k > 0$ é um valor de referência, $a_t > 0$ monitora um aumento na sobrevivência e $a_t < 0$ monitora uma redução. O papel do valor de referência é compensar pequenos desvios positivos (no sentido de uma alternativa unilateral) que podem ocorrer aleatoriamente no log-rank mas que não devem ser evidência suficiente para a rejeição da hipótese nula. O ideal é determinar um valor de k que seja grande o suficiente para amenizar os alarmes falsos, mas sem comprometer o poder de detecção de uma diferença significativa.

Recentemente, Wu & Xiong (2017) discutiram o fato de haver desbalanceamento no tamanho amostral de cada grupo quando se aplica o teste log-rank. Os autores propõem um procedimento diferente para auxiliar o pesquisador a saber o número de indivíduos que devem ser observados para que se possa realizar uma comparação dos grupos com o teste log-rank em uma situação com proporção muito maior de participantes em um determinado grupo. Eles mostram que essa expressão é mais coerente, em comparação com a usual proposta originalmente por Schoenfeld (Schoenfeld, 1981),

$$d(\tau) = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{\omega_1 \omega_2 [\log(RR)]^2}, \quad (3.10)$$

em que τ é o tempo de duração do estudo, α é o nível de significância (teste unilateral), $1 - \beta$ é o poder, ω_1 e $\omega_2 = 1 - \omega_1$ são as probabilidades de que um indivíduo seja do grupo 1 ou 2, respectivamente, e RR é a razão de riscos do grupo 1 em relação ao grupo 2. Desta forma, $d(\tau)$ define a quantidade de eventos que devem ter sido observados até o momento τ para atender às características do teste log-rank desejadas. Seja $p_i(\tau)$ a probabilidade de que um indivíduo do grupo i apresente um evento no período de acompanhamento $[0, \tau]$, para $i = 1, 2$, e seja $P(\tau) = \omega_1 p_1(\tau) + \omega_2 p_2(\tau)$ a probabilidade de se observar um evento para determinado indivíduo aleatoriamente escolhido dos dois grupos com probabilidades ω_1 e ω_2 , respectivamente. Assim, o número total de pacientes necessários para o experimento é $n = d(\tau)/P(\tau)$, que conduz ao tamanho amostral:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{\omega_1 \omega_2 [\log(RR)]^2 P(\tau)}. \quad (3.11)$$

Wu & Xiong (2017) abordam uma extensão dessa fórmula, que eles mostram ser mais precisa

em situações com desbalanceamento do número de indivíduos em cada grupo. A expressão para o tamanho amostral é dada por:

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 P(\tau)}{\omega_1 \omega_2 [\log(RR)]^2 p_1(\tau) p_2(\tau)}. \quad (3.12)$$

Com isso, pode-se calcular o tamanho da amostra e o número de eventos em cada grupo i , $n_i = n\omega_i$ e $d_i(\tau) = n_i p_i(\tau)$, $i = 1, 2$, respectivamente. Para o cálculo das probabilidades de eventos $p_i(\tau)$, os autores simplificam ao considerar que não há perda de acompanhamento dos indivíduos, e que estes dão entrada no estudo de acordo com uma distribuição uniforme no período de recrutamento t_a e obtêm:

$$p_i(\tau) = 1 - \frac{1}{t_a} \int_{t_f}^{\tau} S_i(t) dt, \quad (3.13)$$

com $S_2(t) = [S_1(t)]^{1/RR}$ e $\tau = t_a + t_f$, em que t_f é o período de acompanhamento dos pacientes. Eles ainda sugerem que a integral na expressão pode ser aproximada satisfatoriamente pela regra de Simpson, e com isso se tem:

$$p_i(\tau) = 1 - \frac{1}{6} \{S_i(t_f) + 4S_i(0.5t_a + t_f) + S_i(t_a + t_f)\}, \quad (3.14)$$

em que $S_1(t)$ pode ser uma estimativa de Kaplan-Meier, uma distribuição paramétrica de sobrevivência ou uma curva de sobrevivência *spline* dos dados históricos do grupo de controle.

Para o nosso propósito de estudo, entretanto, não se pretende alocar pacientes a um grupo controle ou experimental. Em vez disso, consideramos que há um grupo controle já definido previamente, e que ao longo do tempo desejamos acompanhar novos pacientes que são considerados como grupo de monitoramento, a ser comparado pelo log-rank com o grupo de referência. Logo, o tamanho da amostra do grupo controle é fixado, e resta somente definir um tamanho amostral mínimo para o grupo monitorado que se deve alcançar para que o log-rank possa ser aplicado com segurança.

Assim, devemos calcular n_2 em função do tamanho total da amostra necessário para utilizar o teste. Fazendo $A = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{[\log(RR)]^2}$ na expressão (3.12), e admitindo que a probabilidade de pertencer a um determinado grupo $j = 1, 2$ pode ser calculada como $\omega_j = n_j/n$, obtém-se o valor de n_2 , como desenvolvido na equação (3.15). Com isso, fixado um nível de significância α e um poder $1 - \beta$ para o teste, e um tamanho amostral do grupo de controle, pode-se obter o tamanho de amostra mínimo desejável para o grupo 2.

Outro ponto importante para definir o gráfico CUSUM é utilizar um valor de referência k ideal para detectar uma determinada mudança. Hawkins & Olwell (1998) discutem como definir o valor de k , com ênfase para a situação em que a variável monitorada tem distribuição normal. Neste caso, define-se um valor de $k = \Delta/2$, com Δ especificado para monitorar uma mudança na média de tal forma que $\mu_1 = \mu_0 + \Delta\sigma_0$, em que μ_1 é a média fora de controle e μ_0 e σ_0 são, respectivamente, a média e desvio padrão sob controle. Os autores discutem a determinação deste valor para outras distribuições, e sugerem um método para calcular k quando a distribuição é da família exponencial.

$$\begin{aligned}
n_2 + n_1 &= (\omega_1 \omega_2)^{-1} A \frac{P(\tau)}{p_1(\tau) p_2(\tau)} \\
n_2 + n_1 &= \left(\frac{n_1}{n_1 + n_2} \frac{n_2}{n_1 + n_2} \right)^{-1} A \left(\frac{w_1 p_1(\tau) + w_2 p_2(\tau)}{p_1(\tau) p_2(\tau)} \right) \\
n_2 + n_1 &= \frac{(n_1 + n_2)^2}{n_1 n_2} A \left(\frac{w_1}{p_2} + \frac{w_2}{p_1} \right) \\
n_2 + n_1 &= \frac{(n_1 + n_2)^2}{n_1 n_2} A \left(\frac{n_1}{n_1 + n_2} p_2^{-1} + \frac{n_2}{n_1 + n_2} p_1^{-1} \right) \\
n_2 + n_1 &= \frac{(n_1 + n_2)^2}{n_1 n_2} A \left(\frac{n_1 p_2^{-1} + n_2 p_1^{-1}}{n_1 + n_2} \right) \\
1 &= \frac{A}{n_1 n_2} (n_1 p_2^{-1} + n_2 p_1^{-1}) \\
n_1 n_2 &= A n_1 p_2^{-1} + A n_2 p_1^{-1} \\
n_2 (n_1 - A p_1^{-1}) &= A n_1 p_2^{-1} \\
n_2 &= \frac{A n_1 p_2^{-1}}{n_1 - A p_1^{-1}}.
\end{aligned} \tag{3.15}$$

Se a variável aleatória T tem distribuição pertencente à família exponencial, é possível escrever a densidade $f(t)$ como:

$$f(t|\theta) = \exp\{a(t)b(\theta) + c(t) + d(\theta)\}, \tag{3.16}$$

em que $a(t)$ e $c(t)$ são funções que não dependem de parâmetros desconhecidos e $b(\theta)$ e $d(\theta)$ dependem do parâmetro de interesse θ e nenhuma quantidade desconhecida. Para o caso em que $T \sim \text{Exp}(\theta)$, temos:

$$f(t) = \theta e^{-\theta t} = \exp\{-\theta y + \log(\theta)\}. \tag{3.17}$$

Assim, $a(t) = t, b(\theta) = -\theta, c(t) = 0$ e $d(\theta) = \log(\theta)$. Conforme Hawkins & Olwell (1998), o valor de referência k para monitorar uma mudança em um parâmetro θ da distribuição de um valor θ_0 para θ_1 é dado por:

$$k = -\frac{d(\theta_1) - d(\theta_0)}{b(\theta_1) - b(\theta_0)}. \tag{3.18}$$

Para o caso da distribuição exponencial, desejamos monitorar uma mudança em θ de θ_0 para $\theta_1 = \xi \theta_0$, e, assim, obtém-se:

$$\begin{aligned}
k &= -\frac{\log(\theta_1) - \log(\theta_0)}{-\theta_1 - (-\theta_0)} \\
&= -\frac{\log(\xi \theta_0) - \log(\theta_0)}{-(\theta_1 - \theta_0)} \\
&= \frac{\log(\xi)}{(\xi - 1)\theta_0}.
\end{aligned} \tag{3.19}$$

A expressão (3.19) está vinculada ao valor de θ_0 , o que pode retornar valores muito grandes de k

sem necessidade. Uma alternativa à esta abordagem é considerar o aumento na média, de modo similar ao caso da distribuição normal.

Sabe-se que $\mu_0 = E(T_0) = 1/\theta_0$ e $\mu_1 = E(T_1) = 1/(\xi\theta_0) = \xi^{-1}\mu_0$. Além disso, $\sigma_0 = \sqrt{\text{Var}(T_0)} = \sqrt{1/\theta_0^2} = 1/\theta_0 = \mu_0$, então temos:

$$\begin{aligned}\mu_1 &= \xi^{-1}\mu_0 = \mu_0 + (\xi^{-1} - 1)\mu_0 \\ \mu_1 &= \mu_0 + (\xi^{-1} - 1)\sigma_0 \\ \mu_1 &= \mu_0 + \Delta\sigma_0.\end{aligned}\tag{3.20}$$

Desta forma, podemos considerar $k = \Delta/2$, como feito na abordagem para dados normais, o que retorna $k = \frac{\xi^{-1}-1}{2}$. Esta fórmula depende apenas da intensidade de mudança no parâmetro da distribuição e retorna, em geral, valores pequenos para k , o que pode ter bom desempenho no CUSUM. Por exemplo, para $\xi = 0.9$, que é uma redução leve no risco de falha (aumento leve na sobrevivência), essa fórmula determina $k = 0.05$, enquanto que a outra retorna 6.45, o que é impraticável para o nosso método.

Para o cálculo do ARL de forma teórica, investigamos a abordagem de cadeia de Markov para definir probabilidades de transição do CUSUM para diferentes estados. Hawkins & Olwell (1998) indicam que o CUSUM tem a propriedade de Markov, isto é, dado o t -ésimo valor do CUSUM, os valores anteriores não têm efeito no $(t+1)$ -ésimo, uma vez que há independência entre as observações e pela definição recursiva do CUSUM. Assim, seja o CUSUM definido em (3.8), temos que

$$P(C_t = c_t | C_0 = c_0, C_1 = c_1, \dots, C_{t-1} = c_{t-1}) = P(C_t = c_t | C_{t-1} = c_{t-1}).\tag{3.21}$$

Como o CUSUM tem espaço de estados contínuo, utiliza-se uma abordagem de discretização, ao fazer partições do intervalo de decisão $(0, h)$ em M subintervalos de mesmo comprimento, e assim cada um desses intervalos se torna um estado discreto no processo de Markov, junto com o estado inicial ($C_t = 0$) e outro para ($C_t > h$). Esses $M+2$ estados permitem então considerar o CUSUM como uma cadeia de Markov.

Considerando uma partição com largura $L = h/M$, temos os seguintes estados discretos:

$$\begin{array}{ll}\text{Estado } 0 & C_t = 0 \\ \text{Estado } i & C_t \in ((i-1)L, iL], i = 1, 2, \dots, M \\ \text{Estado } M+1 & C_t > h.\end{array}\tag{3.22}$$

A matriz de probabilidades de transição é dada por P , de forma que:

$$p_{ij} = P(C_t \in E_j | C_{t-1} \in E_i), \quad i = 0, 1, \dots, M+1,\tag{3.23}$$

em que E_i representa o estado i da cadeia de Markov. A transição de E_{M+1} para E_0 ocorre com probabilidade 1 (o CUSUM deve ser reiniciado após uma detecção). Com isso, calcula-se ARL como:

$$ARL = \frac{1}{\pi_{M+1}} - 1,\tag{3.24}$$

em que, π_{M+1} é a probabilidade estacionária de visitar o estado $M+1$ no processo de Markov.

Para calcular as probabilidades de transição, considere que em cada momento, deve-se avaliar se o valor de CUSUM está em um determinado intervalo, daqui que o valor anterior está também em determinado intervalo. Assim, as probabilidades são da forma:

$$p_{ij} = P(a < C_t < b | c < C_{t-1} < d). \quad (3.25)$$

Se $\mu(x)$ é a distribuição de C_{t-1} condicional a $c < C_{t-1} < d$ e $F(w) = P(w_t \leq w)$ então:

$$p_{ij} = \int_c^d \{F(b-s+k) - F(a-s+k) d\mu(s)\}. \quad (3.26)$$

Como $\mu(x)$ é desconhecida, aproxima-se a integral pela regra de Simpson, e então:

$$p_{ij} = \frac{1}{6} \{ [F(b-c+k) + 4F(b-m+k) + F(b-d+k)] - [F(a-c+k) + 4F(a-m+k) + F(a-d+k)] \}, \quad (3.27)$$

em que $m = (c+d)/2$. A distribuição estacionária Π é dada por $\lim_{n \rightarrow \infty} P^n = (\pi_1, \pi_2, \dots, \pi_{M+1})'$, e pode ser aproximada ao tomar sucessivos quadrados na matriz P até se conseguir uma estabilização nas probabilidades.

Para nossa aplicação os escores w_t são incrementos na estatística log-rank, $w_t = Z_t - Z_{t-1}$. Sob a hipótese nula, a estatística log-rank Z_t tem distribuição normal padrão assintoticamente, e sob a alternativa de que o risco de falha difere entre os grupos, a distribuição de Z_t é normal com média $\log(RR)\sqrt{V_t}$ e variância 1, em que RR é o risco relativo do grupo 1 em comparação com o 2 e $V_t = \sum_{i=1}^{L(t)} \frac{r_{i1}(t)r_{i2}(t)d_i(t)(r_i(t)-d_i(t))}{r_i^2(t)(r_i(t)-1)}$ é a variância do numerador no momento t , a ser utilizada no denominador do log-rank (ver Wu & Xiong, 2017). Na expressão, $L(t)$ define o total de tempos de falha distintos $y_1, y_2, \dots, y_{L(t)}$ observados no momento t e $r_i(t)$ é o total de indivíduos em risco no tempo de falha t_i , $r_{ij}(t)$ é o total de indivíduos em risco em t_i para o grupo j , e $d_{ij}(t)$ e $d_i(t)$ são as correspondentes quantidades de falhas equivalentes a t_i no grupo j e no total, respectivamente, no momento t .

Tsiatis (1981) mostra que a sequência de estatísticas log-rank converge para um processo com incrementos independentes, o que indica:

$$Cov(Z_{t-1}, Z_t - Z_{t-1}) = Cov(Z_{t-1}, w_t) = 0. \quad (3.28)$$

A variável w_t , no entanto, é uma diferença entre variáveis com distribuição normal, mas que são correlacionadas. Considerando que w_t e Z_{t-1} são independentes, sob H_0 tem-se

$$\begin{aligned} Cov(Z_t, Z_{t-1}) &= Cov(Z_{t-1} + w_t, Z_{t-1}) \\ &= Cov(Z_{t-1}, Z_{t-1}) + Cov(w_t, Z_{t-1}) \\ &= Var(Z_{t-1}) + 0 = 1. \end{aligned} \quad (3.29)$$

Desta forma, como resultado de distribuição da diferença de normais correlacionadas, w_t teria também distribuição normal, mas com variância dada por $\sigma_{Z_t}^2 + \sigma_{Z_{t-1}}^2 - 2r\sigma_{Z_t}\sigma_{Z_{t-1}}$, em que r é a correlação entre Z_t e Z_{t-1} e σ_{Z_t} e $\sigma_{Z_{t-1}}$ são os desvios-padrão de Z_t e Z_{t-1} . O resultado dessa variância é zero, o que teoricamente indica ser uma variável aleatória degenerada no valor zero, que

é sua esperança. Percebe-se com isso que há uma inconsistência na aplicação dos resultados teóricos neste caso. A utilização da teoria de martingais pode ser aplicada para resolver este problema.

Notamos que o desenvolvimento teórico para obtenção dos limites de controle ótimos está cercado por desafios e então não será o enfoque deste trabalho. Consideramos que um método computacional pode ser uma solução a se recorrer. Na Seção 3.5 apresentamos um estudo de simulação que propõe uma alternativa para obtenção dos limites de controle otimizados com respeito à medida de desempenho ARL.

3.5 Estudo de simulação

Consideramos um caso simples em que, sob controle, a variável aleatória T tem distribuição exponencial com parâmetro θ_0 e fora de controle há uma mudança apenas no parâmetro que se torna $\theta_1 = \xi \theta_0$. Nesta notação, ξ^{-1} é a razão entre o risco de falha no grupo controle e o risco de falha do grupo monitorado, conforme apresentado na discussão sobre o tamanho amostral do teste log-rank. Os tempos de censura também são gerados de uma distribuição exponencial com parâmetro θ_{0C} no grupo controle e θ_{1C} no grupo monitorado. E os tempos entre chegadas de pacientes da mesma forma são gerados de uma distribuição exponencial com parâmetros θ_E . Os tempos de falha e censura gerados da exponencial são considerados em anos, e os tempos entre chegadas de pacientes são em dias.

Os cenários considerados na simulação levam em conta estimativas destes parâmetros nos dados do InCor, de forma a gerar uma situação similar ao problema prático que tem sido estudado. A saber, o tamanho do grupo controle é considerado como $n_0 = 975$ e os parâmetros são $\theta_0 = 0.1631$, $\theta_{0C} = 0.1376$, $\theta_{1C} = 0.3143$ e $\theta_E = 5$ (o que representa uma taxa de entrada de 5 pacientes por dia). O tamanho amostral do grupo monitorado é tomado como $n_1 = 975$, e o número mínimo de indivíduos no grupo 2 para iniciar o monitoramento é calculado com a expressão (3.15), em que $\alpha = 0.05$, $\beta = 0.40$ e $\omega_1 = 0.5$. Aqui, consideramos 60% de poder inicial para o teste de forma que o número mínimo de indivíduos calculado para começar o monitoramento seja aplicável, mas este poder deve aumentar na prática à medida que o grupo monitorado cresce.

A Tabela 3.2 apresenta medidas-resumo calculadas com os valores de RL (*run length*) dos gráficos CUSUM, e também o desvio-padrão (DP) obtidas com 1000 réplicas. Foram escolhidos 4 possíveis valores de referência k (0, 0.001, 0.005, 0.010), e com estes se obteve os limites de controle ótimos h para atingir 3 possíveis ARL_0 desejados (500, 1000 e 2000), em dias, o que é retratado no bloco da tabela com $\xi = 1$ (dados sob controle). Foram testadas duas possibilidades de mudanças para avaliar o ARL_1 , $\xi = 0.5$ que representa uma redução considerável no risco de falha e $\xi = 0.8$ que reflete um aumento mais sutil na sobrevivência.

A obtenção de limites ótimos é feita para cada valor de referência fixado, fazendo-se busca direta em um intervalo. Utilizamos o método de busca *golden ratio* que consiste em uma busca binária com subdivisões otimizadas que permitem maior rapidez na obtenção do valor ótimo. A ideia é obter um valor de h que minimiza a função $f(h) = |ARL(h) - ARL_0|$, em que ARL_0 é um valor desejado para o ARL sob controle e $ARL(h)$ é a estimativa desta medida que é calculada como a média dos valores de RL em cada réplica com o limite h fixado. Sabe-se que essa estimativa é não decrescente em relação ao valor de h (para um k fixo). Assim, se o intervalo escolhido para realização da busca de h contém o valor ótimo que aproxima esta estimativa do valor alvo, isso sugere haver um mínimo

para a função $f(h)$ definida, já que esta f deve ser monótona e não crescente à esquerda do valor ótimo e também monótona mas não decrescente à direita do mesmo. Desta forma, minimizar f permite obter um limite de controle ótimo. Foram consideradas nesta simulação buscas entre 0.5 e 10, o que retornou valores de h cujas estimativas ARL se aproximaram do valor ARL_0 solicitado.

Observamos que o valor do limite ótimo para atender determinado ARL_0 tende a diminuir para maiores valores de referência k , e com base no desempenho do método em relação ao ARL_1 é que se deve escolher uma combinação de k e h que retorne a detecção mais ágil. Por exemplo, para mudanças de intensidade $\xi = 0.5$ observamos valores de ARL_1 entre 26 e 67. Se considerarmos apenas os casos em que a combinação de k e h retornam um $ARL_0 = 1000$, por exemplo, esta variação fica de 31 a 54, e qualquer uma das escolhas é satisfatória, já que indica não mais do que 2 meses de espera para se detectar uma mudança. A escolha de $k = 0.010$ e $h = 0.9810$, neste caso, retornaria o menor ARL_1 (31), mas podemos também considerar $h = 1.0656$ que aumenta um pouco essa medida (para 39), só que esta proporciona um ARL_0 muito maior (1986). Ou seja, perde-se bem pouco no poder de detecção, mas permitindo reduzir muito o alarme falso do gráfico.

Para a situação em que a mudança tem uma intensidade mais leve, $\xi = 0.8$, já se nota maior dificuldade de detecção para alguns casos. Por exemplo, não seria recomendado tomar a mesma combinação de h preferida na situação anterior, pois embora o ARL_0 seja próximo de 2000, o ARL_1 fica muito grande (1153). Nesta situação, acreditamos que é melhor considerar um valor de referência menor, por exemplo, com $k = 0.001$ e $h = 1.6059$ se consegue o maior ARL_0 testado e um $ARL_1 = 132$, ou seja, leva-se em média pouco mais que 4 meses para detectar uma redução de magnitude 20% no risco de falha e bem mais que 5 anos até à ocorrência de um alarme falso ($ARL_0 = 1985$ dias). Se é do interesse, contudo, contentar-se com uma frequência maior de alarmes falsos para se conseguir uma detecção muito mais rápida, a sugestão seria, por exemplo, considerar $h = 1.4616$, e assim a espera é de pouco mais que 2 meses (73 dias) até uma detecção e 998 dias até um alarme falso (algo próximo de 3 anos).

A Tabela 3.3 apresenta os resultados da simulação em dados com censura (que é em torno de 45% no grupo 1, sob controle). Notamos um comportamento similar no desempenho do gráfico. Nesta situação, para detectar uma mudança de intensidade $\xi = 0.5$, uma boa escolha é $k = 0.005$ e $h = 1.4061$, que retorna um $ARL_0 = 2000$ e $ARL_1 = 51$, isto é, por volta de 2 meses, em média, para conseguir detectar a mudança, mas mais que 5 anos até se observar um alarme falso. Para $\xi = 0.8$, se a intenção é ter um ARL_0 máximo, sugere-se considerar a combinação $k = 0$ e $h = 1.9613$, para a qual se tem $ARL_1 = 253$ (por volta de 8 meses). Caso um ARL_0 em torno de 1000 seja aceitável, pode-se tomar $k = 0$ e $h = 1.7286$, o que reduz essa espera até detectar uma mudança para 5 meses ($ARL_1 = 148$). Neste cenário, notamos que um valor de referência maior que zero já pode retardar consideravelmente a detecção se queremos um ARL_0 grande. Se, contudo, um ARL_0 em torno de 500 for considerável, é possível ter $ARL_1 = 73$ com $k = 0.001$ e $h = 1.3506$.

Um contraponto deste método é a necessidade de aguardar um tempo suficiente para alcançar um número satisfatório de indivíduos no grupo de monitoramento. Para as simulações apresentadas o tempo médio necessário que se aguarda até atingir o tamanho de amostra mínimo é de 212 dias. Assim, o tempo médio até detecção de mudanças que foi obtido em cada situação deve ser somado a este valor. Desta forma, por exemplo, se o tempo esperado até a detecção é de 5 meses, aguarda-se algo em torno de 1 ano (7 meses até o início do monitoramento + 5 meses até a detecção), para conseguir identificar a melhora na sobrevivência.

ξ	k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
1	0.000	1.4972	4	27	50	506	113	24109	2682.0
		1.6261	4	33	62	999	164	24818	3906.2
		1.8609	5	44	102	2007	382	31378	5415.3
	0.001	1.3334	4	22	37	508	77	24109	2852.2
		1.4616	4	26	48	998	112	24818	4029.2
		1.6059	4	33	64	1985	194	31378	5641.2
	0.005	1.1043	3	17	26	489	52	22604	2851.8
		1.1676	3	19	31	1018	65	28921	4185.5
		1.2813	4	23	41	2008	99	28921	5730.4
	0.010	0.9170	1	13	21	506	39	22604	2913.2
		0.9810	3	15	25	1008	52	28921	4189.0
		1.0656	3	18	31	1986	70	31378	5788.6
	0.000	1.4972	4	26	37	46	58	171	28.1
		1.6261	4	30	44	54	70	200	32.0
		1.8609	5	39	57	67	87	317	38.1
	0.001	1.3334	4	22	30	39	47	169	26.0
		1.4616	4	26	36	46	58	186	29.0
		1.6059	4	30	45	55	72	206	33.3
	0.005	1.1043	3	17	24	32	38	162	23.3
		1.1676	3	19	26	36	43	172	25.9
		1.2813	3	23	33	44	55	196	30.9
	0.010	0.9170	1	13	19	26	31	175	20.6
		0.9810	3	15	22	31	37	337	27.9
		1.0656	3	19	27	39	47	351	32.7
0.5	0.000	1.4972	3	26	44	69	86	670	70.2
		1.6261	3	32	55	90	109	816	97.2
		1.8609	4	42	83	134	161	1868	154.5
	0.001	1.3334	3	22	34	54	67	526	56.8
		1.4616	3	26	43	73	88	944	86.8
		1.6059	3	32	56	132	114	17751	792.1
	0.005	1.1043	2	17	26	147	48	29481	1656.3
		1.1676	3	19	29	197	56	29481	1913.5
		1.2813	3	23	38	496	82	31378	3141.1
	0.010	0.9170	1	13	20	394	35	29481	2896.2
		0.9810	2	15	24	622	47	31378	3636.6
		1.0656	2	18	30	1153	63	31378	4902.8

Tabela 3.2: Medidas-resumo do RL dos gráficos ZDiff CUSUM em dados com mudanças de intensidade ξ e sem censura, fixados os valores de referência (k) e limites ótimos (h) obtidos.

Contudo, outro aspecto que vale destacar é que para se detectar aumentos na sobrevivência há a necessidade intrínseca de ter maior tempo de acompanhamento, e, portanto, já é algo que naturalmente requer um tempo razoável até que qualquer método possa perceber uma melhora. Realizamos apenas simulações no sentido de aumento na sobrevida, mas acreditamos que este gráfico seja ainda melhor para identificar reduções na sobrevivência. Isso pode ser inferido tendo em vista que para identificar aumento na sobrevivência já é necessário naturalmente aguardar mais tempo para constatar os tempos mais longos e ter evidência sobre o crescimento da sobrevida, enquanto que reduções na sobrevivência fazem com que eventos sejam observados mais rapidamente (tempos menores), e assim é possível ter evidência da queda na sobrevida de forma mais breve.

ξ	k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
1	0.000	1.5227	6	27	48	501	123	14192	1884.9
		1.7286	7	37	76	996	251	14192	2631.4
		1.9613	8	53	134	1996	808	14192	3632.2
	0.001	1.3506	5	22	37	499	75	14192	1946.4
		1.5436	6	30	54	1008	171	14192	2776.1
		1.7505	7	40	88	2001	427	14192	3685.0
	0.005	1.1438	4	18	28	499	55	14192	2017.1
		1.2336	4	21	35	996	75	14192	2809.7
		1.4061	6	28	53	2000	222	14192	3746.9
	0.010	0.9615	4	14	21	502	40	14192	2061.2
		1.0494	4	17	28	997	59	14192	2818.4
		1.1674	4	22	40	1999	166	14192	3769.6
0.5	0.000	1.5227	3	27	36	45	54	204	28.2
		1.7286	3	35	47	58	71	318	34.0
		1.9613	3	44	61	73	90	396	42.7
	0.001	1.3506	3	22	29	38	44	180	24.4
		1.5436	3	29	39	49	59	320	31.7
		1.7505	3	37	51	62	77	400	38.3
	0.005	1.1438	3	18	25	32	38	196	24.0
		1.2336	3	21	29	38	44	323	28.3
		1.4061	3	27	38	51	60	484	41.2
	0.010	0.9615	3	14	21	27	33	233	23.1
		1.0494	3	17	25	60	40	9089	455.6
		1.1674	3	22	32	115	53	10148	774.5
0.8	0.000	1.5227	6	27	44	83	81	9039	306.7
		1.7286	7	36	61	148	128	11191	608.9
		1.9613	9	49	91	253	192	11375	939.5
	0.001	1.3506	6	22	34	73	63	10293	434.6
		1.5436	7	29	48	154	90	11191	833.7
		1.7505	8	38	69	337	149	11191	1372.9
	0.005	1.1438	5	18	26	177	47	10293	1077.0
		1.2336	5	21	32	315	62	14192	1556.9
		1.4061	6	27	48	725	100	14192	2383.0
	0.010	0.9615	4	14	21	197	39	14192	1240.6
		1.0494	5	17	26	502	49	14192	2055.1
		1.1674	5	22	37	1039	79	14192	2864.3

Tabela 3.3: Medidas-resumo do RL dos gráficos ZDiff CUSUM em dados com mudanças de intensidade ξ e percentual de censura 45%, em média, no grupo 1, fixados os valores de referência (k) e limites ótimos (h) obtidos.

3.6 Modelo de ponto de mudança

As abordagens de monitoramento apresentadas até então concentram-se na detecção de uma determinada intensidade de mudança no processo em observação, previamente estabelecida. Contudo, é de interesse prático obter estimativas para o momento em que tal alteração ocorre nos dados e também de sua intensidade, de forma a fornecer maior suporte ao pesquisador para identificar as causas especiais relacionadas e fazer o reparo necessário, em caso de deterioração de qualidade. Em seguida são discutidos alguns métodos presentes na literatura para identificação do ponto e da intensidade de mudança em uma distribuição de dados.

A teoria de modelos de ponto de mudança envolve problemas que avaliam uma sequência de variáveis aleatórias $X_1, X_2, \dots$, em que as primeiras τ observações $X_1, X_2, \dots, X_\tau$ seguem uma distribuição F_0 , e a partir de então, as novas observações $X_{\tau+1}, X_{\tau+2}, \dots$ seguem uma distribuição F_1 diferente. Considera-se por simplicidade que as distribuições F_0 e F_1 são as mesmas e diferem apenas no valor de algum dos parâmetros, isto é, os primeiros indivíduos seguem uma distribuição $F_{\theta_0}(X)$ e após τ observações se tem $F_{\theta_1}(X)$, com $\theta_0 \neq \theta_1$. Sendo assim, é de interesse detectar ao longo do tempo, e o mais breve possível, o momento em que ocorre uma alteração no comportamento dos dados em determinado estudo, com a obtenção de estimativas para o ponto de mudança τ e os diferentes valores dos parâmetros associados à distribuição θ_0 e θ_1 .

Diversos autores descrevem essa abordagem para detecção de alteração no valor do parâmetro de uma determinada distribuição. Sen & Srivastava (1975) consideram problemas em que se deseja identificar mudanças na média de uma distribuição normal, apresentam diferentes estatísticas de teste presentes na literatura para o caso em que a média inicial μ é conhecida ou desconhecida e comparam com estatísticas de máxima verossimilhança correspondentes obtidas. Lorden (1971) trata de problemas deste tipo mas com um enfoque mais geral, e propõe testes sequenciais de razão de verossimilhanças, de modo similar ao que foi proposto por Page (1954) na fórmula do CUSUM. Propriedades assintóticas dessa metodologia são discutidas.

No contexto da análise de sobrevivência, contudo, deve-se considerar que as observações são acompanhadas ao longo do tempo e entram no estudo em momentos possivelmente distintos, e são acompanhados até a ocorrência de uma falha ou censura. Com isso, se há uma mudança de tratamento proposta em certo instante, alguns dos indivíduos que deram início antes desse ponto e que ainda estão em acompanhamento após a substituição do tratamento podem ser beneficiados por esse efeito ao optarem por abandonar o antigo. Assim, supõe-se que o conjunto de indivíduos organizados cronologicamente pode ser subdividido em 3 diferentes grupos, sendo o primeiro com pacientes que só utilizaram o tratamento antigo durante todo o acompanhamento, $X_1, X_2, \dots, X_{\tau_1}$, o segundo com pacientes que iniciaram o estudo no tratamento tradicional, e em algum instante trocam pelo método novo sugerido, $X_{\tau_1+1}, X_{\tau_1+2}, \dots, X_{\tau_1+\tau_2}$, e outro conjunto de pacientes, $X_{\tau_1+\tau_2+1}, X_{\tau_1+\tau_2+2}, \dots$, que já inicia o acompanhamento no novo tratamento e o mantém.

Convém, então, assumir que o problema de identificar o ponto de mudança no comportamento dos dados de uma distribuição $F_{\theta_0}(X)$ para $F_{\theta_1}(X)$ é melhor abordado se considerar que os indivíduos afetados pelo período de transição possuem uma contribuição diferente na verossimilhança. Por exemplo, pode-se admitir que estes apresentam uma mistura das duas distribuições testadas. Com isso, o problema de identificação da alteração de distribuições deve ser visto então como a detecção de dois pontos de mudança τ_1 e τ_2 , que representam, respectivamente, a mudança para o período de transição e para o novo estado de qualidade.

Como ilustração do desempenho desse modelo, vamos considerar a seguir um caso paramétrico com um ponto de mudança, em que dados oriundos da distribuição exponencial são monitorados. Seja então uma amostra aleatória $X_1, X_2, \dots, X_\tau, X_{\tau+1}, X_{\tau+2}, \dots, X_n$, em que as primeiras τ observações são geradas da distribuição exponencial de parâmetro $\theta = \theta_0$, e as seguintes com $\theta = \theta_1$. Desta forma, tem-se uma estrutura de mudança no parâmetro da distribuição de um valor θ_0 para θ_1 , que ocorre após o ponto τ . Considere ainda a estrutura de censura à direita, e que na prática observa-se uma amostra $T_1, T_2, \dots, T_\tau, T_{\tau+1}, T_{\tau+2}, \dots, T_n$, com $T_i = \min\{X_i, C_i\}$, em que C_i é o tempo de censura associado à observação i , e $\delta_1, \delta_2, \dots, \delta_\tau, \delta_{\tau+1}, \delta_{\tau+2}, \dots, \delta_n$ são os indicadores de censura

correspondentes. Deseja-se com base na amostra observada estimar os 3 parâmetros relacionados com a mudança na distribuição de X .

A verossimilhança para o modelo é dada por:

$$\begin{aligned} L(\tau, \theta_0, \theta_1 | \mathbf{T}, \delta) &= \prod_{i=1}^{\tau} (\theta_0 e^{-\theta_0 T_i})^{\delta_i} (e^{-\theta_0 T_i})^{1-\delta_i} \prod_{i=\tau+1}^n (\theta_1 e^{-\theta_1 T_i})^{\delta_i} (e^{-\theta_1 T_i})^{1-\delta_i} \\ &= \prod_{i=1}^{\tau} \theta_0^{\delta_i} e^{-\theta_0 T_i} \prod_{i=\tau+1}^n \theta_1^{\delta_i} e^{-\theta_1 T_i}. \end{aligned} \quad (3.30)$$

Desta forma, a log-verossimilhança é:

$$\ell(\tau, \theta_0, \theta_1 | \mathbf{T}, \delta) = \sum_{i=1}^{\tau} \delta_i \log(\theta_0) - \theta_0 T_i + \sum_{i=\tau+1}^n \delta_i \log(\theta_1) - \theta_1 T_i \quad (3.31)$$

As derivadas parciais, com relação à θ_0 e θ_1 , e os correspondentes estimadores de máxima verossimilhança para estes parâmetros são:

$$\begin{aligned} \frac{\partial \ell(\tau, \theta_0, \theta_1 | \mathbf{T}, \delta)}{\partial \theta_0} &= \sum_{i=1}^{\tau} \frac{\delta_i}{\theta_0} - T_i = 0 \Rightarrow \frac{\sum_{i=1}^{\tau} \delta_i}{\hat{\theta}_0} = \sum_{i=1}^{\tau} T_i \Rightarrow \hat{\theta}_0 = \frac{\sum_{i=1}^{\tau} \delta_i}{\sum_{i=1}^{\tau} T_i} \\ \frac{\partial \ell(\tau, \theta_0, \theta_1 | \mathbf{T}, \delta)}{\partial \theta_1} &= \sum_{i=\tau+1}^n \frac{\delta_i}{\theta_1} - T_i = 0 \Rightarrow \frac{\sum_{i=\tau+1}^n \delta_i}{\hat{\theta}_1} = \sum_{i=\tau+1}^n T_i \Rightarrow \hat{\theta}_1 = \frac{\sum_{i=\tau+1}^n \delta_i}{\sum_{i=\tau+1}^n T_i} \end{aligned} \quad (3.32)$$

O ponto de mudança τ , contudo, não pode ser obtido por solução analítica, e assim uma alternativa é considerar uma malha de pontos, e para cada valor fixado de τ obter a estimativa de máxima verossimilhança para os outros parâmetros. A estimativa de τ é então a que fornecer o maior valor da log-verossimilhança. Outra forma é utilizar rotinas de otimização, como o *optim* e estimar os 3 parâmetros em conjunto, maximizando a expressão (3.31). Estes dois métodos são utilizados em uma breve simulação, em que consideramos $\theta_0 = 1$ o valor do parâmetro antes da mudança, e $\theta_1 = 1.5, 1.3$ ou 1.1 o valor depois da mudança. O caso em que não há mudança ($\theta_1 = 1$) é também investigado. Nesta situação há uma divisão da amostra completa em dois conjuntos, mas toda ela é proveniente da mesma distribuição.

Além do método de estimação por verossimilhança para o ponto de mudança, propomos também um método simples que é maximizar a estatística de teste log-rank para comparação de duas amostras. Feito isso, os parâmetros θ_0 e θ_1 são obtidos teoricamente por máxima verossimilhança ao fixar a estimativa de τ encontrada. Foram feitas 10 mil réplicas do procedimento em cada situação, com diferentes tamanhos de amostra e percentuais de censura, e os resultados dessa simulação são apresentados nas tabelas a seguir.

Depara-se com estimativas discrepantes em algumas situações, e para contornar isto, foram considerados diferentes cortes do conjunto de busca, no qual desconsidera-se uma quantidade fixa dos primeiros e últimos valores possíveis para o ponto de mudança. A justificativa para tal é garantir que haja pelo menos 10 indivíduos em cada subamostra comparada, a fim de que a estimativa de máxima verossimilhança dos parâmetros θ_0 e θ_1 seja calculada de forma plausível, e que a comparação entre as duas subamostras seja razoável.

Amostra	Parâmetro	Método de Estimação	Corte	Mín.	Q1	Q2	Média	Q3	Máx.
$n_0 = 500$ $n_1 = 1000$ $\tau = 500$ $\theta_0 = 1$ $\theta_1 = 1.5$	τ	MV teórica	0	212.00	487.00	498.00	497.40	506.00	893.00
			10	212.00	487.00	498.00	497.40	506.00	893.00
			50	212.00	487.00	498.00	497.40	506.00	893.00
			90	212.00	487.00	498.00	497.40	506.00	893.00
		MV <i>optim</i>	-	10.03	468.50	493.50	486.80	505.90	874.30
			0	1.00	489.00	499.00	517.10	510.00	1499.00
			10	242.00	489.00	499.00	501.00	508.00	1113.00
			50	242.00	489.00	499.00	501.00	508.00	1113.00
			90	242.00	489.00	499.00	501.00	508.00	1113.00
		log-rank máximo	0	0.7961	0.9641	0.9942	0.9951	1.0250	1.2140
			10	0.7961	0.9641	0.9942	0.9951	1.0250	1.2140
			50	0.7961	0.9641	0.9942	0.9951	1.0250	1.2140
			90	0.7961	0.9641	0.9942	0.9951	1.0250	1.2140
	θ_0	MV teórica	-	0.6257	0.9656	0.9958	0.9973	1.0280	1.2160
			0	0.8235	0.9663	0.9972	8.9580	1.0310	16510.0000
			10	0.8235	0.9652	0.9954	0.9965	1.0270	1.2420
			50	0.8235	0.9652	0.9954	0.9965	1.0270	1.2420
		MV / log-rank	90	0.8235	0.9652	0.9954	0.9965	1.0270	1.2420
			0	1.3380	1.4720	1.5040	1.5060	1.5380	1.7170
			10	1.3380	1.4720	1.5040	1.5060	1.5380	1.7170
			50	1.3380	1.4720	1.5040	1.5060	1.5380	1.7170
			90	1.3380	1.4720	1.5040	1.5060	1.5380	1.7170
		MV <i>optim</i>	-	1.2830	1.4610	1.4960	1.4960	1.5300	1.7170
			0	1.2120	1.4720	1.5060	18.9300	1.5420	47940.0000
			10	1.3380	1.4730	1.5050	1.5070	1.5390	1.8100
			50	1.3380	1.4730	1.5050	1.5070	1.5390	1.8100
		MV / log-rank	90	1.3380	1.4730	1.5050	1.5070	1.5390	1.8100
$n_0 = 200$ $n_1 = 300$ $\tau = 200$ $\theta_0 = 1$ $\theta_1 = 1.5$	τ	MV teórica	0	1.00	186.00	198.00	199.00	207.00	499.00
			10	11.00	187.00	198.00	198.80	207.00	489.00
			50	51.00	187.00	198.00	198.80	207.00	448.00
			90	91.00	187.00	198.00	199.30	207.00	407.00
		MV <i>optim</i>	-	4.71	188.50	198.30	197.80	208.60	491.60
			0	1.00	188.00	200.00	216.50	216.00	499.00
			10	11.00	189.50	199.00	205.90	211.00	489.00
			50	51.00	189.00	199.00	204.10	211.00	449.00
			90	91.00	190.00	199.00	203.70	211.00	409.00
		log-rank máximo	0	0.1197	0.9336	0.9830	1.3810	1.0350	1962.0000
			10	0.4361	0.9337	0.9830	0.9853	1.0350	2.4820
			50	0.6695	0.9349	0.9836	0.9862	1.0350	1.3170
			90	0.7452	0.9365	0.9848	0.9878	1.0360	1.3130
	θ_0	MV <i>optim</i>	-	0.3780	0.9568	1.0040	1.0070	1.0540	4.3980
			0	0.2476	0.9427	0.9961	9.7690	1.0630	16510.0000
			10	0.4361	0.9387	0.9880	0.9942	1.0400	4.2400
			50	0.6783	0.9390	0.9878	0.9915	1.0400	1.3170
		MV / log-rank	90	0.7452	0.9400	0.9886	0.9921	1.0400	1.3130
			0	0.4794	1.4580	1.5200	8.3600	1.5900	42920.0000
			10	0.5479	1.4580	1.5200	1.5320	1.5890	6.1960
			50	0.9689	1.4580	1.5200	1.5280	1.5880	2.7870
			90	1.2120	1.4590	1.5200	1.5260	1.5880	2.3390
		MV <i>optim</i>	-	0.8977	1.4340	1.4920	1.4950	1.5530	2.1900
			0	0.6155	1.4620	1.5320	16.9000	1.6170	42920.0000
			10	0.7728	1.4640	1.5260	1.5490	1.5980	6.1960
			50	0.9689	1.4630	1.5250	1.5350	1.5950	2.7870
		MV / log-rank	90	1.2260	1.4630	1.5240	1.5320	1.5940	2.4660

Tabela 3.4: Medidas-resumo das estimativas do ponto de mudança e dos parâmetros das distribuições antes e depois, para dados gerados da distribuição exponencial, com parâmetro $\theta_0 = 1$, $\theta_1 = 1.5$, ponto de mudança $\tau = 500$ ou $\tau = 200$ e sem censura na amostra completa, com $n_0 = 500$ e $n_1 = 1000$ ou $n_0 = 200$ e $n_1 = 300$.

Amostra	Parâmetro	Método de Estimação	Corte	Mín.	Q1	Q2	Média	Q3	Máx.
$n_0 = 500$ $n_1 = 1000$ $\tau = 500$ $\theta_0 = 1$ $\theta_1 = 1.5$	τ	MV teórica	0	1.00	480.00	500.00	503.10	520.00	1499.00
			10	11.00	480.00	500.00	503.20	520.00	1482.00
			50	51.00	480.00	500.00	503.50	520.00	1447.00
			90	91.00	480.00	500.00	503.30	520.00	1400.00
		MV <i>optim</i>	-	85.35	471.50	499.50	498.70	525.40	1441.00
			0	1.00	483.00	502.00	550.70	534.00	1499.00
			10	11.00	485.00	501.00	516.80	527.00	1489.00
			50	51.00	485.00	501.00	515.10	527.00	1449.00
			90	91.00	485.00	501.00	514.80	527.00	1409.00
		log-rank máximo	0	0.0839	0.9353	0.9821	1.1300	1.0310	1303.0000
			10	0.0839	0.9353	0.9821	0.9836	1.0310	4.4040
			50	0.4018	0.9355	0.9822	0.9837	1.0320	1.4110
			90	0.4671	0.9356	0.9822	0.9839	1.0320	1.3370
$n_0 = 200$ $n_1 = 300$ $\tau = 200$ $\theta_0 = 1$ $\theta_1 = 1.5$	θ_0	MV <i>optim</i>	-	0.6343	0.9473	0.9928	0.9952	1.0410	1.3550
			0	0.4018	0.9432	0.9913	9.5700	1.0490	16510.0000
			10	0.4018	0.9401	0.9859	0.9901	1.0360	4.6360
			50	0.4018	0.9401	0.9859	0.9887	1.0360	1.4110
		MV / log-rank	90	0.4903	0.9401	0.9859	0.9887	1.0360	1.3430
			0	0.7881	1.4640	1.5100	1.7500	1.5570	1275.0000
			10	0.7881	1.4640	1.5100	1.5130	1.5570	4.9070
			50	0.7881	1.4640	1.5100	1.5120	1.5570	2.9700
			90	1.2480	1.4640	1.5100	1.5120	1.5570	2.2000
		MV <i>optim</i>	-	1.1330	1.4550	1.5020	1.5020	1.5480	1.9390
			0	0.7881	1.4640	1.5100	1.7500	1.5570	1275.0000
			10	0.7881	1.4640	1.5100	1.5130	1.5570	4.9070
			50	0.7881	1.4640	1.5100	1.5120	1.5570	2.9700
		MV / log-rank	90	1.2480	1.4640	1.5100	1.5120	1.5570	2.2000
$n_0 = 200$ $n_1 = 300$ $\tau = 200$ $\theta_0 = 1$ $\theta_1 = 1.5$	τ	MV teórica	0	1.00	177.00	200.00	211.10	229.00	499.00
			10	11.00	178.00	200.00	209.10	227.00	489.00
			50	51.00	180.00	200.00	207.40	225.00	449.00
			90	91.00	182.00	200.00	208.40	225.00	409.00
		MV <i>optim</i>	-	1.41	188.70	200.20	198.80	211.40	491.30
			0	1.00	182.00	204.00	239.30	279.20	499.00
			10	11.00	186.00	203.00	225.10	242.00	489.00
			50	51.00	186.00	202.00	217.80	235.00	449.00
			90	91.00	187.00	202.00	215.80	233.00	409.00
		log-rank máximo	0	0.0835	0.8705	0.9575	5.3810	1.0480	16510.0000
			10	0.0835	0.8706	0.9567	0.9648	1.0440	4.9800
			50	0.3329	0.8769	0.9600	0.9622	1.0420	1.9820
			90	0.5477	0.8860	0.9645	0.9681	1.0450	1.7500
$n_0 = 200$ $n_1 = 300$ $\tau = 200$ $\theta_0 = 1$ $\theta_1 = 1.5$	θ_0	MV <i>optim</i>	-	0.3842	0.9268	0.9998	1.0060	1.0770	5.1770
			0	0.0000	0.9006	0.9995	10.1800	1.1690	16510.0000
			10	0.0000	0.8870	0.9725	0.9946	1.0640	4.9800
			50	0.4176	0.8892	0.9708	0.9755	1.0550	1.9820
		MV / log-rank	90	0.5477	0.8940	0.9732	0.9769	1.0530	1.7500
			0	0.1331	1.4510	1.5460	10.2400	1.6490	42920.0000
			10	0.1331	1.4510	1.5440	1.5770	1.6430	6.3070
			50	0.6689	1.4530	1.5420	1.5550	1.6350	2.9840
			90	0.8415	1.4560	1.5420	1.5510	1.6330	2.4590
		MV <i>optim</i>	-	0.9224	1.4180	1.4970	1.4990	1.5780	2.6550
			0	0.0000	1.4600	1.5780	17.6700	1.7380	42920.0000
			10	0.3013	1.4650	1.5600	1.6340	1.6720	6.3070
			50	0.6689	1.4630	1.5530	1.5720	1.6530	2.9840
		MV / log-rank	90	0.8415	1.4620	1.5490	1.5610	1.6450	2.4590

Tabela 3.5: Medidas-resumo das estimativas do ponto de mudança e dos parâmetros das distribuições antes e depois, para dados gerados da distribuição exponencial, com parâmetro $\theta_0 = 1$, $\theta_1 = 1.5$, ponto de mudança $\tau = 500$ ou $\tau = 200$ e em torno de 50% de censura na amostra completa, com $n_0 = 500$ e $n_1 = 1000$ ou $n_0 = 200$ e $n_1 = 300$.

Amostra	Parâmetro	Método de Estimação	Corte	Mín.	Q1	Q2	Média	Q3	Máx.
$n_0 = 500$ $n_1 = 1000$ $\tau = 500$ $\theta_0 = 1$ $\theta_1 = 1.1$	τ	MV teórica	0	1.00	236.00	506.00	646.90	1092.00	1499.00
			10	11.00	293.00	507.00	635.10	974.20	1489.00
			50	51.00	345.00	512.00	629.70	899.00	1449.00
			90	91.00	368.00	515.00	627.20	862.00	1409.00
		MV <i>optim</i>	-	1.02	434.50	481.40	492.10	530.40	1460.00
			0	1.00	142.80	527.00	702.10	1431.00	1499.00
			10	11.00	314.00	520.00	659.90	1061.00	1489.00
			50	51.00	359.00	523.00	646.60	945.00	1449.00
			90	91.00	380.00	526.00	641.70	900.00	1409.00
		log-rank máximo	0	0.1072	0.9277	0.9860	8.4760	1.0560	16510.0000
			10	0.3946	0.9267	0.9799	1.0130	1.0390	3.8430
			50	0.6244	0.9315	0.9785	0.9862	1.0280	1.8000
			90	0.7190	0.9361	0.9791	0.9846	1.0250	1.4760
	θ_0	MV <i>optim</i>	-	0.2307	0.9574	0.9884	0.9904	1.0210	1.5120
			0	0.1253	0.9510	1.0260	10.7600	1.0900	16510.0000
			10	0.3946	0.9341	0.9866	1.0330	1.0460	4.2400
			50	0.6244	0.9354	0.9818	0.9907	1.0310	1.8000
		MV / log-rank	90	0.7126	0.9390	0.9814	0.9872	1.0260	1.5120
			0	0.1077	1.0780	1.1210	12.6400	1.1750	35150.0000
			10	0.4187	1.0810	1.1200	1.1600	1.1640	4.0400
			50	0.6880	1.0860	1.1210	1.1300	1.1600	2.0040
			90	0.7389	1.0880	1.1210	1.1250	1.1570	1.6320
		MV <i>optim</i>	-	0.8878	1.0800	1.1050	1.1060	1.1310	1.5370
			0	0.1077	1.0780	1.1290	15.8000	1.3110	35150.0000
			10	0.4187	1.0830	1.1240	1.1870	1.1750	4.0400
			50	0.6880	1.0870	1.1230	1.1360	1.1640	2.0040
		MV / log-rank	90	0.7389	1.0890	1.1230	1.1280	1.1600	1.6320
$n_0 = 200$ $n_1 = 300$ $\tau = 200$ $\theta_0 = 1$ $\theta_1 = 1.1$	τ	MV teórica	0	1.00	50.00	211.00	240.80	447.00	499.00
			10	11.00	84.00	212.00	238.40	405.00	489.00
			50	51.00	125.00	213.00	235.80	354.00	449.00
			90	91.00	151.00	218.00	236.10	326.00	409.00
		MV <i>optim</i>	-	1.03	186.80	198.30	199.90	211.30	498.00
			0	1.00	25.00	232.00	253.00	485.00	499.00
			10	11.00	85.00	222.00	244.60	417.00	489.00
			50	51.00	130.00	222.00	241.10	361.00	449.00
			90	91.00	156.00	223.00	240.00	331.00	409.00
		log-rank máximo	0	0.1072	0.8953	1.0040	9.2660	1.0980	16510.0000
			10	0.3946	0.8911	0.9812	1.0330	1.0730	4.2400
			50	0.6391	0.9052	0.9726	0.9927	1.0520	1.8000
			90	0.7090	0.9191	0.9739	0.9904	1.0440	1.5450
	θ_0	MV <i>optim</i>	-	0.5727	0.9506	0.9988	1.0040	1.0520	3.5830
			0	0.1197	0.9446	1.0410	10.8800	1.2020	16510.0000
			10	0.3946	0.9040	0.9919	1.0550	1.0810	4.2400
			50	0.6391	0.9114	0.9772	0.9972	1.0560	1.8000
		MV / log-rank	90	0.7090	0.9220	0.9761	0.9925	1.0460	1.5450
			0	0.1233	1.0530	1.1420	10.9800	1.2960	31480.0000
			10	0.4018	1.0540	1.1400	1.1920	1.2470	4.5440
			50	0.6262	1.0650	1.1410	1.1440	1.2180	2.0440
			90	0.7340	1.0710	1.1400	1.1350	1.2050	1.8080
		MV <i>optim</i>	-	0.4267	1.0580	1.1020	1.1050	1.1480	2.9140
			0	0.1238	1.0590	1.1610	13.4400	1.7180	31480.0000
			10	0.4018	1.0580	1.1480	1.2190	1.2700	4.5440
			50	0.6262	1.0670	1.1440	1.1480	1.2240	2.0440
		MV / log-rank	90	0.7340	1.0720	1.1420	1.1370	1.2070	1.8080

Tabela 3.6: Medidas-resumo das estimativas do ponto de mudança e dos parâmetros das distribuições antes e depois, para dados gerados da distribuição exponencial com parâmetro $\theta_0 = 1$, $\theta_1 = 1.1$, ponto de mudança $\tau = 500$ ou $\tau = 200$ e sem censura na amostra completa, com $n_0 = 500$ e $n_1 = 1000$ ou $n_0 = 200$ e $n_1 = 300$.

Amostra	Parâmetro	Método de Estimação	Corte	Mín.	Q1	Q2	Média	Q3	Máx.
$n_0 = 500$ $n_1 = 1000$ $\tau = 500$ $\theta_0 = 1$ $\theta_1 = 1.1$	τ	MV teórica	0	1.00	134.00	569.00	713.50	1377.00	1499.00
			10	11.00	194.00	568.00	703.40	1294.00	1489.00
			50	51.00	280.00	572.00	689.80	1155.00	1449.00
			90	91.00	326.00	579.00	688.90	1103.00	1409.00
		MV <i>optim</i>	-	1.00	449.50	496.80	518.80	558.90	1485.00
			0	1.00	62.00	616.50	742.80	1465.00	1499.00
			10	11.00	219.00	602.00	725.50	1320.00	1489.00
			50	51.00	303.00	598.00	709.50	1185.00	1449.00
			90	91.00	344.00	602.00	704.90	1125.00	1409.00
		log-rank máximo	0	0.0814	0.8828	0.9985	8.9960	1.0880	16510.0000
			10	0.0814	0.8728	0.9797	1.0100	1.0720	4.7210
			50	0.3969	0.8868	0.9723	0.9838	1.0570	2.0770
			90	0.4671	0.8953	0.9714	0.9818	1.0510	1.7370
	θ_0	MV <i>optim</i>	-	0.0000	0.9316	0.9794	0.9905	1.0340	22.5500
			0	0.0000	0.9328	1.0400	10.8600	1.1450	16510.0000
			10	0.0000	0.8947	0.9944	1.0570	1.0810	4.7900
			50	0.3969	0.8970	0.9788	0.9942	1.0620	2.0770
		MV / log-rank	90	0.4671	0.9014	0.9760	0.9874	1.0530	1.7370
			0	0.0803	1.0550	1.1280	13.5700	1.2430	35150.0000
			10	0.0803	1.0500	1.1240	1.1690	1.2120	5.6440
			50	0.4666	1.0610	1.1280	1.1340	1.2010	2.5080
			90	0.5232	1.0670	1.1310	1.1290	1.1960	1.8580
		MV <i>optim</i>	-	0.5532	1.0690	1.1090	1.1090	1.1490	2.9110
			0	0.0000	1.0640	1.1430	15.8600	1.5730	35150.0000
			10	0.0000	1.0610	1.1340	1.2250	1.2430	6.3930
			50	0.4666	1.0650	1.1340	1.1460	1.2120	2.5080
		MV / log-rank	90	0.5232	1.0690	1.1340	1.1350	1.2030	1.8580
$n_0 = 200$ $n_1 = 300$ $\tau = 200$ $\theta_0 = 1$ $\theta_1 = 1.1$	τ	MV teórica	0	1.00	44.00	227.00	245.90	457.00	499.00
			10	11.00	69.00	228.00	244.30	428.00	489.00
			50	51.00	118.00	231.00	244.40	374.00	449.00
			90	91.00	150.00	236.00	245.00	342.00	409.00
		MV <i>optim</i>	-	1.17	187.30	200.20	202.50	213.60	495.90
			0	1.00	23.00	248.00	254.80	483.00	499.00
			10	11.00	78.00	241.00	252.00	434.00	489.00
			50	51.00	124.00	238.00	248.60	379.00	449.00
			90	91.00	154.00	239.00	247.30	345.00	409.00
		log-rank máximo	0	0.0626	0.8247	1.0020	9.3510	1.1540	16510.0000
			10	0.0626	0.8090	0.9706	1.0190	1.1160	4.9800
			50	0.3329	0.8468	0.9585	0.9873	1.1020	2.0020
			90	0.5363	0.8694	0.9594	0.9866	1.0920	1.7640
	θ_0	MV <i>optim</i>	-	0.0000	0.9204	0.9954	1.0020	1.0750	4.3400
			0	0.0000	0.9154	1.0500	10.9100	1.4530	16510.0000
			10	0.0000	0.8492	0.9904	1.0690	1.1370	4.9800
			50	0.3329	0.8566	0.9645	0.9953	1.1060	2.0020
		MV / log-rank	90	0.5363	0.8732	0.9626	0.9898	1.0940	1.7640
			0	0.0890	1.0040	1.1450	11.3900	1.3610	31480.0000
			10	0.0890	0.9925	1.1400	1.1840	1.2960	4.9700
			50	0.4357	1.0060	1.1510	1.1440	1.2680	2.2670
			90	0.5898	1.0130	1.1530	1.1350	1.2520	1.9540
		MV <i>optim</i>	-	0.0723	1.0380	1.1050	1.1070	1.1710	6.9480
			0	0.0000	1.0330	1.1740	13.4400	1.9320	31480.0000
			10	0.0000	1.0100	1.1590	1.2410	1.3510	4.9700
			50	0.4357	1.0100	1.1580	1.1530	1.2800	2.3160
		MV / log-rank	90	0.5898	1.0150	1.1560	1.1390	1.2570	1.9540

Tabela 3.7: Medidas-resumo das estimativas do ponto de divisão e dos parâmetros das distribuições antes e depois, para dados gerados sem mudança da distribuição exponencial, com parâmetro $\theta_0 = 1$, $\theta_1 = 1.1$, ponto de mudança $\tau = 500$ ou $\tau = 200$ e percentual de censura em torno de 50% na amostra completa, com $n_0 = 500$ e $n_1 = 1000$ ou $n_0 = 200$ e $n_1 = 300$.

Amostra	Parâmetro	Método de Estimação	Corte	Mín.	Q1	Q2	Média	Q3	Máx.
$n_0 = 500$ $n_1 = 1000$ $\tau = 500$ $\theta_0 = 1$ $\theta_1 = 1$	τ	MV teórica	0	1.00	57.00	728.00	746.50	1446.00	1499.00
			10	11.00	114.00	730.00	744.50	1378.00	1489.00
			50	51.00	207.00	713.50	736.50	1274.00	1449.00
			90	91.00	269.80	734.00	744.10	1224.00	1409.00
		MV <i>optim</i>	-	1.74	444.50	495.00	517.60	560.60	1495.00
			0	1.00	20.00	772.00	756.10	1483.00	1499.00
			10	11.00	112.00	749.00	751.50	1390.00	1489.00
			50	51.00	211.00	728.00	742.80	1281.00	1449.00
		log-rank máximo	90	91.00	271.80	737.00	746.30	1226.00	1409.00
			0	0.1072	0.9578	1.0080	9.2630	1.1250	16510.0000
			10	0.3946	0.9487	1.0050	1.0640	1.0840	4.2400
			50	0.6244	0.9453	1.0030	1.0160	1.0710	1.8000
	θ_0	MV <i>optim</i>	90	0.7126	0.9471	1.0030	1.0100	1.0660	1.5450
			-	0.2854	0.9597	1.0000	1.0040	1.0440	8.9780
			0	0.1253	0.9762	1.0190	11.0200	1.4790	16510.0000
			10	0.3946	0.9567	1.0080	1.0870	1.0990	4.2400
		MV / log-rank	50	0.6244	0.9493	1.0040	1.0210	1.0740	1.8000
			90	0.7126	0.9493	1.0030	1.0120	1.0680	1.5450
			0	0.0979	0.9571	1.0070	12.6500	1.1170	31960.0000
			10	0.3806	0.9477	1.0020	1.0560	1.0760	3.6730
		MV <i>optim</i>	50	0.6255	0.9481	1.0020	1.0160	1.0650	1.8220
			90	0.6717	0.9473	1.0010	1.0090	1.0620	1.4840
			-	0.4993	0.9729	0.9995	0.9999	1.0260	1.4810
			0	0.0979	0.9746	1.0180	14.4800	1.5100	31960.0000
	MV / log-rank	10	0.3806	0.9551	1.0060	1.0800	1.0970	3.6730	
		50	0.6255	0.9503	1.0030	1.0210	1.0710	1.8220	
		90	0.6717	0.9490	1.0020	1.0120	1.0640	1.4840	
		θ_1	MV teórica	0	1.00	29.00	248.00	248.90	470.00
	10			11.00	62.00	246.50	248.30	435.00	489.00
	50			51.00	112.00	245.00	247.40	382.00	449.00
	90			91.00	145.00	249.00	248.60	352.00	409.00
	MV <i>optim</i>		-	2.20	187.60	199.40	202.10	212.50	497.60
			0	1.00	12.00	245.00	248.30	488.00	499.00
			10	11.00	58.00	244.00	246.80	436.00	489.00
			50	51.00	111.00	246.00	247.80	383.00	449.00
	log-rank máximo		90	91.00	146.00	250.00	249.40	353.00	409.00
0			0.1072	0.9331	1.0180	9.5210	1.2140	16510.0000	
10			0.3946	0.9164	1.0100	1.0730	1.1350	4.2400	
50			0.6087	0.9154	1.0040	1.0190	1.1070	1.8000	
$n_0 = 200$ $n_1 = 300$ $\tau = 200$ $\theta_0 = 1$ $\theta_1 = 1$	θ_0	MV teórica	90	0.6927	0.9229	1.0040	1.0120	1.0940	1.5450
			-	0.3777	0.9519	1.0010	1.0060	1.0550	4.1520
			0	0.1072	0.9602	1.0340	11.0300	1.6180	16510.0000
			10	0.3946	0.9272	1.0130	1.0960	1.1580	4.2400
		MV <i>optim</i>	50	0.6087	0.9193	1.0060	1.0220	1.1100	1.8000
			90	0.7126	0.9243	1.0050	1.0140	1.0950	1.5450
			0	0.1126	0.9275	1.0140	10.3900	1.2040	28610.0000
			10	0.3653	0.9119	1.0060	1.0680	1.1320	4.1310
		MV teórica	50	0.5693	0.9160	1.0040	1.0190	1.1050	1.8580
			90	0.6673	0.9213	1.0020	1.0110	1.0930	1.6440
			-	0.4139	0.9604	1.0010	1.0030	1.0450	3.3470
			0	0.1182	0.9563	1.0350	12.2100	1.6130	28610.0000
MV <i>optim</i>	10	0.3653	0.9243	1.0110	1.0910	1.1530	4.1310		
	50	0.5693	0.9188	1.0050	1.0230	1.1080	1.8580		
	90	0.6673	0.9220	1.0020	1.0120	1.0940	1.6440		

Tabela 3.8: Medidas-resumo das estimativas do ponto de divisão e dos parâmetros das distribuições antes e depois, para dados gerados sem mudança da distribuição exponencial, com parâmetro $\theta_0 = 1$, $\theta_1 = 1$, ponto de divisão $\tau = 500$ ou $\tau = 200$ e sem censura na amostra completa, com $n_0 = 500$ e $n_1 = 1000$ ou $n_0 = 200$ e $n_1 = 300$.

Amostra	Parâmetro	Método de Estimação	Corte	Mín.	Q1	Q2	Média	Q3	Máx.
$n_0 = 500$ $n_1 = 1000$ $\tau = 500$ $\theta_0 = 1$ $\theta_1 = 1$	τ	MV teórica	0	1.00	58.00	754.00	748.90	1444.00	1499.00
			10	11.00	106.00	748.00	746.70	1392.00	1489.00
			50	51.00	211.00	745.00	746.10	1282.00	1449.00
			90	91.00	271.00	751.00	747.90	1223.00	1409.00
		MV <i>optim</i>	-	1.09	435.50	490.50	514.80	558.70	1495.00
			0	1.00	24.75	732.50	746.00	1477.00	1499.00
			10	11.00	104.00	735.00	743.80	1391.00	1489.00
			50	51.00	211.00	738.00	744.20	1282.00	1449.00
			90	91.00	267.00	745.00	745.50	1221.00	1409.00
		log-rank máximo	0	0.0814	0.9253	1.0070	9.4230	1.1530	16510.0000
			10	0.0814	0.9060	1.0010	1.0510	1.1080	4.7210
			50	0.3969	0.9146	1.0000	1.0150	1.1000	2.0250
			90	0.4671	0.9165	0.9989	1.0080	1.0930	1.7370
	θ_0	MV <i>optim</i>	-	0.0000	0.9331	0.9993	1.0090	1.0700	35.3400
			0	0.0000	0.9575	1.0250	11.0200	1.5490	16510.0000
			10	0.0000	0.9290	1.0100	1.1050	1.1500	4.7900
			50	0.3969	0.9210	1.0040	1.0250	1.1120	2.0770
		MV / log-rank	90	0.4671	0.9203	1.0010	1.0140	1.1010	1.7370
			0	0.0803	0.9261	1.0080	12.6400	1.1560	31960.0000
			10	0.0803	0.9054	0.9974	1.0480	1.1020	6.0890
			50	0.4016	0.9120	0.9975	1.0130	1.0950	2.3440
			90	0.4805	0.9154	0.9986	1.0080	1.0910	1.7210
		MV <i>optim</i>	-	0.0208	0.9589	0.9990	1.0010	1.0420	2.0230
			0	0.0000	0.9592	1.0260	14.4600	1.5840	31960.0000
			10	0.0000	0.9276	1.0050	1.1000	1.1420	6.0890
			50	0.4016	0.9198	1.0020	1.0230	1.1050	2.3440
		MV / log-rank	90	0.4805	0.9212	1.0020	1.0140	1.0990	1.7210
$n_0 = 200$ $n_1 = 300$ $\tau = 200$ $\theta_0 = 1$ $\theta_1 = 1$	τ	MV teórica	0	1.00	33.00	244.00	247.80	465.00	499.00
			10	11.00	59.00	245.00	247.80	440.00	489.00
			50	51.00	115.00	250.00	250.80	388.00	449.00
			90	91.00	147.00	250.00	250.10	352.00	409.00
		MV <i>optim</i>	-	1.32	187.30	199.50	201.70	212.80	499.20
			0	1.00	15.00	245.00	248.20	483.00	499.00
			10	11.00	59.00	248.00	248.90	439.00	489.00
			50	51.00	114.00	251.00	251.00	388.00	449.00
		log-rank máximo	90	91.00	148.00	250.00	250.40	352.00	409.00
			0	0.0626	0.8756	1.0160	9.6020	1.2550	16510.0000
			10	0.0626	0.8469	0.9999	1.0590	1.1720	4.9800
			50	0.3329	0.8614	0.9972	1.0140	1.1470	2.0020
	θ_0	MV <i>optim</i>	90	0.4713	0.8746	0.9958	1.0090	1.1350	1.7640
			-	0.0004	0.9228	0.9996	1.0040	1.0810	1.9500
			0	0.0000	0.9272	1.0450	11.0000	1.7550	16510.0000
			10	0.0000	0.8750	1.0130	1.1080	1.2170	4.9800
		MV / log-rank	50	0.3329	0.8696	1.0010	1.0220	1.1570	2.0020
			90	0.5363	0.8777	0.9988	1.0120	1.1380	1.7640
		MV teórica	0	0.0802	0.8747	1.0140	10.5900	1.2330	28610.0000
			10	0.0802	0.8472	0.9994	1.0530	1.1650	4.7780
			50	0.3758	0.8622	1.0010	1.0150	1.1500	2.1080
			90	0.5105	0.8747	1.0020	1.0090	1.1350	1.8170
		MV <i>optim</i>	-	0.0631	0.9377	0.9992	1.0030	1.0640	8.8040
			0	0.0000	0.9270	1.0430	12.2000	1.7250	28610.0000
			10	0.0000	0.8716	1.0120	1.1010	1.2080	4.7780
		MV / log-rank	50	0.3758	0.8685	1.0060	1.0230	1.1590	2.1080
			90	0.5105	0.8779	1.0040	1.0120	1.1380	1.8170

Tabela 3.9: Medidas-resumo das estimativas do ponto de divisão e dos parâmetros das distribuições antes e depois, para dados gerados sem mudança da distribuição exponencial, com parâmetro $\theta_0 = 1$, $\theta_1 = 1$, ponto de divisão $\tau = 500$ ou $\tau = 200$ e percentual de censura em torno de 50% na amostra completa, com $n_0 = 500$ e $n_1 = 1000$ ou $n_0 = 200$ e $n_1 = 300$.

É possível perceber que o corte de apenas 10 dos valores de cada extremidade do intervalo de busca para o ponto de mudança já é capaz de amenizar o erro de estimação dos parâmetros. Para o método de máxima verossimilhança otimizada computacionalmente, no entanto, isso foi desconsiderado, mas ainda resultou em estimativas próximas. Para tanto, o chute inicial foi fixado em um valor próximo ao verdadeiro. Na prática, pode-se dar como chute inicial para o ponto de mudança o valor $n/2$ e as estimativas dos parâmetros de um modelo usual em cada metade da amostra podem ser tomadas como valor inicial dos parâmetros monitorados para considerar na otimização.

Com relação ao método mais básico que utiliza a estatística de teste log-rank como critério para estimar o ponto de mudança, observa-se também resultados satisfatórios. Assim, acredita-se que esse método possa ser empregado em situações práticas nas quais a distribuição verdadeira dos dados é desconhecida. Neste caso é identificada uma mudança na sobrevida entre os grupos comparados, mas não se define exatamente a intensidade desta mudança (ao menos que uma distribuição seja assumida para os dados). Contudo, uma simples análise das curvas de sobrevida estimadas pode auxiliar a perceber se esta mudança é para melhor ou pior, o que permite conclusões importantes para o pesquisador.

Resultados dessa simulação com percentual de censura 30% e para outra intensidade de mudança (de $\theta_0 = 1$ para $\theta_1 = 1.3$) foram omitidos aqui, mas o comportamento dos métodos é similar ao que está apresentado, ou seja, apresenta eficiência na estimação dos parâmetros do modelo de ponto de mudança. Este método utiliza um modelo menos usual e não será o enfoque deste trabalho, mas nesta exploração inicial já notamos que é bastante promissor.

Capítulo 4

Monitoramento da Sobrevivência de Populações Heterogêneas

Neste capítulo são apresentados métodos CUSUM para monitorar tempo de sobrevivência considerando um modelo estatístico que leva em conta a heterogeneidade da população. Consideramos o modelo de Cox para captar a influência de características dos indivíduos no risco de falha e com isso são determinados escores ajustados para o gráfico CUSUM. Desenvolvemos inicialmente um gráfico CUSUM que utiliza a razão de verossimilhanças completas do modelo de Cox para identificar alteração no risco de falha *baseline* e em seguida gráficos CUSUM com base em resíduos para este modelo.

4.1 Monitoramento com verossimilhança completa do modelo de Cox

Como já discutido em diversos artigos apresentados neste manuscrito, é importante em estudos de sobrevivência levar em conta que cada indivíduo pode ter um risco de falha influenciado por características particulares. Isto incentiva, portanto, desenvolver gráficos de controle ajustados ao risco, isto é, que avaliam uma mudança no perfil da sobrevida com a devida ponderação pelo efeito da heterogeneidade no estudo.

Uma das vantagens de se utilizar um modelo de sobrevivência para calcular escores CUSUM, além de levar em conta a heterogeneidade das unidades amostrais, é que a variável tempo de sobrevivência conta com mais informação do que os processos de contagem utilizados no teste log-rank, por exemplo. Enquanto na comparação das curvas de sobrevida por estes métodos não paramétricos se requer um maior tempo de acompanhamento dos pacientes, de modo a se contar com mais tempos de falha na amostra para que o teste seja eficiente, um teste feito com a variável tempo de sobrevivência é menos afetado por esse atraso na obtenção de informações para cálculo das estatísticas.

Ao considerar o modelo usual de Cox para construir um monitoramento feito com a função de verossimilhança há que se atentar para as diferenças entre a fórmula completa e a parcial, que é a utilizada de fato na obtenção de estimativas para o vetor de parâmetros β . Admita que fora de controle haja uma mudança proporcional na função de risco *baseline* que pode ir de $\alpha_0(t)$ para $e^\mu \alpha_0(t)$, com $\mu \neq 0$ (e $\mu = 0$ significa que não há mudança). Ou seja, temos que, fora de controle,

a função de risco para um determinado indivíduo é da forma:

$$\alpha_1(t_i|\mathbf{x}_i) = e^\mu \alpha_0(t_i) \exp(\mathbf{x}'_i \beta) = \alpha_0(t_i) \exp[\mu + \mathbf{x}'_i \beta]. \quad (4.1)$$

Assim, é possível definir que, controlado o efeito individual de covariáveis no risco, há num instante qualquer t um risco de falha baseline $\alpha_1(t) = \alpha_0(t) \exp(\mu)$ fora de controle. Com a suposição de que o efeito das covariáveis é bem estimado e fixado, uma mudança na qualidade pode se referir a uma função verossimilhança maior para um valor de $\mu \neq 0$. Com isso, há um interesse em testar hipóteses do tipo $H_0: \mu = 0$ contra $H_1: \mu = \log \xi_1$, em que $\xi_1 > 1$ representa uma intensidade de aumento no risco de falha, e $\xi_1 \in (0, 1)$ (redução no risco) reflete uma melhoria na sobrevivência em comparação com o estado de referência sob controle.

Com este propósito, de modo similar ao que é discutido por Gandy et al. (2010), um escore individual para gráfico CUSUM com base no logaritmo da razão de verossimilhanças, conforme exibido em (2.7), pode ser definido para este teste em termos da função de verossimilhança completa do modelo de Cox nesta situação, conforme (4.2).

$$\begin{aligned} w_i &= \log \left\{ \frac{[\alpha_0(t_i) \exp(\log \xi_1 + \mathbf{x}'_i \beta)]^{\delta_i} \exp[-\xi_1 A_0(t_i) \exp(\mathbf{x}'_i \beta)]}{[\alpha_0(t_i) \exp(\mathbf{x}'_i \beta)]^{\delta_i} \exp[-A_0(t_i) \exp(\mathbf{x}'_i \beta)]} \right\} \\ &= \log \left\{ \xi_1^{\delta_i} \exp[(1 - \xi_1) A_0(t_i) \exp(\mathbf{x}'_i \beta)] \right\} \\ &= \delta_i \log \xi_1 + (1 - \xi_1) A_0(t_i) \exp(\mathbf{x}'_i \beta) \\ &= \delta_i \log \xi_1 + (1 - \xi_1) A(t_i|\mathbf{x}_i, \beta), \end{aligned} \quad (4.2)$$

em que $A_0(t_i)$ é a função de risco acumulada baseline definida pelo modelo de Cox no instante t_i , e $A(t_i|\mathbf{x}_i, \beta) = A_0(t_i) \exp(\mathbf{x}'_i \beta)$ é a estimativa do risco acumulado até o instante t_i para o indivíduo i especificamente, sob a hipótese nula. Na prática se utiliza a estimativa $\hat{\beta}$ obtida pela função de verossimilhança parcial do modelo de Cox nos dados de referência sob controle, e também se considera uma estimativa $\hat{A}_0(t_i)$ para a função de taxa de falha acumulada baseline.

Esta expressão do CUSUM é parecida com a obtida em Gandy et al. (2010), porém considera escores para cada indivíduo (em vez de considerar um escore global com a função de verossimilhança completa da amostra), e é diretamente comparável com a abordagem apresentada em Oliveira et al. (2016), tendo em vista que o procedimento CUSUM é definido por observações e no contexto retrospectivo. Entretanto, neste estudo concentramos em procedimentos CUSUM definidos de forma prospectiva na escala do tempo de calendário. Para utilizar nessa perspectiva, é possível adaptar o método com as informações disponíveis diariamente, como já discutido anteriormente. Por exemplo, a cada dia pode se considerar o escore CUSUM como a média dos valores de w_i dos indivíduos monitorados.

No entanto, apesar de o modelo semiparamétrico de Cox ser muito difundido, um desafio com que se depara para aplicação deste método é a necessidade de estimar a função de risco baseline ($\alpha_0(t)$), para obter o risco acumulado ou a sobrevivência correspondente. Na literatura são encontrados alguns métodos sem utilização de covariáveis como o de Breslow que sugere estimar a função de sobrevivência baseline $S_0(t) = \exp\{-\Lambda(t)\}$, em que $\Lambda(t)$ é o estimador de Nelson-Aalen para a função de risco acumulada. Outra abordagem é o estimador de Breslow que utiliza contribuição das covariáveis (mais detalhes em Cox & Oakes, 1984). A ideia é calcular uma estimativa pontual

nos tempos de falha distintos ordenados $(t_{(1)}, \dots, t_{(k)})$ para a função de risco baseline $\alpha_i = \alpha_0(t_{(i)})$ que leva em conta as variáveis

$$\hat{\alpha}_i = \frac{d_i}{\sum_{j \in R_i} e^{\mathbf{x}'_j \hat{\beta}}}, \quad (4.3)$$

em que R_i é o conjunto de índices dos indivíduos em risco no tempo de falha $t_{(i)}$ e d_i é o total de falhas neste instante. Com isso eles sugerem estimar a função de risco acumulada e correspondente sobrevivência como:

$$\hat{\Lambda}_0(t) = \sum_{i: t_{(i)} < t} \hat{\alpha}_i \quad \text{e} \quad \hat{S}_0(t) = e^{-\hat{\Lambda}_0(t)}. \quad (4.4)$$

Note que a estimativa da função de risco acumulada só é diferente de zero se o valor de t em que é avaliada não for inferior ao menor tempo de falha observado que foi utilizado para construção dessa estimativa.

Alguns métodos propõem parametrizar o risco acumulado baseline como uma função do tipo escada com saltos nos tempos de falha, e com isso estima-se mais parâmetros para compor a função $A_0(t)$. Esta abordagem, no entanto, pode ser computacionalmente difícil, pois há mais um parâmetro a ser estimado para cada tempo de falha observado na amostra. Para o desenvolvimento deste trabalho, devemos considerar a estimativa de $A_0(t)$ como definido em (4.4). O gráfico CUSUM definido com escore apresentado em (4.2) será denotado por Cox LHR CUSUM (Cox LikeliHood Ratio CUSUM).

4.2 Simulação: Cox LHR CUSUM

Nesta seção mostramos um estudo de simulação com obtenção dos valores de referência e limites de controle otimizados para o gráfico Cox LHR CUSUM. Foram feitas 1000 réplicas de dados gerados por um modelo exponencial com uma variável dicotômica associada ao risco de falha. Ou seja, os tempos de falha são gerados de uma distribuição exponencial com parâmetro $\theta_i = \theta_0 \exp(x\beta)$, em que θ_0 é o risco de falha baseline e β é o efeito da variável categórica x no risco de falha.

Os parâmetros considerados foram $\theta_0 = 0.40$, $\beta = 0.4626$ (que foi a estimativa do efeito da variável Sexo calculada nos dados de referência do InCor). A unidade de tempo considerada aqui é em anos. Além disso, tempos de censura também foram gerados da distribuição exponencial com parâmetros variados de modo a obter um percentual de censura médio desejado. Os tempos entre chegadas de indivíduos, em dias, foram também gerados de uma exponencial com parâmetro 5, para simular uma situação prática em que a taxa de entrada é 5 pacientes por dia.

A Tabela 4.1 apresenta os limites de controle ótimos h obtidos para diferentes valores de k que pretendem atender a 3 possíveis ARL_0 (4000, 5000 ou 6000), informando as medidas-resumo dos RL's (mínimo, máximo, primeiro e terceiro quartis, mediana, média e desvio-padrão (DP)) em cada cenário. Neste caso são considerados apenas dados sem censura. Observa-se que para alguns valores de referência k não se consegue obter um limite de controle que atenda ao ARL_0 desejado. Foram testados neste caso uma sequência de valores de k no intervalo $[0, 0.02]$, começando em zero e com incrementos de 0.0001, o que compreende 201 valores, dos quais apenas alguns são mostrados na Tabela 4.1.

Constatou-se, por exemplo, que valores de k a partir de 0.0050 já não retornam limites ótimos

que aproximam satisfatoriamente todos os valores de ARL_0 solicitados. O intervalo de busca utilizado para h foi $[0.01, 10]$, e para alguns casos seria necessário um limite de controle menor que 0.01 para atender ao ARL_0 desejado, o que acreditamos ser impraticável. Em alguns casos o valor ótimo obtido foi próximo ao limite superior do intervalo de busca (para $k = 0.0100$, por exemplo). Nestes casos observa-se que o gráfico CUSUM não cresce o suficiente para conseguir ultrapassar um limite de controle buscado (devido ao valor de referência que é grande e zera os escores, provavelmente). Com isso, o gráfico segue até o final do monitoramento sem apresentar um sinal, o que leva a RL's muito grandes e prejudica a otimização pelo algoritmo de busca. Visto isso, entende-se que na prática não se deve considerar k muito grande, mas um valor pequeno que sirva para controlar melhor os alarmes falsos.

Além disso, percebe-se que para muitas combinações diferentes de k e h é possível alcançar um ARL_0 específico, e, portanto, é preciso avaliar o desempenho do ARL_1 para decidir qual par de valores é mais adequado, fixada uma quantidade ARL_0 . A avaliação do ARL_1 para detectar mudanças de intensidade $\xi_1 = 0.8$ é apresentada na Tabela 4.2 para algumas das combinações de k e h . Novamente, percebe-se uma variedade de possibilidades, e um critério que pode ser considerado para escolher uma combinação ideal é fixar um valor de ARL_0 desejado e para este tomar o valor de referência e limite de controle que o atendem e proporcionam o menor valor de ARL_1 . A Figura 4.1 ilustra essa variação do ARL_1 para alguns diferentes valores de referência (com seus respectivos limites de controle ótimos) na análise com dados sem censura.

Para esta situação, foram obtidos os seguintes pares que minimizam ARL_1 em cada cenário de ARL_0 : $k = 0.0009$ e $h = 0.024978$, que retornam $ARL_0 = 3996.848$ e $ARL_1 = 132.675$; $k = 0.0018$ e $h = 0.017246$ ($ARL_0 = 4996.365$ e $ARL_1 = 179.393$); $k = 0.0020$ e $h = 0.051069$ ($ARL_0 = 5998.046$ e $ARL_1 = 255.661$). Esta análise alerta para o fato de que se deve na prática obter um valor de referência e limite de controle otimizado para cada situação. Na primeira combinação o tempo médio até ocorrer alarme falso é aproximadamente 4000 dias (quase 11 anos), mas em pouco mais de 4 meses (132 dias), em média, já se pode detectar uma redução de 20% no risco de falha. Na segunda situação, espera-se em torno de 6 meses (179 dias) para detectar a mudança, mas o tempo esperado até um alarme falso é de quase 14 anos (5000 dias). No terceiro cenário, deve-se aguardar em média 8 meses e meio (256 dias) para conseguir identificar um aumento na sobrevivência, mas o tempo até alarme falso é maior que 16 anos (6000 dias).

Dentre estes cenários, parece razoável utilizar a terceira das configurações de parâmetros para o gráfico CUSUM, uma vez que esta atrasa a detecção de mudança apenas em 5 meses, aproximadamente, em comparação com a primeira situação, mas se ganha mais que 5 anos na espera até um alarme falso. De qualquer forma, as três possibilidades são recomendáveis e atestam para a eficácia do método Cox LHR CUSUM em dados sem censura.

Para a situação com censura, os resultados também foram favoráveis ao método. A Tabela 4.3 apresenta os limites de controles ótimos para cada valor de referência considerado para alcançar os valores de ARL_0 desejados (1000, 2000 ou 3000) com o gráfico Cox LHR CUSUM em dados com censura em torno de 50%. As medidas-resumo dos RL's são também informadas. Para esta análise foram considerados diferentes valores de ARL_0 como alvos, pois o fato de ter censura levou a tempos de acompanhamento bem menores, o que impossibilita obter RL's mais longos. Novamente, observa-se uma variedade de pares (k, h) que levam aos valores de ARL_0 desejados. Neste caso, para $k > 0.005$ já se começa a notar inconsistência na obtenção do h ótimo para $ARL_0 = 1000$ (que

precisaria de um valor menor que limite inferior de busca), e para $k > 0.083$ não é possível atender aos valores de ARL_0 solicitados (com o algoritmo finalizando no limite superior).

A Figura 4.2 exibe o comportamento do ARL_1 para alguns valores de referência k (com respectivos limites de controle ótimos h). Para esta situação os pares que minimizam o ARL_1 são: $k = 0.0002$ e $h = 0.014604$, que retornam $ARL_0 = 997.241$ e $ARL_1 = 98.654$; $k = 0.0002$ e $h = 0.078384$ ($ARL_0 = 2000.68$ e $ARL_1 = 218.418$); $k = 0.0005$ e $h = 0.642684$ ($ARL_0 = 3000.988$ e $ARL_1 = 543.25$). Em geral, nota-se que o método consegue detectar a mudança em menos de um ano (até menos que 9 meses), com uma espera até alarme falso de pelo menos 5 anos (excluindo o pior caso de 1000 dias, que são também quase 3 anos).

(a) $ARL_0 = 4000$

(b) $ARL_0 = 5000$

(c) $ARL_0 = 6000$

Figura 4.1: Comparação de ARL_1 do gráfico Cox LHR CUSUM em dados sem censura para valores fixos de ARL_0 (a) 4000 (b) 5000 e (c) 6000.

(a) $ARL_0 = 1000$ (b) $ARL_0 = 2000$ (c) $ARL_0 = 3000$

Figura 4.2: Comparação de ARL_1 do gráfico Cox LHR CUSUM em dados sem censura para valores fixos de ARL_0 (a) 1000 (b) 2000 e (c) 3000.

k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
0.0000	0.070652	26	68	5817	3998	6993	12487	3622.1
	0.137636	38	200	6340	4997	7382	12487	3420.5
	0.294491	70	5823	6614	5997	7620	12487	2820.4
0.0001	0.064804	25	66	5831	4003	6998	12487	3620.0
	0.128583	37	199	6340	4996	7375	12487	3419.5
	0.274908	68	5836	6612	6001	7620	12487	2814.4
0.0002	0.058393	24	64	5817	4002	6998	12487	3631.6
	0.116325	36	187	6340	4999	7382	12487	3419.7
	0.255073	65	5823	6614	6000	7620	12487	2815.5
0.0003	0.052137	23	60	5798	3992	7004	12487	3640.2
	0.102745	34	185	6347	5005	7382	12487	3419.0
	0.236307	62	5823	6614	6000	7628	12487	2817.3
0.0004	0.046542	21	59	5817	4000	6995	12487	3633.6
	0.095013	33	183	6347	5001	7388	12487	3424.4
	0.217793	58	5823	6614	6002	7628	12487	2816.6
0.0010	0.021520	15	50	5817	3997	7006	12487	3650.7
	0.051069	27	183	6351	5002	7382	12487	3425.7
	0.143484	48	5836	6623	6004	7620	12487	2812.7
0.0011	0.019131	15	50	5817	4003	7006	12487	3650.4
	0.044813	26	172	6356	5006	7375	12487	3428.9
	0.133110	47	5823	6623	5998	7620	12487	2819.3
0.0012	0.016586	15	49	5822	4017	7015	12487	3649.7
	0.040035	25	172	6351	5004	7375	12487	3429.7
	0.121920	45	5823	6623	5998	7620	12487	2819.2
0.0013	0.013788	14	48	5798	3996	7022	12487	3661.3
	0.035508	24	170	6347	4996	7362	12487	3431.9
	0.111390	43	5836	6626	6000	7628	12487	2822.3
0.0014	0.012059	14	48	5822	4018	7031	12487	3659.7
	0.029913	23	161	6356	5006	7375	12487	3425.4
	0.102337	42	5823	6626	5994	7628	12487	2828.2
0.0015	0.010330	13	45	5847	4043	7036	12487	3655.8
	0.026455	20	160	6356	5011	7375	12487	3423.6
	0.092624	41	5836	6626	5994	7628	12487	2828.2
0.0020	0.010330	15	119	6274	4819	7274	12487	3484.8
	0.012467	16	152	6336	4990	7347	12487	3421.2
	0.051069	34	5836	6642	5998	7642	12487	2841.1
0.0025	0.010330	18	5309	6530	5501	7524	12487	3195.9
	0.010330	18	5309	6530	5501	7524	12487	3195.9
	0.028844	30	5848	6643	6004	7642	12487	2842.9
0.0050	0.010330	46	6200	6849	6861	7742	12487	1823.1
	0.010330	46	6200	6849	6861	7742	12487	1823.1
	0.010330	46	6200	6849	6861	7742	12487	1823.1
0.0100	9.999670	4879	6276	6896	7143	7773	12487	1219.6
	9.999670	4879	6276	6896	7143	7773	12487	1219.6
	9.999670	4879	6276	6896	7143	7773	12487	1219.6
0.0200	9.999670	4879	6276	6896	7143	7773	12487	1219.6
	9.999670	4879	6276	6896	7143	7773	12487	1219.6
	9.999670	4879	6276	6896	7143	7773	12487	1219.6

Tabela 4.1: Limites ótimos e medidas-resumo dos RL's obtidos com o gráfico Cox LHR CUSUM em dados sob controle sem censura, para atender $ARL_0 = 4000, 5000$ ou 6000 .

k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
0.0000	0.070652	23	46	76	136	157	1557	161.0
	0.137636	38	76	125	184	221	1617	182.0
	0.294491	62	134	196	262	295	6702	292.0
0.0001	0.064804	23	45	76	137	158	1562	162.9
	0.128583	37	76	125	185	222	1624	183.0
	0.274908	60	132	194	262	298	6702	295.2
0.0002	0.058393	22	43	75	136	158	1566	163.1
	0.116325	36	73	124	183	224	1630	183.8
	0.255073	58	130	194	265	298	6702	347.6
0.0003	0.052137	21	41	74	135	158	1572	163.6
	0.102745	34	69	120	181	222	1636	184.6
	0.236307	56	127	192	264	297	6702	348.3
0.0004	0.046542	20	40	74	134	157	1577	164.1
	0.095013	33	69	120	181	223	1650	185.5
	0.217793	54	124	191	264	298	6702	352.5
0.0005	0.041103	19	39	73	134	156	1585	164.7
	0.085300	31	67	118	181	222	1960	194.8
	0.204874	53	122	192	264	298	6702	353.6
0.0010	0.021520	15	32	68	136	160	6260	252.3
	0.051069	25	62	115	185	229	6260	273.5
	0.143484	46	116	194	267	308	6702	361.1
0.0011	0.019131	15	32	69	137	162	6260	252.6
	0.044813	23	60	114	183	225	6260	272.6
	0.133110	45	115	196	268	309	6702	361.8
0.0012	0.016586	15	32	70	138	165	6260	255.1
	0.040035	22	59	114	184	229	6260	273.7
	0.121920	43	114	194	268	309	6702	363.6
0.0013	0.013788	14	30	68	137	163	6260	256.2
	0.035508	21	59	113	183	228	6260	274.8
	0.111390	42	112	196	268	310	6702	364.2
0.0014	0.012059	14	31	68	138	166	6260	256.9
	0.029913	20	56	109	180	226	6260	275.1
	0.102337	41	111	194	269	313	6702	364.9
0.0015	0.010330	13	30	67	139	172	6260	257.4
	0.026455	20	55	109	180	226	6260	275.2
	0.092624	40	109	194	267	313	6702	365.3
0.0020	0.010330	15	48	101	175	222	6260	274.8
	0.012467	16	52	112	183	229	6260	277.9
	0.051069	33	100	188	256	308	6260	309.3
0.0025	0.010330	18	67	142	216	269	6260	305.8
	0.010330	18	67	142	216	269	6260	305.8
	0.028844	29	99	191	266	315	6702	375.4
0.0030	0.010330	21	88	184	264	314	6702	389.8
	0.010330	21	88	184	264	314	6702	389.8
	0.014604	23	102	199	275	327	6702	391.5

Tabela 4.2: Medidas-resumo dos RL's obtidos com o gráfico Cox LHR CUSUM em dados fora de controle ($\xi_1 = 0.8$) sem censura, considerando os limites ótimos obtidos para atender $ARL_0 = 4000, 5000$ ou 6000 .

k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
0.0000	0.019791	12	18	40	1006	2722	5157	1480.4
	0.094761	32	95	2759	2002	3216	7072	1565.4
	0.876645	176	2757	3054	3000	3428	7072	862.6
0.0001	0.016994	11	17	36	998	2716	5157	1479.6
	0.087029	30	94	2758	1998	3211	7072	1564.9
	0.819529	168	2760	3056	3004	3430	7072	863.1
0.0002	0.014604	11	17	35	997	2726	5157	1478.9
	0.078384	29	92	2759	2001	3211	7072	1565.3
	0.783569	165	2762	3055	3000	3428	7072	867.4
0.0003	0.012720	10	16	34	1004	2734	5157	1483.5
	0.071312	28	91	2760	2002	3212	7072	1562.3
	0.735507	160	2762	3055	3000	3428	7072	868.3
0.0004	0.010738	10	16	32	1003	2734	5157	1484.2
	0.064804	27	87	2760	2003	3216	7072	1564.4
	0.687696	154	2762	3058	2999	3424	7072	871.6
0.0005	0.010330	10	16	38	1039	2765	5157	1498.3
	0.058141	26	84	2759	1999	3216	7072	1565.6
	0.642684	148	2764	3063	3001	3428	7072	871.7
0.0010	0.010330	12	23	132	1458	3041	5157	1604.0
	0.030982	21	68	2762	2003	3222	7072	1571.7
	0.494222	130	2766	3065	3002	3429	7072	872.9
0.0015	0.010330	14	42	2640	1842	3179	7072	1612.0
	0.014857	18	59	2760	1999	3250	7072	1590.0
	0.338027	112	2764	3065	3000	3429	7072	878.9
0.0020	0.010330	17	93	2842	2174	3270	7072	1532.9
	0.010330	17	93	2842	2174	3270	7072	1532.9
	0.225777	90	2762	3066	2999	3429	7072	887.5
0.0025	0.010330	20	2287	2917	2437	3340	7072	1409.3
	0.010330	20	2287	2917	2437	3340	7072	1409.3
	0.138957	75	2764	3067	3001	3432	7072	890.2
0.0030	0.010330	23	2638	3003	2691	3391	7072	1247.3
	0.010330	23	2638	3003	2691	3391	7072	1247.3
	0.069331	55	2768	3076	3002	3442	7072	899.7
0.0040	0.010330	33	2766	3076	2982	3442	7072	940.5
	0.010330	33	2766	3076	2982	3442	7072	940.5
	0.013128	36	2769	3078	2999	3443	7072	914.7
0.0050	0.010330	45	2803	3095	3095	3458	7072	756.4
	0.010330	45	2803	3095	3095	3458	7072	756.4
	0.010330	45	2803	3095	3095	3458	7072	756.4
0.0083	0.010330	292	2833	3108	3190	3487	7072	561.7
	0.010330	292	2833	3108	3190	3487	7072	561.7
	0.010330	292	2833	3108	3190	3487	7072	561.7
0.0084	9.999670	2201	2840	3110	3209	3494	7072	522.0
	9.999670	2201	2840	3110	3209	3494	7072	522.0
	9.999670	2201	2840	3110	3209	3494	7072	522.0
0.0200	9.999670	2201	2840	3110	3209	3494	7072	522.0
	9.999670	2201	2840	3110	3209	3494	7072	522.0
	9.999670	2201	2840	3110	3209	3494	7072	522.0

Tabela 4.3: Limites ótimos e medidas-resumo dos RL's obtidos com o gráfico Cox LHR CUSUM em dados sob controle com censura em torno de 50%, para atender $ARL_0 = 1000, 2000$ ou 3000 .

k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
0.0000	0.019791	12	17	26	100	71	4232	332.2
	0.094761	30	57	98	219	199	5907	480.6
	0.876645	148	269	350	550	510	5907	658.3
0.0001	0.016994	11	17	25	100	70	4232	332.8
	0.087029	29	56	98	219	199	5907	481.4
	0.819529	143	264	345	547	506	5907	663.7
0.0002	0.014604	11	16	23	99	69	4232	333.2
	0.078384	28	54	96	218	198	5907	482.2
	0.783569	140	261	344	548	508	5907	669.8
0.0003	0.012720	10	16	23	99	70	4232	333.6
	0.071312	27	52	96	225	203	5907	503.2
	0.735507	135	257	340	548	505	5907	677.0
0.0004	0.010738	10	15	21	99	68	4232	343.9
	0.064804	26	51	95	232	206	5907	522.6
	0.687696	130	252	336	545	504	5907	681.1
0.0005	0.010330	10	16	25	104	72	4232	348.0
	0.058141	25	49	94	235	206	5907	535.2
	0.642684	124	248	332	543	501	5907	685.1
0.0006	0.010330	10	17	30	110	81	4232	349.4
	0.051477	24	47	92	234	203	5907	536.2
	0.608861	120	246	330	545	499	5907	693.9
0.0007	0.010330	11	18	34	117	93	4232	359.9
	0.045630	23	46	92	234	204	5907	537.0
	0.589026	119	246	330	551	508	5907	700.2
0.0010	0.010330	12	21	48	166	127	4232	467.1
	0.030982	21	42	90	240	207	5907	565.6
	0.494222	109	240	324	556	506	5907	718.1
0.0015	0.010330	14	30	74	232	196	5907	575.2
	0.014857	16	37	88	254	217	5907	599.1
	0.338027	92	220	312	557	500	5907	754.6
0.0020	0.010330	16	45	109	292	240	5907	634.3
	0.010330	16	45	109	292	240	5907	634.3
	0.225777	78	212	306	568	494	5907	792.2
0.0025	0.010330	20	64	151	365	303	5907	715.2
	0.010330	20	64	151	365	303	5907	715.2
	0.138957	62	201	297	569	489	5907	808.5
0.0030	0.010330	23	92	198	441	379	5907	777.5
	0.010330	23	92	198	441	379	5907	777.5
	0.069331	48	188	285	569	480	5907	826.2
0.0035	0.010330	26	124	242	514	435	5907	829.1
	0.010330	26	124	242	514	435	5907	829.1
	0.032710	38	177	279	587	480	5907	866.7
0.0040	0.010330	29	167	273	598	479	5907	903.8
	0.010330	29	167	273	598	479	5907	903.8
	0.013128	31	177	283	617	496	5907	915.0
0.0045	0.010330	34	205	310	687	552	5907	967.3
	0.010330	34	205	310	687	552	5907	967.3
	0.010330	34	205	310	687	552	5907	967.3

Tabela 4.4: Medidas-resumo dos RL's obtidos com o gráfico Cox LHR CUSUM em dados fora de controle ($\xi_1 = 0.8$) com censura em torno de 50%, considerando os limites ótimos obtidos para atender $ARL_0 = 1000$, 2000 ou 3000.

4.3 Monitoramento com resíduo martingal e deviance

Propomos nesta seção um gráfico de controle CUSUM que considera como escore o resíduo martingal calculado a partir do modelo de Cox (e posteriormente o resíduo *deviance* obtido como transformação deste). Considere o cenário de monitoramento prospectivo que estamos tratando. A cada dia de acompanhamento dos pacientes é possível obter o resíduo de cada observação monitorada de acordo com um modelo de Cox ajustado para os dados de referência sob controle.

O resíduo martingal do modelo de Cox sem covariáveis dependentes do tempo, conforme apresentado em Therneau et al. (1990), é definido como:

$$\hat{M}_i = \delta_i - \hat{A}_0(t_i)e^{X_i'\hat{\beta}}. \quad (4.5)$$

Um contraponto deste resíduo é o fato de ter um comportamento muito assimétrico no lado negativo (seu valor máximo é 1 e mínimo pode ser $-\infty$). Com isso, uma transformação do mesmo para obter uma distribuição mais simétrica é incentivada na prática. Therneau et al. (1990) mostram os resíduos *deviance* obtidos para o modelo de Cox com covariáveis não dependentes do tempo, que são expressos por:

$$\hat{d}_i = \text{sin}(\hat{M}_i) \left[-2 \{ \hat{M}_i + \delta_i \log(\delta_i - \hat{M}_i) \} \right]^{1/2}, \quad (4.6)$$

em que $\hat{M}_i$ é o resíduo martingal, como definido em (4.5) e δ_i é o indicador de falha. Nota-se que o resíduo deviance é zero somente se $\hat{M}_i = 0$. O resíduo martingal pode ser interpretado como uma diferença entre o número observado de falhas e seu valor esperado.

Do ponto de vista da análise prospectiva, temos que a cada dia é possível calcular o resíduo martingal (e deviance) para cada observação monitorada nos dados. Assim, vamos definir um gráfico CUSUM diário, que tem como escore um valor que depende da média dos resíduos calculados na amostra monitorada. Se esse valor é constantemente positivo, é um indício de que o número de falhas que se observa para os novos indivíduos está maior do que o esperado, e assim há um aumento no risco de falha e redução no perfil da sobrevivência. Se por outro lado esse valor se mantém negativo, isto indica que o número observado de falhas é menor que o esperado, e assim há uma melhora na sobrevivência e redução do risco de falha. O gráfico Cox martingal CUSUM pode ser definido para detectar um aumento ou uma redução na sobrevivência ao se considerar o escore da seguinte forma:

$$w_t = a_t \tilde{M}_t, \quad (4.7)$$

em que a_t é um fator para definir a direção que se quer detectar a mudança na sobrevivência ($a_t < 0$ para identificar aumento e $a_t > 0$ para monitorar redução) e $\tilde{M}$ é a média dos resíduos martingal calculados para todas as observações da amostra monitorada no dia t . Neste caso t representa o dia do monitoramento (na escala de tempo diária) e $w_0 = 0$ por convenção do método CUSUM.

Da mesma forma, define-se o gráfico Cox deviance CUSUM, com escore $w_t = a_t \tilde{d}_t$, no qual $\tilde{d}_t$ é a média dos resíduos deviance obtidos para as observações da amostra de monitoramento no dia t . Para o desenvolvimento deste trabalho, é avaliado por simulação apenas o caso $a_t = -1$ (detectar aumento na sobrevivência) por ser de interesse para a aplicação em dados reais, mas entende-se que o aumento (redução) da magnitude desse valor deve apenas aumentar (reduzir) a intensidade da detecção pelo método CUSUM. Concentramos então em fixar esse valor em uma unidade e otimizar apenas outros parâmetros para o gráfico, ou seja, o valor de referência k e o limite de controle h .

4.4 Simulação: Cox Martingal CUSUM e Cox Deviance CUSUM

Nesta seção apresentamos um estudo de simulação mais detalhado realizado para avaliar o comportamento dos gráficos Cox Martingal CUSUM e Cox Deviance CUSUM, com escores definidos a partir desses resíduos do modelo de Cox. Foram feitas 1000 réplicas de dados gerados por um modelo exponencial, de modo similar ao estudo conduzido na avaliação do gráfico Cox LHR CUSUM que foi apresentado na seção anterior. Para esta análise, $\theta_0 = 0.25$ foi considerado para gerar os tempos de falha sob controle.

A Tabela 4.5 apresenta os limites de controle ótimos h obtidos para diferentes valores de k de modo a atender 3 possíveis ARL_0 (500, 1000 ou 2000), informando as medidas-resumo dos RL's (mínimo, máximo, primeiro e terceiro quartis, mediana, média e desvio-padrão (DP)) em cada cenário com o gráfico martingal CUSUM. Observa-se que para alguns valores de referência k não se consegue obter um limite de controle que atenda ao ARL_0 desejado, o que reforça a importância de utilizar um método de otimização desta escolha, para não prejudicar a detecção ao impor um valor de k muito grande, por exemplo. Para valores de k a partir de 0.0050 os limites obtidos não aproximam satisfatoriamente todos os valores de ARL_0 solicitados.

A avaliação do ARL_1 para detectar mudanças de intensidade $\xi_1 = 0.9$ é apresentada na Tabela 4.6 para algumas das combinações de k e h . Novamente, percebe-se uma variedade de possibilidades, que é ilustrada na Figura 4.3. Para esta situação, foram obtidos os seguintes pares que minimizam ARL_1 em cada cenário de ARL_0 : $k = 0.0029$ e $h = 0.01661$, que retornam $ARL_0 = 494.59$ e $ARL_1 = 131.838$; $k = 0.0054$ e $h = 0.02404$ ($ARL_0 = 998.686$ e $ARL_1 = 235.59$); $k = 0.0058$ e $h = 0.36511$ ($ARL_0 = 1999.832$ e $ARL_1 = 533.862$). Esta análise alerta para o fato de que se deve na prática obter um valor de referência e limite de controle otimizado para cada situação.

Na primeira combinação o tempo médio até ocorrer alarme falso é aproximadamente 500 dias (pouco mais que 1 ano), mas em pouco mais de 4 meses (131 dias), em média, já se pode detectar uma redução de 10% no risco de falha. Na segunda situação, espera-se em torno de 8 meses (236 dias) para detectar a mudança, mas o tempo esperado até um alarme falso é de quase 3 anos. No terceiro cenário, deve-se aguardar em média 1 ano e meio (534 dias) para conseguir identificar uma melhora leve na sobrevivência, mas o tempo até alarme falso é de mais de 5 anos (2000 dias).

Na prática, por exemplo, se o pesquisador está disposto a arcar com o custo de alarmes falsos para que seja detectada com rapidez uma mudança leve no risco de falha, a primeira combinação obtida é recomendada. Se, no entanto, é razoável aguardar mais para identificar mudanças leves no perfil da sobrevivência para que ocorram alarmes falsos com menor frequência, então a terceira escolha de parâmetros para o gráfico CUSUM é indicada. O segundo cenário pode ser considerado como um meio termo entre essas escolhas extremas e é também interessante.

A Tabela 4.7 apresenta os limites de controles ótimos para cada valor de referência considerado para alcançar os valores de ARL_0 desejados (500, 1000 ou 2000) com o gráfico deviance CUSUM. As medidas-resumo dos RL's são também informadas. Foi notado um comportamento diferente nos escores obtidos com o resíduo deviance, que têm magnitude maior em comparação com o resíduo martingal, como é esperado pela sua construção. Desta forma, os valores de referência k utilizados na otimização são maiores, e o intervalo de busca para o h foi também ampliado (k é uma sequência de 0 a 0.02 com incrementos de 0.001 e o intervalo de busca para h foi de 0.01 a 200).

Figura 4.3: Comparação de ARL_1 do gráfico Cox martingal CUSUM em dados com censura 50% para valores fixos de ARL_0 (a) 500 (b) 1000 e (c) 2000.

Figura 4.4: Comparação de ARL_1 do gráfico Cox deviance CUSUM em dados com censura 50% para valores fixos de ARL_0 (a) 500 (b) 1000 e (c) 2000.

Neste caso, para $k > 0.160$ já se começa a notar inconsistência na obtenção do h ótimo para $ARL_0 = 500$ (que precisaria de um valor menor que limite inferior de busca), e para $k > 0.184$ não é possível atender aos valores de ARL_0 solicitados (com o algoritmo finalizando no limite superior). Como já discutido na simulação do gráfico Cox LHR CUSUM, isto ocorre devido ao valor de referência ser muito grande, e com isso o gráfico CUSUM não cresce o suficiente para ocorrer alarmes falsos. Por exemplo, para $k = 0.185$, o RL mínimo é 3652 e o algoritmo encerrou no limite superior, então não deve ter sido estimado um ARL próximo de 500, 1000 ou 2000, que foram os valores buscados.

A Figura 4.4 exibe o comportamento do ARL_1 para alguns valores de referência k (com respectivos limites de controle ótimos h). Nota-se novamente que valores de referência muito pequenos ou muito grandes podem prejudicar um pouco o poder de detecção. Para esta situação os pares ótimos que minimizam o ARL_1 são: $k = 0.156$ e $h = 0.06802$, que retornam $ARL_0 = 500.806$ e $ARL_1 = 139.219$; $k = 0.163$ e $h = 0.42736$ ($ARL_0 = 998.674$ e $ARL_1 = 271.146$); $k = 0.104$ e $h = 30.84293$ ($ARL_0 = 2000.117$ e $ARL_1 = 629.417$). Observamos um valor de h grande neste caso, em comparação com outros cenários, e que não é comumente aplicável na prática. Considerando apenas cenários cujo valor de h é menor que 10, uma outra solução é $k = 0.147$ e $h = 9.24551$ ($ARL_0 = 1996.266$ e $ARL_1 = 703.761$).

Com esta análise, podemos observar um comportamento em geral não discrepante entre os dois métodos. Pelo fato de o escore deviance ter valores mais inflacionados, isto pode ter ocasionado alguns resultados inesperados, mas pode-se constatar que o método é eficiente na detecção de mudanças pequenas na sobrevivência. Resultados da simulação para dados sem censura não são exibidos aqui, mas se mostraram similares no desempenho da detecção pelos gráficos de controle.

k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
0.0000	0.0722	12	19	28	501	104	7812	1414.9
	0.2476	25	48	110	999	464	11494	1900.8
	1.6124	92	253	562	1999	4516	11494	2298.1
0.0001	0.0702	11	19	28	498	103	7812	1412.4
	0.2414	25	47	111	999	464	11494	1900.8
	1.5780	91	252	562	1998	4516	11494	2298.5
0.0002	0.0688	11	19	29	502	108	7812	1412.6
	0.2320	24	46	111	999	465	11494	1901.0
	1.5299	88	250	558	2001	4524	11494	2302.1
0.0003	0.0664	11	18	28	500	106	7812	1412.4
	0.2238	24	46	110	999	466	11494	1902.3
	1.5057	88	250	562	1996	4516	11494	2299.5
0.0004	0.0640	11	18	28	498	104	7812	1412.0
	0.2178	23	45	111	1000	467	11494	1902.5
	1.4723	86	249	565	2000	4524	11494	2301.9
0.0037	0.0104	6	15	21	511	89	7812	1460.0
	0.0633	15	32	93	998	448	11494	1919.9
	0.6973	59	220	530	2002	4552	11494	2318.4
0.0038	0.0104	7	15	21	518	97	7812	1468.2
	0.0598	15	32	92	998	441	11494	1920.2
	0.6774	58	217	522	1996	4549	11494	2318.1
0.0039	0.0104	7	15	22	529	100	7812	1479.1
	0.0570	15	31	92	997	442	11494	1920.4
	0.6579	57	216	520	2001	4550	11494	2321.5
0.0040	0.0104	7	16	22	545	105	7812	1497.5
	0.0539	15	31	92	997	442	11494	1920.6
	0.6396	57	214	519	1999	4550	11494	2322.4
0.0050	0.0104	8	19	36	677	183	7812	1631.7
	0.0310	13	28	85	995	443	11494	1921.0
	0.4692	50	201	505	1999	4552	11494	2334.1
0.0100	0.0104	16	62	290	1597	4137	11494	2248.2
	0.0104	16	62	290	1597	4137	11494	2248.2
	0.0579	26	142	470	2000	4633	11494	2377.2
0.0150	0.0104	27	248	869	2452	4888	11494	2468.8
	0.0104	27	248	869	2452	4888	11494	2468.8
	0.0104	27	248	869	2452	4888	11494	2468.8
0.0170	0.0104	30	315	1394	2749	5035	11494	2488.8
	0.0104	30	315	1394	2749	5035	11494	2488.8
	0.0104	30	315	1394	2749	5035	11494	2488.8
0.0190	0.0104	33	410	3908	2940	5085	11494	2456.8
	0.0104	33	410	3908	2940	5085	11494	2456.8
	0.0104	33	410	3908	2940	5085	11494	2456.8
0.0200	0.0104	34	449	4108	3030	5115	11494	2436.9
	0.0104	34	449	4108	3030	5115	11494	2436.9
	0.0104	34	449	4108	3030	5115	11494	2436.9

Tabela 4.5: Limites ótimos e medidas-resumo dos RL's obtidos com o gráfico martingal CUSUM em dados sob controle com censura em torno de 50%, para atender $ARL_0 = 500, 1000$ ou 2000 .

k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
0.0000	0.0722	12	18	25	133	75	6488	542.6
	0.2476	25	44	84	259	188	8821	749.9
	1.6124	87	183	264	547	434	8821	982.3
0.0001	0.0702	11	18	25	134	75	6488	542.8
	0.2414	25	43	84	258	188	8821	750.1
	1.5780	85	182	263	546	433	8821	982.7
0.0002	0.0688	11	18	25	135	79	6488	543.3
	0.2320	24	43	84	260	188	8821	758.2
	1.5299	84	179	261	544	431	8821	983.3
0.0003	0.0664	11	18	25	135	78	6488	543.5
	0.2238	24	42	83	260	190	8821	758.5
	1.5057	83	178	262	544	431	8821	983.6
0.0004	0.0640	11	18	24	135	78	6488	544.0
	0.2178	24	42	83	260	190	8821	758.7
	1.4723	82	177	260	543	430	8821	984.0
0.0005	0.0621	11	18	24	135	79	6488	544.4
	0.2072	23	41	80	259	189	8821	759.2
	1.4427	81	176	260	543	429	8821	984.5
0.0010	0.0505	10	17	24	137	78	6488	556.7
	0.1733	21	39	78	255	188	8821	751.6
	1.3187	77	172	258	543	430	8821	987.1
0.0015	0.0392	9	17	22	135	73	6488	557.7
	0.1558	21	38	79	256	190	8821	753.0
	1.1795	73	167	256	544	431	8821	998.4
0.0020	0.0303	9	16	22	134	69	6488	558.0
	0.1322	20	37	78	256	192	8821	753.9
	1.0593	68	161	253	553	431	8821	1034.0
0.0025	0.0217	8	15	20	132	64	6488	558.2
	0.1087	18	35	74	258	192	8821	773.4
	0.9333	65	157	249	555	428	8821	1051.6
0.0030	0.0159	7	15	20	132	64	6488	558.5
	0.0909	17	34	74	259	193	8821	773.9
	0.8357	62	153	246	554	424	8821	1061.2
0.0035	0.0116	7	14	20	137	63	6488	586.6
	0.0711	16	32	70	252	188	8821	760.0
	0.7365	59	148	243	548	428	8821	1054.8
0.0040	0.0104	7	15	22	144	72	6488	592.7
	0.0539	15	30	68	247	187	8821	754.6
	0.6396	56	141	240	548	428	8821	1062.5
0.0045	0.0104	8	17	24	152	87	6488	593.8
	0.0407	14	28	65	249	182	8821	769.9
	0.5466	53	132	234	545	428	8821	1064.4
0.0050	0.0104	8	18	30	162	99	6488	594.8
	0.0310	13	26	62	239	179	6488	721.4
	0.4692	50	129	231	546	426	8821	1074.5

Tabela 4.6: Medidas-resumo dos RL's obtidos com o gráfico martingal CUSUM em dados fora de controle ($\xi_1 = 0.9$) com censura em torno de 50%, considerando os limites ótimos obtidos para atender $ARL_0 = 500$, 1000 ou 2000.

k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
0.000	68.3248	284	369	406	500	459	6488	607.0
	92.9224	368	495	556	1001	671	11494	1403.6
	131.3223	509	741	886	2000	4109	11494	1995.6
0.001	67.8650	283	368	406	500	459	6488	607.2
	92.1549	366	493	554	1000	670	11494	1404.0
	129.8558	506	734	880	1998	4114	11494	1999.8
0.002	67.4011	283	367	406	500	459	6488	607.3
	91.2795	364	491	552	1002	668	11494	1410.1
	128.5475	503	730	876	2000	4124	11494	2006.1
0.003	66.8788	282	367	405	500	458	6488	607.5
	90.2649	362	488	549	1000	664	11494	1414.3
	127.1092	500	724	870	2000	4138	11494	2011.0
0.004	66.1664	280	365	403	498	456	6488	607.8
	89.3797	360	486	547	998	662	11494	1414.9
	126.0218	498	721	868	2000	4138	11494	2012.8
0.160	0.0111	22	59	89	502	151	7812	1407.2
	0.7057	48	110	160	1000	247	11494	1934.9
	4.9219	115	227	299	2000	4711	11494	2445.7
0.161	0.0104	22	60	90	525	153	7812	1438.0
	0.5964	46	108	158	1002	243	11494	1938.2
	4.6022	112	223	295	2002	4722	11494	2448.7
0.162	0.0104	22	60	92	532	159	7812	1443.5
	0.4947	43	106	154	1000	242	11494	1940.0
	4.3604	110	221	294	2000	4722	11494	2448.8
0.163	0.0104	23	62	94	567	165	7812	1489.1
	0.4274	39	103	152	999	242	11494	1940.4
	4.1340	108	219	292	2000	4722	11494	2450.1
0.164	0.0104	23	64	96	584	172	7812	1508.6
	0.3670	38	102	152	999	239	11494	1944.3
	3.8773	106	217	289	2002	4733	11494	2450.8
0.165	0.0104	23	65	99	624	177	7812	1558.4
	0.3087	37	100	150	998	239	11494	1944.8
	3.6300	103	215	287	1998	4722	11494	2449.4
0.170	0.0104	26	73	115	811	202	7812	1764.3
	0.0863	31	91	141	1001	229	11494	1961.1
	2.4632	89	201	274	1999	4756	11494	2461.6
0.180	0.0104	29	94	154	1226	266	11494	2135.0
	0.0104	29	94	154	1226	266	11494	2135.0
	0.7661	61	169	248	2000	4764	11494	2486.6
0.184	0.0104	31	102	176	1482	4078	11494	2295.0
	0.0104	31	102	176	1482	4078	11494	2295.0
	0.3853	53	159	238	1998	4769	11494	2491.5
0.185	199.9996	3652	4623	5062	5235	5700	11494	867.1
	199.9996	3652	4623	5062	5235	5700	11494	867.1
	199.9996	3652	4623	5062	5235	5700	11494	867.1
0.200	199.9996	3652	4623	5062	5235	5700	11494	867.1
	199.9996	3652	4623	5062	5235	5700	11494	867.1
	199.9996	3652	4623	5062	5235	5700	11494	867.1

Tabela 4.7: Limites ótimos e medidas-resumo dos RL's obtidos com o gráfico deviance CUSUM em dados sob controle com censura em torno de 50%, para atender $ARL_0 = 500, 1000$ ou 2000 .

k	h	Mínimo	Quartil 1	Mediana	Média	Quartil 3	Máximo	DP
0.000	68.3248	275	343	370	383	404	4391	136.8
	92.9224	352	445	484	514	536	6488	266.9
	131.3223	473	613	677	746	772	6488	401.6
0.001	67.8650	274	342	370	382	403	4391	136.9
	92.1549	351	443	482	512	535	6488	267.1
	129.8558	470	609	673	742	768	6488	402.4
0.002	67.4011	274	341	369	382	403	4391	137.0
	91.2795	349	441	480	510	533	6488	267.2
	128.5475	467	605	669	739	765	6488	403.6
0.003	66.8788	273	341	369	381	402	4391	137.1
	90.2649	347	439	478	508	530	6488	267.4
	127.1092	463	601	665	735	761	6488	404.8
0.004	66.1664	271	339	367	380	401	4391	137.2
	89.3797	345	437	476	506	528	6488	267.5
	126.0218	461	599	663	733	758	6488	406.9
0.160	0.0111	20	55	79	141	119	6271	477.9
	0.7057	47	98	132	277	189	6488	790.5
	4.9219	111	198	239	723	307	8821	1453.6
0.161	0.0104	20	55	81	146	123	6271	493.9
	0.5964	44	96	129	275	187	6488	791.0
	4.6022	108	194	237	720	304	8821	1455.1
0.162	0.0104	22	56	83	148	124	6271	493.9
	0.4947	41	93	126	272	184	6488	791.4
	4.3604	106	192	236	727	302	8821	1468.0
0.163	0.0104	23	57	85	159	128	6271	538.3
	0.4274	39	92	125	271	183	6488	791.6
	4.1340	104	190	234	730	301	8821	1473.0
0.164	0.0104	23	58	86	166	129	6271	560.3
	0.3670	38	90	124	280	182	6488	820.9
	3.8773	102	187	232	738	299	8821	1490.0
0.165	0.0104	23	59	86	168	132	6271	560.3
	0.3087	37	89	121	278	181	6488	821.2
	3.6300	99	184	229	741	297	8821	1498.2
0.170	0.0104	26	66	95	221	149	6488	726.0
	0.0863	31	80	111	289	170	6488	880.6
	2.4632	84	171	218	726	284	8821	1498.7
0.180	0.0104	29	81	120	374	188	6488	1065.0
	0.0104	29	81	120	374	188	6488	1065.0
	0.7661	60	140	193	766	255	8821	1598.1
0.181	0.0104	29	84	124	393	194	6488	1095.9
	0.0104	29	84	124	393	194	6488	1095.9
	0.6716	58	138	192	777	253	8821	1621.2
0.182	0.0104	30	86	126	411	199	6488	1125.7
	0.0104	30	86	126	411	199	6488	1125.7
	0.5608	56	134	189	775	250	8821	1622.1
0.183	0.0104	30	88	130	427	201	6488	1152.9
	0.0104	30	88	130	427	201	6488	1152.9
	0.4709	54	131	186	781	248	8821	1631.6

Tabela 4.8: Medidas-resumo dos RL's obtidos com o gráfico deviance CUSUM em dados fora de controle ($\xi_1 = 0.9$) com censura em torno de 50%, considerando os limites ótimos obtidos para atender $ARL_0 = 500$, 1000 ou 2000.

Capítulo 5

Simulação Comparativa

Neste capítulo fazemos uma breve comparação dos métodos discutidos. Devemos considerar novamente 1000 réplicas de amostras geradas por um modelo exponencial, mas desta vez com duas covariáveis influenciando no risco de falha. Consideramos uma variável dicotômica (de influência similar ao sexo nos dados do InCor) e outra numérica (parecido com a idade nos dados do InCor). A ideia é saber qual dos gráficos CUSUM apresentados consegue ter melhor desempenho na identificação de reduções no risco de falha. Os três métodos Cox CUSUM (LHR, martingal e deviance) serão então avaliados e comparados.

Para esta simulação consideramos um método alternativo para estimar o ARL com base em $m = 1000$ réplicas. É comum ao gerar dados sob controle para obtenção de limites ótimos que para algumas das réplicas seja difícil observar um alarme falso, por menor que seja o h considerado, o que pode resultar em algumas réplicas não apresentarem alarme em todo o período de acompanhamento prospectivo.

Em simulações anteriores, consideramos uma estimativa pessimista, na qual era anotado como RL da réplica o valor correspondente ao sucessor do período total de acompanhamento dos indivíduos (como se o alarme fosse ocorrer no dia seguinte). Isso ocasiona um viés na obtenção desta estimativa e então para a comparação dos 3 métodos aqui será considerada uma estimativa de ARL que leva em conta uma correção para as amostras que terminam o monitoramento sem apresentar alarme.

A ideia é admitir uma distribuição geométrica para a variável RL, que é o número de inspeções até que se observe um alarme, e neste caso pode ser entendido como o número de dias monitorados até a ocorrência de um alarme. Os casos em que a amostra encerra o período de monitoramento sem apresentar um alarme serão então considerados como censuras da variável RL, e assim o estimador da média de uma geométrica censurada é considerado como estimativa do ARL.

Desta forma, as informações provenientes das m réplicas simuladas são $y_1, y_2, \dots, y_m$, os valores observados da variável RL ou de seu valor censurado, que corresponde ao tempo máximo de acompanhamento dos indivíduos, e $\delta_1, \delta_2, \dots, \delta_m$, as respectivas variáveis indicadoras de alarme (1 se o valor observado de RL corresponde de fato a um alarme e 0 caso houve censura no alarme por término do monitoramento). Com isso,

$$ARL(h) = \frac{\sum_{j=1}^m y_j}{\sum_{j=1}^m \delta_j} \quad (5.1)$$

é a estimativa do ARL das m réplicas, que considera a possibilidade de censura nos alarmes. No

caso em que não se observam censuras nos alarmes, essa estimativa corresponde à própria média amostral usual. Se há censura, o denominador da expressão é reduzido, o que produz uma estimativa um pouco maior, o que é coerente no sentido de que se houvesse mais tempo de monitoramento o alarme de fato poderia ser observado em um valor maior. Assim, esse ajuste pode ser útil para contornar essa dificuldade nos aspectos da simulação.

Os dados da simulação foram gerados com um modelo exponencial com alguns parâmetros obtidos a partir de análise dos dados do InCor, de modo a obter dados com comportamento similar para unificar esta comparação. Para os 3 métodos considerados, são informados os limites de controle ótimos h obtidos para atender o ARL_0 desejado (2000 neste caso) em diferentes valores de referência k , as estimativas do ARL calculadas com estas configurações (ARL_h) e correspondentes números de Amostras Sem Alarme (ASA), ou seja, que tiveram um alarme censurado.

Nos gráficos exibidos na Figura 5.1 são apresentadas as estimativas de ARL_h para todos os cenários de k considerados e as respectivas quantidades de ASA (em porcentagem) com os limites de controle otimizados para um $ARL_0 = 2000$. No eixo y é informado o logaritmo natural da razão entre a estimativa de ARL e o ARL_0 alvo desejado, como forma de avaliar a proximidade da estimativa e o valor solicitado (quando essa medida fica em torno de zero), e para que os gráficos possam ser vistos na mesma escala, que é também regulada para valores muito grandes obtidos.

Os valores de referência k foram considerados numa malha de 600 pontos num intervalo adequado para cada método (e também o valor zero). Nota-se que a utilização de um k muito grande pode aumentar o número de amostras sem alarme, o que prejudica a estimação do ARL, e consequentemente a obtenção do limite de controle ótimo para atender a um ARL_0 especificado.

É possível perceber que há um valor máximo que pode ser considerado sem comprometer a obtenção dos parâmetros que levam ao ARL_0 de interesse, que mantém o percentual de amostras sem alarme em um valor pequeno o suficiente para não superestimar o ARL. Percebe-se que algo inferior a 20% das amostras sem alarme não parece dificultar a obtenção dos parâmetros para atender ao ARL_0 . E para valores muito grandes de k fica intolerável aceitar os parâmetros obtidos, pois, possivelmente devido ao total de amostras sem alarme que cresce consideravelmente, a estimativa do ARL fica muito superior ao desejado.

As Figuras 5.2 e 5.3 mostram as mesmas informações para os cenários em que se deseja valores de ARL_0 maiores, respectivamente, 4000 e 8000. O comportamento é similar ao primeiro cenário. Nota-se apenas para o último caso que o percentual de amostras sem alarme é aumentado para algo próximo de 40%, mas isso também não compromete a obtenção dos limites de controle otimizados para estas situações.

De fato, observa-se que mais pontos da malha de valores considerada resultaram em estimativas de ARL próximas do valor desejado, principalmente para o método LHR CUSUM. Isso pode ocorrer devido ao fato de que para um ARL_0 menor desejado espera-se um valor de h menor, e que possivelmente não conseguiu ser alcançado, mesmo com o limite inferior considerado na busca que foi de $1e-6$. No entanto, se o ARL_0 desejado é bem maior, é mais fácil que um valor de h maior que isso consiga se adequar à situação, e por isso para $ARL_0 = 8000$, mais combinações de k e h tiveram estimativas de ARL próximas do desejado.

A Tabela 5.1 apresenta alguns dos resultados da simulação para dados com censura em torno de 35%. São apresentados dois cenários, um com dados sob controle ($\xi = 1$) e outro em que há mudança na sobrevivência da ordem de $\xi = 0.8$ (redução de 20% no risco de falha). Foram utilizados

os limites de controle ótimos obtidos para $ARL_0 = 2000$, os quais são omitidos na tabela. Para os resultados exibidos nota-se que o ARL estimado para os dados sob controle em geral se aproxima do valor alvo 2000. Observa-se que apenas para o método LHR CUSUM algumas amostras não conseguem detectar a mudança, mas que este apresenta valores de ARL estimados bem menores, assim, em média consegue detectar rapidamente que há uma mudança.

LHR Cox CUSUM					Martingal Cox CUSUM					Deviance Cox CUSUM				
k	$\xi = 1$		$\xi = 0.8$		k	$\xi = 1$		$\xi = 0.8$		k	$\xi = 1$		$\xi = 0.8$	
(10^{-4})	ARL_h	ASA	ARL_h	ASA	(10^{-3})	ARL_h	ASA	ARL_h	ASA	(10^{-2})	ARL_h	ASA	ARL_h	ASA
0.0	2047	136	86	2	0.0	1995	112	156	0	0.0	2003	68	845	0
0.1	1959	131	84	2	0.1	1990	112	154	0	0.1	2000	68	842	0
0.2	2010	134	86	2	0.2	2006	113	154	0	0.2	1996	68	840	0
0.3	1960	131	84	2	0.3	1987	112	154	0	0.3	1994	68	837	0
0.4	2065	137	89	2	0.4	2003	113	153	0	0.4	2006	69	835	0
0.5	1943	130	83	2	0.5	2008	113	154	0	0.5	2000	69	831	0
0.6	1978	132	87	2	0.6	2007	113	153	0	0.6	1997	69	828	0
0.7	2065	137	88	2	0.7	2006	113	151	0	0.7	2005	70	824	0
0.8	1943	130	86	2	0.8	1999	113	150	0	0.8	2000	70	820	0
0.9	1997	133	87	2	0.9	1999	113	150	0	0.9	1996	70	816	0
1.0	1943	130	86	2	1.0	1999	113	149	0	1.0	2003	71	812	0
1.5	2066	137	89	2	1.5	1995	114	145	0	1.5	1995	73	788	0
2.0	2080	138	91	2	2.0	2009	115	144	0	2.0	2000	75	770	0
2.5	2059	137	90	2	2.5	1999	115	141	0	2.5	1995	76	754	0
3.0	2060	137	91	2	3.0	1998	115	139	0	3.0	2000	78	736	0
3.1	2112	140	93	2	3.1	1993	115	139	0	3.1	1999	78	734	0
3.2	1872	126	86	2	3.2	2002	115	141	0	3.2	2006	79	730	0
3.3	1940	130	87	2	3.3	2004	115	141	0	3.3	2002	79	726	0
3.4	1991	133	91	2	3.4	2007	115	142	0	3.4	1998	79	723	0
3.5	2026	135	92	2	3.5	1995	114	143	0	3.5	1994	79	720	0
3.6	1922	129	86	2	3.6	2010	115	143	0	3.6	2005	80	717	0
3.7	1957	131	88	2	3.7	2007	115	142	0	3.7	2000	80	713	0
3.8	2008	134	91	2	3.8	1999	115	140	0	3.8	1997	80	710	0
3.9	2026	135	92	2	3.9	2002	115	141	0	3.9	1993	80	707	0
4.0	2111	140	94	2	4.0	1995	115	139	0	4.0	1989	80	704	0
4.1	2146	142	95	2	4.1	2011	116	139	0	4.1	2000	81	701	0
4.2	1871	126	81	2	4.2	2003	116	137	0	4.2	1996	81	697	0
4.3	1889	127	84	2	4.3	2006	116	137	0	4.3	1992	81	694	0
4.4	1922	129	84	2	4.4	2000	116	136	0	4.4	1988	81	691	0
4.5	1973	132	85	2	4.5	2000	116	136	0	4.5	1998	82	688	0
4.6	2024	135	89	2	4.6	2007	116	137	0	4.6	1994	82	684	0
4.7	2076	138	92	2	4.7	2023	117	137	0	4.7	2003	83	680	0
4.8	2128	141	93	2	4.8	1969	114	135	0	4.8	2000	83	678	0
4.9	2181	144	95	2	4.9	2041	118	137	0	4.9	1995	83	674	0
5.0	2233	147	96	2	5.0	1966	114	135	0	5.0	2003	84	669	0

Tabela 5.1: Estimativas de ARL (ARL_h) e respectivas quantidades de ASA dos três métodos comparados sob controle ($\xi = 1$) e com mudança ($\xi = 0.8$) para diferentes valores de referência k , considerando limites de controle ótimos obtidos para atender $ARL_0 = 2000$ em dados com censura em torno de 35%.

A Figura 5.4 apresenta graficamente os resultados de ARL para dados fora de controle, com a inclusão de todos os 601 pontos testados. Neste caso os parâmetros utilizados k e h tentam atender a um $ARL_0 = 2000$. Aparentemente o ARL estimado começa em um valor que é reduzido à medida que k aumenta, e a partir de determinado valor para k ele começa a crescer (como efeito das dificuldades de obtenção dos valores ótimos já discutidas), o que fica mais nítido para os métodos martingal e deviance. É possível então estabelecer uma combinação de k e h que retorna o menor ARL_1 possível para essa situação.

Os métodos LHR e Deviance CUSUM apresentam maior dificuldade de se perceber graficamente qual o ponto ótimo, mas nota-se que o melhor ARL_1 deve ser menor que 100 para o LHR (80), para

o método Martingal fica um pouco maior que 100 (133) e para o Deviance foi registrado um valor de 215. Foram, porém, constatadas 2 réplicas para os quais o LHR CUSUM não apresentou alarme e 1 para o deviance, e o martingal CUSUM conseguiu detectar a mudança em todas as simulações. Este comportamento é similar para os outros cenários de ARL_0 , como pode ser visto nas Figuras 5.5 ($ARL_0 = 4000$) e 5.6 ($ARL_0 = 8000$). A diferença está no valor de referência k ideal que é maior para estes casos, visto que deve atender a maiores valores de ARL_0 , e que o ARL_1 estimado como melhor alternativa é um pouco maior que no primeiro cenário.

Contudo, nota-se que mesmo para a situação em que $ARL_0 = 8000$, este valor não ultrapassa 400 para o gráfico martingal CUSUM (378), é pouco maior que 200 para o LHR CUSUM (230) e para o deviance foi obtido um valor de 332. Porém, foram observadas 5 réplicas que não apresentaram detecção pelo LHR CUSUM e 13 para o deviance, enquanto que o martingal CUSUM conseguiu detectar a mudança em todas as simulações, embora, em média, mais tardiamente. Podemos creditar aos métodos uma boa capacidade de detecção, visto que neste caso só demoram no máximo pouco mais que 1 ano para identificar uma mudança de apenas 20% no risco de falha dos indivíduos e quase 22 anos (8000 dias), em média, para apresentar um alarme falso.

Percebe-se que o método LHR CUSUM é muito eficiente na detecção de uma mudança quando se sabe a verdadeira intensidade para utilizar na construção do teste inerente ao gráfico. Mas com esta análise entendemos que os métodos CUSUM com resíduos do modelo de Cox são uma boa alternativa na prática para monitoramento da sobrevivência de pacientes, capazes de alertar sobre possíveis alterações no comportamento esperado sob controle.

Os resultados para as simulações com dados sem censura são também apresentados de forma gráfica. Nas Figuras 5.7, 5.8 e 5.9 são mostradas as estimativas de ARL para dados sob controle comparadas ao valor desejado, respectivamente para $ARL_0 = 2000$, $ARL_0 = 4000$ e $ARL_0 = 8000$. O percentual de amostras sem alarme é também informado. Nas Figuras 5.10, 5.11 e 5.12 pode ser avaliado o comportamento do ARL estimado para dados com mudança de intensidade $\xi = 0.8$, para os respectivos cenários de ARL_0 desejados (2000, 4000 e 8000).

Para o primeiro cenário, nota-se que o que ARL_1 estimado para o LHR CUSUM é inferior a 100 dias (70), para o martingal CUSUM este valor é um pouco maior que 100 (141) e para o deviance CUSUM foi obtido um valor de 215. Somente para o deviance, neste caso, foi observada uma réplica que não conseguiu detectar a mudança. Para o cenário com $ARL_0 = 8000$, o menor ARL_1 estimado para os gráficos foi de 208 para o LHR (3 amostras sem alarme), 422 para o martingal (todas as réplicas detectam a mudança) e 327 para o deviance (12 amostras sem alarme).

Em comparação com a simulação de dados com censura, neste caso há menor quantidade de réplicas que não conseguem identificar a mudança na sobrevivência e essa detecção ocorre mais rápido, em média, exceto para o método martingal que apresenta valores um pouco maiores. Ainda assim, vemos que os métodos são eficientes ao demorar em geral menos que 1 ano para identificar uma mudança, mas com mais que 20 anos, em média, para se observar um alarme falso.

Vale ressaltar que embora os dados sejam gerados sem censura na sobrevivência, isto se refere apenas ao momento final do estudo, em que todas as falhas já têm sido observadas. Durante o acompanhamento prospectivo é esperado que o percentual de censuras seja muito alto no período inicial, principalmente, e que este se mantém em um valor alto por bom tempo, então entende-se que o método já enfrenta naturalmente a dificuldade da alta quantidade de censura nos tempos de vida que sabe-se ser importante em análise de sobrevivência.

(a) *ARL estimado | LHR Cox CUSUM* (b) *ASA (%) | LHR Cox CUSUM* (c) *ARL estimado | Martingal Cox CUSUM* (d) *ASA (%) | Martingal Cox CUSUM* (e) *ARL estimado | Deviance Cox CUSUM* (f) *ASA (%) | Deviance Cox CUSUM*

Figura 5.1: Estimativas do ARL dos 3 métodos comparados em dados sob controle com censura em torno de 35% considerando as configurações de k e h obtidas para $ARL_0 = 2000$ (figuras à esquerda) e respectivos percentuais de amostras sem alarme na simulação (figuras à direita).

(a) *ARL estimado | LHR Cox CUSUM* (b) *ASA (%) | LHR Cox CUSUM* (c) *ARL estimado | Martingal Cox CUSUM* (d) *ASA (%) | Martingal Cox CUSUM* (e) *ARL estimado | Deviance Cox CUSUM* (f) *ASA (%) | Deviance Cox CUSUM*

Figura 5.2: Estimativas do ARL dos 3 métodos comparados em dados sob controle com censura em torno de 35% considerando as configurações de k e h obtidas para $ARL_0 = 4000$ (figuras à esquerda) e respectivos percentuais de amostras sem alarme na simulação (figuras à direita).

(a) *ARL estimado | LHR Cox CUSUM* (b) *ASA (%) | LHR Cox CUSUM* (c) *ARL estimado | Martingal Cox CUSUM* (d) *ASA (%) | Martingal Cox CUSUM* (e) *ARL estimado | Deviance Cox CUSUM* (f) *ASA (%) | Deviance Cox CUSUM*

Figura 5.3: Estimativas do ARL dos 3 métodos comparados em dados sob controle com censura em torno de 35% considerando as configurações de k e h obtidas para $ARL_0 = 8000$ (figuras à esquerda) e respectivos percentuais de amostras sem alarme na simulação (figuras à direita).

(a) *ARL estimado | LHR Cox CUSUM* (b) *ARL estimado | Martingal Cox CUSUM* (c) *ARL estimado | Deviance Cox CUSUM*

Figura 5.4: Estimativas do ARL dos 3 métodos comparados em dados com mudança de intensidade $\xi = 0.8$ e censura em torno de 35% na amostra sob controle, considerando as configurações ótimas de k e h obtidas para atender um $ARL_0 = 2000$.

(a) *ARL estimado | LHR Cox CUSUM* (b) *ARL estimado | Martingal Cox CUSUM* (c) *ARL estimado | Deviance Cox CUSUM*

Figura 5.5: Estimativas do ARL dos 3 métodos comparados em dados com mudança de intensidade $\xi = 0.8$ e censura em torno de 35% na amostra sob controle considerando as configurações ótimas de k e h obtidas para atender um $ARL_0 = 4000$.

(a) *ARL estimado | LHR Cox CUSUM* (b) *ARL estimado | Martingal Cox CUSUM* (c) *ARL estimado | Deviance Cox CUSUM*

Figura 5.6: Estimativas do ARL dos 3 métodos comparados em dados com mudança de intensidade $\xi = 0.8$ e censura em torno de 35% na amostra sob controle, com as configurações ótimas de k e h obtidas para atender $ARL_0 = 8000$.

(a) *ARL estimado | LHR Cox CUSUM* (b) *ASA (%) | LHR Cox CUSUM* (c) *ARL estimado | Martingal Cox CUSUM* (d) *ASA (%) | Martingal Cox CUSUM* (e) *ARL estimado | Deviance Cox CUSUM* (f) *ASA (%) | Deviance Cox CUSUM*

Figura 5.7: Estimativas do ARL dos 3 métodos comparados em dados sob controle e sem censura com as configurações de k e h obtidas para $ARL_0 = 2000$ (figuras à esquerda) e respectivos percentuais de amostras sem alarme na simulação (figuras à direita).

(a) *ARL estimado | LHR Cox CUSUM* (b) *ASA (%) | LHR Cox CUSUM* (c) *ARL estimado | Martingal Cox CUSUM* (d) *ASA (%) | Martingal Cox CUSUM* (e) *ARL estimado | Deviance Cox CUSUM* (f) *ASA (%) | Deviance Cox CUSUM*

Figura 5.8: Estimativas do ARL para as configurações de k e h obtidas para $ARL_0 = 4000$ em dados sob controle e sem censura (figuras à esquerda) e respectivos percentuais de amostras sem alarme na simulação (figuras à direita) para os 3 métodos comparados.

(a) *ARL estimado | LHR Cox CUSUM* (b) *ASA (%) | LHR Cox CUSUM* (c) *ARL estimado | Martingal Cox CUSUM* (d) *ASA (%) | Martingal Cox CUSUM* (e) *ARL estimado | Deviance Cox CUSUM* (f) *ASA (%) | Deviance Cox CUSUM*

Figura 5.9: Estimativas do ARL para as configurações de k e h obtidas para $ARL_0 = 8000$ em dados sob controle e sem censura (figuras à esquerda) e respectivos percentuais de amostras sem alarme na simulação (figuras à direita) para os 3 métodos comparados.

(a) *ARL estimado | LHR Cox CUSUM* (b) *ARL estimado | Martingal Cox CUSUM* (c) *ARL estimado | Deviance Cox CUSUM*

Figura 5.10: Estimativas do ARL com as configurações ótimas de k e h obtidas para atender um $ARL_0 = 2000$ em dados com mudança de intensidade $\xi = 0.8$ e sem censura para os 3 métodos comparados.

(a) *ARL estimado | LHR Cox CUSUM*

(b) *ARL estimado | Martingal Cox CUSUM*

(c) *ARL estimado | Deviance Cox CUSUM*

Figura 5.11: Estimativas do ARL com as configurações ótimas de k e h obtidas para atender um $ARL_0 = 4000$ em dados com mudança de intensidade $\xi = 0.8$ e sem censura para os 3 métodos comparados.

(a) *ARL estimado | LHR Cox CUSUM* (b) *ARL estimado | Martingal Cox CUSUM* (c) *ARL estimado | Deviance Cox CUSUM*

Figura 5.12: Estimativas do ARL com as configurações ótimas de k e h obtidas para atender um $ARL_0 = 8000$ em dados com mudança de intensidade $\xi = 0.8$ e sem censura para os 3 métodos comparados.

Capítulo 6

Aplicação dos Métodos nos Dados do InCor

Neste capítulo os gráficos CUSUM que foram apresentados serão aplicados nos dados do InCor. Com isso analisaremos a competência dos métodos em detectar uma alteração na sobrevivência dos pacientes ao longo do tempo. Apresentamos inicialmente os métodos de monitoramento para populações heterogêneas (gráficos Cox CUSUM), que são similares no aspecto de levar em conta um modelo para explicar o risco de falha dos pacientes.

Análises prévias deste conjunto de dados sugerem que há uma melhoria na sobrevivência dos pacientes ao longo do tempo, que pode ter sido ocasionada por mudanças nos tratamentos/procedimentos realizados no hospital ou mesmo pelas características dos pacientes (Marçula et al., 2011). Desta forma, a utilização de um modelo para prever o risco específico associado a cada indivíduo pode ser importante para evitar que um método de monitoramento sinalize uma melhora que não é devida à melhoria na qualidade de tratamento oferecida, mas sim a fatores aleatórios.

Dentre as variáveis observadas como significantes nas análises e as que os pesquisadores consideraram importante para explicar o risco de falha dos pacientes, a fração de ejeção do ventrículo esquerdo é posta em destaque. Assim, primeiramente é realizado um monitoramento que considera um modelo apenas com esta variável explanatória, para investigar se ela é capaz de explicar sozinha o risco de falha dos indivíduos e permitir o monitoramento. Uma característica presente nos dados é a grande frequência de dados omissos (NA), e desta forma, exceto pela idade e sexo, as variáveis foram categorizadas, incluindo a categoria NA como um dos fatores possíveis. Para o caso da fração de ejeção, por exemplo, são ao todo 5 categorias, ao contar com a NA. Esta estratégia permite que nenhuma observação precise ser retirada da análise, e acreditamos que consiga captar com variáveis categóricas o efeito das informações do indivíduo no seu risco de falha.

Para o gráfico Cox LHR CUSUM, que utiliza a razão de verossimilhanças, deve-se definir a intensidade de mudança que se deseja monitorar. Testamos alguns valores de ξ_1 inferiores a 1 (que indicam redução no risco de falha / aumento na sobrevivência). Além disso uma série de valores para k foi considerada (101 valores de 0 a 0.1). Para a maioria é observado um comportamento similar do gráfico e exibimos aqui algumas que valem destacar. Ao testar redução no risco de falha em 10% ($\xi_1 = 0.9$) o gráfico é firme ao acumular escores e indica que há uma melhora na sobrevivência dos pacientes.

A Figura 6.1 mostra o gráfico para diferentes valores de referência k . O período de monitora-

mento dos dados compreende os dias 2859 a 6905 do acompanhamento dos pacientes. Observa-se que para $k = 0$ o CUSUM cresce muito rápido, de forma linear. Para $k = 0.03$, ele continua acumulando escores e indicando então que há mudança, mas os valores são menores, e demoram um pouco mais a apresentar o crescimento constante linear como no cenário anterior. Para k a partir de 0.05 já se observa um comportamento diferente, em que o CUSUM não consegue crescer o suficiente para indicar que há mudança em quase todo o período do monitoramento, e começa a apresentar crescimento só a partir de um pouco antes do dia 6000 em diante. Para $k = 0.058$ o CUSUM sai do zero em alguns pontos, mas não alcança quantidades expressivas, e para valores de referência maiores que isso os gráficos CUSUMs são completamente anulados, sem mostrar nenhuma evidência de que ocorre a mudança monitorada.

Figura 6.1: Gráfico Cox LHR CUSUM aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e $\xi_1 = 0.9$ para diferentes valores de referência (a) $k = 0$ (b) $k = 0.03$ (c) $k = 0.05$ e (d) $k = 0.058$.

Neste caso, pode-se questionar se realmente há uma redução de 10% no risco de falha. Contudo, para os mesmos valores de referência, o gráfico CUSUM para detectar redução de 20% no risco

($\xi_1 = 0.8$) se mantém crescente e sugere que há de fato uma melhoria na sobrevivência dos pacientes, até mesmo para o valor de referência $k = 0.06$, no qual começa a mostrar crescimento dos escores já pouco antes do dia 4000. Isso indica então que esta mudança provavelmente deve ser mais intensa que a investigada no teste anterior, e desta forma o CUSUM é mais consistente para esta situação. O CUSUM mencionado está apresentado na Figura 6.2. Nota-se neste caso que o CUSUM não foi anulado para os valores de referência testados, tendo apenas os escores reduzidos e o tempo de espera maior para apresentar crescimento constante em alguns cenários.

Figura 6.2: Gráfico Cox LHR CUSUM aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e $\xi_1 = 0.8$ para diferentes valores de referência (a) $k = 0$ (b) $k = 0.06$ (c) $k = 0.08$ e (d) $k = 0.1$.

Uma das dificuldades na aplicação deste método CUSUM está na escolha do valor de ξ_1 para monitorar, pois para cada valor considerado há um gráfico CUSUM diferente e com outra interpretação. Se sabemos exatamente qual a intensidade da mudança que acontece no risco de falha dos pacientes, então se espera que o CUSUM monitorando esta mudança deve ser eficiente ao sinalizar a alteração. No entanto, na prática, não conhecemos a intensidade da mudança que ocorre (e nem

mesmo sabemos se há de fato uma alteração na sobrevida dos pacientes), porém é válido testar vários valores e avaliar em qual o gráfico mantém mais forte a evidência de que há uma mudança (caso haja), e com isto possivelmente decidir se há alteração na sobrevivência.

Sabe-se também da importância de considerar um valor de referência k adequado para não prejudicar o poder detecção do gráfico e que o limite de controle h deve ser obtido por otimização para a situação que queremos monitorar. Nesta aplicação não estamos utilizando um algoritmo de otimização para escolha de k e h , mas é possível notar que independente do limite de controle adotado, o gráfico não teria dificuldades em identificar uma mudança para a maioria dos valores de k considerados e para várias intensidades de mudanças (ξ_1) testadas, o que nos leva a crer que esta melhoria na sobrevida (redução do risco de falha) é significativa.

Figura 6.3: Gráfico Cox martingal CUSUM, com $a_i = -1$, aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e diferentes valores de referência (a) $k = 0$ (b) $k = 0.3$ (c) $k = 0.5$ e (d) $k = 0.59$.

Outro aspecto prático importante a ser considerado é a característica prospectiva do monitoramento. Aqui apresentamos a análise dos dados do InCor em visão prospectiva, mas já conhecemos

toda a informação do acompanhamento dos pacientes por um longo período, o que nos permite ver como o gráfico se comporta muito tempo depois de o CUSUM ter acumulado o suficiente para indicar uma mudança. Na prática, no monitoramento prospectivo, não temos acesso a essa informação, o que pode dificultar na tomada de decisão. Entretanto, esse exemplo de utilização do gráfico nos possibilita entender o mecanismo deste método para ajudar nas escolhas dos parâmetros para aplicação em novos estudos de forma prospectiva.

Para o método Cox martingal CUSUM (com $a_t = -1$, para monitorar redução do risco de falha) são apresentados os gráficos na Figura 6.3. Nota-se que até $k = 0.3$ o CUSUM consegue acumular escores e indicar que há melhora na sobrevivência dos pacientes na maior parte do acompanhamento. Para $k = 0.5$ o gráfico já não suporta esta evidência na maior parte do tempo, e o mesmo ocorre para valores de referência maiores, até o último valor testado $k = 1$. Neste caso para valores de referência a partir de 0.6 o CUSUM já tem seus escores anulados. Entendemos que o valor de referência maior que 0.3 já pode comprometer o poder de detecção nesta situação, e que com este valor o gráfico pode seguramente indicar que há uma redução no risco de falha dos pacientes ao longo do tempo.

A Figura 6.4 apresenta o monitoramento dos dados do InCor pelo método Cox deviance CUSUM, que procura identificar se há melhoria na sobrevivência dos pacientes. Aqui notamos que este CUSUM consegue manter escores bem maiores que o método martingal. E mesmo para $k = 0.8$ ele apresenta uma instabilidade inicial, mas finaliza com um crescimento próximo ao dia 6000 do acompanhamento, como pode ser visto acima. Como sabemos que a mudança do tratamento foi bem antes do dia 3000, entendemos que talvez este valor de referência seja muito grande para a situação. Com valores de k a partir de 0.85 o CUSUM tem escores anulados, e para os valores entre 0.80 e 0.85 observou-se na maior parte escores nulos com alguns pontos de instabilidade em que o gráfico cresce e logo volta para o zero. É notável neste caso que o método consegue detectar seguramente e de forma rápida que há uma redução no risco de falha dos pacientes ao longo tempo para 60% dos valores de referência k testados (sequência com 101 números de 0 até 1), o que nos leva a crer que esta mudança é significativa.

Uma das diferenças principais entre este método CUSUM que considera resíduos do modelo de Cox e o método da razão de verossimilhanças está na não necessidade de especificar uma intensidade de mudança para monitorar, deve-se apenas escolher uma direção, ou seja, aumento no risco de falha ($a_t = 1$) ou redução ($a_t = -1$). Isto simplifica o monitoramento, no sentido de que há apenas um gráfico para considerar (fixado o valor de k), mas o contraponto está justamente no fato de não sabermos a intensidade da mudança na sobrevivência detectada pelo gráfico quando este apresenta um alarme. Contudo este método ainda pode ser muito útil na prática, pois é capaz de detectar que há alguma mudança (que pode ser de qualquer magnitude), e a partir disso é possível facilmente utilizar algum método para estimar os riscos de falha em cada grupo para ter ideia desta diferença.

Na prática, por exemplo, pode-se utilizar uma combinação dos dois métodos apresentados. Acompanha-se apenas dois gráficos Cox martingal CUSUM (um pra cada direção), e se há um alarme produzido ao monitorar os pacientes com algum destes gráficos, então pode-se aplicar o método Cox LHR CUSUM com diferentes intensidades ξ_1 (coerentes com a direção de mudança que foi indicada) para saber qual alteração é mais provável ter ocorrido, o que é mais simples do que controlar a princípio um gráfico diferente para várias possibilidades de mudança só com o método Cox LHR CUSUM.

As análises realizadas consideraram um modelo de Cox com a variável categórica fração de ejeção

Figura 6.4: Gráfico Cox deviance CUSUM, com $a_t = -1$, aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e diferentes valores de referência (a) $k = 0$ (b) $k = 0.6$ (c) $k = 0.8$ e (d) $k = 0.84$.

para explicar o risco de falha específico dos pacientes. A seguir avaliamos os mesmos gráficos com um modelo mais completo, que considera variáveis significativas e apontadas pelos pesquisadores em análises prévias, a saber, idade, IMC, classe funcional, fração de ejeção do ventrículo esquerdo, diâmetro do átrio esquerdo, diâmetro diastólico do ventrículo esquerdo, colesterol HDL, sódio sérico e creatinina. Todas as variáveis tiveram suas medidas originais categorizadas para contemplar a categoria NA e não perder a informação do paciente no modelo.

A Figura 6.5 mostra o gráfico Cox LHR CUSUM monitorando uma mudança de intensidade $\xi_1 = 0.8$ com este modelo completo. Nota-se que o gráfico consegue crescer significativamente a maior parte do tempo, após o dia 4000 do monitoramento para $k = 0.03$, o que levaria em torno de 1000 dias para conseguir identificar a mudança. Para valores de referência menores ele consegue apresentar este crescimento bem antes, podendo identificar a mudança em até 1 ano de monitoramento, o que é satisfatório.

Figura 6.5: Gráfico Cox LHR CUSUM aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e $\xi_1 = 0.8$ para diferentes valores de referência (a) $k = 0$ (b) $k = 0.03$ (c) $k = 0.06$ e (d) $k = 0.07$, considerando um modelo com várias variáveis.

Na Figura 6.6 é apresentado o gráfico Cox LHR CUSUM que monitora $\xi_1 = 0.9$ com o modelo completo. Para este monitoramento, percebe-se menos evidência de mudanças pelo gráfico CUSUM, que já não apresenta crescimento firme para $k = 0.03$, por exemplo, onde vemos apenas um crescimento no final do acompanhamento dos pacientes. Para $k = 0.039$ ou valores maiores, o CUSUM é completamente anulado, e não consegue então detectar que há uma mudança na intensidade monitorada.

A Figura 6.7 mostra o gráfico Cox martingal CUSUM que considera o modelo completo. Observa-se comportamento similar aos já observados, de crescimento significativo, para valores de referência k até 0.2, para o qual é possível identificar uma mudança pouco depois do dia 4000, a qual é bem mais evidente após o dia 5000. Para valores de referência a partir de 0.039 o gráfico é anulado e então não apresenta sinal de mudanças. Em geral, pode-se concluir que há alguma

mudança no risco de falha dos pacientes, que é reduzido ao longo do tempo, e que os gráficos com diferentes valores k (até 0.2) são capazes de sinalizar esta evidência em pouco mais que 1000 dias de monitoramento, e a depender do valor k , este alarme pode ocorrer bem antes, até menos que 1 ano de acompanhamento.

Figura 6.6: Gráfico Cox LHR CUSUM aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e $\xi_1 = 0.9$ para diferentes valores de referência (a) $k = 0$ (b) $k = 0.02$ (c) $k = 0.03$ e (d) $k = 0.038$, considerando um modelo com várias variáveis.

Para o gráfico Cox deviance CUSUM com modelo completo o comportamento é parecido, mas com maior evidência acumulada pelo gráfico, que consegue sustentar o crescimento ainda para $k = 0.3$, e mesmo para $k = 0.6$ ele não é completamente anulado como observado nos anteriores. A Figura 6.8 mostra este gráfico para diferentes valores k .

Figura 6.7: Gráfico Cox martingal CUSUM, com $a_i = -1$, aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e diferentes valores de referência (a) $k = 0$ (b) $k = 0.2$ (c) $k = 0.3$ e (d) $k = 0.39$, considerando um modelo com várias variáveis.

Esta análise permite constatar que a utilização de um modelo mais completo para construir os escores dos gráficos CUSUM faz com que estes cresçam mais devagar, o que pode ser reflexo da explicação do risco de falha por mais fatores específicos dos pacientes. A princípio isso parece ser mais aceitável na prática pois há menos chance de haver uma detecção de melhora associada aos pacientes e não à qualidade do tratamento em si. Com esse modelo foi notada uma melhora na sobrevivência dos pacientes em todos os métodos testados, e que esta é da ordem de 20% menos risco de falha, como visto pelo método da verossimilhança. Outros testes com mudanças mais intensas (por exemplo, 50%) também conseguiram identificar de forma eficiente evidências de mudança pelo gráfico Cox LHR CUSUM.

Figura 6.8: Gráfico Cox deviance CUSUM, com $a_t = -1$, aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e diferentes valores de referência (a) $k = 0$ (b) $k = 0.3$ (c) $k = 0.5$ e (d) $k = 0.6$, considerando um modelo com várias variáveis.

O monitoramento pelo método log-rank com gráfico ZDiff CUSUM é apresentado como uma alternativa simples, sem utilização dos modelos de sobrevivência, mas que é também eficiente na identificação de mudanças no risco de falha. A Figura 6.9 apresenta o ZDiff CUSUM aplicado aos dados do InCor com amostra de referência $n_0 = 975$, considerando $a_t = 1$ (que procura identificar melhoria na sobrevivência). Como já constatado na avaliação deste gráfico em capítulo anterior, os escores calculados, que são os incrementos da estatística de teste log-rank, tendem a ser valores pequenos, e daí entende-se que o valor de referência k não deve ser muito grande.

A vantagem do CUSUM é conseguir acumular estes pequenos escores ao longo do tempo e detectar que há um aumento significativo da estatística que indica uma alteração no risco de falha. Os gráficos apresentados com $k = 0, 0.001, 0.002$ mostram um crescimento considerável já nos primeiros dias de monitoramento, o que é ainda mais evidente para os dois primeiros valores

Figura 6.9: Gráfico ZDiff CUSUM, com $a_t = 1$, aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e diferentes valores de referência (a) $k = 0$ (b) $k = 0.001$ (c) $k = 0.002$ e (d) $k = 0.003$.

testados. O monitoramento inicia no dia 2859 de acompanhamento dos pacientes e para $k = 0.002$ observa-se um crescimento mais expressivo após o dia 4500, o que a depender do limite de controle, só pode identificar a mudança por volta dos 4 anos e meio (4500 – 2859 dias) de inspeção. Com os outros gráficos, em pouco mais que 1000 dias (menos que 3 anos) de monitoramento é possível detectar esta mudança.

Assim como nos gráficos CUSUM com resíduos do modelo de Cox, para o ZDiff CUSUM esta mudança também é identificada e não é possível afirmar qual a sua intensidade ou a partir de qual momento ela começou a valer. De todo modo, o método é útil para alertar sobre alteração no risco de falha dos pacientes, o que pode levar a equipe médica a buscar possíveis causas que estejam comprometendo a qualidade do atendimento, em caso de ser detectada uma piora na sobrevida. Para valores de referência $k = 0.003$ ou maiores o gráfico não consegue manter o crescimento expressivo, apesar de alcançar escores altos em alguns pontos. Outra característica observada é o

Figura 6.10: Gráfico ZDiff CUSUM, com $a_t = -1$, aplicado aos dados do InCor com amostra de treinamento $n_0 = 975$ e diferentes valores de referência (a) $k = 0$ (b) $k = 0.01$ (c) $k = 0.02$ e (d) $k = 0.03$.

decremento do CUSUM ao final do monitoramento, mesmo quando $k = 0$. Entende-se que isto é um efeito provocado pela finalização do estudo, em que não se tem mais informações atualizadas sobre os pacientes, algo que na prática em um monitoramento prospectivo não deve ser afetado. Vale ressaltar que esta aplicação reflete a visão prospectiva de dados disponíveis somente no período de 1991 a 2007.

A Figura 6.10 apresenta o ZDiff CUSUM com $n_0 = 975$ e $a_t = -1$ (que detecta redução na sobrevivência). Nota-se um crescimento após o dia 6000 que se mostra permanente, embora só alcance valores em torno de 3. Novamente acreditamos que este efeito de crescimento ao final se dá pela interrupção do estudo e possivelmente pela característica do log-rank que é afetado pela não observação de novas falhas. Esse comportamento se mantém para valores de k entre 0 e 0.01, mas para $k = 0.01$ este crescimento tem magnitude ainda menor, e para valores de referência maiores já não se observa mais o crescimento, assim, não há evidências suficientes pelo gráfico sobre uma piora na sobrevida dos pacientes.

Capítulo 7

Conclusões e Perspectivas

Os objetivos da tese foram aprimorar a proposta do gráfico de controle CUSUM que faz uso da estatística log-rank (pensando na formulação geral do teste com a estatística linear de postos), com discussão sobre a determinação dos limites de controle adequados. Foram desenvolvidos gráficos CUSUM que comportam a heterogeneidade por meio do modelo semiparamétrico de Cox usual, os quais consideraram o método Breslow para estimação do risco baseline que é definido no modelo para utilizar na razão de verossimilhanças completas e nos resíduos.

Como conclusões do estudo destacamos o potencial observado nos métodos novos propostos sem utilização de modelos paramétricos, que são enfatizados em alguns dos trabalhos apresentados na revisão da literatura do ponto de vista retrospectivo. No caso não paramétrico, foi visto que o gráfico ZDiff CUSUM é capaz de identificar prospectivamente e com eficiência mudanças na sobrevida de populações homogêneas.

No que se refere a populações heterogêneas, os gráficos CUSUM com modelo de Cox também se mostraram eficientes. Foi visto que o gráfico Cox LHR CUSUM é bastante ágil na identificação de uma mudança na sobrevivência, considerando que a hipótese testada utiliza o valor conhecido desta mudança. Por outro lado, os gráficos Cox Martingal CUSUM e Cox Deviance CUSUM, embora não tão ágeis, conseguem detectar também rapidamente que há alteração na sobrevida, e são mais simples por não dependerem de um valor específico para a intensidade de mudança a ser testada. Em especial o método com resíduo martingal teve um desempenho bem satisfatório.

Na aplicação, todos os métodos conseguem detectar aumento na sobrevida dos pacientes. O método da razão de verossimilhanças sugere que a intensidade desta mudança é de pelo menos 20% de redução no risco de óbito. Isso vai então de acordo com a especulação dos pesquisadores, que acreditam ter conseguido ao longo da evolução nos procedimentos proporcionar maior tempo esperado de vida aos pacientes atendidos.

Em relação a perspectivas futuras de estudo decorrentes deste trabalho podem ser destacados diversos pontos. Por exemplo, na abordagem baseada no modelo de Cox foi utilizada a verossimilhança completa, que necessita da estimação da função de risco basal. Pode ser de interesse também considerar um monitoramento com base na função de verossimilhança parcial, que é isenta desta necessidade. Pode-se investigar ainda medidas alternativas à razão de verossimilhanças, como a razão de sobrevivências relativa, para identificar alterações no perfil da sobrevida.

Também com relação ao método de avaliação do desempenho da detecção pelo gráfico, foi utilizado apenas o ARL como medida de desempenho. Outras formas de avaliação da eficiência podem ser utilizadas. E mesmo com relação ao ARL, foi utilizado um modelo discreto geométrico censu-

rado para permitir a estimação desta medida, vista como o número médio de amostras/inspeções até que o gráfico apresente um sinal. Alternativamente, é possível avaliar o tempo médio até o gráfico sinalizar a mudança e assim considerar uma modelagem contínua para esta variável.

Para os métodos que utilizam a estatística linear de postos, concentramos no teste log-rank em particular, com ênfase para o gráfico ZDiff CUSUM, devido a instabilidades observadas nos métodos que consideram a estatística qui-quadrado. Sugere-se em estudos posteriores investigar melhor este método e possivelmente reparar esses problemas com a utilização de pesos adequados nos escores. Também considerar outras estatísticas, como a de Wilcoxon, pode ser de interesse, tendo em vista que pode haver situações nas quais o log-rank tem desempenho inferior.

Outro ponto de interesse é a identificação do ponto de mudança, que não é enfoque principal da maioria dos trabalhos discutidos e não foi desenvolvido neste trabalho. Os gráficos conseguem sinalizar que há uma alteração no processo observado, mas isso ocorre bem depois do momento em que a mudança de fato acontece. Estimar este instante é importante para auxiliar os pesquisadores a descobrir causas especiais que provocaram a alteração na sobrevida de pacientes. Sobre este tópico, foi feito um estudo de simulação para uma abordagem simples com a estatística log-rank, e que pode ser melhor desenvolvida em pesquisas futuras. Além disso, pode-se considerar influência de covariáveis com diferentes modelos de sobrevivência estruturados na modelagem de ponto de mudança.

Finalmente, um tópico que é relevante refere-se aos modelos de longa duração ou fração de cura. Os dados utilizados como motivação para este trabalho apresentam tal característica e então é interessante desenvolver gráficos CUSUM com esta modelagem para investigar mudanças na sobrevida. Além disso, por ser um modelo mais flexível, é possível que sua utilização proporcione resultados ainda mais eficientes em comparação com modelos usuais da análise de sobrevivência.

Apêndice A

Processos Pontuais

Seja T_j o tempo de falha do indivíduo j e C_j seu tempo de censura. Define-se, Z_j , δ_j , $N_j(t)$ e $Y_j(t)$, que são, respectivamente, o tempo observado, o indicador de censura, o processo de contagem e o indicador de risco, como:

$$\begin{aligned} Z_j &= \min\{T_j, C_j\} \\ \delta_j &= I(T_j \leq C_j) \\ N_j(t) &= I(Z_j \leq t; \delta_j = 1) \\ Y_j(t) &= I(Z_j \geq t). \end{aligned} \tag{A.1}$$

Desta forma, tem-se:

$$dN_j(t) = \begin{cases} 1 & \text{se } T_j \in [t, t + dt) \\ 0 & \text{caso contrário.} \end{cases} \tag{A.2}$$

O processo de intensidade é a caracterização do processo de contagem, e é definido como

$$d\Lambda_j(t) = E[dN_j(t) | \mathcal{F}_{t-}]. \tag{A.3}$$

em que $\mathcal{F}_{t-}$ representa toda a informação já conhecida até o instante imediatamente anterior a t .

Se $Y_j(t) = 0$ (o indivíduo não está em risco) $\Rightarrow E[dN_j(t) | \mathcal{F}_{t-}] = 0$.

Se $Y_j(t) = 1 \Rightarrow$

$$\begin{aligned} E[dN_j(t) | \mathcal{F}_{t-}] &= 0 \cdot P\{dN_j(t) = 0 | \mathcal{F}_{t-}\} + 1 \cdot P\{dN_j(t) = 1 | \mathcal{F}_{t-}\} \\ &= P\{dN_j(t) = 1 | \mathcal{F}_{t-}\} \\ &= P\{dN_j(t) = 1 | T_j \geq t, C_j \geq t\} \\ &= P\{T_j \in [t, t + dt) | T_j \geq t, C_j \geq t\} \\ &= P\{T_j \in [t, t + dt) | T_j \geq t\} \end{aligned}$$

$$\begin{aligned} \alpha_j(t) &= \lim_{dt \rightarrow 0} \frac{P\{T_j \in [t, t + dt) | T_j \geq t\}}{dt} \\ dt \downarrow 0 \Rightarrow \alpha_j(t) &\approx \frac{P\{T_j \in [t, t + dt) | T_j \geq t\}}{dt} \Rightarrow dt \alpha_j(t) \approx P\{T_j \in [t, t + dt) | T_j \geq t\} \end{aligned}$$

Se $Y_j(t) = 1$,

$$d\Lambda_j(t) = E\{dN_j(t)|\mathcal{F}_{t-}\} = P\{T_j \in (t, t+dt]|T_j \geq t\} \approx \alpha_j(t) dt$$

$$Y_j(t) = 0 \Rightarrow d\Lambda_j(t) = E\{dN_j(t)|\mathcal{F}_{t-}\} = 0$$

$$Y_j(t) = 1 \Rightarrow d\Lambda_j(t) \approx \alpha_j(t) dt$$

Em geral,

$$d\Lambda_j(t) = Y_j(t) \alpha_j(t) dt$$

$$\Lambda_j(t) = \int_0^t d\Lambda_j(s) = \int_0^t Y_j(s) \alpha_j(s) ds$$

Apêndice B

Fórmula do CUSUM

A decisão de tomar uma ação ocorre se

$$C_j - \min_{0 \leq i \leq j-1} C_i \geq h,$$

com $h > 0$ representando o limite de controle e o escore acumulado é $C_j = \sum_{k=1}^j w_k$. Define-se o CUSUM como

$$C'_j = \max(C'_{j-1} + w_j, 0),$$

com $C'_0 = 0$. Note que $w_j = C_j - C_{j-1}$, então, definindo $S_0 = 0$, temos que:

$$C'_1 = \max(C'_0 + w_1, 0) = \max(w_1, 0) = \max(C_1 - C_0, 0).$$

Com isso,

$$\begin{aligned} C'_1 \geq h &\Leftrightarrow \max(C_1 - C_0, 0) \geq h \\ &\Leftrightarrow C_1 - C_0 \geq h. \end{aligned}$$

Para $j = 2$, temos

$$C'_2 = \max(C'_1 + w_2, 0) = \max(C'_1 + C_2 - C_1, 0) = \max[C_2 - (C_1 - C'_1), 0].$$

Com isso,

$$\begin{aligned} C'_2 \geq h &\Leftrightarrow \max[C_2 - (C_1 - C'_1), 0] \geq h \\ &\Leftrightarrow C_2 - (C_1 - C'_1) \geq h \\ &\Leftrightarrow C_2 - [C_1 - \max(C_1 - C_0, 0)] \geq h. \end{aligned}$$

Se $C_1 < C_0 \Rightarrow [C_1 - \max(C_1 - C_0, 0)] = C_1$; do contrário $[C_1 - \max(C_1 - C_0, 0)] = C_1 - C_1 + C_0 = C_0$. Portanto, $[C_1 - \max(C_1 - C_0, 0)] = \min_{0 \leq i \leq 1} S_i$, e temos que:

$$C'_2 \geq h \Leftrightarrow C_2 - \min_{0 \leq i \leq 1} C_i \geq h.$$

De forma análoga, segue que

$$\begin{aligned}
C'_3 \geq h &\Leftrightarrow C'_2 + w_3 \geq h \\
&\Leftrightarrow C_3 - (C_2 - C'_2) \geq h \\
&\Leftrightarrow C_3 - \{C_2 - \max[C_2 - (C_1 - C'_1), 0]\} \\
&\Leftrightarrow C_3 - \{C_2 - \max[C_2 - (C_1 - \max(C_1 - C_0, 0)), 0]\} \\
&\Leftrightarrow C_3 - \{C_2 - \max[C_2 - \min(C_1, C_0), 0]\}.
\end{aligned}$$

Se $C_2 < \min(C_1, C_0) \Rightarrow C_2 - \max[C_2 - \min(C_1, C_0), 0] = C_2$; caso contrário, temos que $C_2 - \max[C_2 - \min(C_1, C_0), 0] = C_2 - C_2 + \min(C_1, C_0) = \min(C_1, C_0)$. Resume-se então a expressão como $C_2 - \max[C_2 - \min(C_1, C_0), 0] = \min\{C_2, C_1, C_0\} = \min_{0 \leq i \leq 2} C_i$. Logo,

$$C'_3 \geq h \Leftrightarrow C_3 - \min_{0 \leq i \leq 2} C_i \geq h.$$

Sucessivamente, conclui-se que

$$C'_j \geq h \Leftrightarrow C_j - \min_{0 \leq i \leq j-1} C_i \geq h.$$

Apêndice C

Cálculos de Razão de Chances

No exemplo mostrado para monitoramento de uma variável binária é definida uma razão entre a chance de ocorrer complicação para o indivíduo no período de 30 dias após a realização da cirurgia e a chance pré-operativa de ocorrer complicação para o mesmo indivíduo. Desta forma, sendo p_{Ai} a probabilidade de complicação pós-operativa para o paciente i , sob a hipótese alternativa, obtém-se:

$$\begin{aligned}
 \frac{\frac{p_{Ai}}{1-p_{Ai}}}{\frac{p_i}{1-p_i}} &= R_A \\
 \frac{p_{Ai}}{1-p_{Ai}} &= R_A \frac{p_i}{1-p_i} \\
 p_{Ai} &= R_A \frac{p_i}{1-p_i} (1-p_{Ai}) \\
 p_{Ai} &= R_A \frac{p_i}{1-p_i} - R_A \frac{p_i}{1-p_i} p_{Ai} \\
 p_{Ai} \left(1 + R_A \frac{p_i}{1-p_i} \right) &= R_A \frac{p_i}{1-p_i} \\
 p_{Ai} \left(\frac{1-p_i + R_A p_i}{1-p_i} \right) &= R_A \frac{p_i}{1-p_i} \\
 p_{Ai} &= \frac{R_A p_i}{1-p_i + R_A p_i}
 \end{aligned} \tag{C.1}$$

De modo análogo, tem-se $p_{0i} = \frac{R_0 p_i}{1-p_i + R_0 p_i}$ a probabilidade de complicação pós-operativa sob H_0 . Com isso, a razão de verossimilhanças é dada por:

$$\frac{L_1}{L_0} = \frac{(p_{Ai})^{y_i} (1-p_{Ai})^{1-y_i}}{(p_{0i})^{y_i} (1-p_{0i})^{1-y_i}}. \tag{C.2}$$

Se $y_i = 1$, esta razão é dada por:

$$\frac{L_1}{L_0} = \frac{p_{Ai}}{p_{0i}} = \frac{\frac{R_A p_i}{1-p_i + R_A p_i}}{\frac{R_0 p_i}{1-p_i + R_0 p_i}} = \frac{(1-p_i + R_0 p_i) R_A}{(1-p_i + R_A p_i) R_0} \tag{C.3}$$

Se $y_i = 0$, esta razão é dada por:

$$\frac{L_1}{L_0} = \frac{1-p_{Ai}}{1-p_{0i}} = \frac{\frac{1-p_i + R_A p_i - R_A p_i}{1-p_i + R_A p_i}}{\frac{1-p_i + R_0 p_i - R_0 p_i}{1-p_i + R_0 p_i}} = \frac{\frac{1-p_i}{1-p_i + R_A p_i}}{\frac{1-p_i}{1-p_i + R_0 p_i}} = \frac{1-p_i + R_0 p_i}{1-p_i + R_A p_i} \tag{C.4}$$

Referências Bibliográficas

- Bashirov, A. E., Kurpinar, E. M. & Özyapıcı, A. (2008). Multiplicative calculus and its applications, *Journal of Mathematical Analysis and Applications* **337**(1): 36–48.
- Biswas, P. & Kalbfleisch, J. D. (2008). A risk-adjusted cusum in continuous time based on the cox model, *Statistics in Medicine* **27**: 3382–3406.
- Chen, M.-H., Ibrahim, J. G. & Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction, *Journal of the American Statistical Association* **94**(447): 909–919.
- Cox, D. R. & Miller, H. D. (1965). *The Theory of Stochastic Processes*, Chapman & Hall, London.
- Cox, D. R. & Oakes, D. (1984). *Analysis of survival data*, Vol. 21, CRC Press.
- Gandy, A., Kvaløy, J. T., Bottle, A. & Zhou, F. (2010). Risk-adjusted monitoring of time to event, *Biometrika* **97**(2): 375–388.
- Grigg, O. & Farewell, V. (2004). An overview of risk-adjusted charts, *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **167**(3): 523–539.
- Harrington, D. P. & Fleming, T. R. (1982). A class of rank test procedures for censored survival data, *Biometrika* **69**(3): 553–566.
- Hawkins, D. M. & Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*, Springer Science & Business Media.
- Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, Wiley, New York.
- Klein, J. P. & Moeschberger, M. L. (2005). *Survival analysis: techniques for censored and truncated data*, Springer Science & Business Media.
- Lawless, J. F. (2003). *Statistical models and methods for lifetime data*, John Wiley, New York.
- Li, Z. & Kong, Z. (2015). A generalized procedure for monitoring right-censored failure time data, *Quality and Reliability Engineering International* **31**(4): 695–705.
- Li, Z., Zhou, S., Sievenpiper, C. & Choubey, S. (2012). Statistical monitoring of time-to-failure data using rank tests, *Quality and Reliability Engineering International* **28**(3): 321–333.
- Lorden, G. (1971). Procedures for reacting to a change in distribution, *The Annals of Mathematical Statistics* pp. 1897–1908.
- Maller, R. A. & Zhou, X. (1996). *Survival analysis with long-term survivors*, Wiley, New York.
- Marçula, M., Cuoco, M., Yamada, A., Freitas, H., De Paula, R., Uemura, R., Ferreira, P., Alencar, A., Barretto, A. & Mansur, A. (2011). Comparison of prognostic variables between two cohorts of outpatients with heart failure in the last two decades, *Paris European Heart Journal* **32**: 459–459.
- Montgomery, D. C. (2007). *Introduction to statistical quality control*, John Wiley & Sons.

- Oliveira, J. W., Valença, D. M., Medeiros, P. G. & Marçula, M. (2016). Risk-adjusted monitoring of time to event in the presence of long-term survivors, *Biometrical Journal* **58**(6): 1485–1505.
- Page, E. S. (1954). Continuous inspection schemes, *Biometrika* **41**(1/2): 100–115.
- Peto, R. & Peto, J. (1972). Asymptotically efficient rank invariant test procedures, *Journal of the Royal Statistical Society. Series A (General)* pp. 185–207.
- Rodrigues, J., Cancho, V. G., de Castro, M. & Louzada-Neto, F. (2009). On the unification of long-term survival models, *Statistics & Probability Letters* **79**(6): 753–759.
- Schoenfeld, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions, *Biometrika* pp. 316–319.
- Sego, L. H., Reynolds, M. R. & Woodall, W. H. (2009). Risk-adjusted monitoring of survival times, *Statistics in medicine* **28**(9): 1386–1401.
- Sen, A. & Srivastava, M. S. (1975). On tests for detecting change in mean, *The Annals of statistics* pp. 98–108.
- Shewhart, W. A. (1931). *Economic control of quality of manufactured product*, Macmillan, New York.
- Steiner, S. H., Cook, R. J. & Farewell, V. T. (2001). Risk-adjusted monitoring of binary surgical outcomes, *Medical Decision Making* **21**(3): 163–169.
- Steiner, S. H., Cook, R. J., Farewell, V. T. & Treasure, T. (2000). Monitoring surgical performance using risk-adjusted cumulative sum charts, *Biostatistics* **1**(4): 441–452.
- Therneau, T. M., Grambsch, P. M. & Fleming, T. R. (1990). Martingale-based residuals for survival models, *Biometrika* **77**(1): 147–160.
- Tsiatis, A. A. (1981). A large sample study of cox's regression model, *The Annals of Statistics* pp. 93–108.
- Wald, A. (1947). *Sequential analysis*, Wiley, New York.
- Woodall, W. H. (2006). The use of control charts in health-care and public-health surveillance, *Journal of Quality Technology* **38**(2): 89–104.
- Wu, J. & Xiong, X. (2017). Group sequential survival trial design and monitoring using the log-rank test, *Statistics in Biopharmaceutical Research* **9**(1): 35–43.